Evaluation under contract. Government pressure and the production of policy research[†].

The LSE GV314 Group[*]

Abstract

Do governments lean on researchers who evaluate their policies to try to get them to produce politically useful results? Do researchers buckle under such pressure? This paper, based on a survey of 205 academics who have recently completed commissioned research for government, looks at the degree to which British government departments seek to produce research that is designed to provide "political ammunition", above all making them "look good" or minimizing criticism of their policies. Looking at different stages in the research process – from deciding which policies to evaluate, shaping the nature and conduct of inquiry and writing the results – the paper finds evidence of government sponsors making significant efforts to produce politically congenial results. The reaction of the researchers appears to be that of more or less successful resistance to these efforts, though the evidence base (researchers own accounts of their work) suggests that this conclusion be treated with some caution.

[†] We are grateful to Philip Cowley (Nottingham), Tony Travers (LSE) and Tim Newburn (LSE) for their help on this project. We are also indebted to the anonymous referees for their useful comments on earlier versions of this paper.

[*]The LSE GV314 Group consists of staff and students in the Department of Government at the London School of Economics and Political Science following the undergraduate course 'Empirical Research in Government' (course code GV314). For this project it included Taha Abdel Dayem, Hera Barbullushi, George Barnes, Christopher Cubitt, Joseph Hayward, Rui Jin, Jonathan Kay, Annabel Lee, Edward C Page, Nikesh Patel, Hannah Payne, Joana Pelicano Santos, Lucy Perry and Athina Valyraki.

Political pressure and government research.

Do governments lean on researchers who evaluate their policies to try to get them to produce politically useful results? Governments might be expected to seek to generate good publicity or at least avoid bad publicity from evaluations. Even away from the glare of mass publicity, apparent endorsement or challenge by a scientific evaluation can make a policy or programme substantially stronger or more vulnerable to its opponents or competitors. As Bovens, t'Hart and Kuipers (2006: 321) write, "[i]t is only a slight exaggeration to say, paraphrasing Clausewitz, that policy evaluation is nothing but the continuation of politics by other means". Yet governments might also be expected to refrain from leaning on their researchers. Evaluations provide evidence and insights on which the development of current and future policies may be based, and producing inaccurate or biased research defeats the whole object of evaluation. Moreover, professional social researchers might resist such pressures and any attempt to exert them could, if made public, produce even worse publicity than a negative evaluation itself.

For their part, researchers might be expected to succumb to being leaned on, whether through enthusiastic acceptance of government constraints and cues or through acquiescence. Government contracts offer, if not always substantial personal enrichment, career benefits, access to research material and the prestige that comes of having one's research credentials endorsed by government. By the same virtue, researchers might even give their sponsor what they think it wants to hear even without it asking. Alternatively researchers might resist such pressures. Academic reputations are built on independence and can be destroyed by evidence or suspicions that their expert judgment can be bought such that researchers might be expected to

defend their integrity at all costs.

Even if government “leans”, does this matter? Government, like any organization, is bound to be involved closely in the policy research it purchases and its intervention is to be expected, especially if the researchers show any signs of missing what it sees as the point of the evaluation exercise (see Weiss 1970: 62). Moreover, as Weiss (1970: 58) suggests, "evaluation has always had explicitly political overtones. It is designed to yield conclusions about the worth of programs and, in so doing, is intended to affect the allocation of resources". By their nature, government-commissioned evaluations are in this sense political (see also Chelimsky 2007). For many, even the selective use of research to support a "predetermined position" reflects a "worthy" political model of research utilization, as long as the evidence is made available to all (Weiss 1979: 429; see also Davies, Nutley and Walter 2010: 205; Powell 2011: 15-16). One may well wish or need to be sanguine about the value of the "political model" of research utilization (see Bulmer 1982: 157 and Beyer and Trice 1982: 600 for notes of dissent). However, judgments about propriety and legitimacy aside, if evidence is generated under pressure from sponsors intending to produce results that suit them, then the nature of this pressure and the responses of those facing it are at the very least relevant to our assessment of the character of that evidence.

Despite the interest in “evidence based policy”, and the importance of political values in the use of evidence (see, for example, MacGregor 2013; Boswell 2008; Hope 2004; Fischer 2003; Sabatier 1978) there has been rather scant attention devoted to understanding the impact of political constraints on the production of research under contract for government. The question of the effect that “policy relevant” research

might have on reshaping academic priorities across whole disciplines has been raised in a range of different contexts (see, for example, Manzi and Smith-Bowers 2010; Walters 2007; Allen 2005; Hillyard, Sim, Tombs and Whyte 2004; Slaughter and Leslie 1997; Kogan and Henkel 1992; Cohen 1988: 23). However, the degree to which specific research conclusions are shaped by the process of commissioning, and the consequent interaction between government and researchers in its production, has been left largely to discussion of one case or few cases (see O'Brien et al 2008; Allen 2005; Hope 2004; for exceptions see Salisbury *et al* 2010; Walker 2001). This paper explores the question of political pressure in the production of policy research on the basis of an empirical study of academic researchers who have recently conducted evaluations of public policy for government in the UK.

This paper explores what kind of influence government can have on shaping the research it commissions. We first set out the broad strategy for assessing government attempts to steer research towards producing politically favourable results. We examine government-researcher interactions at three distinct stages in the production of policy evaluations: research commissioning, managing and report drafting. We then go on to outline our data and use it to look at the different sources of political pressure within government. In the discussion and conclusion we explore the implications of our findings, above all from the perspective of whether it matters much whether government leans on its researchers and on the consequences for character of evidence based policy.

Leaning, political advocacy and evaluation

Government may either lean on researchers, that is, it can exert pressure on them to provide politically useful reports, or it may not. The researcher's reaction to any pressure might be to buckle under the pressure and produce the kind of politically supportive report the government wants or it might stand firm against such pressure. With these two variables, leaning/not leaning and buckling/not buckling, we can derive four broad types of relationship. Where government leans and researchers buckle we have a "servant/master relationship" (Manzi and Smith-Bowers 2010: 143), a relationship of servitude where "strong control [is] exercised by ministers and civil servants due to concerns about the political implications of the findings". Researchers could, of course, also buckle without government leaning on them directly, producing a second type of relationship which we can term "docility". For example, Allen's confessional (2005: 1004) invokes Foucault to show how "my constant exposure to the disciplinary power of research funding bodies, since 1993, and my submission to its 'inspecting gaze', moulded me into a 'docile' researcher intuitively oriented to producing satisfied funders". A third type of relationship, where government leans but researchers do not buckle, can be described as "resistance". A fourth kind is where government does not lean and researchers do not buckle – "both parties have equivalent objectives and are committed to the goal of truth seeking" (Manzi and Smith-Bowers 2010: 141) – which we may term "collegiality". As we will see, it is not quite so easy to fit the empirical reality into four categories as it is to derive a two-by-two table, yet the typology will provide us with a set of expectations about what we might find in the empirical research and a terminology for discussing the results.

How do we tell whether government is leaning or exercising reasonable diligence and support in supervising research? An operational distinction might be based on

whether government officials and/or politicians make intentional efforts to shape the conduct of the research so as to increase the chances that it reaches conclusions useful to them politically; efforts intended to provide good "political ammunition" (Bulmer 1982: 155). As a *mens rea* issue of intention it becomes very difficult to investigate. Not only is it impossible to see inside the heads of others, there are also many individuals involved in commissioning and managing research and they may differ in their inclination to lean and in the particular objectives they seek in so doing. They may decide to lean at one stage in the process and not at another; they may be cross-pressured, undecided or even just confused about exactly what they want from the research (see Salisbury et al 2010: 119). Our approach to the question of the degree to which the government leans on researchers by pursuing "political ammunition" objectives as opposed to "scientific relevance and quality" objectives is based on a set of expectations about how government handles its relationship with research and researchers at different stages in the development and management of a project. Some types of government behaviour are, we argue, are more consistent with the intention to produce political ammunition, others are more consistent with the pursuit of scientific relevance and quality.

In this paper we divide the process into three stages, effectively chronological, at which we expect to find evidence of political ammunition or scientific relevance and quality objectives. In a first stage, commissioning and designing the research, we would expect the quest for political ammunition to be reflected in governments disproportionately selecting topics for evaluation and designing research questions which have the potential for making the government look good and its opponents bad (see, for example, Ham, Hunter and Robinson 1995: 71), and choosing evaluators

more likely to provide favourable results. In a second stage, managing the research, one would expect a government as client pursuing scientific relevance and quality objectives to orient its continuing relationship with the researchers to the provision of advice and guidance in maintaining the policy focus of the research concerned and providing data and support in gathering evidence. At this stage the pursuit of political ammunition would lead sponsors to try and steer the research so that it reaches desired conclusions, usually positive (or at least not negative), about the policy. In a third stage, writing up, we would expect direct attempts to make the conclusions of the research more positive than they would otherwise have been as evidence of political ammunition objectives.

How do we tell whether researchers buckle or not? Our evidence on this issue is of two types. First there is direct evidence derived from specific questions about respondents' reactions to pressures at different stages in the production process from those that commissioned them. Second, we look at the degree to which the conclusions of research were critical of government as evidence that researchers were following or resisting government wishes. Norris (1995: 274) writes "Just as civil servants can never be seen to be critical of ministers or government policy, researchers funded by government risk similar strictures. Criticism is a problem". Even if one does not accept that government is quite so intolerant of criticism, or that critical reports, say of predecessors or rivals, cannot constitute welcome political ammunition, on the whole we might still expect the degree to which a report remains critical of policy despite government leaning to be a broad indicator of resisting pressure to produce political ammunition.

The study

Empirical data

Our analysis is based on a survey of academic researchers who have been involved in commissioned research. We examined all the reports on government-commissioned research from 2005 onwards we could find online in autumn 2011. These included reports on evaluating schemes for providing "handypersons" to elderly and disabled people (DCLG 2011), an "action research" evaluation of a new system of managing criminal offenders (PA Consulting 2005) and a quantitative evaluation of new methods aimed at dealing with fraud in Jobseekers' Allowance, a major unemployment benefit (Middlemass 2006). We noted the names of the authors of these 215 reports, where named. We included only academic researchers working for a university or an organisation based in a university for practical reasons. In our early approaches we received no replies from researchers working for commercial organisations; almost all academics replied to our emails. At this early stage before the questionnaire was drawn up we interviewed 22 researchers. Academics were authors or co-authors of 95 of the 215 reports and these yielded 251 unique names for which we could find valid email addresses. We supplemented this list with an additional 106 names of academic researchers whose university websites indicated they had worked on government commissioned evaluations since 2005. After two reminders we received 204 questionnaire responses from the 357 emails that were successfully sent, indicating a response rate of 57 per cent. Exactly half of our respondents chose to write comments in the space we left for the purpose at the end of the questionnaire, writing on average 94 words. These will be referred to as "free comments" when quoted.

The questionnaire was completed under guarantee of anonymity. This restricted some of the questions we could reasonably ask. Because some departments had only few reports we could not look at variation across government departments without threatening anonymity. One free comment gets across the broad importance that many respondents and interviewees attach to departmental variation:

> Government departments vary considerably … ... I have very good experiences of Department W in commissioning research but much poorer experiences with respect to Department X, Department Y and Department Z (the latter by repute - I wouldn't touch them with a bargepole personally).

For similar reasons we could not look at the precise nature of the topic being researched even though we might expect that some evaluations are politically more sensitive than others and might attract government leaning more than less sensitive issues.

Our study was conducted at a time in which we caught many projects that were commissioned under a Labour Government and reported when the Conservative-Liberal Democrat Coalition was in office. Some of the free comments argued that things had deteriorated under the Coalition. For instance one wrote that her/his most recent report, commissioned under Labour but completed under the Coalition, "was 'buried' along with others in historical research that was effectively disowned on the Departmental website. ... Since the change of government, in [the Department I often work for] at least, it appears that there is no longer any interest in commissioning research or developing an evidence base".

We do not know how many projects were started and finished under the Coalition as we only asked about the year in which the research was completed. However, we know that 56 respondents referred to research completed under Labour before 2010, 61 to research completed in 2010 and thus likely to have been commissioned by Labour, and we expect that a large proportion of the 85 respondents whose research was completed in 2011 and 2012 were also commissioned under Labour. However, there were no noticeable differences that approached statistical significance in the responses for any of the questions based on these three time periods.

Commissioning and designing research

Does government play it safe with commissioning?

It is difficult to offer a clear picture of the degree to which government evaluation research tends to concentrate on "safe" topics, i.e. those less likely to produce results that will embarrass it. According to the most direct evidence on selectivity in commissioning from the survey, government does appear in general play it safe: 31 per cent of respondents agreed with the statement that "Government is often prepared to take risks in commissioning research that might produce results that highlight shortcomings in policies and programmes" while 49 per cent disagreed and 20 per cent neither agreed nor disagreed (N=185). A larger proportion of respondents felt government played it safe.

Some of our free comments highlighted wider political goals in the selection of research topics. For example, one respondent wrote that "Evaluation tends to be funded to support rhetorical assumptions, and findings are buried if the assumptions

are not supported". However, in the selection of topics for commissioned research we should not necessarily conclude that playing safe is a particularly dominant feature of commissioning behaviour. While more respondents believed government played it safe than did not, the margin (at 49 per cent to 31 per cent) was not overwhelming. Moreover very few academics (10 per cent) "disagreed strongly" that government was prepared to take risks. This was much larger than the 2 per cent who "agreed strongly", but is still rather more suggestive of a clear tendency to caution in selecting topics than an attempt to cook the books to provide political ammunition. Many programmes, especially large ones, have to be evaluated as part of the deal (above all with the Treasury) which led to their development.

Other aspects of the commissioning behaviour of government departments caution against the view that the quest for commissioning favourable evaluations is a dominant concern in deciding which policies to evaluate. Of course, to offer a satisfactory answer to our question we would need some tally of the total number of policies eligible for evaluation and compare it with what was actually evaluated. In the absence of any such statistics we must rely on more indirect measures.

Government, according to our coding of the 215 recent research reports, unsurprisingly tended to focus its attention on recent policy innovations: 27 per cent were evaluations of pilot projects and 44 per cent evaluated policies that had been in existence for less than three years, with a further 27 per cent being studies of older programmes or general surveys of policy areas. So insofar as taking risks is concerned, the evidence is that governments evaluate policies that are likely to reflect on their own performance, not that of a predecessor. Moreover government did not

shy away from commissioning research that gave judgments on policy success. We examined whether policies were summative (i.e. formed judgements about the overall impact of research) or formative (looking at ways of improving the way the policy was being implemented or delivered). Only a minority of reports, 22 per cent, confined themselves to formative goals (sometimes termed a "process evaluation"), 46 per cent were summative and a further 30 per cent mixed summative and formative evaluations. This is broadly consistent with Salisbury *et al's* findings (2010: 25-6), although it is important to note their suggestion that *intentions* to provide summative evaluations may not be matched by the research design and form of evidence used to develop them. Thus the quality of evidence may prevent a study intending to be summative from adequately filling that objective (see also Salisbury et al 2011: 222-3).

[Table 1 Here]

Were academics more likely to produce reports critical of government than anyone else? Existing evidence rather suggests so (see also Salisbury et al 2010: 109). Our coding of the available reports since 2005 agrees with this. It suggests they are not much more likely to be critical, though the numbers involved are too small to be statistically significant. Using the data from our coded reports database, academics (at 52 per cent) were marginally less likely to produce a positive report than average (54 per cent, Table 1). While reports produced in house, i.e. by researchers employed by government organisations, tend to be the most favourable (62 per cent positive), governments do not use them to produce published reports as frequently as the

slightly more critical private consultants (58 per cent positive) or even more critical academics (52 per cent supportive).

Does government use the research specification to produce positive reports?

Governments might try and skew the research to produce favourable reports by defining carefully the research questions to be pursued and the methods used to pursue them. One does not have to look very far in the free comments and interviews to learn find such evidence of the pursuit of political ammunition objectives. One respondent wrote:

> More often however the real place where research is politically managed is in the selection of topics/areas to be researched and then in the detailed specification. It is there that they control the kinds of questions that are to be asked. This gives plenty of opportunity to avoid difficult results.

Another pointed out:

> I have worked on two projects from two different government departments and the problems were very similar. In both cases, we were doing an experiment and the level of interference was high in the design of the experiment, such that in the end the answer for one of the studies was not worth having as we were not allowed to compare the government supported intervention against a realistic control group. We had to fight continually to maintain the integrity of the research design. Government needs to be hands off when it funds research projects.

A third suggested

> I was shocked at the level of interference of civil servants at certain points in the progress of the research. Specifically, they intervened at the sampling stage, changing entirely the case-study sample … [It]meant that those most sympathetic to government values and most closely aligned with the thinking behind the policy were more likely to be represented.

One interviewee offered another specific example "The sample … was not random but involved targeting various groups ..[believed to] have benefited most from the scheme". Another was very explicit about the kind of direction received: "We were told by the [department] to evaluate how cost effective the project [was]… [W]e were pretty much told at the beginning the purpose of the report was to show that the programme was cost effective, i.e. we had to try to show it in a good light".

When we asked how "the initial research questions you were to address in the project developed", 45 per cent (N=190) answered that "the government organisation(s) had a clear idea of the precise questions the research should examine", 27 per cent suggested that "the government organisation(s) had a broad idea of what the research should examine and left the definition of the precise research questions to me/us" and 26 per cent that the "precise research questions were developed jointly by me/us and the government organisation(s)" (2 per cent did not know).

Thus, nearly half the time, government gave clear instructions over the research questions. However, such instructions might have been for scientific quality and relevance rather than political ammunition objectives. Of those writing reports that were based on questions devised by researchers alone (i.e. the questions were "left to us", N=43), only 23 per cent were supportive of policy compared with 50 per cent of

those writing reports (N=80) where governments set the research questions. This suggests that where the evaluation project is specified in some detail by government it is more than twice as likely to be supportive of the policy or programme than where academics set the research questions. Where researchers worked together with government to devise the research questions (N=44) 57 per cent of them produced supportive reports. When government was involved in developing research questions, whether alone or in conjunction with researchers, their reports were substantially less likely to be critical, suggesting some success in pursuing political ammunition objectives.

Budgetary constraints can also serve to support political ammunition objectives in commissioned research. Salisbury et al (2010; 2011) show how the constraints set by research design features specified by commissioning departments -- the budgets, the timelines as well as the specification of the methods to be used -- can prevent the generation of clear judgments of how well or badly a policy is working. Instead, for example, of the research design allowing some form of "control group/experimental group" comparison that might be expected to determine with some precision the effects of a programme, they show how service provider comments and general indicators of satisfaction are mandated -- these generally tend to give more flattering accounts of the impact of policies than harder evaluative techniques such as Randomised Control Trials. We could not examine this feature of the commissioning behaviour of government directly, but our evidence suggests it is important. When we asked our respondents whether they agreed with the statement "Keeping costs down plays too large a role when government organisations are deciding what type of

research they want", 47 per cent agreed (including 16 per cent who "strongly agreed") while 23 per cent disagreed (including 2 per cent who "strongly disagreed").

The Management of Research

When research is commissioned by the department, an official or officials from the department is/are usually given responsibility for managing the research -- making sure that it is progressing, dealing with any questions that arise, handling data provision and access to other government organisations among other things. A handful of free comments and passages in our interviews indicated the potential for the imposition of political objectives at this stage. In a free comment, for instance, one respondent said that the experience of working for government "is not one I would wish to repeat. Civil servants kept a very close eye on the research and the research process and there was a real risk that academic freedom would be compromised as a result". An interviewee complained: "The liaison with the [departmental] team that was managing the programme … was difficult. The evaluation was micro-managed despite a lack of research knowledge in the [departmental] project management team".

However, while examples of this (possibly political) micro-management by officials can be found, our evidence does not suggest it is particularly frequently experienced by academic researchers. Certainly officials managing research tend, according to our academic respondents, to show significant interest in the work of their researchers; 60 per cent (N=185) agreed that "officials in government organisations responsible for monitoring research generally maintain a high level of interest in it" while only 16 per cent disagreed (24 per cent neither agreed nor disagreed). A

substantial contributor to perceptions of lack of interest might be the common view that "civil servants move from one post to another so often that there is too little continuity in its management of research projects" with which 59 per cent (N=184) agreed and 20 per cent disagreed (21 per cent neither disagreed nor agreed, see also Salisbury et al 2011: 222). Respondents who reported a lack of interest were significantly more likely (70 per cent, $p<.05$ using a Chi Square test) than others (51 per cent) to agree that civil servants moved around too much.

However, interest in research cannot be equated with micro-management, still less that directed towards providing political ammunition for government. When asked specific questions about official involvement in their most recent evaluation report, respondents did not present a picture of officialdom as a channel for political steering of their research during its progress. 17 per cent of respondents (N=190) felt there was "too much contact" in the course of the project[1], 4 per cent that there was too little but the large majority, 73 per cent, felt the amount of contact was "just right". Moreover the fact that the amount of contact did not appear to be substantially related to how positive the report produced turned out to be (40 per cent of those with "too much" contact produced broadly supportive reports compared with 48 per cent of those with "just right" amounts of contact, a statistically insignificant result) suggests that routine supervision of research does not appear to be accompanied by successful pressure to produce politically acceptable results. We will come back to this question later since, as we shall see, we have reason to believe that those officials with the

[1] The question we asked was "Thinking of all the contact you had with officials in the commissioning organisation over the entire project, face-to-face as well as through phone, mail, email and other electronic forms, would you say that there was: Too much contact; Too little contact; About the right amount of contact".

more routine contacts with researchers might be less likely to press political ammunition over scientific quality and relevance objectives on researchers.

Drafting the Report

It is conventional for researchers to discuss a draft of a report with the client before submitting it for publication. At this stage we might expect the bureaucrats supervising the research to seek to change or scale down critical content within it. When we asked whether the government sought to make substantial changes to the draft report[2], 46 per cent said they were not asked to make any changes, 33 per cent were asked to make "one or two" changes and 19 per cent more than two substantial changes. This suggests that we might expect government to propose changes that affect the interpretation of findings or the weight given to them slightly more often than not.

The free responses in the survey as well as the interview material gives ample evidence of those sponsoring the research seeking to shape the way the results are reported. If one were to judge solely on the basis of frequency of mentions of such attempted influence, this would be the single most important stage in the transaction between departments and researchers at which one would find an attempt to sustain political goals in commissioned research. A few quotes will give a flavour of the comments we received:

---

[2] The wording of the question was "When you were finalising your report, did the commissioning organisation(s) ask you to make any substantial changes to your findings? *By '"substantial" is meant any changes that affect the interpretation of your findings or the weight that you might give them*"

There were some requests for wording changes to the report so that it would be in keeping with the political agendas.

There was a huge pressure to spin findings to fit their agenda.

There was a lot of dialogue back and forth at the end between us and the Department … before it was published to ensure they did not look bad. They wanted certain wording changed so that it was most beneficial from a PR and marketing point of view; and they wanted certain things emphasised and certain things omitted.

The individual who had demanded (despite our advice) [that we provide a particular set of figures because he felt it would make the programme look better] then wanted this analysis to be removed from both the Executive Summary and the main report - he argued that this would not put the programme in the 'best-light'. As I had been asked to provide this [analysis] four days before the final report (and had already been pulling 120 hour weeks), I refused. However, the whole [of the Department's] project management team continued to send emails and make phone calls (approximately four or five a day) demanding its removal. In the end, this data was removed from the Executive Summary ...

And the people working with the Departments will come and change the wording of your results without explanation; the secretary/assistant of the

> Department would edit and remove bits and pieces…. "the minister does [not] want this or that"….

Moreover, our survey data suggests that when asked to make some changes to the final report the academics tend either to oblige or meet their sponsors half way. When we asked what their reaction to the requests were, of those asked to make changes (N=89) only 5 per cent made no changes and 20 per cent simply made the changes asked for, 74 per cent reached a compromise with the government sponsors (1 per cent did not know).

There is plenty of evidence that government asks its researchers to make changes, but less evidence that this helps produce supportive reports. The handful of researchers who were asked to make changes and simply made them were less likely to produce positive reports (15 per cent) than those not asked to make any changes at all (46 per cent, see Table 2). Certainly the researchers who were asked to make changes and did so on the basis of a compromise with government were the most likely (52 per cent) to produce positive results, but only marginally so, and of the four respondents who refused to make any changes, two produced supportive reports and two did not make any changes in response to government requests were no more likely to produce supportive reports than those who compromised. None of these findings approach statistical significance.

[TABLE 2 HERE]

These survey findings are somewhat puzzling and at first appearance entirely at odds with the evidence provided by the free comments and interviews of powerful

government editing and sanitising research findings. The free comments threw up many examples of attempts to change the way the research report was produced. Most of them, however, also went on to suggest, broadly, that they gave in on the cosmetic changes and held firm on the substantial. As one respondent put it: "There were some requests for wording changes to the report so that it would be in keeping with the political agendas - if this went contrary to the data, we did not make the changes. When suggested wording was not contrary to the data, we did make changes".

Forms of "government" influence

We have been talking about "government" as a whole as a conduit for asserting political ammunition objectives in commissioned research. Several respondents in interviews and free comments made distinctions between different audiences for the research within government. There are at least four distinct groups within the sponsoring department that potentially have different relationships with researchers and different interests in pursuing political ammunition or programme evaluation objectives. Not all four groups will be involved in every research project, but they are nevertheless worth setting out. First there are the officials who are responsible for research, possibly because they are researchers themselves. Where mentioned, these seem to have the best relationships with researchers and appear most likely to share a belief in the importance of programme evaluation objectives in research. As one suggested

> A key distinction in my experience is between commissioners of research and their policy counterparts. It's the latter who are often the trickier to handle, whilst the former sometimes even see themselves as protecting research

> integrity against the demands of the policy people. This was certainly my experience of doing work … in a politically contentious area.

The second group, as this quote suggests, are the "policy people" -- officials with the task of looking after policy within the department, whether to amend, defend, expand or contract it. Usually these are middle ranking officials but such policy work can include senior ranks (see Page and Jenkins 2005). Another survey respondent makes a similar point about the political sensitivity of the policy officials: "the research manager places a lot of emphasis on research integrity, whereas the policy teams may have their own ideological or policy motives".

The third group is the ministerial political leadership, already discussed in this paper. Our survey evidence suggests unsurprisingly that they are highly likely, in the view of our respondents, to downvalue programme evaluation objectives of research as only 4 per cent of respondents (N=182) agreed with the proposition that "Ministers are prepared to act on evidence provided by social research even when it runs counter to their views" (71 per cent disagreed and 25 per cent neither agreed nor disagreed). This is consistent with one written comment that was also reflected in others: "my area is a sensitive one, and civil servants generally are interested in research results whereas politicians have a keen eye on how the public will respond (especially The Daily Mail)". One researcher described his perception of the civil servants being on the side of the researchers against the minister: "often our results are critical of policy interventions but I find civil servants open and accepting of the findings even while they are having to think how to square the results with ministerial expectations".

A fourth group is that provided by the professionals and service providers in the programmes being evaluated. These can be brought into contact with researchers more or less directly. Less directly their views might be sought on an *ad hoc* basis as the research develops. A particularly direct incorporation of such officials in commissioned research is through their inclusion as "stakeholders" or "experts" in steering groups. Several respondents and interviewees mentioned the role of service providers as a source of constraint on their research findings. One argued

> We met regularly with the Head of Research in the [Department] and also occasionally with their policymaking colleagues. One difficulty with these meetings was that they insisted representatives of the [organisation running programme being evaluated] attended. This made it quite difficult to discuss the report openly because these peoples' livelihoods depended on the scheme. My part of the report was critical of the [programme] and I thought it inappropriate for the [department] to invite these people along. I felt it hindered honest and open discussion.

Steering group involvement was particularly awkward for one respondent

> The steering groups often set up to 'steer the research' usually have their own agendas and are usually upset when for example, … the research is evaluating a policy that steering group members may have some role in, and the findings appear to challenge the policy being implemented. This can make for difficult steering group meetings where the research team is often forced to justify their credibility and the rigour of the study, and also the findings which they invariably question (if they do not confirm their own findings) when they read the draft reports. Finally, in writing up the research findings you sometimes feel

that you have no say as so much is questioned and/or you are asked to rephrase particular sentences so as to put things in a more positive light.

The importance of this distinction between the four groups is that each has a different set of interests that they may be expected to pursue. If, as Salisbury *et al* (2011: 221) suggest, the original objectives behind conducting any particular evaluation are often no more precise than the rhetorical commitment to evaluation being a "good thing" to do, we might expect the precise objectives government officials pursue in their transactions with the researchers to be diverse.

Discussion

Can we combine our evidence on "leaning" and "buckling" to produce indicators of the degree to which the relationships between government and academic researchers suggest servitude, docility, collegiality and resistance, the categories discussed at the outset? We might take as our overall indicator of compliance with government wishes the degree to which respondents claim their reports were supportive of government policy or not. As our measure of government pressure we may use answers to some of the key questions suggestive of pressure at different stages in the production of research: whether government was involved in developing the research questions, had too much contact with researchers in the development of the project and asked for changes to be made in the draft. Around one in six respondents (30 out of 173) reported none of these pressures from government. If we relax our criteria and consider respondents who report one or fewer such pressures as facing lower pressure and those with two or more as facing higher pressure (each group 50 per cent of the 173) we might classify our respondents in a two-by-two table similar to the typology of docility, collegiality, servitude and resistance. The results are set out in

table 2; of the 50 per cent of respondents who experienced lower government pressure, 21 per cent produced supportive reports while 29 per cent did not; of the 50 per cent of respondents who experienced pressure, an equal proportion produced supportive as unsupportive results. While the unpressured critics were a marginally larger group than any of the others, Table 3 hardly suggests that overall "leaning" on researchers produces supportive reports -- the (Chi Square) relationship between the two variables in table 3 is marginal and does not approach statistical significance (p=0.34).

[Table 3 Here]

If we were to do a direct read across from our figures to the two-by-two table, this would place 21 per cent of researchers as suggesting their relationship was one of docility, 29 per cent as collegiality, 25 per cent as in a position of servitude and 25 per cent as resistant[3]. While this would be a wholly misleading conclusion, it is worth considering why. One central assumption, among others, the equation makes is that researchers could not produce supportive results unless they had been leant on by government or were willing to do anything to please their funders. This is plausible only if one believes that all government policies are bad and that proper scientific enquiry would reveal this. Some researchers appear to come close to making this assumption: Huitema et al (2010: 188) suggest in their European cross national study that climate change “policy evaluations [that] questioned official policy goals" could be categorized as "reflexive” and those that did not “unreflexive”. However such an

[3] If we were to classify the level of pressure as low only where none of the three main forms of pressure was evident then the percentages would be servitude 42 per cent, resistant 42 per cent, docile 12 per cent and collegial 4 per cent; the equivalent Chi square statistic for table 3 is also insignificant (p=0.25)

assumption is unwarranted as we would expect some level of support for policy irrespective of any government pressure. While it may be reasonable to take as an indicator of successful government pressure the degree to which it produces reports less critical of government policy than might otherwise have been expected, it is not reasonable to expect all reports that support government policy do so as a result of servitude or docility.

So where does this leave us with our four categories of relationship? Here the limits of our evidence in casting light on the *mens rea* issues of researcher and government intentions become apparent. Nevertheless, the pattern of our results suggests we have not entirely drawn a blank here. Working backwards through our stages, the writing up process seems to be characterised above all by resistance in some form -- while four respondents asked to make changes made none, three-quarters compromised, and the fact that the compromises made did not end up producing more reports supportive of government, as discussed above, suggests significant resistance was the most common reaction. In the management of projects stage the low chances of being leaned on suggests either docility or collegiality. The signs of docility are not strong, however. If docility were important we might expect experienced researchers to be less likely to be leaned on as they become trusted to produce reports supportive of government policy. Indeed, this is even the case to some degree; 84 per cent of the 103 most experienced researchers with over five government contracts under their belt were left alone to get on with the research (i.e. there was not "too much" contact with government at this stage) compared with 71 per cent of those 21 respondents for whom the research was the first government contract, and 80 per cent for moderately experienced researchers (between two and five contracts). This relationship is,

however, not statistically significant, and the fact that government left nearly three quarters of the novices largely to their own devices appears to offer stronger evidence for the view that the relationship is predominantly collegial at this stage.

The issue of whether government leaning skews research results through producing servitude or docility seems to focus strongly on the very early stage of the research process. The statistical relationship between researchers being left alone by government to develop the research specification and eventually producing a report supportive of government policy is negative and significant at the $p<.001$ level. After this stage the evidence of "leaning" making much difference is small. Our best indicator of docility discussed above (based on the expectation that those who have done more commissioned research might be expected to be more docile) has found little evidence of researchers being eager to please government by writing positive things about their policies. This points to a significant degree of servitude at the early stage in the relationship with government, and at no other.

Conclusion

There is sufficient evidence here to suggest that governments do lean on researchers. This is not a great leap forward in our knowledge on its own, not least because it surprises nobody. We have been able to pin down a range of mechanisms through which governments might seek to influence the outcome of evaluations for what might be described as the provision of "political ammunition", some obvious, some less so, and to assess their importance. Most of these mechanisms appear to have little systematic importance in shaping the nature of the conclusions that researchers reach. The most effective constraint appears to be found when government specifies

the nature of the research to be done at the outset. No other form of constraint has as powerful an effect on the degree to which the overall conclusions the researchers reach support government policy.

We did not directly test or examine the relative contribution to the pursuit of political ammunition in different bits of "government": policy officials, research managers, officials responsible for services under investigation and politicians. Yet the pattern of government involvement is at least consistent with the notion that these groups become more involved in the research process at some stages than others; with those expected to be more sensitive to the political ammunition aspects of the research, above all the policy officials and politicians, taking part in the design of research questions at the beginning and the writing up and reporting at the end and those more committed to the programme evaluation, above all researchers and research managers, in the conduct of the research once under way. This is similar to the findings of in respect of the use of ex ante policy evaluations (see Hertin et al 2009: 14).

Are such ammunition gathering objectives pursued strongly enough to undermine the value of government-sponsored research as scientific evidence? One view might be that any intervention by government threatens the scientific value of research (Davis and Portis 1982), whereas another is that only outright falsification of results renders it valueless as evidence (see Weiss 1979: 429). Head (2010: 87) believes that "politicized context" behind the "realities of research funding" steer research towards evaluations likely to be supportive of political priorities, and this is one of the most important "systematic obstacles to the wider impact of EPB [Evidence Based Policy] approaches". The threat to the integrity of research and researchers does not, from our

evidence appear to be likely to be of such a scale to justify Head's conclusion here. The academics' ability to resist pressure to steer the results appears to be substantial.

However, we have to bear in mind that this assessment is based on academics' own perceptions. While we have no special reason to think their reported perceptions of government influence and other features associated with commissioned research are erroneous or misleading, one cannot ignore the fact that academic reputations in part rest on the ability to maintain, or at least appear to maintain, scientific standards against government interference. Academics not only have an interest in resisting political pressure, but also an interest in appearing to be able to resist it even if they cannot or do not. Yet the evidence suggests that the whole process of government commissioning and managing research by academics leaves sufficient space for academics to produce critical reports and that governments can and do tolerate and manage such reports without losing face or credibility.

Are our findings capable of extension to other jurisdictions? From the limited available evidence it appears clear that national governments vary substantially in their propensity to commission policy evaluations. Huitema et al (2011) suggest that climate change policy evaluations are conducted less commonly in some European jurisdictions, such as Portugal and Poland, than others such as Germany, and they single out the UK as "taking the lead" in the number of evaluations. In France, Pinson (2011: 204) points to a traditional pattern of frustrated attempts to engage academic researchers in the "intellectual interests and concerns" of bureaucratic elites; instead officials encountered "a social scientific community that was constantly trying to challenge their world views and reformulate the research issues in different ways".

The kinds of evaluations, whether produced for audit institutions, parliaments or other public bodies also appears to vary substantially cross-nationally (see the contributions in Furubo, Rist and Sandahl 2002). Moreover the type of research set-up that characterises much of the evaluation work reported here – "departmental initiatives" where individual ministries or agencies seek to evaluate the performance of their own policies -- appeared to Derlien and Rist (2002: 441-2; 449) as somewhat unusual and they cite the UK, USA, Germany, Denmark, Norway, Canada, Sweden and the Netherlands as examples of countries where evaluations take this form. To such variations we need to add other potential forms of variability such as those arising from the structure of scientific research professions (Molas-Gallart 2012), the consultancy industry as well as variation in institutional structures such as audit institutions and subnational constitutional arrangements. Moreover, a proneness to self-criticism and openness to external scientific criticism, might even be part of a "kinder, gentler" approach to government supposed to be found in some countries rather than others (Lijphart 1999).

While the direct application of our findings is likely to be limited to the UK, it does hold implications for other jurisdictions. The UK findings suggest that there are two main forces that reduce the impact of government "leaning" on the character and quality of the research reports. The first is the persistence of disincentives within academic career structures to compromise scientific integrity for the sake of securing government contracts. Our findings point to this, but we have also noted the shortcomings of our evidence base in this respect. The second is the existence within government of a body of research administrators given significant responsibility for developing and managing research and coming in between policy officials and

politicians on the one hand and the researchers on the other. In the absence of such a body of research administrators, the pressures on researchers to produce political ammunition are likely to be far stronger. Without the disincentives to compromise scientific integrity, one would have to have serious concern for the value of commissioned research.

References

Allen, C (2005) "On the Social Relations of Contract Research Production: Power, Positionality and Epistemology in Housing and Urban Research" *Housing Studies* 20(6): 989–1007.

Beyer, JM and Trice, HM (1982) The Utilization Process: A Conceptual Framework and Synthesis of Empirical Findings *Administrative Science Quarterly*, 27(4): 591-622.

Boswell, C (2008) "The political functions of expert knowledge: knowledge and legitimation in European Union immigration policy" *Journal of European Public Policy* 15(4): 471-488.

Bovens, M., P. 't Hart, and S. Kuipers (2006) 'The Politics of Policy Evaluation' in M. Moran, M. Rein and R.E. Goodin. eds. *The Oxford Handbook of Public Policy*, Oxford: Oxford University Press, pp. 319-335.

Bulmer, M (1982) *The Uses of Social Research* (London: Allen and Unwin)

Chelimsky, E (2009) "Integrating evaluation units into the political environment of

government: The role of evaluation policy" *New Directions for Evaluation* 123 (Autumn): 51-66.

Cohen, S (1988) *Against Criminology* (New Brunswick, NJ: Transaction Books)

Davies, H. T. O. , Nutley, S. & Walter, I. (2010) "Using evidence: how social research could be better used to improve public service performance" in Walshe, K., Harvey, G. & Jas, P. (eds.). *Connecting knowledge and performance in public services: from knowing to doing*. (Cambridge University Press)

Davis DF and Portis, EN (1982) "A Categorical Imperative for Social Scientific Policy Evaluation" *Administration & Society* August 1982 14(2):175-193

DCLG (2011) "Independent evaluation of the Department for Communities and Local Government (DCLG)'s handyperson programme" (London: Department for Communities and Local Government)

Derlien, H-U and Rist, R (2002) "Policy Evaluation in International Comparison" in Furubo, J. , Rist, R. C, Sandahl, Rolf, eds. (2002), *International Atlas of Evaluation*, (New Brunswick and London: Transaction).

Fischer, F (2003) *Reframing Public Policy. Discursive Politics and Deliberative. Practice*s. (Oxford: Oxford University Press).

Furubo, J. , Rist, R. C, Sandahl, Rolf, eds. (2002), *International Atlas of Evaluation*, (New Brunswick and London: Transaction).

Ham, C Hunter, DJ and Robinson, R "Evidence based policymaking" *British Medical Journal* January 310 (6972): 71-2.

Head, BW (2010) "Reconsidering evidence-based policy: Key issues and challenges" *Policy and Society* 29(2): 77-94.

Hertin, J , Jordan, A , Turnpenny, J , Nilsson, M, Russel, D and Björn, N (2009) "Rationalising the Policy Mess? Ex Ante Policy Assessment and the Utilisation of Knowledge in the Policy Process" (March 17, 2009). Available at SSRN: http://ssrn.com/abstract=1361519 or http://dx.doi.org/10.2139/ssrn.1361519

Hillyard, P., Sim, J., Tombs, S. and Whyte, D. (2004) "Leaving a 'stain upon the silence': contemporary criminology and the politics of dissent", *British Journal of Criminology*, 44, 369-390.

Hope, T (2004) "Pretend it works: Evidence and governance in the evaluation of the Reducing Burglary Initiative" *Criminal Justice* 4(3): 287-308.

Huitema D, Jordan A, Massey E, Rayner , T, van Asselt H. Haug, C, Hildingsson, R, Monni, S, Stripple J (2011) "The evaluation of climate policy: theory and emerging practice in Europe" *Policy Sciences* 44:179–198

Kogan, M and Henkel, M (1992) “Constraints on the Individual Researcher” in Whiston, TG and Geiger, RL *Research and Higher Education. The United Kingdom and the United States* Buckingham: Open University Press)

Lijphart, A. (1999) *Patterns of democracy*. New Haven: Yale University Press

Manzi, T and Smith-Bowers, B (2010) "Partnership, Servitude or Expert Scholarship? The Academic Labour Process in Contract Housing Research" in C. Allen and R. Imrie (eds.) *The Knowledge Business: Critical Perspectives on Contract Research* (London: Ashgate).

MacGregor, S (2013) "Barriers to the influence of evidence on policy: Are politicians the problem?" *Drugs: Education, Prevention, and Policy* 20(3): 225–233.

Middlemas, J (2006) Jobseekers Allowance intervention pilots quantitative evaluation. (London: Department for Work and Pensions. Research Report No 382).

Molas-Gallart, J (2012) "Research Governance and the Role of Evaluation: A Comparative Study" *American Journal of Evaluation* 33: 583-598.

Norris, N. (1995) "Contracts, control and evaluation", *Journal of Educational Policy*, 1: 271-285.

O'Brien, M; Clayton, S; Varga-Atkins, T; Qualter, A (2008) "Power and the theory-and-practice conundrum: the experience of doing research with a local authority" *Evidence and Policy* 4(4): 371-390.

PA Consulting (2005), Action Research Study of the Implementation of the National Offender Management Model in the North West Pathfinder (London: PA Consulting)

Page, EC and Jenkins, B (2005) Policy Bureaucracy. Government with a Cast of Thousands (Oxford:Oxford University Press).

Pinson, G (2010) "The Knowledge Business and the Neo-Managerialism of Research and Academia in France" in C. Allen and R. Imrie (eds.) *The Knowledge Business: Critical Perspectives on Contract Research* (London: Ashgate).

Powell, M (2011) "The Policy Process" in J Glasby (ed) *Evidence, Policy and Practice: Critical Perspectives in Health and Social Policy* Bristol: Policy Press.

Sabatier, P (1978) "The Acquisition and Utilization of Technical Information by Administrative Agencies" *Administrative Science Quarterly* 23(3): 396-417.

Salisbury, C, Stewart, J, Purdy, S Thoro, H, Cameron, A, Lart R, Peckham, S amd Calnan, M (2011) Making the most of evaluation: a mixed methods study in the English NHS *Journal of Health Services Research and Policy* 16(4): 218-225.

Salisbury, C; Stewart, K, Cameron, A, Peckham, S Calnan, M Lart, R Purdy, S Thorp, H (2010) "Making the Most of Policy Evaluations: Overview and synthesis of evaluations of the White Paper 'Our health, our care, our say'. Final Report" (Commissioned for Department of Health)(Bristol: University of Bristol)

Slaughter, S and Leslie, LL (1997) *Academic Capitalism: Politics, Policies, and the Entrepreneurial University* (Baltimore, MD: Johns Hopkins University Press).

Walker, R (2001) " Great Expectations: Can Social Science Evaluate New Labour's Policies?" *Evaluation* 7(3): 305–330

Weiss, C. H. (1970) "The Politicization of Evaluation Research" *Journal of Social Issues* 26(4): 57-68.

Weiss, C.H. (1979) "The many meanings of research utilisation" Public Administration Review, 39(5):426-31.

Walters, R (2007). Embedded Criminology. In: Brannigan, A. and Pavlich, G. eds. *Governance and Regulation in Social Life: Essays in Honour of W.G. Carson*. (London: Routledge).

Table 1 Positive findings in evaluation reports by authorship

Author	*N*	*% positive**
University-based academics	58	52
Private consultants	58	58
Mix with academics	35	57
In-house researchers	40	62
Mix without academics	7	53
Unclear	8	47
Total	206	54

*Coded as "positive" or "very positive"
Source: coded reports (see text)

Table 2 Positive findings in evaluation reports by response to requests for draft changes

	% positive	*N*
Not asked to make changes	46	80
Asked and made changes	15	13
Asked but did not make changes	50	4
Asked and compromised	52	58