Assessing an experimental approach to industrial policy evaluation: Applying RCT+ to the case of Creative Credits

Hasan Bakhshi[a], John S. Edwards[b] , Stephen Roper[c], Judy Scully[b], Duncan Shaw[d], Lorraine Morley[e] and Nicola Rathbone[b]

a Nesta, 1 Plough Place, London, EC4A 1DE
Hasan.Bakhshi@nesta.org.uk

b Aston Business School, Aston University, Birmingham, B4 7ET
j.s.edwards@aston.ac.uk; j.w.scully@aston.ac.uk; rathbnj1@aston.ac.uk

c Enterprise Research Centre and Warwick Business School, University of Warwick, Coventry, CV4 7AL, UK.
stephen.roper@wbs.ac.uk

d Manchester Business School, Booth Street West, Manchester, M15 6PB, UK.
duncan.shaw-2@mbs.ac.uk

e Warwick Business School, University of Warwick, Coventry, CV4 7AL, UK.
lorraine.morley.09@mail.wbs.ac.uk

Corresponding author: Stephen Roper, stephen.roper@wbs.ac.uk

Abstract: Experimental methods of policy evaluation are well-established in social policy and development economics but are rare in industrial and innovation policy. In this paper we consider the arguments for applying experimental methods to industrial policy measures, and propose an experimental policy evaluation approach (which we call RCT+). This approach combines the randomised assignment of firms to treatment and control groups with a longitudinal data collection strategy incorporating quantitative and qualitative data (so-called mixed methods). The RCT+ approach is designed to provide a causative rather than purely summative evaluation, i.e. to assess both 'whether' and 'how' programme outcomes are achieved. In this paper we assess the RCT+ approach through an evaluation of Creative Credits – a UK business-to-business innovation voucher initiative intended to promote new innovation partnerships between SMEs and creative service providers. The results suggest the potential value of the RCT+ approach to industrial policy evaluation, and the benefits of mixed methods and longitudinal data collection.

1

Assessing an experimental approach to industrial policy evaluation: Applying RCT+ to the case of Creative Credits

1. Introduction

Although unusual in terms of industrial policy, experimental methods of policy evaluation are well established in social policy and development economics. Burtless (1995, p. 63), for example, cites Greenberg and Schroder (1991) who 'identified more than 90 separate field trials involving a wide range of distinctive research areas including health insurance, prisoner rehabilitation, labour supply, worker training and housing subsidies'. Banerjee and Duflo (2008, p. 32) also describe the 'recent surge in experimental work' in development economics. Typically such evaluations involve individual human subjects facing some common socio-economic problem, and random allocation of subjects to a treatment and control group. Differences in outcomes between the treatment and control groups are then attributed to the effect of the policy intervention. In terms of industrial policy, however, experimental policy evaluation approaches remain marginal, with non-experimental, ex post policy evaluations remaining the norm. In the context of small business policy evaluation, Potter and Storey (2007), for example, provide an extensive review of best practice in OECD countries without any mention of either the application or potential for experimental methods. Similarly, UK government guidance on industrial policy evaluation focuses entirely on non-experimental ex post evaluation approaches (BIS 2009). Related observations might be made in relation to the evaluation of R&D and innovation policies: despite the increasing importance of evaluation as part of the process of development of technology policy, evaluation approaches remain almost universally ex post and non-experimental (Laredo 1997).

The methodological and practical advantages of experimental and non-experimental evaluation methods have been widely debated in the context of social policy interventions (Burtless 1995; Heckman and Smith 1995; Bratberg, Grasdal, and Risa 2002; Banerjee and Duflo 2011; Deaton 2010). Experimental methods based on randomised allocation have – at least in theory – the advantage of transparency and may be more convincing to policy-makers than the results of more complex econometric evaluation approaches (Burtless 1995). In small samples, however, perhaps less than 300, randomisation may be ineffective at ensuring the homogeneity of control and treatment groups (Bratberg, Grasdal, and Risa 2002), though

robust experiments can still be run with smaller samples if the intervention is sufficiently powerful (Bloom et al. 2011). Implementing experimental approaches may also lead to other specific biases (Heckman and Smith 1995), while small-scale experimental studies may fail to replicate the likely macro-impacts of a scheme which is implemented at national level – that is, they lack external validity (Garfinkel, Manski, and Michalopoulos 1990). Perhaps the key advantage of experimental approaches, however, and the central issue with non-experimental evaluation approaches, relates to potential selection biases. In particular, if subjects which are allocated to a treatment group have a higher preponderance of some characteristic which is correlated with outcomes this will lead to bias in the estimation of treatment effects. Such biases may be significant where policy interventions are targeted at particular groups of subjects, where support is allocated on the basis of routinized decision rules, or where there is an element of self-selection into an experiment. One recent study, for example, illustrates how funding allocations in the Norwegian Research Council are based on ex ante project rankings generating a potential selection bias when evaluating the Research Council's funding decisions (Bremnes, Bergem, and Nesset 2011). In terms of policy for innovation or small and medium enterprises (SMEs), similar selection biases might arise where a policy initiative seeks to back winners or is focussed on firms which have an established track record of growth or innovation. In this sense the receipt of public support may itself need to be treated as endogenous (Garcia and Mohnen 2010).

As many methodological and implementation issues arise in industrial policy evaluations, which focus on firms, as they do in social policy, where the subjects are likely to be individuals. The selection biases are likely to be as great. However, arguably some of the ethical issues which arise in implementing experimental approaches to social policy may be seen as less significant in industrial policy interventions. It may be more ethically acceptable, for example, to randomly allocate public support among firms rather than adopting a similar allocation rule to the distribution of support among financially disadvantaged individuals. Nevertheless, the random allocation among firms needs to be carefully considered to ensure money is not simply thrown at applicants in the hope that they do something useful with it. Application criteria, an evidenced business need, project governance and, potentially, co-funding can raise confidence in the ethical aspects of investing in companies' innovation potential. This makes it all the more surprising that experimental evaluation approaches in industrial policy are not more common.

In this paper we propose and assess an experimental approach to the evaluation of new industrial policy interventions. Our approach (which we label RCT+ or Randomised Control Trial plus) takes advantage of randomisation, but combines this with a longitudinal and mixed methods data collection strategy to provide causative insights rather than only summative policy assessment. In other words, RCT+ seeks to evaluate the validity of an underlying logic model rather than simply generating point estimates of policy impact (Donaldson and Gooler 2003), and seeks to explain why these results are observed (Denzin and Lincoln 2005; Ludwig, Kling, and Mullainathan 2011).

The aim of this paper is to assess the potential value of the RCT+ approach when applied to industrial policy. We make three main contributions. First, we provide an assessment of the value of experimental evaluation approaches to industrial policy initiatives, suggesting an alternative enhanced approach to the development of industrial policy. Second, through an application of the RCT+ approach we extend standard (quantitative) experimental evaluation approaches beyond the summative to provide causal explanations for policy outcomes – i.e. to identify the 'why, how, and at what cost' an intervention may have worked (White, 2008, p. 98). Adopting this type of rigorous qualitative approach provides potentially frame-breaking insights (Eisenhardt, 1989), and may enable the conceptualisation of the context in which an intervention was implemented, facilitating the generalisation of results (White, 2008). Third, our application of RCT+ shows how a longitudinal approach can provide a time profile of policy outcomes, without which policy-makers can make incomplete, and potentially misleading, inference. We apply the RCT+ approach to Creative Credits, a UK-based business-to-business innovation voucher programme designed to foster new innovative partnerships between SMEs and creative service providers.

The remainder of the paper is organised as follows. In Section 2 we discuss the rationale for industrial policy evaluation and review previous debates about the relative merits of experimental and non-experimental evaluation approaches. Subsequent sections of the paper report the application of the RCT+ evaluation approach to the Creative Credits experiment. Section 3 reports the application of the RCT+ approach and its outcomes. Section 4 discusses

the implications of the RCT+ experiment and the strengths and limitations of experimental methods in industrial policy evaluation. Parallels with meta-analytical approaches in medicine are also considered. Section 5 concludes the paper.

2. Evaluating industrial policy initiatives

Potter and Storey (2007) identify five reasons why industrial policy evaluation might be undertaken: to establish the impact of industrial policy; to inform the allocation of funding to alternative policy measures; to demonstrate value for money; to stimulate debate about forms of public intervention; and, to contribute to improvements in the design and administration of policy interventions. In each case the problem of causal inference is the same, i.e. that the treated and non-treated outcomes for any single firm are never observed (Holland 1986). The analytical problem this raises is how to estimate the difference between the actual realised outcomes and the potential outcomes if no treatment had been administered. Ideally, the substitute for the unobserved (un-treated) outcome needs to meet two criteria: (i) it should be observable to the researcher; and, (ii) it should be an 'internally' valid substitute for the set of un-treated outcomes. Internal validity in this sense requires that 'the only difference between the member of the control group and the member of the treated group corresponds to the fact that the latter is treated and the first one is not' (Reiner, 2011, p. 18).

More comprehensively, Imbens and Wooldridge (2009) outline three situations which describe the allocation of subjects to a control and treatment group. The first, and simplest, is the classical experimental situation of randomised allocation in which allocation is unrelated to outcomes. The second allocation mechanism – 'un-confounded allocation' – occurs where assignment is independent of outcomes but may be related to subject characteristics. Here, where the assignment mechanism is either observable or discoverable, sampling and/or statistical approaches can be used to minimise any systematic differences between the characteristics of the treatment and control groups and provide a valid estimate of treatment effects (Burtless 1995). In practice, evaluations of SME policy vary in the sophistication of their approach to un-confoundedness. Potter and Storey (2007), for example, cite evaluation studies which compare the performance of treated firms with control groups of 'typical' firms (Lambrecht and Pirnay 2005), and studies which use 'matched' control groups based on

treatment and control groups with similar baseline characteristics (Lerner 1999). However, despite careful matching or selection of control groups, the potential remains for bias in terms of the background characteristics of the two groups (Bratberg, Grasdal, and Risa 2002).

This has led to the development and widespread application of a third group of econometric approaches which can 'control' ex post for potential selection biases by either implicitly or explicitly modelling the probability that a firm will be in the treatment rather than the control group, and then estimating the impact of the treatment 'controlling' for any selection biases (Bratberg, Grasdal, and Risa 2002; Imbens and Wooldridge 2009). Implementing this type of non-experimental approach, however, often involves making assumptions about the underlying causal process which determines the allocation of firms to the treatment and non-treated groups (Burtless 1995). Specific approaches may also pose challenges in terms of identification, requiring, for example, the use of a variable or group of variables which influence allocation but which have no influence on subsequent outcomes. Early comparisons of non-experimental econometric estimators with experimental estimators tended to favour experimental estimators (Heckman and Hotz, 1989), and highlighted the sensitivity of the results of non-experimental studies to the choice of econometric approach (Bratberg, Grasdal, and Risa 2002). This led Burtless (1995) to conclude that: 'the classical experimental estimator still has a major advantage over non-experimental estimators for users who care about the statistical precision of the estimates they use. But the more important advantage is that the validity of the experimental estimator depends upon assumptions that are ordinarily much easier to evaluate – and to believe' (Burtless 1995, p. 73). More recently, however, developments in propensity scoring, matching estimators and instrumental variables provide alternative statistical approaches to dealing with un-confoundedness in different contexts (Imbens and Wooldridge 2009). This type of evaluation, allowing for potential selection biases, has been emphasised by the OECD as best practice in ex post evaluation and has been widely applied in recent years in the context of SME support measures (Potter and Storey 2007).

Thus, in non-experimental evaluations either un-confounded or confounded allocation – where allocation has some dependence on potential outcomes – can have profound implications for a researcher's ability to obtain reliable estimates of treatment effects. In

experimental approaches, by contrast, random assignment to the treatment and control group should avoid such biases and allow more accurate estimation of treatment effects (Burtless 1995). A number of potential implementation issues arise, however, even with experimental designs involving randomised allocation (Reiner 2011). First, in small samples randomised allocation may fail to eliminate differences in the characteristics of firms in the treatment and control groups, influencing the internal validity of the experiment. As a result Bruhn and McKenzie (2009) suggest that in small samples (30-100 observations in their simulations) the similarity of treatment and control groups is better where pair-wise matching, stratification or re-randomisation approaches are used rather than simple random allocation. For larger sample sizes (>300), however, their simulation analysis suggests that the choice of randomisation approach is a much less significant factor. A second threat to internal validity in experimental studies highlighted by Heckman and Smith (1995) is the potential for substitution bias where members of a control group are able to substitute alternative forms of support for the focal treatment effect. If significant, this contamination may undermine the estimate of the treatment derived from the difference in outcomes between the treatment and control groups. In the context of an industrial policy initiative where the subjects are firms rather than individuals it is also possible to envisage a related 'signalling bias' where firms in the treatment group are able to attract additional investment because they are in the treated group. Meuleman and Maeseneire (2012), for example, find that Belgian SMEs which obtained an R&D subsidy were better able to access long-term debt than other firms. This type of effect may again influence the scale of any measured treatment effect.

Finally, it is not clear how readily the results of any single policy experiment involving a small proportion of a target population can be generalised to the entire target population, i.e. whether the results of small scale policy experiments are 'externally' valid. Garfinkel et al (1990), for example, focus on the 'macro' or society-wide effects which would occur if an intervention was implemented at macro-level but may not be replicated in small scale studies. These include: policy effects on the economic equilibrium; widespread diffusion of information about the programme; and, social interaction and norm formation which might influence programme participation or other related decisions. External validity may also be reduced if the characteristics of those firms participating in a programme differ significantly from the wider target group (Burtless 1995). Applicants for innovation support measures, for example, may be more strongly innovation-oriented than firms in general, a capability which

has been linked to above average business performance (Rosenbusch, Brinckmann, and Bausch 2011). External validity may also be impacted by 'randomisation bias', the idea that the adoption of random allocation itself might either induce or discourage some types of firms from applying for a scheme (Heckman and Smith 1995).

While internal and external validity have been widely discussed in the summative, ex post evaluation literature, similar themes also emerge in respect of evaluations which adopt a constructivist or qualitative approach. Guba and Lincoln (1989), for example, emphasise trustworthiness, the nature of the evaluation process and authenticity as key attributes of robust qualitative evaluation approaches. They describe trustworthiness as a group of robustness criteria (i.e. credibility, transferability, dependability and confirmability) that are 'parallel' to the experimental criteria of internal/external validity, reliability and objectivity; parallel due to ontological and epistemological differences (p. 233). Guba and Lincoln (1989) describe a second criterion, which they call process, as evaluating how the data was accumulated and the rigour used to ensure it accurately reflects the context (p. 244). However, because the trustworthiness criterion does not ensure that stakeholders' constructions have been collected and faithfully represented within the messiness of research (p. 245), and the process criterion does not provide explicit evidence, there is a third criterion − authenticity. Authenticity provides balancing aspects of: the fairness of integration of all viewpoints; the extent to which (through the process of the evaluation) stakeholders expand their own constructions and their constructions of others; the action facilitated by the evaluation; and, the empowerment of stakeholders to act. Attending to these three criteria helps to develop constructivist evaluations that are rigorous and from which translatable results can emerge.

The statistical and interpretative arguments favouring experimental evaluation approaches may in some circumstances be reinforced by economic arguments related to the potential for adverse selection in more traditional evaluation approaches, and the relative costs of basing allocation on randomisation and, say, peer group evaluation. For example, in SME and innovation programmes, public support is often allocated on the basis of some ex ante screening or peer evaluation process (which GPrix (2012) calls 'cream-skimming'). As in equity investments, this situation involves asymmetric information with the firm having an

incentive to over-estimate the strength of its project proposal. This creates the potential for adverse selection and the backing of weaker projects (Reiner, 2011). Evaluation or screening of programme applications is also resource intensive, and may be uneconomic particularly where the public support being offered to each firm is relatively small (Takalo and Tanayama 2010). In the case of innovation voucher schemes, for example, where the support available to each firm is typically small, undertaking detailed due diligence on each applicant company and project is not cost-effective, providing an additional economic rationale for randomised allocation.

3. Applying RCT+

3.1 Experiment rationale, design and conduct

We apply the RCT+ methodology to the Creative Credits programme, an innovation voucher programme intended to stimulate new partnerships between SMEs and creative service providers. Such partnerships may influence the innovation performance of firms through stimulating creativity, enhancing product quality and providing reputational benefits which signal the quality of firms' innovation activities (Powell 1998). Developing external partnerships may pose particular challenges for SMEs, however, associated with resource constraints (van de Vrande et al. 2009) and potentially behavioural failures on the part of managers such as inertia, excessive risk aversion, and myopia (Potts and Morrison 2009). Policy-makers have responded with instruments such as innovation vouchers to overcome these constraints with an EU survey in 2009 identifying 23 such schemes across Europe at regional and national level (DG ENTR-Unit D2 2009) although the number of schemes has probably increased since that point[1]. Such schemes have proved very popular with SMEs, having low entry barriers and playing a role in enabling small firms to access the knowledge necessary for innovation. The revenue stream provided by innovation vouchers has also encouraged knowledge providers to engage more actively with SMEs (Good and Tiefenthaler 2011; Biggar Economics 2009).

---

[1] Perhaps the best known innovation voucher scheme remains that operated in the Netherlands which, like many such schemes, aimed to overcome or lower the barriers which SMEs face in developing innovative partnerships with universities and colleges (Cornet, van der Steeg, and Vroomen 2007; Cornet, Vroomen, and van der Steeg 2006).

The logic model for the Creative Credits scheme is shown in Figure 1 and links the justification for public intervention in the relationship between SMEs and creative service providers, the scheme's objectives, the process by which the scheme operates, its immediate outputs and intended longer term outcomes (Donaldson and Gooler 2003)[2]. In the short term, scheme outputs are measured primarily in terms of increased levels of project additionality, i.e. the frequency of interactions between SMEs and creative businesses. In the longer term, we anticipate that this might generate three types of scheme outcome: output additionality, as the innovations developed impact on sales and growth (Hewitt-Dundas and Roper 2009); behavioural additionality, as organisational learning takes place and SMEs learn to work with and value more highly creative inputs to innovation (OECD 2006; Buiseret, Cameron, and Georgiou 1995; Georghiou 2004); and, network additionality, as SMEs extend their future collaborative networks (Consult 2006). The logic model also highlights that these medium- to long-term outcomes are contingent on other factors, most notably the business environment. Applying RCT+ allows us to focus on the 'causative' elements of the logic model. In other words, our evaluation objective is not just to assess scheme outcomes but also to consider whether these outcomes are being achieved through the mechanisms envisaged in the logic model (Chen 1990). This requires a theoretically grounded analysis of process and causal mechanisms alongside the evaluation of outcomes (Ludwig, Kling, and Mullainathan 2011).

The policy experiment was conducted in the Manchester City Region of North-West England between September 2009 and October 2010 with firms' applications being sought in two waves. The first wave opened for applications in September 2009 and the second in February 2010. The Creative Credits scheme was promoted and marketed through a number of media channels, business networks and tele-marketing. In promoting the scheme care was taken to minimise any selection biases. For example, companies tele-marketing the scheme used random methods of identifying which SMEs to call to promote the scheme (Bakhshi et al. 2013, p. 43). Over two thousand firms made some form of enquiry about the scheme with a total of 672 SMEs making eligible applications: 312 in the first wave and 501 in the second, 141 firms applied in both waves[3]. On receipt, applications from SMEs were first checked for

---

[2] Supporting evidence is found in Bakhshi and McVittie (2009), Stam, de Jong and Marlet (2008) and Muller, Rammer and Truby (2009) which all report that businesses which make proportionately greater use of creative services introduce more innovations.
[3] This was equivalent to around 1:8 of the eligible business population (Bakhshi et al. 2011, p. 12).

eligibility[4], Creative Credits were then randomly distributed across applicant firms. In total one hundred and fifty Creative Credits were 'awarded' in two equal groups roughly six months apart. Each Creative Credit had a face value of £4,000 with recipient firms also required to contribute a minimum of £1,000 to the cost of the project. Subsequent to the award, SMEs were encouraged to identify a new creative partner and develop a collaborative project proposal[5]. Once a partnership was formed, all projects were required to be completed within five months.

We examined the effectiveness of randomisation by comparing some of the key characteristics of the control and treatment firms at the time Creative Credits were awarded (Table 1). As anticipated, no significant differences were evident in the ownership status, exporting behaviour or prior innovation activity of the treatment and control firms. Firms in the treatment group were, however, significantly *less* likely to have engaged in R&D prior to their Creative Credits project than firms in the control group. As R&D is generally linked positively to innovation (Roper and Arvanitis 2012; Forsman 2011; Hoffman 1998), we might anticipate ceteris paribus that future innovation might also be lower among the treatment group (biasing downwards estimates of output additionality) (Bakhshi et al. 2013, p. 49-50). To compare the characteristics of scheme applicants and the population of eligible firms we conducted a small ancillary survey of non-applicant firms (Bakhshi et al. 2011). This suggested that applicants were similar in size to the eligible population but were less likely to be exporters, more likely to have a high proportion of graduate employees and engaged in prior innovation (Bakhshi et al. 2011, pp. 15-16). As discussed later in the paper these differences have implications for the external validity and generalizability of the results from our RCT+ evaluation.

Data collection for the quantitative element of the RCT+ evaluation comprised four sequential surveys of the treatment and control groups. Survey 1 was a baseline survey undertaken at the time firms were awarded a Creative Credit and allocated to the treatment

---

[4] Eligibility details are outlined in detail in Bakhshi et al. (2013). They included criteria relating to location, size and legal status.
[5] To help with this process a web-based marketplace – a Creatives Gallery – of eligible creative firms was designed and made available to all eligible SMEs in the treatment and control groups. The aim of creating the online Gallery was to explore the potential for a minimal brokerage model and reduce the burden of administrative costs of the pilot project. SMEs were not allowed to work with creative companies which they had previously worked with, however.

and control groups. Survey 2 was undertaken around six months after Survey 1, at a point just after the firms in the treatment group had completed their projects. Questions in Survey 2 related primarily to project additionality. Surveys 3 and 4 undertaken six month and twelve months later focussed on output, behavioural and network additionality. Over the course of the four surveys, and despite the payment of small cash incentives to firms to encourage continued participation in the data collection, significant attrition in response was experienced. By Survey 4, response numbers in the control group had fallen to 157, 52.2 per cent of those firms responding to Survey 1. In the treatment group attrition was less severe, with 78.0 per cent of Survey 1 respondents also responding to Survey 4. A key issue in terms of the internal validity of the experiment was whether the characteristics of the respondents to all of the four surveys and those firms which dropped out were similar, or whether attrition was systematically related to some respondent characteristic. Comparing the starting characteristics (from Survey 1) of stayers and those firms which did drop out, however, suggested no systematic differences to the initial control and treatment groups. This suggests that despite significant attrition this aspect of the internal validity of the experiment was maintained. Annex 1 provides a more detailed discussion of attrition and our tests for the representativeness of the longitudinal sample of firms which responded to all four surveys.

Data collection for the qualitative element of the RCT+ evaluation adopted a Cresswell (2011) explanatory mixed methods approach to bring rigour and provide insight to the quantitative phases through its integration with emerging qualitative results. Our qualitative research comprised of 4 stages of longitudinal data collection that involved 25 case studies with SMEs and their creative partners – recruited through a randomised process. The timing of the 4 stages aligned with the quantitative surveys. The case studies included 106 semi-structured interviews, multi-organisational workshops, and ethnographic direct observations of SME practices. Alongside Survey 1, 43 semi-structured interviews were undertaken with SMEs (24) and their creative partners (19). For Survey 2, a further 36 semi-structured interviews were conducted with the same organisations (22 SMEs and 14 creative partners), and Survey 4 added 27 interviews (19 SMEs and 8 creative partners). Interview questions explored core issues from the survey in depth to provide meaning, and probed to enable respondents to give their views on the scheme and their working relationships with their partner SMEs. Survey 3 was accompanied by two group workshops facilitated by members of the research team which included fourteen firms in the treatment group (Eden and

Ackermann 1998). These workshops focussed on themes arising from the qualitative data gathered alongside Surveys 1 and 2, and contributed to the development of topics and questions for the Stage 4 quantitative survey and qualitative interviews[6]. Attrition in the qualitative sample, due to firms ceasing trading or refusing to continue to participate in interviews, resulted in longitudinal qualitative data being available from 11 pairings plus the 'surviving' member of five other partnerships.

Qualitative data collection and analysis was guided by the constant comparative method (Strauss and Corbin 1990) which is central to grounded theory methodology (Glaser and Strauss 1967; Glaser 1998; Katz 1983), and complements longitudinal explanatory designs (Cresswell and Plano Clark 2011). This allowed substantive concepts and trends to emerge from a comparison of the data collected over the four stages. Qualitative data analysis involved a process of open coding to identify concepts and then axial coding to establish higher-level categories and identify relationships between the open codes (Glaser 1998). Then selective coding refined codes until clear relationships between them were identified, leading to the development of a theory about the data. Themes identified from each stage informed subsequent stages of both the qualitative and quantitative data collection, providing theoretical sampling (Glaser and Strauss, 1967). The longitudinal aspect of the interviews provided opportunities to confront the emerging theory with further data, and move towards theoretical saturation. Rigour was ensured through three types of triangulation (Yin 2003). Data triangulation corroborated findings in different types and sources of qualitative data. Investigator triangulation ensured that analyses were conducted on data by more than one researcher to avoid investigator bias (Bryman 2004). Methodological triangulation involved different data (interviews, workshops, observations) being analysed using different techniques (manual and software-supported approaches) to search for generic findings (Bryman 2001). Furthermore, as respondents agreed to be taped and all interviews were transcribed, data was analysed in its totality several times (Gillham 2000). As such the application of thematic data analysis, contributes towards protecting the confidentiality and anonymity of the 25 case study respondents.

---

[6] Where results from the qualitative data are reported in the form of quotations or conclusions from interviews, each business has been allocated a letter and number to signify SMEs in the treatment group (S), their creative partners (C) and their participation in wave 1 or wave 2 of the experiment. So, for example, W1S1 relates to wave 1, SME 1. Any unattributed quotes come from the 'free format comments' boxes in the quantitative surveys and therefore cannot be attributed under the non-disclosure terms of those surveys.

3.1 Evaluation results

The first question considered in the RCT+ evaluation was the extent of project additionality, i.e. the extent to which the award of a Creative Credit increased the probability that an SME entered into a new relationship with a creative business (Figure 1). Modelling suggests that the Creative Credits treatment increased the probability that firms went ahead with their project within the five months since allocation of the Creative Credits by around 84 per cent (Table 3) [7], a level of project additionality very similar to that reported for the pilot Dutch innovation vouchers scheme (Cornet, Vroomen, and van der Steeg 2006)[8]. Evaluations of the Swiss and Austrian innovation voucher schemes suggest similarly high levels of project additionality (Good and Tiefenthaler 2011).

The quantitative Dutch, Swiss and Austrian innovation voucher evaluations explained their results summatively but not causally. With RCT+, our qualitative data identified two factors that underpinned the robust quantitative estimates of project additionality. First, SMEs described how their project fed operational plans and how they anticipated it would enable them to market their company offering more widely (in line with the stated objectives of the scheme). Second, a number of companies indicated how the Creative Credit had helped them to 'accelerate things' (respondent from *W*ave 1 *S*ME 8 – W1S8): 'I'd have waited until we'd accumulated more money. And then probably programmed it in for sort of the back end of this year' (W1S10). In only a minority of cases included in the interviews did Creative Credits instigate projects that would not otherwise have taken place: 'couldn't have afforded [the project, so] wouldn't have done it but it has made a real difference' (W1S9).

Output additionality should be reflected in identifiable differences in the sales and innovation profiles of the treatment and control groups in the period shortly after the completion of the treatment. In the RCT+ evaluation, survey-based comparisons were made 6 months and 12

---

[7] Of the 301 firms in the control group which responded to our baseline survey, 36 firms (12 per cent) went ahead anyway with their projects. Among the group of 150 firms which were assigned Creative Credits 144 (96.0 per cent) actually commissioned projects.

[8] We modelled project additionality using both simple OLS and as a robustness check a treatment model allowing for potential sample selection. As expected, selection effects proved insignificant. As part of the same exercise we also investigated whether the small cash incentives provided to firms to help encourage survey response had biased these results. No evidence of any significant bias could be identified Details of these models and those referred to later in this section can be found in Bakhshi et al. (2013).

months after the completion of the Creative Credits project. At each point we measured firms' innovation activity in the previous six months using three binary measures: whether the firm had undertaken any product/service innovation; whether they had undertaken any new to the market innovation; and, whether they had undertaken any process innovation. Six months after the end of the treatment, firms in the treatment group were significantly more likely to be undertaking product/service and process innovation, and had (at the 10 per cent level) a significantly more positive distribution of sales growth rates than firms in the control group (Table 3). These positive output additionality effects were short-lived, however, with no significant differences between the treatment and control groups in product/service or process innovation or sales growth distribution evident 12 months after the end of the treatment (Table 3). To test the robustness of the RCT+ results we used multivariate models allowing for potential selection effects which confirmed the RCT+ results (Bakhshi et al. 2013)[9].

Inevitably there were exceptions but, in the short-term, the qualitative element of the RCT+ analysis suggested a largely positive feeling about the Creative Credits scheme. The majority of innovations involved website enhancements and/or the development of marketing collateral and SMEs had generally positive expectations of the impact of these on their business. However longer-term, the qualitative data supported the absence of any boost to sales growth. Many SMEs, like their wider markets, experienced a drop in sales over the period which tended to push them towards focusing on winning contracts or obtaining support from banks, taking attention away from their innovation, so undermining the value of the project. As one SME reported: 'like sticking a Band Aid on a much bigger problem' (W1S3). For a few SMEs the impact was low because they were dissatisfied with their innovation outcome: 'If I was to pay that out of my own account I would just not be happy at all'.

Behavioural additionality is measured here primarily in terms of firms' innovation intentions. Firms in the treatment and control groups were asked: 'Over the next three years how likely

---

[9] These sales effects may under-estimate the longer sales benefits of Creative Credits. Twelve months after the end of the project – in Survey 4 – firms were asked how long they expected the benefits to persist. On average firms indicated that the benefits of their projects were likely to persist for a further 2.5 years, or 3.5 years after the end of the project. To capture these longer-term effects we calculated a net present value for the scheme which we estimated at 1.37 depending on the assumptions made (Bakhshi et al. 2013, Annex 5).

are you to develop new or significantly improved …'. Six different dimensions of innovation were considered: goods or services, processes, corporate strategy, advanced management techniques, organisational structure and marketing. We found no significant differences in future innovation intentions between the treatment and control groups after either 6 or 12 months (Table 4). This RCT+ finding was again confirmed in a multivariate analysis (Bakhshi et al. 2013). Despite this, significant proportions of firms in the treatment group reported enhancements to their innovative capacity as a result of their Creative Credits project, though this result was not testable in comparison with the control group. This result was echoed in the qualitative analysis which revealed significant behavioural change. In previous innovations many SMEs had 'muddled along' (W2S11) alone, either delivering creativity in-house without the required skills, outsourcing to cheap (unprofessional) alternatives, or outsourcing to a professional supplier without a proper budget (e.g. W2S11, W2S2). Creative Credits provided a learning opportunity for SMEs, and some SMEs received knowledge on the creative process: 'to see how it's all put together … how they do it' (W1S3). In some cases this was used to educate senior managers who preferred (low quality) in-house creative supply that 'sometimes you outsource' and bring 'fantastic agencies on board' and help them realise the added value by 'open[ing] ... people's eyes' (W2S11) to the need for changing behaviour. Skills transfer was designed into some contracts (e.g. media training (W1S1) and website management (W2S2)), while other SMEs received reusable knowledge/skills as a by-product, for example on: writing marketing collateral (W1S3); web site design (W1S4); non-traditional advertising (W1S5). However, this learning was sometimes painful (W1S11) and complex, 'a complete can of worms' (W1S10), which hindered inter-organisational learning, limiting behavioural additionality. For example, the inaccessible/technical language which suppliers and customers use (W1S2) delayed benefits and disrupted/jeopardised the innovation.

Some creative companies aimed to encourage behavioural additionality by transferring knowledge and tools (e.g. respondent from *Wave 2 Creative 14* (W2C14) and W2C17), however, this was burdensome (W2C12): 'small client syndrome, education time is high... a lot of hand-holding, it's the worst part of the job' (W2C18).

16

Network additionality was measured using a similar question to that for behavioural additionality. Six and twelve months after the end of the Creative Credits project firms were asked: Over the next three years how likely are you to co-operate with the following innovation partners. Eight different partner types were then identified (Table 5). Perhaps unsurprisingly given our results on behavioural additionality, we also found no evidence of any significant network additionality, i.e. no significant differences between the treatment and control groups in the propensity to co-operate on future innovation (Table 5). Again this RCT+ result was confirmed using multivariate models allowing for potential selection bias (Bakhshi et al. 2013). In this respect our results are again similar to those in the Dutch innovation voucher scheme, where follow-up after 18 months suggested no evidence of persistent additionality effects either in the formation of new partnerships or the development of new products or processes (Cornet, van der Steeg, and Vroomen 2007). One possible reason for the lack of longer-term effects suggested in relation to the Austrian and Swiss innovation voucher evaluations, is the small size of many voucher recipients and the difficulty which they have in establishing follow-up partnerships (Good and Tiefenthaler 2011).

In interviews, findings on network additionality pointed to three considerations. First, for some SMEs their Creative Credits project had been based on a 'transactional' relationship with their creative service providers: 'a one-off hit where they got something for free' (W1C7). Indeed, fifteen of the nineteen SMEs interviewed at the time of Survey 4 referred to having already worked with other creative businesses prior to the Creative Credits scheme. For these firms, Creative Credits encouraged them to work with a new creative partner, but generated little new organisational learning about effective partnering with creative businesses in general. Consequently, most SMEs ended their working relationship with their Creative Credits partner when their project had completed, consistent with their previous working relationships with creative businesses (W1S1, W1S10, W2S8, W2S6, W2S7, W2S12, W1S12). A second reason for the lack of sustained network additionality was dissatisfaction with their creative partner (W2S2, W1S7, W1S8, W1S11, W1S3). Sometimes this related to not building lasting personal relationships with the supplier. At other times, the dissatisfaction related to the attitude of their creative partner: 'it did feel that they were in a sense had bigger fish to fry than our project' (W2S4) but SMEs 'trusted in them to do it all

really' but more SME involvement could have ensured it went 'in the direction I wanted it to go in' (W1S7). Creative suppliers verified these difficulties, also emphasising communication issues, different interpretations of project objectives, not receiving input data, client dithering, and differing priorities: 'We had to hold their hand all the way through it' (W2C12). The third reason for the lack of sustained networking additionality was the lack of brokerage or assistance with identifying/managing the creative partner. SMEs advised that additional networking could help by 'actually go[ing] and meet[ing] a few of these other companies … I would get heaps more out of the project'.

4. Discussion

The aim of the evaluation reported here is to assess the potential value of experimental approaches to industrial policy evaluation. The RCT+ experimental evaluation approach we apply has three main elements: randomised assignment of firms to the treatment and control groups; a longitudinal approach to data collection; and, the use of a mixed-methods data collection strategy. In this section we assess the value of each of these elements of the evaluation in terms of internal and external validity. We also consider the value and interpretation of evidence from single policy experiments such as that conducted here building on approaches from medicine.

Central to all experiments is the randomised allocation of subjects to the treatment and control groups to ensuring internal validity (Burtless 1995). Randomisation also enables the evaluator to avoid issues of sample selection rather than having to control for these ex post. Randomisation itself, however, has some limits. In large samples randomisation will result in no significant differences between the characteristics of the treatment and control groups. In our RCT+ trial, however, in relatively small samples we identified some significant differences between the treatment and control groups at the time that Creative Credits were awarded (Table 1). These systematic and significant differences between the characteristics of the treatment and control group illustrate the potential limitations of randomised allocation, even where samples are above the critical threshold (c. 300 observations)

highlighted by Bruhn and McKenzie (2009), imparting a bias of unknown extent to any evaluation results.

Another aspect of internal validity is the appropriateness of the timescale over which policy effects are measured and additionality interpreted. Behavioural or network benefits from measures such as Creative Credits, for example, may take some time to translate both into new innovation and innovation outcomes (Cunningham, Gök, and Laredo 2013). A priori, it is also difficult to assess how long the delays may be between any specific policy initiative and its impacts. Adopting a single impact period – as in the majority of innovation evaluations – may therefore either miss or under-estimate the true effect of any initiative. To address this problem our RCT+ approach, which aims to capture both short-term impacts and longer-term outcomes, adopts a longitudinal data collection strategy, measuring impacts at different times after project completion. In the case of Creative Credits this proved important in revealing significant differences in the profile of output additionality after six months and twelve months. After six months significant innovation and sales effects were evident, effects which had lost their significance after twelve months. Measuring output additionality at either single point would therefore have provided an incomplete and potentially misleading assessment of output additionality (Roper 2013). Adopting a longitudinal data collection approach in industrial policy experiments therefore helps to avoid potentially misleading inference and assess any additionality-sustainability trade-offs which may arise in policy outcomes (Hewitt-Dundas and Roper 2011).

Understanding the reasons why a particular outcome profile results from a policy initiative requires qualitative insights, the third element of the RCT+ methodology. In particular, although many SMEs in the Creative Credits treatment group highlighted positive aspects of working with their creative partner, our longitudinal case-studies highlighted issues which may have contributed to the lack of sustained output additionality e.g.: the transactional nature of the relationship with their creative partners; communication and co-ordination issues; and, in some cases, unsatisfying collaborations. Such data provide insights into the causal processes underlying additionality outcomes and potentially useful suggestions for subsequent implementation of any scheme beyond the initial policy experiment.

Industrial policy experiments conducted using an RCT+ or similar methodology therefore have significant advantages over a standard industrial policy pilot with ex post evaluation. The advantages are three-fold: substantive – providing potentially more robust evaluation results; resourced-based – in terms of more cost-effective evaluation; and, accessibility – in terms of the ease of communication of evaluation results. In substantive terms, randomisation helps the evaluator to avoid sample selection bias, while a longitudinal approach reduces the risk of incorrectly estimating policy effects due to the choice of an inappropriate impact period. We found important insights from later parts of the experiment which would be lost if it were not for the staged design of RCT+. The translation of short-term gains into longer term behavioural or network additionality was not significantly different in treatment group versus the control group, evident as the tendency of recipient firms to continue existing (or establish new) relationships with creative companies was not found. The incentive to initiate a policy on innovation partnerships, therefore, needs to be based on the investment having sufficient direct payback as behavioural by-products from the experience may be limited by exogenous factors.

Mixed-methods also provide a useful complement, with qualitative approaches providing rich causal insights. Thus, while econometric analysis can shed light on the return on investment, behavioural and network additionalities, the use of interviews and ethnography can provide deeper levels of understanding allowing insight to the complexity of policy evaluation in social systems. Cross-methodological approaches require a multi-disciplinary research team that can best combine quantitative and qualitative expertise with reference to the paradigm and technical debates that characterise mixed methods using two or more modes of investigation (Tashakorri and Teddie 1998; Cresswell and Plano Clark 2011). RCT+ type evaluation approaches may also be cost-effective as they have more limited information requirements than the complex econometric models necessary to control for sample selection. However, the longitudinal element of RCT+ puts added pressure on the research team to find ways to maintain response rates over time and over different methods. This therefore needs a sympathetic group of respondents who have to bear with two (or more) modes of research investigation and a stronger focus on mid-term feedback to engage with respondents. This additional investment brings risks of higher drop-out rates due to fatigue but also brings

added benefits in providing more engaged respondents, data that can explain the correlations found, and insights to new questions to ask in later stages of the longitudinal survey design. Finally, evaluations based on randomisation, and the control group-treatment group comparison, generate more transparent and easily communicated policy messages than evaluations based on more complex econometric approaches.

Beyond the internal validity and insights generated by the Creative Credits experiment, the other primary issue is external validity, i.e. the generalisability of the evaluation results to the wider population of firms (Garfinkel, Manski, and Michalopoulos 1990). To assess external validity we conducted a baseline survey of the characteristics of non-applicants to the Creative Credits scheme. This suggested that members of the treatment and control groups were more focused on innovation than non-applicants to the Creative Credits scheme, and also were more likely to have previously worked with external partners and had higher internal skill levels than non-applicants (Bakhshi et al. 2011, Annex 2)[10]. As each of these factors is likely to influence scheme outcomes this clearly limits the external generalisability of our evaluation results to the general population of SMEs. Two other limitations to external validity are also worth noting. First, even where applicant and population characteristics are similar short-term (and small-scale) experimental studies are unlikely to capture the macro and social effects highlighted by Garfinkel et al. (1990). One additional 'macro' effect not considered by Garfinkel et al. (1990) is the potential for industrial policy initiatives to generate potential synergies between schemes[11]. In the case of the Creative Credits scheme, for example, a number of scheme applicants had already received support from another innovation voucher scheme operating in the region which focussed on developing SME-university linkages. Such synergies are likely to be positive in the medium-term but are unlikely to be evident during any short-term experimental evaluation.

---

[10] See Nightingale and Coad (Nightingale and Coad 2014) for a recent discussion of heterogeneity in the population of small firms and some of the policy implications.
[11] Heckman and Smith (1995) consider a related point in their discussion of substitution bias. This relates to a situation where schemes are direct substitutes, however, rather than complementary to each other.

One other element of external validity which is of particular importance in industrial policy experiments relates to timing and the potential impact of the broader business climate. Our policy experiment took place between September 2009 and October 2010, a period when the UK economy was either contracting or experiencing only marginal growth. For example, one firm in the treatment group commented on their Creative Credits project that 'it was useful at the time, but because of the recession and all the rest of it, I don't believe we've really seen the benefit yet, but I suspect we will do'. Another SME could not fully implement the outcome of their project because it had not been able to secure bank funding, saying that 'We are working our guts out, I have to try and raise finance'. Such comments were relatively typical, emphasising the strongly contingent nature of outcomes from single policy experiments. This type of consideration relating to the validity (either internal or external) of individual trials have suggested the potential value of triangulation or meta-analytical reviews across a number of RCTs. In medicine this approach is evident in the Cochrane Library which has established clear protocols for the conduct of individual RCTs and the undertaking of 'reviews' or meta-analyses of individual RCTs.

Two recent Cochrane reviews illustrate this approach. One Cochrane review considering treatments for frozen shoulder synthesised the results from nineteen separate RCTs (Page et al. 2014). Another considered three RCTs in a treatment for early stage throat cancer (Yuan et al. 2014). In each case the reviews compare and contrast the results from the studies and identify generalised findings based on the principle that any one study may either reflect methodological or operational anomalies which may influence the results. In the context of medicine, however, interventions are primarily mechanical: a treatment which works with one human being is very likely to work with another independent of the context in which they are living. In terms of industrial support measures, however, the situation is arguably more complex given the heterogeneity of small firms, the social and interactive nature of business and the importance (and diversity) of the contexts in which firms operate. Internal and external validity may therefore be *more* difficult to maintain in industrial policy RCTs than in medicine. Arguably this means that standards of evidence should also be higher, requiring

consistent evidence from multiple RCTs from different contexts, before robust conclusions should be drawn[12].

5. **Conclusions**

The aim of this paper was to propose and assess the RCT+ evaluation approach. The RCT+ approach combines the random allocation of scheme applicants to control and treatment group with a mixed methods and longitudinal data collection strategy. The combination of the mixed methods and longitudinal aspects of the RCT+ approach provided valuable insights into the validity of the Creative Credits logic model. The longitudinal element of the evaluation method proved important in highlighting the time profile of additionality, while the mixed methods approach suggested some of the underlying causal factors. Overall, this suggests the value and feasibility of experimental methods for industrial policy evaluation, especially where quantitative analyses are complemented by structured qualitative data collection. Two other aspects of our evaluation experiment suggest, however, that generalisations from such small-scale experiments need to be made with care. First, it is important to consider the external validity of the evaluation in terms of both the representativeness of scheme applicants and the contingent nature of scheme outcomes. Both potentially limit the generalisability of results from the experimental evaluation beyond the immediate spatial and temporal context. Second, our evaluation is subject to the micro to macro biases discussed in Garfinkel et al. (1990) and Heckman and Smith (1995). Overcoming these limitations requires replication in different economic and market contexts and then synthesis along the lines of the approach adopted by the Cochrane Library in medicine.

Experimental approaches to industrial policy evaluation remain rare despite the long tradition of experimental evaluation in social policy and development studies (although see Georgiadis and Pitelis, 2015). Hopefully, the RCT+ approach described here will encourage others to combine experimental and qualitative approaches to industrial policy evaluation. Policy makers may of course argue that random allocation of public subsidy is inefficient and fails

---

[12] Innovation and management studies come closest to this approach in meta-analyses (Mueller, Rosenbusch, and Bausch 2013; Rosenbusch, Brinckmann, and Bausch 2011).

to maximise the potential benefits of public investment. In the short term this is most probably correct as random allocation is likely to mean that some poor projects will be supported. The counter argument – and in our view it is a powerful one – is that randomisation permits more effective policy evaluation and testing, and in the longer run may help avoid poor (and expensive) policy decisions. In the UK, at least, this is a message which policy makers have accepted with the establishment of a series of large-scale industrial policy experiments relating to Growth Vouchers, mentoring for high growth firms and university-industry co-operation[13].

---

[13] For details of the Growth Vouchers programme see: https://www.gov.uk/apply-growth-vouchers.

Annex 1: Attrition and the representativeness of survey respondents

Four surveys were conducted with the treatment and control groups. These were: the initial baseline Survey 1, conducted immediately after the award of the Creative Credits; Survey 2 at the time projects were completed; Survey 3, six months after project completion; and, Survey 4, twelve months after project completion. Attrition was evident in each survey with response rates given in Table A1.1. Attrition was, unsurprisingly, worse among the control group as they had relatively little commitment to the Creative Credits initiative with respondent numbers by Survey 4 falling to 157, 52.2 per cent of those firms initially responding to Survey 1. In the treatment group attrition was less severe, with 78.0 per cent of Survey 1 respondents also responding to Survey 4.

Table A1.1: Survey response and attrition by survey, wave and recipient

	Survey				Survey			
	1	2	3	4	1	2	3	4
	No.	No.	No.	No.	%	%	%	%
Treatment	150	136	129	117	100.0	90.7	86.0	78.0
Control	301	247	175	157	100.0	82.1	58.1	52.2
Total	451	383	304	274	100.0	84.9	67.4	60.8

During the course of the study considerable efforts were made to reduce attrition and ensure adequate response during the later surveys. Surveys were conducted online with firms being contacted initially by email and then (often repeatedly) by telephone to encourage response. Small cash incentives (£30) were also paid to firms in the control group to encourage their continued participation in the project. In Survey 2, for example, 167 of the 247 control businesses responding to the survey were incentivized in this way. One possibility is that these incentives might have biased the pattern of responses by control group companies. Investigation of respondent characteristics at the time of Survey 2, however, suggested that incentive payments had no impact on estimates of short-term additionality. Following Survey 4 we also tested econometrically for the impact of these cash incentives on project additionality and found no evidence that incentives had any effect on the control group – treatment group comparison (Bakhshi et al. 2013, p. 59).

Given the level of attrition it is of considerable interest to consider whether those firms which responded to all four surveys were similar to those who dropped out of the group of respondents (Table A1.2). Reassuringly, for both the treatment and control groups, we see little systematic difference in the characteristics of respondents and non-respondents, suggesting that the propensity to drop-out of the project was not systematically related to firms' key characteristics.

Table A1.2: Characteristics drop-outs and responders

	N	Responders	Drop-outs	t-statistic	Signif.
A. Treatment					
Limited company	117/33	90.6	84.8	0.83	0.41
Exporting firm	116/33	39.7	30.3	1.00	0.32
Internal R&D	113/31	69.9	54.8	1.50	0.14
External R&D	114/28	21.9	25.0	-0.33	0.74
Innovation	113/32	79.6	78.1	0.18	0.86
Radical innovation	83/20	55.4	55.0	0.03	0.97
B. Control					
Limited company	157/144	94.9	91.0	1.32	0.19
Exporting firm	150/134	31.3	37.3	-1.06	0.29
Internal R&D	151/138	79.5	75.4	0.83	0.41
External R&D	150/138	36.0	32.6	0.60	0.55
Innovation	148/135	75.0	74.8	0.04	0.97
Radical innovation	106/98	62.3	50.0	1.77*	0.08

Note: Characteristics were measures in Survey 1. Sample numbers in first column relate to respondents to responders/drop outs. * denotes significance at the 10 per cent level; ** at 5 per cent and *** at the 1 per cent level.

Table 1: Comparison of characteristics of treatment and control groups

	Treatment Group		Control Group			
	n	%	n	%	t-statistic	Signif.
Limited company (%)	150	89.33	301	93.02	-1.26	0.21
Exporting firm (%)	149	37.58	284	34.15	.70	0.48
R&D inside the firm (%)	144	66.67	289	77.51	-2.33**	0.02
R&D external to the firm (%)	142	22.54	288	34.38	-2.63***	0.01
Product or service innovation (%)	145	79.31	283	74.91	1.04	0.30
New to the market innovation (%)	103	55.34	204	56.37	-.17	0.86

Notes: Table is based on the baseline survey. All measures are binary. Differences in response numbers reflect survey responses to each question and those selecting the "Don't know' response. ** denotes significance at the 5 per cent level and *** at the 1 per cent level.

Table 2: Project Additionality - The likelihood of firms undertaking their innovation project

Dependent variable: Whether or not firms undertook their project

Number of observations   451

Adjusted R-squared   0.653

Variable	Coefficient	Std. Err.	t-statistic	Signif
Creative Credit	0.840	0.028	29.11	0.000***
Constant term	0.119	0.017	7.18	0.000***

Notes: Analysis is based on respondents to the initial baseline survey undertaken immediately after the allocation of Creative Credits. *** denotes significance at the 1 per cent level.

Table 3: Output additionality in terms of the probability of innovation: 6 months and 12 months after the completion of the Creative Credits projects

	N	Control % firms	Treatment % firms	t-statistic	Signif.
A. After 6 months					
Product or service innovation	145/105	55.9	72.4	2.740	0.007***
New to the market innovation	126/92	23.0	35.9	2.089	0.038**
Process innovation	142/105	47.2	63.8	2.618	0.009***
B. After 12 months					
Product or service innovation	154/113	63.0	70.8	1.345	0.180
New to the market innovation	135/97	32.6	40.2	1.192	0.235
Process innovation	153/111	51.0	47.7	0.517	0.606

	N	Control %	Treatment %	$\chi^2(6)$	Signif.
C. Average Sales Growth					
After 6 months	146/107	6.4	7.5	11.5	0.075*
After 12 months	155/114	4.7	7.8	7.7	0.261

Notes: Six and twelve months after the completion of the Creative Credits project firms were asked whether they had undertaken innovation in the previous 6 months, and for a banded estimate of their sales growth. Table is based on the longitudinal sample. Respondent numbers are given as "control group/treatment group"* denotes significance at the 10 per cent level; ** at 5 per cent and *** at the 1 per cent level. The $\chi^2(6)$ test statistic is based on the difference in distribution of sales growth rates between the treatment and control groups.

Table 4: Behavioural additionality - future innovation intentions after completion of Creative Credits projects

	N	Control			Treatment			$\chi^2(2)$	Signif.
		Not Likely %	Quite Likely %	Very Likely %	Not Likely %	Quite Likely %	Very Likely %		
A. After 6 months									
Goods or services	248	4.3	35.5	60.3	5.6	27.1	67.3	2.036	0.361
Processes	245	7.9	34.5	57.6	8.5	33.0	58.5	.074	0.964
Strategy	235	15.7	41.8	42.5	12.9	38.6	48.5	.912	0.634
New technologies	204	38.5	36.8	24.8	41.4	42.5	16.1	2.321	0.313
Organisation	216	39.5	28.7	31.8	44.8	32.2	23.0	1.984	0.371
Marketing	231	10.4	40.7	48.9	11.5	45.8	42.7	.864	0.649
B. After 12 months									
Goods or services	257	6.7	40.9	52.3	7.4	35.2	57.4	.876	0.645
Processes	260	9.3	42.4	48.3	11.9	36.7	51.4	1.059	0.589
Strategy	249	14.6	45.8	39.6	18.1	38.1	43.8	1.583	0.453
New technologies	206	47.6	29.8	22.6	45.1	32.9	22.0	.224	0.894
Organisation	221	49.3	36.6	14.2	47.1	32.2	20.7	1.676	0.433
Marketing	235	18.4	38.2	43.4	15.2	48.5	36.4	2.464	0.292

Notes: Six and twelve months after the end of the Creative Credits project firms were, for example, asked: 'Thinking about the next three years. How likely are you to develop new or significantly improved goods or services'. Similar questions were asked for other types of innovation. Table is based on the longitudinal sample.
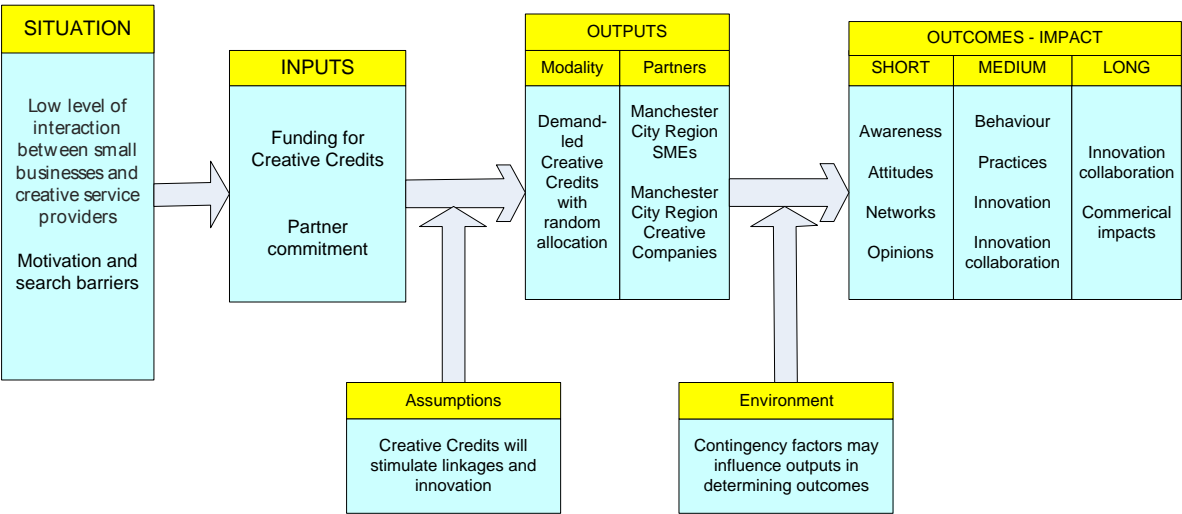* denotes significance at the 10 per cent level; ** at 5 per cent and *** at the 1 per cent level.

Table 5: Network Additionality – probability of innovation cooperation after completion of Creative Credits projects

	N	Control % firms	Treatment % firms	t-statistic	Signif.
A. After 6 months					
Other group companies	79/73	27.8	31.5	-0.490	0.625
Suppliers	81/74	58.0	52.7	0.662	0.509
Creative service suppliers	80/75	53.8	58.7	-0.613	0.540
Customers	81/74	58.0	55.4	0.327	0.744
Competitors	79/73	24.1	9.6	2.429	0.016
Higher Education Institutes	81/75	22.2	20.0	0.338	0.736
Public Laboratories	81/74	12.3	10.8	0.297	0.767
B. After 12 months					
Other group companies	93/75	33.3	37.3	-0.536	0.593
Suppliers	95/77	52.6	63.6	-1.458	0.147
Creative service suppliers	95/77	52.6	64.9	-1.637	0.104
Customers	95/77	65.3	71.4	-0.864	0.389
Competitors	96/77	24.0	32.5	-1.228	0.221
Higher Education Institutes	93/78	29.0	26.9	0.305	0.761
Public Laboratories	94/76	11.7	18.4	-1.204	0.231

Notes: Six and twelve months after the completion of the Creative Credits project firms were asked 'Over the next three years how likely are you to co-operate with the following types of innovation partners?'. Table is based on the longitudinal sample. Respondent numbers are given as "control group/treatment group". * denotes significance at the 10 per cent level; ** at 5 per cent and *** at the 1 per cent level.

Figure 1: Logic model for Creative Credits

SITUATION	INPUTS	OUTPUTS		OUTCOMES - IMPACT		
		Modality	Partners	SHORT	MEDIUM	LONG
Low level of interaction between small businesses and creative service providers Motivation and search barriers	Funding for Creative Credits Partner commitment	Demand-led Creative Credits with random allocation	Manchester City Region SMEs Manchester City Region Creative Companies	Awareness Attitudes Networks Opinions	Behaviour Practices Innovation Innovation collaboration	Innovation collaboration Commerical impacts

Assumptions

Creative Credits will stimulate linkages and innovation

Environment

Contingency factors may influence outputs in determining outcomes

References

Bakhshi, H. 2009. Creative Supply Chain Linkages and Innovation: Do the Creative Industries Stimulate Business Innivation in the Wider Economy? *Innovation: Management, Policy & Practice* **11:169-189**.

Bakhshi, H.; J. Edwards; S. Roper; J. Scully; and D. Shaw. 2011. Creating innovation in small and medium-sized enterprises: Evaluating the short-term effects of the Creative Credits pilot. London: Nesta.

Bakhshi, H.; J. Edwards; S. Roper; J. Scully; D. Shaw; L. Morley; and N. Rathbone. 2013. Creative Credits - a randomized controlled industrial policy experiment. London Nesta.

Banerjee, A.V. and E. Duflo. 2008. Title. NBER working paper 14467.

———. 2011. Poor Economics: A Radical Rethinking of the Way to Fight Global Poverty. *Public Affairs*.

BIS. 2009. Research to improve the assessment of additionality, ed. D.f.B.I.a. Skills. London.

Biggar Economics, B. 2009. Interim evaluation of the innovation voucher project - a final report to North West Development Agency Manchester.

Bloom, N.; B. Eifert; A. Mahajan; D. McKenzie; and J. Roberts. 2011. Does Management Matter? Evidence from India. *NBER working paper*.

Bratberg, E.; A. Grasdal; and A.E. Risa. 2002. Evaluating social policy by experimental and nonexperimental methods. *Scandinavian Journal of Economics* **104:147-171**.

Bremnes, H.; B. Bergem; and E. Nesset. 2011. Title. Working Paper 2011:3, Molde University College.

Bruhn, M. and D. McKenzie. 2009. In Pursuit of Balance: Randomization in Practice in Development Field Experiments. *American Economic Journal-Applied Economics* **1:200-232**.

Bryman, A. 2001. *Social Research Methods*. Oxford: Oxford University Press.

———. 2004. Triangulation. In *The Sage Encyclopedia of Social Science Research Methods*, ed. M. Lewis-Beck; A. Bryman; and T. Liao. California: Sage Publications Inc

Buiseret, T.; H.M. Cameron; and L. Georgiou. 1995. What differences does it make? Additionality in the public support of R&D in large firms. *International Journal Of Technology Management* **10:587-600**.

Burtless, G. 1995. The case for randomised field trials in economic and policy research *Journal of economic perspectives* **9:63-84**.

Chen, H.-T. 1990. *Theory-Driven Evaluations*. London: Sage

Consult, I. 2006. Study and Evaluation of the Behavioural Additionality of R&D subsidies. Brussels.

Cornet, M.; M. van der Steeg; and B. Vroomen. 2007. De effectiviteit van de innovatievoucher 2004 en 2005 Effect op innovatieve input en innovatieve output van bedrijven, ed. C.D.P.N. 140.

Cornet, M.; B. Vroomen; and M. van der Steeg. 2006. Do innovation vouchers help SMEs to cross the bridge towards science? In *No 58 CBP Discussion Paper* CBP.

Cresswell, J. and V. Plano Clark. 2011. *Designing and Conducting Mixed Method Research*. Los Angeles; London; New Delhi; Singapore; Washington DC: Sage.

Cunningham, P.; A. Gök; and P. Laredo. 2013. The Impact of Direct Support to R&D and Innovation in Firms. Compendium of Evidence on the Effectiveness of Innovation Policy. London: NESTA.

Deaton, A. 2010. Instruments, Randomization and Learning about Development. *Journal Of Economic Literature* **48:424-455**.

Denzin, K.D. and Y.S. Lincoln. 2005. *The SAGE Handbook of Qualitative Research*. Sage Publications.

DG ENTR-Unit D2. 2009. Availability and Focus on Innovation Voucher Schemes in European Regions, ed. EU. Brussels

Donaldson, S.I. and L.E. Gooler. 2003. Theory-driven evaluation in action: lessons from and $20 million statewide Work and Health Initiative. *Evaluation and Program Planning* **26:355-366**.

Duflo, E.; R. Glennerster; and M. Kremer. 2006. Title. Technical Working Paper 333, NBER

Eden, C. and F. Ackermann. 1998. *Making Strategy: The Journey of Strategic Management*. London: Sage.

Forsman, H. 2011. Innovation capacity and innovation development in small enterprises. A comparison between the manufacturing and service sectors. *Research Policy* **40:739-750**.

Garcia, A. and P. Mohnen. 2010. Impact of government support on R&D and Innovation In *UNU-Merit Working Paper 2010-034*.

Garfinkel, I.; C.F. Manski; and C. Michalopoulos. 1990. Title. Institute for research on poverty, discussion papers Dp 922-90, University of Wisconsin-Madison.

Georgiadis, A and Pitelis, C.N. 2015. The impact of employees' and managers' training on the performance of small and medium-sized enterprises: Evidence from a randomised natural experiment in the UK service sector, *British Journal of Industrial Relations*, forthcoming.

Georghiou, L. 2004. Evaluation of Behavioural Additionality - Concept Paper. In *European Conference on Good Practice in Research and Evaluation and Indicators*.

Gillham, B. 2000. *Case Study Research Methods*. London: Continuum.

Glaser, B. 1998. *Doing Grounded Theory: Issues and Discussions*. Mill Valley: The Sociology Press.

Glaser, B. and A. Strauss. 1967. *The Discovery of Grounded Theory: Strategies for Qualitative Research*. Chicago Aldine Publishing Company.

Good, B. and B. Tiefenthaler. 2011. Innovation voucher- small is beautiful. *Plattform - Forschungs- und Technologieevaluierung***:45-53**.

GPrix. 2012. Good Practices in Innovation Support Measures for SMEs: facilitating transition from the traditional to the knowledge economy. In *Recommendations report by Staffordshire University and Fraunhofer IFF on behalf of the GPrix consortium, FP7*.

Greenberg, D. and M. Shroder. 1991. *Digest of the social experiments*. Madison: Institute for research on poverty, University of Wisconsin.

Guba, E.G. and Y.S. Lincoln. 1989. *Fourth generation evaluation*. Newbury Park: Sage.

Heckman, J.J. and J.A. Smith. 1995. ASSESSING THE CASE FOR SOCIAL EXPERIMENTS. *Journal of economic perspectives* **9:85-110**.

Hewitt-Dundas, N. and S. Roper. 2009. Output Additionality of Public Support for Innovation: Evidence for Irish Manufacturing Plants. *European Planning Studies* **18:107-122**.

Hewitt-Dundas, N. and S. Roper. 2011. Creating advantage in peripheral regions: the role of publicly funded R&D centres. *Research Policy* **40:832-841**.

Hoffman, K., Milady, P., Bessant, J., Perren, L 1998. Small firms R&D, technology and innovation in the UK: a literature review. *Technovation* **18:39-55**.

Holland, P. 1986. Statistics and causal inference *Journal of the American Statistical Association***:945-960**.

Imbens, G.W. and J.M. Wooldridge. 2009. Recent Developments in the Econometrics of Program Evaluation. *Journal Of Economic Literature* **47:5-86**.

Jackson, A. 2001. An evaluation of evaluation: problems with performance measurement in small business loan and grant schemes. *Progress in Planning* **55:1-64**.

Katz, J. 1983. A Theory of Qualitative Methodology: The Social System of Analytical Fieldwork. In *Contemporary Field Research: A Collection of Readings*, ed. R. Emerson. Boston: Little Brown Company.

Lambrecht, J. and F. Pirnay. 2005. An evaluation of public support measures for private external consultancies to SMEs in the Walloon region of Belgium. *Entrepreneurship And Regional Development* **17:89-108**.

Laredo, P. 1997. Evaluation in France: A decade of experience. In *Policy evaluation in innovation and technology: towards best practices*, ed. OECD. Paris.

Lerner, J. 1999. The government as venture capitalist: the long-run impact of the SBIR program. *Journal of Business* **72:285-318**.

Ludwig, J.; J. Kling; and S. Mullainathan. 2011. Title. NBER working paper 17062.

Meuleman, M. and W. De Maeseneire. 2012. Do R&D subsidies affect SMEs' access to external financing? *Research Policy* **41:580-591**.

Muller, K.; C. Rammer; and J. Truby. 2009. The Role of Creative Industries in Innovation. Innovation. *Innovation: Management, Policy & Practice* **11:148-168**.

Nightingale, P. and A. Coad. 2014. Muppets and gazelles: political and methodological biases in entrepreneurship research. *Industrial and Corporate Change* **23:113-143**.

OECD. 2006. Evaluating Government Financing of Business R&D: Measuring Behavioural Additionality – Introduction and Synthesis, ed. DSTI/STP. Sydney.

Potter, J. and D. Storey. 2007. OECD Framework for the Evaluation of SME and Entrepreneurship Policy and Programmes. Paris OECD.

Potts, J. and K. Morrison. 2009. Title. NESTA Working Paper NESTA, Lndon.

Powell, W.W. 1998. Learning from Collaboration: Knowledge and Networks in the Biotechnology and Pharmaceutical Industries. *California Management Review* **40:228-240**.

Reiner, C. 2011. Evaluating innovation policies by chance? The case for randomised R&D-programme evaluation. *Plattform - Forschungs- und Technologieevaluierung* **37:15-28**.

Roper, S. 2013. Gearing up for growth vouchers. Warwick Business School: Enterprise Research Centre

Roper, S. and S. Arvanitis. 2012. From knowledge to added value: A comparative, panel-data analysis of the innovation value chain in Irish and Swiss manufacturing firms. *Research Policy* **41:1093-1106**.

Rosenbusch, N.; J. Brinckmann; and A. Bausch. 2011. Is innovation always beneficial? A meta-analysis of the relationship between innovation and performance in SMEs. *Journal of Business Venturing* **26:441-457**.

Stam, E.; J.P. de Jong; and G. Marlet. 2008. Creative Industries in the Netherlands: Structure, development, innovativeness and the effects on urban growth. *Geografiska Annaler. Series B, Human Geography* **90:119-132**.

Strauss, A. and J. Corbin. 1990. *Basics of Qualitative Research:Grounded Theory Procedures and Techniques*. Newbury Park, Calilifornia Sage.

Takalo, T. and T. Tanayama. 2010. Adverse selection and financing of innovation: is there a need for R&D subsidies? *Journal of Technology Transfer* **35:16-41**.

Tashakorri, A. and C. Teddie. 1998. *Mixed methodology: Combining qualitative and qauntitative approaches*. Thousand Oaks,CA Sage.

van de Vrande, V.; J.P.J. de Jong; W. Vanhaverbeke; and M. de Rochemont. 2009. Open innovation in SMEs: trends, motives and management challenges. *Technovation* **29:423–437**.

White, H. 2008. Of probits and participation: The use of mixed methods in quantitative impact evaluation. *Ids Bulletin-Institute Of Development Studies* **39:98-+**.

Yin, R. 2003. *Case Study Research Design and Methods*. California: Sage Publications Inc.