

Innovation, evaluation design and typologies of professional learning

BOYLAN, Mark <<http://orcid.org/0000-0002-8581-1886>> and DEMACK, Sean <<http://orcid.org/0000-0002-2953-1337>>

Available from Sheffield Hallam University Research Archive (SHURA) at:

<http://shura.shu.ac.uk/22406/>

This document is the author deposited version. You are advised to consult the publisher's version if you wish to cite from it.

Published version

BOYLAN, Mark and DEMACK, Sean (2018). Innovation, evaluation design and typologies of professional learning. *Educational Research*, 60 (3), 336-356.

Copyright and re-use policy

See <http://shura.shu.ac.uk/information.html>

Innovation, evaluation design and typologies of professional learning

Mark Boylan* and Sean Demack

Sheffield Hallam University

*Contact - m.s.boylan@shu.ac.uk

Abstract

Background

Current policy discourses emphasise the importance of evidence in education, including evidencing the impact of teacher professional learning on student outcomes. Randomised controlled trial (RCT) designs are promoted to measure 'impact'. Recent debates about this reflect longer standing methodological disputes. Advocates of comparative approaches contend that these are uniquely capable of establishing causality. However, others dispute this and consider their application in education as often being flawed. Whilst acknowledging the importance of these debates, our concern is how RCTs and similar evaluation designs are specifically used to evaluate innovations in which professional learning is important.

Purpose

Arguably, professional learning is often under-theorised within experimental and quasi-experimental designs. The purpose of this paper is to address this by encouraging developers of innovations and evaluators to consider a proposed typology of professional learning and other important relevant methodological issues. This is so that developers of innovations that involve professional learning are better able to theorise their endeavours and to support more appropriate design of RCTs and other forms of evaluation of innovations.

Sources of evidence

Theoretical and methodological literature from diverse fields is drawn on, namely: descriptions of RCT implementation and process evaluation designs; research on effective professional development; and theoretical models of professional learning. Insights and theories from this literature are used to develop and illustrate the typology and to identify methodological concerns and potential ways to address these.

Main argument

In trials of those innovations that involve professional learning, there is both assessment of the extent to which professional learning occurs and also whether resulting changes in practice improve outcomes. A novel typology of three different ways that professional learning may occur in innovations is proposed. This is related to the centrality (or not) of professional learning to the innovation's success and related to the form and purpose of the professional learning involved. The three analytical categories described are: pedagogical professional learning, technical professional learning and curriculum professional learning. Based on this typology, features of professional learning that are likely to lead to impact on student outcomes are discussed. Tensions are identified between the implementation of experimental and quasi-experimental designs and interpretation of resulting evidence. Further, tensions are identified between the complex and recursive nature of pedagogical professional learning systems and the models of linear pathways in some RCT designs. This is illustrated by discussing examples of innovations and trials.

Conclusion

The proposed typology and greater theorisation of professional learning can support more robust evaluation design. It is important to assess rigorously teacher learning alongside changes in student outcomes.

Key words: professional learning, professional development, RCTs, evaluation, methodology, theory

Introduction

There has been recent growing interest in England and elsewhere in the use of experimental and quasi-experimental research designs as a means of developing evidence (Goldacre 2013; Haynes et al. 2012; Torgerson 2009). The Education Endowment Foundation (EEF), with funding from the UK's Department for Education (DfE), as well as other sources, has funded a large number of efficacy and effectiveness trials of innovations. In addition to EEF trials, the DfE has directly funded a number of randomised controlled trials or quasi-experimental studies of innovations.

We use the term innovation to refer to practices, activities or materials or a combination of these which are different from those in use before the innovation, and where differences are of a type and to an extent that the innovation has the potential to lead to changed

outcomes. In reports on RCTs and similar evaluations in education, often innovations are referred to as 'interventions' or, if more medicalised discourse is used, as 'treatments'. Given our aim of addressing both evaluation researchers and designers of professional learning, the term 'innovation' is used as it is more neutral and meaningful outside of experimental and quasi-experimental evaluation. However, in discussing accounts of specific trials we use the terms the evaluators adopt. Generally, we use 'intervention' in relation to the group or sample that is experiencing or participating in the innovation (when more medicalised discourse is used, the term 'treatment' group is sometimes used as an equivalent). Using 'innovation' and 'intervention' in these ways helps to avoid confusion between changed practice or the programme being evaluated (the innovation) and the group experiencing the programme (the intervention group).

Many of the innovations recently funded in England have been either directly focused on teacher professional development programmes or have involved some element of teacher professional learning in them. Others have included the professional learning of teaching assistants. Innovations may vary in the extent to which educator professional learning is: the key component of the innovation; an important aspect of it; or a more peripheral element. An example of the latter is where professional learning is an emergent phenomena rather than an essential feature planned for within the innovation.

The use of RCTs in education is contentious. Some argue that the evidence they provide is more useful and more secure than evidence from any other form of research (Goldacre 2013; Haynes et al. 2012); colloquially they are described as 'the gold standard'. However, others dispute their value and urge caution about the potential misapplications or negative consequences of this approach in terms of how learning and teaching are understood (see, for example, Biesta 2010; Harrison and McCaig 2016; Lehman 2015; Watson 2015).

It is beyond the scope of this paper to consider these important debates fully. However, whilst sympathetic to those who dispute the 'gold standard' arguments, the position we take is that, in principle, a well-designed RCT (CONSORT 2010) may be able to identify if participation in a teacher professional learning programme results in (or causes) a positive or negative 'impact' with respect to a measurable and quantified outcome (such as pupil attainment).

However, causation is necessarily probabilistic, given that an RCT design is based on reducing the possibility of false results within given probabilities. An RCT can provide empirical evidence of causation within a probabilistic framework determined by the specific research design. The use of randomisation and the size/scale of an RCT are used to limit both the probability of falsely concluding no impact (a false negative) and the probability of falsely concluding an impact (a false positive) to a specified level (commonly 0.20 and 0.05 respectively). These probabilistic limits to drawing false-negative and false-positive conclusions assume zero attrition/drop out over the time period of an RCT. Missing data

undermines the effect of randomisation balancing other factors that may influence outcomes (see below for discussion of randomisation and confounding variables). This weakens the validity of drawing causal conclusions. Even with zero attrition, a single RCT would not be sufficient to draw causal conclusions with complete certainty. Replication - that is, testing the programme in further trials - is needed in order to build confidence that a teacher professional learning programme results in a genuine positive or negative impact on intended outcomes. Moreover, the level of confidence in trial outcomes depends not only on the size of the trial - larger trials, in general, tend to reduce the risk of false results - but also on the size of the effect or level of impact of the innovation being evaluated.

Further, whilst internal validity in RCTs following recognised procedures can be high, this approach can lack external validity (Rothwell 2005): that is, outcomes may not be applicable outside of the trial conditions. An example of this in education is that schools applying to join trials may have particular characteristics that are not shared with all schools - such as the capacity to engage in the innovation - and so outcomes may not apply to schools in general.

Further, the claim for probabilistic causation is more convincing if an underlying change mechanism is identified that makes sense, theoretically, of the findings. Change mechanisms may be identified by considering patterns in those findings beyond a narrow focus on positive effect sizes. An example of this, in relation to professional learning, is the observation that different types of professional development such as coaching, lesson study or action research are effective when the pedagogical focus of the development activity is closely related to teachers' immediate concerns, including professional dilemmas. In such an instance, this common feature resonates with observation of patterns in adult learning in which a relationship between dilemmas and transformative learning has been identified (for example, Mezirow 1991).

To claim causality, some argue that the identification of mechanism is essential (see for example, Pawson 2005; Pawson and Tilly 1997). If this view is accepted, then one response to the 'gold standard' argument for RCTs is that the identification of a change mechanism is likely to rest on other research methods, often those that are part of a process evaluation. Whilst RCTs can identify whether an effect has occurred and, potentially, causality, other methods are needed in order to identify why this causal relationship exists. Even if the argument about change mechanisms and causality is not accepted, understanding the reason why an effect is found as part of an RCT, in any case, requires rich research on the implementation of the programme being evaluated. Thus, as part of RCT design, rigorous implementation and process evaluations are advised (see for example, Humphrey et al. 2016).

Regardless of general debates about the value of RCTs in education, there are a number of specific issues to consider with regard to evaluations of educational innovations that involve

significant levels of teacher professional learning. In the next section of the paper, we briefly discuss further the logic of RCTs. We then consider theorisation of varieties of forms and purposes of professional development. Drawing on this discussion, we propose a novel typology of three ways in which professional learning may be involved in educational innovations. The examples that are used to illustrate this heuristic are impactful programmes, evaluated using experimental or quasi-experimental designs. Further, we consider the evidence base on effective professional development. We argue that there are a number of tensions between what is effective in professional development programmes and what RCT designs require. We go on to discuss the implications of the arguments for the design and conduct of trials of innovations that include professional learning and curriculum development projects that involve professional development.

The logic of RCTs

In this section, we briefly recap the logic of RCTs - that is, the basis for the method and the claims to causality of findings. Within evaluation research, change over time can be observed using a longitudinal design that collects data before, during and after a teacher professional learning programme. However, whilst evidence of positive or negative change might be observed, using such a single-sample longitudinal design, it is not valid to conclude that the observed change was caused by the programme being evaluated. This is because the change might have happened anyway and regardless of the programme. That the 'change might have happened anyway' is known as the counterfactual. RCTs are considered as a 'gold standard' because, it is claimed, they have the capacity to isolate the causal impact of something (such as a teacher professional learning programme) on a measurable and quantified outcome (such as pupil attainment).

RCTs require at least two samples; one to experience the programme being evaluated and one to act as a control group. The purpose of the control group is to capture the counterfactual. If the observed change over time for the 'intervention' group is greater or smaller than the change observed in the control group, a positive or negative impact can be concluded. Further, randomisation potentially removes the possibility of systematic differences between the intervention and control groups. This is true for potentially important variables, many of which are not observed. In RCT methodology, 'observed' is used to refer to variables that are measured through a variety of data collection methods. Unobserved variables such as school, teacher or pupil characteristics may dampen, amplify or otherwise mediate the innovation. Such variables are often described as confounding as they mean that observed changes may not solely be a cause of the innovation.

Randomisation procedures are the principal method used to create intervention and control groups that can be treated as equivalent. If the sample is large enough then any confounding variables should, it is argued, be evenly distributed - that is, the samples will be

balanced. A consequence of this is that it is not necessary to know, in advance, what all the important variables are.

To claim causality for an innovation, it is also important that there is consistency of implementation with design, generally referred to as 'fidelity'. Also it must be established that the control group does not receive the innovation or a similar one. Thus, there is a sense in which both the intervention and control group are 'controlled'. The logic of control relates both to the intervention group that follows clearly specified protocols for the innovation and the use of the comparison group that does not experience the innovation. The cause of any observed positive or negative impact can then be attributed to the intended difference between the two groups – i.e. that one of the groups participated in an educational programme.

A single RCT will be unable to identify conclusively whether or not a particular programme had a causal impact. To do this requires multiple RCTs that evolve through stages. The starting point is a pilot stage, where a programme is developed / honed to assess whether an RCT is a feasible method that might lead to an 'efficacy trial'. During the efficacy trial, the impact of a programme is evaluated under 'optimal conditions' informed by the pilot stage. If an impact is identified at the efficacy stage, an 'effectiveness trial' is needed to see whether this impact remains if a programme is rolled out to a larger (ideally nationally representative) sample (see Glasgow, Lichtenstein, & Marcus 2003). Because the RCT can only identify whether a programme had a probabilistic causal impact, replication, of the effectiveness stage, is needed in order to build confidence that the observed impact is genuine. It is not until the effectiveness stage that external validity becomes a central concern for RCTs.

Varieties of professional development and learning

In this section, we discuss analyses of forms of professional development and learning. A wide variety of terms is used for professional learning, development and teacher change. The terms used entail commitments to particular views on how teacher learning happens, should happen and what outcomes are considered (see O'Brien and Jones 2014, for discussion). As important as these debates are, in this paper, for simplicity, we do not address them directly. We define key terms as follows: by *professional development*, we refer to the activity or experience that teachers engage in - in other words, it is short for professional development event, activity or programme. To refer to the outcome of professional development we use the term *professional learning*. Any exceptions to this usage are when we are referring to or citing other authors.

It is recommended that innovations should be described in detail in reports of trials (Hoffman et al. 2014; Humphry et al. 2016) – this includes describing the materials, procedures, activities, processes, goals, theory and rationale of the innovation. However, we contend that often the theory and rationale, if provided, will focus on the potential impact

on pupils and not on the theory or purpose of a professional development component or on professional learning.

Kennedy (2005, 2014) offers a much-cited categorisation of forms of professional development. She describes nine categories of professional development (PD) models that vary in the degree of autonomy experienced by teachers. The models form a continuum from most transmissive forms to most transformative forms. Transmissive forms of CPD are associated with equipping teachers with the skills and knowledge to implement policy or recommended practices; transformative forms of CPD support the development of criticality about policy and practices (Kennedy 2005). Considering the nine categories listed below, there are increasing levels of teacher autonomy from the more transmissive to the more transformative.

In summary, these models are:

- *Training* - skills and competence-based, with a pre-determined agenda/content, 'delivered' to the teacher by an 'expert', with the teacher a passive recipient, commonly delivered off-site
- *Award-bearing* - usually awards from Higher Education (HE) or validated by HE, accredited Continuing Professional Development (CPD)
- *Deficit* - focused on perceived weakness usually of individual teachers linked to performance management
- *Cascade* - teachers receiving CPD are also trained to disseminate information to colleagues
- *Standards-based* - content of PD is focused on achieving competencies or standards, teaching as composite of competencies
- *Coaching/mentoring* - usually a one-to-one relationship, can be peer but usually hierarchical - can be more or less transformative
- *Community of practice (CoP)* - collective engagement in groups (not mentioned by Kennedy, but 'lesson study' would be a clear example of a CoP approach)
- *Action research* - teacher as researcher/enquirer
- *Transformative* - not a model in itself, but a way of using other models to effect PD that includes addressing issues of agendas and power.

In a related categorisation, Sachs (2011) proposes a fourfold schema, professional development as variously, and herein summarised as:

- *Retooling* - development of new skills, where immediate application is the prime objective; positions teacher as manager of student learning, and a technical approach to teaching
- *Remodelling* - this also does not challenge orthodoxies, is based on experts 'telling' teachers and building on previous practices and occurs in two variations -
 - teaching as performance, emphasising how to engage/entertain students
 - enhancement of teacher's content and pedagogical knowledge
- *Revitalizing* - marked by teacher renewal, reflective practice in and on action, and the development of collaborative and collegial internal and external networks

- *Reimagining* - "transformative in its intent and practice" (p. 160); this is described as dialogic, authentic professional learning that "seeks to develop teachers who are creative developers of curriculum and innovative pedagogues" (p.161); values risk-taking, divergence, inquiry approaches; this is deeply collaborative - both between teachers and teachers with students - this allows for reimagining social relations.

Considering these purposes and characteristics of professional development is important. Not least because the features identified in transformative models are found in reviews and meta-analysis of professional learning that is impactful on student outcomes. This is returned to in the next section.

An alternative is to consider conceptions of teaching and the nature of teacher professionalism and, subsequently, the relationship of this to professional learning. Winch, Oancea and Orchard (2015) discuss three conceptions of teaching:

- teacher as a craft worker,
- teacher as an executive technician,
- teacher as professional.

The first conception comes from viewing teaching as a craft, understood in a narrow sense. The emphasis here is on experiential and localised knowledge. The teacher as technician recognises the value of evidence and that teaching can and should be supported by generalisable, applicable knowledge of 'what works', evidence and 'best practice'. The teachers are positioned as the implementers of this knowledge and effective teaching as faithful implementation. The third conception, teacher as professional, emphasises the importance of practical judgement based on theoretical, technical and experiential knowledge; thus critical reflection on practice and on evidence are important. This entails reflection on reasons for action and the value and weight of different types of knowledge. Winch, Oancea and Orchard argue that educational research cannot lead to "clear protocols for action in every circumstance" (p. 211). Thus, as in Kennedy and Sachs typologies, differences are posited in the degree of teacher agency.

The proposers of the professionalism typology are particularly concerned with the relationship of educational research to professional learning. However, the same point can be extended to the relationship between educational knowledge, in general, and forms of professional learning. Professional learning can variously focus on craft (experiential and local aspects) knowledge, technical know-how, or professional critical reflective knowledge.

The three models outlined above - Kennedy, Sachs and Winch et al. - offer heuristics that model recognisable differences between forms of professional development, even though the boundaries between categories are more complex in practice. However, operationalising both the conceptions of transformative learning (Kennedy 2014; Sachs

2011) or professional learning (Winch, Oancea and Orchard 2015) are problematic, particularly in the context of RCTs. This is because the extent to which professional development is, for example, transformative cannot be specified in advance. It depends on teachers' starting points and prior experience, as relevant to the focus of professional learning. These in turn will influence initial responses or orientation to the professional learning activity (Opfer and Pedder 2011). The extent to which transformation happens also depends more obviously on the degree to which professional learning occurs.

Further, what is 'transformative' is dependent on standpoint and educational ideology. Proponents of at least some understandings of 'evidence-based teaching' might consider that teachers' learning would be transformative if they had previously relied on local personal experience and now adopted teaching practices that are evidence-based. This would represent a paradigm change for the teachers. This would, arguably, still be so if the adopted practices could be considered - in Kennedy and Sach's schemas - 'technical'. Put another way, the type of knowledge or skill does not in itself determine whether learning represents a transformation. However, proponents of some understandings of 'evidence-informed teaching' (see Coldwell et al. 2017), in contrast, might consider that learning would not be transformative if it involved adopting recommended practices without careful weighting of different types of evidence in relation to teachers' contexts. Thus 'transformative' is a relative term.

Similarly, it would be confusing - if following Winch, Oancea and Orchard - to refer to 'professional learning' as only denoting one of the types of learning that are ordinarily referred to as professional learning in everyday discourse. That is, what would count in their account as craft learning, technical learning and professional learning would all be referred to as professional learning by teachers in schools and others. In order to overcome potential confusion, below, we use the term 'pedagogical professional learning' to refer to what, for Kennedy, would be more 'transformative' learning, for Sachs 'reimagining' professional development, and Winch, Oancea and Carpenter 'professional' learning. The sense of criticality is included in this concept of pedagogical learning, if we consider it to refer to, firstly, teachers' actions being informed by theory and understandings of students' needs and appraisal of relevant educational and societal situations and systems, and secondly, that teachers are, at least in part, creators of their pedagogies.

It is notable that the three different perspectives - Kennedy, Sachs, and Winch et al.'s - all identify forms of professional development that focus on technical learning, meaning the application by teachers of improved techniques, skills or specific knowledge. Thus, the adoption of the term 'technical' appears more straightforward and this is the approach taken below. The term 'curriculum professional learning' is a neologism in this context and so less ambiguous.

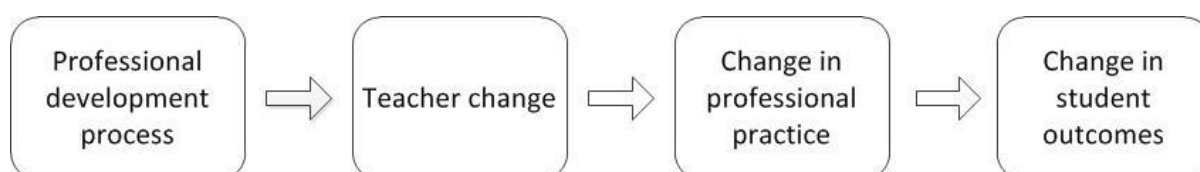
Analysing professional development innovations

Drawing on the analyses of learning and conceptions of teaching discussed above, we propose an alternative analytical distinction, suitable for use in RCTs, between pedagogical professional learning and technical professional learning, as well as a hybrid form found in innovations involving curriculum design. In this section, each of these three forms are illustrated with an abstracted theory of change diagram (Weiss 1997), and illustrated with examples from reported randomised controlled trials or other quasi-experimental studies. In selecting these examples, we draw on those identified as leading to impact on professional learning in a meta-review and assessed as meeting standards of high quality evidence (Yoon et al. 2007).

Form 1: Pedagogical professional learning

In this form, professional learning itself is the focus for innovation. Professional development activities are intended to lead to changes in professional practice. This will vary from teacher to teacher, as teachers modify their teaching in response to professional learning. Pedagogical professional learning involves experimentation by the teacher rather than the implementation of a predesigned technique. This is illustrated in Figure 1.

Figure 1 Pedagogical professional learning theory of change



Source: authors' original, unpublished Figure

An example of such an innovation is reported by Carpenter et al. (1989). The main professional development activity consisted of a workshop at the start of the year. The focus was on developing teachers' capacity to be aware of children's thinking and problem-solving skills in mathematics, and then apply this knowledge to inform their teaching and their choices of curriculum materials. It was considered that teachers' on-going practice over the following year of teaching, as much as the initial workshop, would constitute 'treatment' (that is, in a professional development context, a changed experience, stimulus or new knowledge or practice). Given this, the collection of classroom observation data was not to assess fidelity of treatment implementation - as this was not applicable. Fidelity of implementation, as noted above, is the consistency of implementation with innovation design. In professional learning contexts this may mean the extent to which PD activities are implemented as intended and opportunities are engaged in. Instead of collecting fidelity data the researchers' purpose was "rather to obtain quantifiable data that would help us understand what the treatment actually was." (p.14). As well as measures of pupil learning and attitudes, secondary measures were gained through: lesson observations; surveys of

teacher beliefs; and measures of teacher knowledge of students' number fact and problem solving strategies compared with strategies as found in student tests.

Form 2: Technical professional learning

In the second type of innovation, professional learning is a mediator and a means by which specific changes in practice are intended to occur. These changes may be reified in the use of new tools or process, for example, different curriculum materials, or a new teaching technique. This is illustrated in Figure 2. For simplicity, we refer to teachers only. However, in practice, some innovations currently promoted or being trialled in England focus on the technical learning of teaching assistants - for example, with catch up or one-to-one programmes in numeracy and literacy. Professional learning in this model has the qualities of retooling (Sachs 2011). The theory of change supposes that it is the applications of new tools or process that will lead to improvements in student outcomes.

Figure 2 Technical professional learning theory of change



Source: authors' original, unpublished Figure

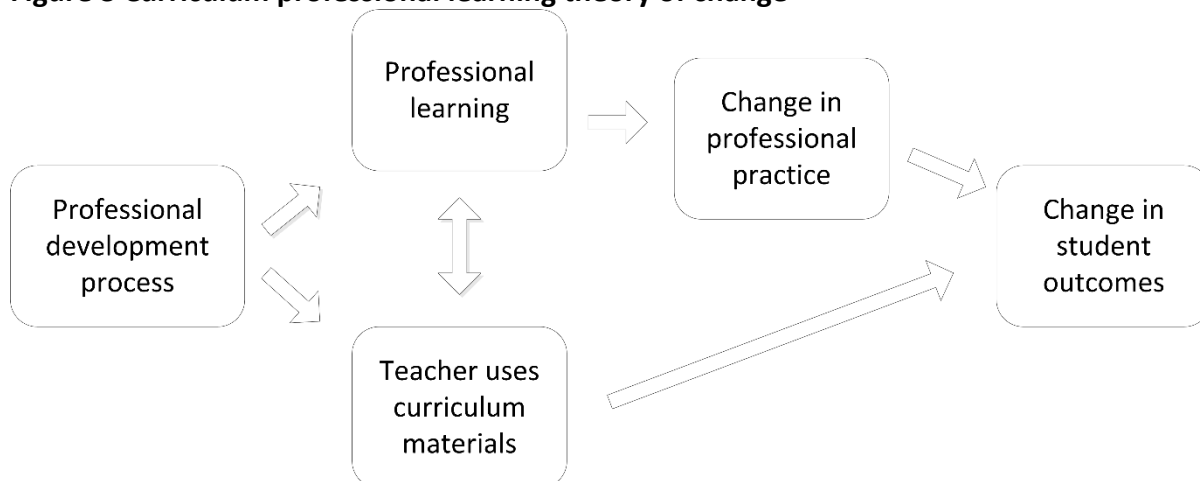
Tienken and Achilles (2003) report an innovation using a specified 'job-embedded' professional development programme to train a small group of teachers in using rubrics to enhance student self-regulation and meta-cognitive thinking about writing and the writing process. Measures were student outcomes on a writing assessment, with the fidelity measure being the extent to which the mode of instruction exemplified in the rubrics was implemented. Such assessment of fidelity not only is meaningful (unlike the previous example), but also necessary to link the impact of professional development to changes in student outcomes.

Form 3: Curriculum professional learning

The third form is a hybrid form in which professional learning occurs to support, and occurs through, curriculum innovation. In general, the relationship between innovation in curriculum materials and professional learning is relatively under-researched (Goldsmith, Doerr and Lewis 2014; Remillard 2005; Remillard and Bryans 2004). However, using new curriculum materials may constitute a professional development opportunity in itself. This is particularly so when the novel curriculum materials are the product of curriculum design research and in which learning theory is embedded (see Remillard 2005; Remillard & Bryans 2004). For innovations of this type, a theory of change is shown in Figure 3. This indicates a dialectical or mutually influencing relationship between use of materials and teacher professional learning. The change model summarises that change in student outcomes may

come both from change in professional practice but also, as in the technical model, from the use of the materials. In some cases, the curriculum materials used directly by students may lead to direct changes in outcomes.

Figure 3 Curriculum professional learning theory of change



Source: authors' original, unpublished Figure

In this form, the extent to which the emphasis is on use of materials as professional development or other activities will vary. Similarly, innovations vary to the extent to which professional development and curriculum materials (tools and processes) are designed together or sequentially. Related to this, it is an empirical question as to how far change in student outcomes results from changes in wider professional practices or from students' experience of new materials or techniques.

Saxe, Gearhart and Suad Nasir (2001) report an innovation, evaluated using quasi-experimental methods, in which teachers who were already trained to use an existing 'reform'- orientated curriculum unit - *Seeing Fractions* - experienced supplementary professional development. Two forms of professional development were compared: one involving research-based professional development focused on teachers' and children's mathematical understanding, in relation specifically to fractions, and children's motivation; the second, utilising facilitated meetings for an equivalent time as the first group, but with teachers setting the agenda and sharing and discussing the materials and sharing practice. Teachers in the two intervention conditions were randomly assigned with stratification by years of experience, prior experience of the materials and additional professional development experiences. Differences in student characteristics were adjusted for during analysis. A control group of teachers was formed for comparison; these teachers committed to using textbooks rather than new materials. Pre- and post-test measurement was through a test of computational and conceptual understanding of fractions. The evaluation design addressed the question of how far it is the use of different curriculum materials or the professional learning experiences that lead to change in student outcomes.

Characteristics of effective professional learning

In this section, we consider characteristics of effective professional learning. We use a narrow definition of 'effective' as meaning leading to improvement in student outcomes within the usual timescales of innovations and evaluations. In the previous section, an example was given of each of the three types of professional learning innovations drawn from a review of studies of professional learning that had evidenced a change in student outcomes. Thus, it is clear that all three forms can be effective, if effect on student attainment is the measure.

However, in reviews and meta-analyses of effective professional development, the features identified are, in general, more aligned with the pedagogical and curriculum forms than the technical form. Before discussing these features, it is important that this proposition is treated tentatively. It may be that innovations involving technical professional learning with positive impact have not been included in such reviews, because the focus of the innovation is on a specific technique, or process, with professional learning as an enabling or mediating feature. Thus, they may have been overlooked because of how the innovations are described in reports.

Nevertheless, professional learning with the greatest effect on student outcomes is consistently reported to have the following features:

- teachers are agentic (Timperley et al. 2007) and encouraged to adapt practices rather than reproduce them (Yoon et al. 2007);
- implementation of professional learning is localised and adaptive and varied (Cordingley 2015; Stoll, Harris and Handscomb 2012; Yoon et al. 2007) and involving professional experimentation (Clarke and Hollingsworth 2002);
- professional development experiences are diverse and tailored to teacher needs (Stoll, Harris and Handscomb 2012; Timperley et al. 2008) - "A didactic model in which facilitators simply tell teachers what to do" (Cordingley 2015, 8) was not effective;
- professional development is shaped by the context in which teachers practice and in relation to student needs (Cordingley, et al. 2015; Timperley et al. 2008);
- professional development is sustained over time and, in the case of professional learning involves using new tools, materials or processes, professional learning may be delayed (Remillard and Bryans 2004);
- professional development is collaborative (Cordingley et al. 2003; Garet et al. 2001; Stoll, Harris and Handscomb 2012; Timperley et al. 2008) or involves communication with others (Garet et al. 2001);
- professional development involves enquiry as key tools (Stoll, Harris and Handscomb 2012; Timperley et al. 2008).

Collaborative professional development can often entail and further amplify variability, given that the professional development experienced by teachers will be influenced by group composition and processes. Specifically, in relation to curriculum professional learning, this, taken together with the other noted features, including enquiry as means of learning, implies a further feature:

- a degree of unpredictability in what happens.

Tensions and conflicts

In this section, we discuss a number of tensions between the logic of RCTs and what might be termed the logic of effective professional development - these are both theoretical and also relate to how RCTs are often enacted. As discussed earlier, key to RCT methodology is the notion of control and this has number of consequences. The first is that teachers are compliant with the protocols of the innovation. This may relate to engagement in professional development activities (for example, attendance at events) as well as in implementing or using new materials or methods. There is a need for consistent implementation in predictable ways. Teachers are expected to adopt new methods. In assessing how far the innovation has been implemented, fidelity is emphasised. Thus, RCT methodology, arguably, is better aligned with technical professional learning than with pedagogical professional learning and curriculum professional learning variants. This is because with the latter forms variation is expected and, moreover, may be desirable.

If the primary focus of an innovation is on professional learning as the causal mechanism for changes in outcomes, there are potentially underlying theoretical conflicts between what is important for powerful and transformative professional learning (for example, agency and variation) and what may be required for an RCT-based evaluation that seeks to claim causality (for example, consistency and uniformity). This is summarised in the Table 1.

Table 1 Comparing RCTs and effective professional development

| Features needed for RCTs (efficacy trial) | Features of effective professional development |
|--|---|
| Teachers compliant with protocols | Teachers agentic and creative |
| Consistent Implementation | Localised Implementation |
| Uniformity | Diversity |
| Predictability | Unpredictable |
| Adoption | Adaptation |
| Fidelity | Variation |

Sources: authors' original unpublished Table

However, these tensions may resolve somewhat in more pragmatic approaches to trial design. Such approaches recognise and anticipate the possibility of variation and adaptability of implementation by teachers. Indeed, this may be particularly so in pilot or efficacy trials as the evaluation may, in part, be focused on understanding what forms of implementation are most effective and what conditions support this. Nevertheless, the meaning of 'fidelity' in such circumstances may need to be understood differently, for example, if part of the purpose of the innovation was to develop teacher agency. The example given above in relation to pedagogical professional learning (Carpenter et al. 1989) provided one instance of how this issue has been addressed in practice.

A further theoretical issue to consider is the divergence between the linear model of teacher change and synthesis of empirical research found in analytical models of professional learning. By analytical model, we mean those that are intended to have wide applicability in contrast to local or specific models or those based on more general social theory (see Boylan, Coldwell, Maxwell and Jordan 2018). Conceptualising or modelling professional learning and developmental process is challenging and it is difficult to develop models that both capture the inherent complexity and are operationalisable in research and evaluation (Boylan, Coldwell, Maxwell and Jordan 2018; Coldwell and Simkins 2011). Many models, for understandable reasons, attempt to simplify processes into pathways that allow for data collection and analysis of the relationship between professional development activities, teacher beliefs, teacher practices and pupil outcomes.

However, models widely used in designing, researching and evaluating professional learning commonly recognise that there is a recursive relationship between outcomes for pupils and teacher learning. Depending on the way learning processes are believed to occur, different views are taken of the relationship between these. We briefly point to three examples here (for a fuller critical appraisal of these models see Boylan, Coldwell, Maxwell and Jordan, 2018). Guskey's (2002) model suggests that the theory of change models, given in Figures 1-3, would need to be amended, including by adding recursive loops. Guskey argues that rather than professional learning preceding a change in pupil outcomes, the latter is a precursor to professional learning. Guskey, based on empirical evidence, proposes that teachers experiencing a professional development stimulus will engage in changed practices. However, these practices will only be integrated and lead to stable changes in beliefs (and so, in Guskey's view, professional learning) if teachers experience a change in pupil outcomes that they view positively. Thus, for Guskey, the causal relationship is reversed to that posited in Figures 1-3.

Clarke and Hollingsworth (2002) propose an interconnected model of teacher professional growth that identifies multiple pathways through four domains - the external, the personal (beliefs and values), outcomes, and practices. Their model recognises that the pathway

posited in Figures 1-3 is one that does occur but also that other causal chains are possible, including the one proposed by Guskey.

Opfer and Pedder (2011), in contrast, draw on complexity theory to conceptualise the process of professional learning as occurring in and through nested systems. Centrally important are, what they term, the 'learning activity system', and teacher orientations to this. Orientations are fluid and changeable. The concept of a learning activity system within wider nested systems contrasts with linear input-output models. Opfer and Pedder's complexity model would require these systems and orientations to be included in the typology models. All these analytical models propose a recursive rather than linear view of teacher learning in terms of its relationship to student outcomes. Thus, they are at variance with how causality is often modelled in theories of change in experimental and quasi-experimental evaluation of innovations involving professional learning.

A further issue is less a theoretical tension, but rather related, contingently, to how RCTs are often currently implemented in education. Although trials may be longitudinal, typically the expectation is that changes in practice will occur relatively soon after professional development experiences. However, there are indications that there may be significant differences between how materials are used or new pedagogical approaches employed over a longer period of time than within a single school year (Remillard and Bryans 2004). This accords with a view of the importance of professional experimentation (Clarke and Hollingsworth 2002) and with changes in student outcomes being as cause as well as consequence of professional learning (Clarke and Hollingsworth 2002; Guskey 2002)

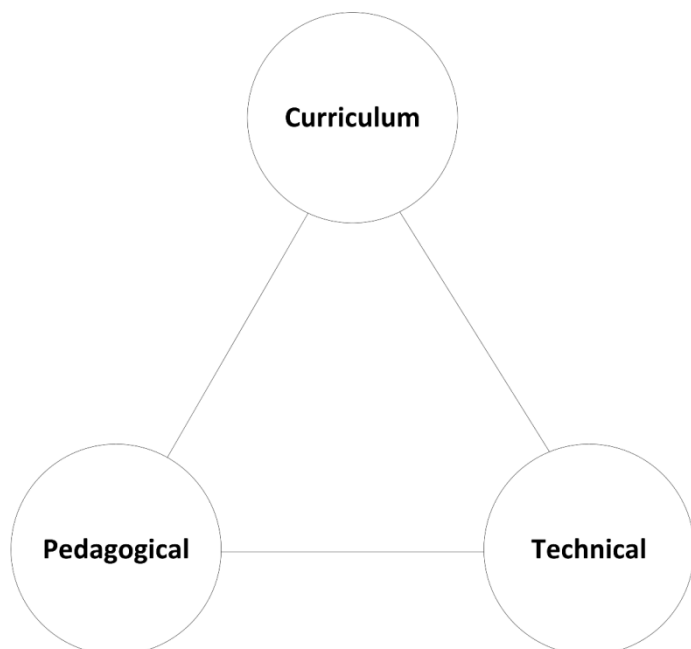
Discussion and implications

The three models we have proposed are a simplified heuristic. The typology is presented below in summary form in Figure 4. The triangle arrangement signifies that, whilst curriculum professional learning is a hybrid form, it has distinctive features. In practice, an innovation may have characteristics, to different degrees, of a combination of the three forms proposed in the typology. The representation in Figure 4 is intended to indicate this. It can be used as a tool or prompt to consider the relationship to the typology, for example, during dialogue between evaluators and developers of innovations to develop theories of change. Thus, a question for evaluators (and designers) is to identify the relationship of the innovation to the three types of professional learning rather than choosing only one of them. The issue is further complicated as whilst the innovation may be designed in a way that accords with a combination of types, it may be implemented in a variety of ways.

To illustrate this we consider the example of professional development for teachers to lead *Philosophy for Children* that promotes a community of enquiry pedagogy (Gorard, Siddiqui and See 2015). The community of enquiry approach may be supported by materials used by the teachers, but not necessarily directly by pupils, and requires following specific protocols which have technical features. However, as a result of engaging in professional

development, the forms of dialogue encouraged in a community of enquiry may be fostered at other times by the teacher and so lead to more general pedagogical change. Further, by so doing, the teacher will practice the approach for longer and this may also augment their learning.

Figure 4 Three forms of professional learning



Source: authors' original, unpublished Figure

The typology is also potentially useful when describing innovations (Hoffmann et al. 2014) and developing programme-specific theories of change, as recommended as part of implementation and process evaluation (Humphrey et al. 2016). Thus, it may support developing measures of fidelity.

Some of the key questions to attend to in applying the typology are:

- To what extent do posited or intended outcomes for students require close adherence by teachers to the recommended use of the novel techniques, methods or materials?
- If curriculum materials or other pedagogical resources are used, then:
 - how far is it the materials or resources themselves that will lead to the intended student outcomes?
 - how important is 'correct' use of materials to intended outcomes?
 - to what extent will use of the curriculum materials lead to professional learning?
- How far is a desired outcome of the innovation teacher adaption, selection, or synthesis with their existing practice of innovative practice?
- To what extent is increasing teacher agency and criticality about practice and evidence an intended outcome of the innovation?

- What types of adaptations that might result from the professional learning experiences are potentially productive, in terms of intended outcomes for students?

However, as we have argued, the linear, uni-directional nature of these models does not fully describe the complex and recursive nature of professional learning. Thus, more general analytical models pointed to above should also be considered alongside these. Further, the importance of materials and artefacts in professional learning may require the use of theories beyond those analytical models. A recent example of this is the application of cultural historical activity theory to lesson study (Wake, Swan and Foster 2016).

We have pointed to greater alignment between innovations that primarily involve technical professional learning and RCT methodology. For innovations involving the pedagogical and curriculum professional learning approaches, there is a risk that the experimental approach may constrain the form of professional learning in innovations to that which is easily measured, or not attempting to measure professional learning at all.

This can be avoided if a wider range of measures and data are used to understand how professional development activity, as well as the implementation of new materials or processes, influences professional learning and practice. For example, we have noted that professional learning that involves the use of curriculum materials as a vehicle and catalyst for learning adds to the complexity of the learning activity system. Therefore, it also adds a further dimension to the evaluation of professional learning. This is because attention needs to be paid not only to engagement and response to professional development events, activities and learning materials, but also to how materials are used and adopted and adapted. In summary, as well as recording how frequently materials are used, implementation evaluation should consider the nature of that use and how far using materials leads to professional learning and of what type(s).

Common to all three is that the posited change in outcomes of learners depends on the outcomes of two related, though distinct, processes. The first is the extent to which professional development leads to the expected professional learning; the second is the extent to which the practices or change that arises from, or are co-emergent with, the professional learning leads to improved outcomes. Across the three models of professional learning, how this happens and the relative importance of the two factors varies. However, the theory of change (Weiss 1997) will involve both aspects - in the sense of the programme theory at least. That there are two distinct processes is important to consider in RCT design.

An implication of this is that if the only impact measure used is of student outcomes, then the inferences that can be made are limited. This is because, if no significant effect is found, it is not possible to make inferences from the student outcome measure alone as to whether the professional development activity has been ineffective or whether the changed practices do not lead to improved outcomes. A recent trial of lessons study in England found

no impact of the programme as implemented (Murphy et al. 2017). However, the question remains as to whether the pedagogical focus of the lesson study - the substantive change in classroom practice - itself may not have been effective practice (Weston 2017). The matrix below illustrates this.

Figure 5 Possible outcomes for professional learning trials

| | | Professional development | |
|--|--|---|---|
| | | Effective - desired PL/teacher change happens | Ineffective - desired PL/teacher change does not happen |
| New pedagogy, curriculum materials or techniques | Effective (better than 'business as usual') | Student impact | No student impact |
| | Ineffective (not better or worse than 'business as usual') | No student impact | No student impact |

Source: authors' original, unpublished Figure

This implies there is a need to measure professional learning outcomes as an impact measure in trials or to address, in a rigorous way, the extent to which professional learning has taken place within the implementation and process evaluation. This will allow for inferences to be made about reasons for no measured outcome on student attainment being found if this happens.

Taking the lesson study trial as an example (Murphy et al. 2017), the pedagogical focus of the trial was on 'Talk for Learning'. This aimed to improve the quality of classroom talk through the use of specific strategies. Some of these strategies were similar to ones already used in some of the schools in the trial. Whilst the evaluation report does report findings on the extent to which Talk for Learning was implemented, quantified data on classroom practice was not collected. The trial might have sought to measure change in relevant teacher practices from base line to the end of the trial and even the quality of pupil talk. The

use of this example is not intended as a criticism of the evaluators. In order to generate reliable data, including such measures would increase evaluation costs considerably and so this may have restricted what was possible. There are also challenges to measuring teacher and classroom practices reliably. However, without such data it is not possible to infer whether it was the pedagogical approach or the professional learning approach that was not effective, or if it was a combination of both of these.

Therefore, within trials in which the innovation involves pedagogical or curriculum professional learning, there is a need to collect richer process evaluation data prior to innovation about contexts, teacher prior learning, disposition and practice. Similarly, during trials, what is often a 'black box' of teacher learning, needs to be opened, using tailored approaches such as those found in trials discussed earlier (Carpenter et al. 1989; Saxe et al. 2001) to help explain student impact findings. Further, the challenge of assessing levels of agency, before and after innovation, needs addressing, at least in pedagogical professional learning innovations. This is because of the importance of teacher agency and agentic practices in reviews of effective professional learning. This would also support the more pragmatic evaluation designs pointed to above; these are ones that anticipate variation, and that can potentially overcome tensions between the logic of RCTs and the logic of professional learning.

Moreover, variation should be planned for, including evaluators anticipating the need to examine sub-samples of the data and reflecting this in trial size. Given expected variability, consideration should be given to extending the analysis beyond comparison of control and intervention groups but also to analyse differences in outcomes within the samples (Hall 2013; Watson 2015). Given the cost of RCTs and the potential richness of data collected, analysis should be undertaken across the data set of associations between differences in practices and differential outcomes, regardless of sample group (Watson 2015).

A further issue relates to meta-analysis and synthesis. We have argued that meta-analysis of findings on the impact of professional learning on student outcomes suggests a set of characteristics which are indicative of pedagogical professional learning. However, as noted earlier, it may be that examples of effective technical professional learning are being overlooked because the importance of professional learning within the innovation is not highlighted in reports. Given the theoretical models and evidence we have pointed to, our view is that this may not be the case, but it is a possibility that requires further reviews of research to address. Thus, the proposed typology is potentially useful in meta-analysis to characterise different forms of innovation that involve professional learning.

Conclusion

In this paper, we have addressed concerns with the use of RCTs in evaluating innovations involving professional development to argue that methodology will be enhanced through greater analytical specificity of the different roles professional development and learning

has in innovations. To address this, we proposed, and illustrated, a novel typology of professional learnings: *pedagogical professional learning*, *technical professional learning*, and a hybrid form that has features of both found in curriculum design innovations - *curriculum professional learning*. We contend that these categorisations form a useful heuristic to inform evaluators and designers in developing theory of change models for specific innovations that involve professional learning. Further, the typology may be useful in evaluation of professional development using other research designs as well.

We have argued that it is important for evaluation designers to consider insights from research on professional learning. We have brought together research findings and designs that are often seen as paradigmatically separate. Tensions were identified between RCT methodology and analytical descriptions of effective professional learning, as well as the recursive relationship between pupil outcomes and teacher learning in contrast to causally linear models of impact often considered in RCTs. Some current discourses of evidence-based teaching, as well as political pressures to develop evidence in relatively short timescales, are also in tension with the value of professional experimentation. Teacher change that will lead to better pupil outcomes may not occur on first implementation by teachers when learning new methods. We have identified a number of implications for evaluation design based on the arguments made.

One of our key purposes in this paper has been to raise a number of methodological issues in the use of RCTs in evaluations of professional learning. Given this, it is important to emphasise that we also consider that, in general, there should be greater use of quantitative methods that assess changes in pupil outcomes as the result of professional development. Such studies are relatively rare in the evaluation of professional development programmes in comparison with the large number of innovations and projects reported and evaluated (Yoon, et al. 2007). Further, a conclusion that follows from our arguments is that quantitative methods should be extended to measure changes in teacher practices and teacher learning, or at least proxies for these, alongside changes in pupil attainment. If this is not done, it can lead to inconclusive outcomes of trials. In some cases, it is not clear whether it is the professional development form that has been ineffective, or teacher change has happened, but the type of change does not lead to improved pupil outcomes on the measures used. Paying attention to this complexity is, we believe, important regardless of the evaluation method that is used. A possible implication of the arguments we have made is that RCTs of innovations in which professional learning is important should be larger in scale and include research in greater depth. Given the costs of RCTs this may mean that fewer RCTs are undertaken. However, the reliability of evidence may increase and interpretation of evidence improved.

Making decisions about what form of professional development to design or to promote is clearly an issue of concern for policy makers, school leaders, researchers and teachers, particularly when resources are limited. So an important question that arises in relation to

professional development (and also curriculum development) is - 'is this worth doing?' One approach to answering the question is to adopt experimental methods to evidence the efficacy of practices. However, if such methods are used, it is important that the complexity of professional learning is properly considered in evaluation design. This requires dialogue between experimental evaluators and researchers of professional learning. The proposed novel typology and related analyses are an outcome of bringing together perspectives from these two areas of educational research that have, in the main, developed separately. The typology and awareness of the identified methodological issues can support the endeavour of improving the design of professional learning programmes as well as the design and interpretation of the evaluation of them.

References

Biesta, G. J. 2010. Why 'what works' still won't work: From evidence-based education to value-based education. *Studies in Philosophy and Education*, 29(5), 491-503.

Boylan, M., Coldwell, M., Maxwell, B., & Jordan, J. 2018. Rethinking models of professional learning as tools: a conceptual analysis to inform research and practice. *Professional Development in education*, 1-20.

Carpenter, T. P., Fennema, E., Peterson, P. L., Chiang, C. P., & Loeff, M. 1989. Using knowledge of children's mathematics thinking in classroom teaching: An experimental study. *American educational research journal*, 26(4), 499-531.

Clarke, D., & Hollingsworth, H. 2002. Elaborating a model of teacher professional growth. *Teaching and teacher education*, 18(8), 947-967.

Coldwell, M., Greany, T., Higgins, S., Brown, C., Maxwell, B., Stiell, B., ... & Burns, H. 2017 *Evidence-informed teaching: an evaluation of progress in England*. Research Report. Department for Education.

CONSORT 2010 Consolidated Standards Of Reporting Trials 2010 statement and checklist available at <http://www.consort-statement.org/>

Cordingley P, Bell M, Rundell B and Evans D 2003 *The impact of collaborative CPD on classroom teaching and learning*. In: *Research Evidence in Education Library*. London: EPPICentre, Social Science Research Unit, Institute of Education.

Cordingley, P., Higgins, S., Greany, T., Buckler, N., Coles-Jordan, D., Crisp, B., Saunders, L., Coe, R. 2015 *Developing Great Teaching: Lessons from the international reviews into effective professional development*. Teacher Development Trust. 2015.

Eldridge, S. & Kerry, S. 2012 *A Practical Guide to Cluster Randomised Trials in Health Services Research*. Wiley-Blackwell

Paper published as BOYLAN, Mark and DEMACK, Sean (2018). Innovation, evaluation design and typologies of professional learning. *Educational Research*, **60** (3), 336-356.
<https://doi.org/10.1080/00131881.2018.1493352>

Glasgow, R. E., Lichtenstein, E., & Marcus, A. C. 2003. Why don't we see more translation of health promotion research to practice? Rethinking the efficacy-to-effectiveness transition. *American journal of public health*, **93**(8), 1261-1267.

Goldacre, B. 2013. *Building evidence into education*.
http://dera.ioe.ac.uk/17530/7/ben%20goldacre%20paper_Redacted.pdf

Goldsmith, L. T., Doerr, H. M., & Lewis, C. C. 2014. Mathematics teachers' learning: A conceptual framework and synthesis of research. *Journal of Mathematics Teacher Education*, **17**(1), 5-36.

Gorard, S., Siddiqui, N., & See, B. H. (2015). Philosophy for Children: Evaluation report and executive summary. Education Endowment Foundation, Millbank, UK.
https://educationendowmentfoundation.org.uk/public/files/Support/Campaigns/Evaluation_Reports/EEF_Project_Report_PhilosophyForChildren.pdf

Guskey, T.R., 2002. Professional development and teacher change. *Teachers and Teaching: theory and practice*, **8**(3), pp.381-391.

Hall, G. E. (2013). Evaluating change processes: Assessing extent of implementation (constructs, methods and implications). *Journal of Educational Administration*, **51**(3), 264-289.

Hargreaves, D. H. 2011. *Leading a self-improving school system*. Nottingham: NCSL.

Harrison, N., & McCaig, C. 2016. Examining the epistemology of impact and success of educational interventions using a reflective case study of university bursaries. *British Educational Research Journal*.

Haynes, L., Service, O., Goldacre, B. & Torgerson, D. 2012 *Test, Learn, Adapt: Developing Public Policy with Randomised Controlled Trials*. London HMSO Cabinet Office Behavioural Insights Team - at <https://www.gov.uk/government/publications/test-learn-adapt-developing-public-policy-with-randomised-controlled-trials>

Hoffmann, T. C., Glasziou, P. P., Boutron, I., Milne, R., Perera, R., Moher, D., ... & Lamb, S. E. 2014. Better reporting of interventions: template for intervention description and replication (TIDieR) checklist and guide. *BMJ*, **348**, g1687.

Humphrey, N., Lendrum, A., Ashworth, E., Frearson, K., Buck, R., & Kerr, K. 2016. *Implementation and process evaluation (IPE) for interventions in educational settings: An introductory handbook*.

Kennedy, A. 2005. Models of continuing professional development: a framework for analysis. *Journal of in-service education*, **31**(2), 235-250.

Paper published as BOYLAN, Mark and DEMACK, Sean (2018). Innovation, evaluation design and typologies of professional learning. *Educational Research*, 60 (3), 336-356.

<https://doi.org/10.1080/00131881.2018.1493352>

Kennedy, A. 2014. Understanding continuing professional development: the need for theory to impact on policy and practice. *Professional development in education*, 40(5), 688-697.

Lehmann, E. R. 2015. What if 'What Works' Doesn't?. *Evaluation*, 21(2), 167-172.

Mezirow, J. 1991 *Transformative dimensions of adult learning*. San Francisco, CA: Jossey-Bass.

Murphy, R., Weinhardt, F., Wyness G. a Rolfe, H. 2017. *Lesson Study: Evaluation report and executive summary*, EEF: London. URL:
https://educationendowmentfoundation.org.uk/public/files/Projects/Evaluation_Reports/Lesson_Study.pdf

O'Brien, J., & Jones, K. 2014. Professional learning or professional development? Or continuing professional learning and development? Changing terminology, policy and practice. *Professional development in education*, 40(5), 683-687.

Opfer, V. D., & Pedder, D. 2011. Conceptualizing teacher professional learning. *Review of educational research*, 81(3), 376-407.

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

Pawson, R., & Tilley, N. 1997. *Realistic evaluation*. Sage.

Remillard, J. T. 2005. Examining key concepts in research on teachers' use of mathematics curricula. *Review of Educational Research*, 75(2), 211-246.

Remillard, J. T., & Bryans, M. B. 2004. Teachers' orientations toward mathematics curriculum materials: Implications for teacher learning. *Journal for Research in Mathematics Education*, 352-388.

Rothwell, P. M. 2005. External validity of randomised controlled trials: "to whom do the results of this trial apply?". *The Lancet*, 365(9453), 82-93.

Sachs, J. 2011. Skilling or emancipating? Metaphors for continuing teacher professional development. In *Rethinking Educational Practice Through Reflexive Inquiry* (pp. 153-167). Springer Netherlands.

Saxe, G. B., Gearhart, M., & Nasir, N. I. S. 2001. Enhancing students' understanding of mathematics: A study of three contrasting approaches to professional support. *Journal of Mathematics Teacher Education*, 4(1), 55-79.

Slavin, R. 2002 Evidence-Based Education Policies: Transforming Educational Practice and Research. *Educational Researcher*, Vol. 31, No. 7, pp. 15-21

Paper published as BOYLAN, Mark and DEMACK, Sean (2018). Innovation, evaluation design and typologies of professional learning. *Educational Research*, **60** (3), 336-356.
<https://doi.org/10.1080/00131881.2018.1493352>

Stoll, L., Harris, A., & Handscomb, G. 2012. Great professional development which leads to great pedagogy: nine claims from research. *National College for School Leadership, Nottingham, United Kingdom*.

Timperley, H., Wilson, A., Barrar, H., & Fung, I. 2008. *Teacher professional learning and development*.

Torgerson, C. 2009 Randomised controlled trials in education research: a case study of an individually randomised pragmatic trial, *Education 3-13: International Journal of Primary, Elementary and Early Years Education*, 37:4, 313-321

Wake, G., Swan, M., & Foster, C. (2016). Professional learning through the collaborative design of problem-solving lessons. *Journal of Mathematics Teacher Education*, 19(2-3), 243-260.

Watson, A. 2015, Why 'what works' doesn't work in practice, and what might work better. In M.A. Hersh and M. Kotecha (eds.) *Proceedings of IMA International Conference on Barriers and Enablers to Learning Maths: Enhancing Learning and Teaching for All Learners*. 10-12 June 2015, University of Glasgow.

Weiss, C. H. 1997. Theory - based evaluation: Past, present, and future. *New directions for evaluation*, 1997(76), 41-55.

Weston, D. 2017. *Does Lesson Study work? A look at the new EEF trial*. Teacher Development Trust <http://tdtrust.org/blog>. Retrieved Dec 2017.

Winch, C., Oancea, A., & Orchard, J. 2015. The contribution of educational research to teachers' professional learning: philosophical understandings. *Oxford Review of Education*, 41(2), 202-216.

Yoon, K. S., Duncan, T., Lee, S. W. Y., Scarloss, B., & Shapley, K. L. 2007. *Reviewing the Evidence on How Teacher Professional Development Affects Student Achievement*. Issues & Answers. REL 2007-No. 033. Regional Educational Laboratory Southwest (NJ1).