Should I take this seriously? A simple checklist for calling bullshit on policy supporting research

Jarl K. Kampen · Peter Tamás

Abstract We propose a simple checklist for the users of policy supporting research in order to decide whether a piece of research begs further study or can be dismissed right away. The checklist focusses on the quality of the research question (is it a research question, and is the research question answerable); the kind of knowledge along with the order, level and quality of data needed for answering the RQ; the methods of analysis used; the degree to which the research results support the conclusions; and whether the conclusions provide an answer to the research question.

Keywords Research methods · Methodological issues · Standards · Evaluation · Decision support · Knowledge assessment · Quality · Policy

May any utterance at all qualify as humbug or bullshit, given that (so to speak) the utterer’s heart is in the right place, or must the utterance have certain characteristics of its own as well?


1 Introduction

Knowledge is a commodity that is ordered, produced and sold. As a consequence, knowledge is increasingly being commercialized (Skyrme 2003). Where knowledge production in the past was largely the monopoly of universities, we live in a world where knowledge is increasingly privatized when at the same time, actual and rhetorical need for and dependency on knowledge by many stakeholders in society is higher than ever and will continue to grow. Increased demand combined with the effects of the worldwide economic crisis of the first decades of the twentyfirst century, has produced fierce competition in the knowledge...
market. Even within universities, academics make attempts to legitimize and claim certain niches as “their” territory. They may, for instance, recast concepts such as “methodology” and “method” that have well established scientific meaning in ways that increase the legitimacy of their own knowledge generation practices. In short, knowledge resources are manipulated in various ways and are subjected “to various influences in an effort to meet the need or take advantage of the opportunity” Holsapple and Joshi (2002, p. 89).

Organizations that commission policy relevant research want useful knowledge. However, producing reports that policymakers recognize as useful, and at the same time, that social scientists recognize as knowledge, can be problematic for three reasons. First, apparently simple policy relevant questions can be very difficult to answer. The gap between questions for which empirical research can generate valid and reliable answers and those that are policy relevant has troubled organizational studies for more than sixty years (e.g., Kieser and Leiner 2009, 2011). Second, as noted in the opening paragraph, the way policy relevant research is commissioned and consumed rewards the delivering of unsubstantiated and overreaching knowledge claims. That is to say, interpretation and presentation of findings in ways that seem to narrow the gap between what is asked and what can be answered produces an atmosphere in which commissioners in accelerating pace request knowledge that cannot possibly be produced when simultaneously researchers, in order to obtain research funding, promise more than they can possibly deliver. Third, policymakers’ questions often require research that combines knowledge products from both the social and natural sciences. There are quite a number of paradigms in the social sciences that use empirical research. However, while a few, like the natural sciences, attempt to generate valid and reliable findings with estimates of confidence, many produce findings that are incompatible with basic conceptions of scientific conduct in the natural sciences. Mixing these fundamentally different conceptions of knowledge can create problems, for instance, because the reader of such a report might assume that the empirical foundation of the social scientific claims is comparable to that of the natural scientific claims.

This essay is written for policymakers who commission and consume empirical social science research. Its purpose is to help policymakers decide how seriously they should take what they are reading. The body of this paper is structured on the model of a decision protocol.
Should I take this seriously?

or checklist used to handle the intake of new cases. It proposes a sequence of questions that can be used in screening reports of policy relevant social science research. This protocol allows the users of policy relevant research to decide whether a piece of research begs further study or can justifiably be summarily dismissed. The checklist focuses on the quality of the research question(s); the kind of knowledge along with the order, level and quality of data needed for answering the RQ; the methods of analysis used; the degree to which the research results support the conclusions; and whether the conclusions provide an answer to the research question.

2 Is it a research question?

Applied research can be motivated by and pretend to answer challenges that are not research questions. For example, while perhaps deeply policy relevant, any question of the general type “How can we [change reality]”, where [change reality] can stand for (a) “bring peace and prosperity to Afghanistan”, (b) “transfer a traditional farm into a biological one”, (c) “involve people in decision-making”, (d) “develop exotic mushrooms as the food of the future”, etc., are not proper research questions. Their direct object is not generating knowledge. Even when the research objective, that is the motive for commissioning the research is the changing of reality, a central research question (CRQ) cannot address that ambition directly. The overstrained situation where policy makers expect ready-made solutions and contract a consultancy firm or university research department to do a study that “tells them what to do” not only puts researchers in a position where they have to satisfy impossible demands, but also occupy the position of policy maker de facto for which they are neither professionally nor legally equipped. Empirical research cannot make decisions. It produces new knowledge that policy makers use to make an informed decision as to what action to undertake to change reality.

Each of the examples given in the previous paragraph can be easily associated with properly formed research questions. The knowledge required may be (a) the main threats of peace and prosperity in Afghanistan, (b) the main differences between running a traditional farm and a biological one, (c) the main obstructions for people’s participation in decision-making, (d) the attitude of potential consumers toward eating exotic mushrooms, etc. The corresponding CRQ are (a) “What are the main threats of peace and prosperity in Afghanistan”, (b) “What are the main differences between running a traditional farm and a biological one”, (c) “What are the main obstructions for people’s participation in decision-making”, (d) “What is the attitude of potential consumers toward eating exotic mushrooms”, etc. Depending on the ranking of the identified threats, differences, obstructions, attitudes etc., the policy maker can make a decision as to which threat requires priority in reducing, which difference requires priority in changing, which obstruction requires priority in removing, which attitude requires priority in changing, etc.

Of course, instead of phrasing fundamental knowledge questions (such as the examples above), that is, from a researchers perspective, it is also possible to phrase the questions from a policy maker’s perspective, for instance, “Which threat of peace and prosperity in Afghanistan requires priority attention”. However, in such cases, the CRQ will invariably produce properly formed secondary research questions, for example, “What are the main threats of [x]”. If a research project fails to reframe policy interests as knowledge-probing research questions, there is good reason to suspect the following research will be of inferior quality.
3 Is it an answerable research question?

There are types of questions that are near impossible empirically to answer. Foremost amongst these are questions and predictive claims that rely on the demonstration of causal relations with social antecedents. Any claim or conclusive explanation of what caused or causes an observed effect must meet three standards. First, the cause and the effect must be logically related, second, alternative explanations must be eliminated and third, the cause must precede the effect.\(^5\) The second of these, identifying and controlling for alternative explanations, is particularly troublesome when the substance of the study is social. Social phenomena are, by definition, social. They cannot meaningfully be abstracted from the context in which they are studied. As such, it is not possible to inventory and separate relevant factors with the rigor required for causal claims (Spicker 2011). Further, causal studies of social phenomena inevitably involve human conduct and causal claims must, therefore, contain an account of what motivates human conduct. As capably demonstrated in a brilliant parody from the mid-1970s, we cannot even agree on how saying ‘please pass the salt’ causes its arrival (Pencil 1976). Policy relevant questions tend to be several orders of magnitude more complex than securing a bit of salt for our soft-boiled eggs.

Another class of predictive questions that is extraordinarily difficult to answer include those which query what an organization should do. In addition to empirically establishing a causal relationship between a particular type of conduct and a particular outcome, should type questions require researchers to make two types of predictions. First, they must predict what an organization will do with the recommendations they receive. On this point empirical studies of what happens to recommendations do not support the hypothesis of a simple connection between submission of recommendations and organizational conduct (see e.g., Beyer 1982; Landry et al. 2001; Haynes et al. 2008), and the level of consensus between these studies appears to be insufficient to predict the fate of any given set of recommendations.\(^6\)

\(^5\) There is one circumstance under which cause need not precede effect. Human conduct can be informed by the anticipation of a future event. In this case while the causal event is subsequent to the effect, the mental act of anticipation does precede the observed effect. Under these circumstances causal studies must account for how the mental act of anticipation determines physical conduct.

\(^6\) There are, reasons to doubt the assumption that the organizations that commission and consume research are rational (Albiek 1995) A reasonable condition for predicting the impact of a given recommendation is an understanding of how organizations make use of recommendations. A selective review of the literature suggests that there is limited consensus on the framework for understanding how organizations use recommendations (c.f. Beyer 1982; Haynes et al. 2008; Landry et al. 2001). While reviews of empirical studies on the use of research in shaping organizational conduct do not arrive at consensus at the level of specifics, they do agree that the link between submission of a report and the conduct of an organization is not simple.

In addition to a framework for understanding how organizations make use of recommendations, predicting the effect of a given recommendation would require an understanding of the factors that shape their uptake by organizations. Findings on this question are, as the following few examples suggest, conceptually and substantively diverse. A regression on nearly 1,000 responses to a questionnaire found that linkage mechanisms, research experience, unit size, and research relevance for the users interacted with individual and organizational variables in explaining reported utilization (Belkhojda et al. 2007). A case study on a significant policy change found that ‘beliefs, reflecting ideas exogenous to the policy process’ were able to shape a significant policy initiative (Hook 2008). A case study on the uptake of research in shaping health policy found organizational motivation, organizational capacity to acquire research findings, organizational capacity to transform research findings and moderating organizational factors all relevant to understanding research uptake (Hamel and Schrecker 2011). Finally, another case study in the health sector noted the relevance to research uptake of the ‘taste’ that policymakers have for research, where this was operationalized as the ability and inclination to seek out and select high quality relevant research (Jewell and Bero 2008).

Discussion of how research shapes organizations’ conduct and the factors that condition this shaping presumes the relevance of research. Recent review articles have noted that “consistent evidence shows that health systems fail optimally to use evidence” (Straus et al. 2011) and that “the literature on knowledge utilization-
Second, should type questions require the researcher to forecast that the future arrangement of all relevant conditions will be such that the observed causal relationship will remain in the future. In summary, establishing causal relations, predicting the fate of a given recommendation and predicting future social states are each individually extraordinarily difficult tasks. Successfully investigating all three of them simultaneously in a single research project is most unlikely.

4 What kind of knowledge is needed to answer the research question?

If a report is structured around an answerable question, the next logical step is to determine if the answers given in the report are empirically well founded. Assessment of those empirical foundations must, however, be done in light of the kind of knowledge that the research was trying to generate. There are three possible kinds of knowledge claims (see Fig. 1). Descriptive research attempts to identify and better understand objects of interest and aims to produce an account about what is; exploratory research tries to find correlational or causal relationships between objects that have been described; and confirmatory projects try and test posited relationships (correlations, causal relationships). The kind of knowledge needed for answering the CRQ in turn, determines the kind of study design (e.g., experiment, cross sectional, longitudinal, case study) suitable for the research.

While a wide range of study designs can produce descriptive findings, the number of designs that can be used shrinks as the research moves towards the confirmation of causal hypotheses. For instance, the research question “Which threat of peace and prosperity in Afghanistan requires priority” presupposes a connection between some identifiable threat and some measure(s) of peace and prosperity. If this connection must be assumed to be causal, so that some threat is assumed to be a cause of lack of peace and prosperity, a time-lag must be supposed between the inception of the cause and of the effect, where the cause should precede the effect. Therefore, in order to detect whether inferences about time-order are correct, multiple measurements in time are required within a period during which the intensity of the causal variable varies. The latter can be prompted by the researcher, in which case the study design is called an experiment, or the change can be “natural” in which case detection requires a longitudinal study design (see e.g., Kampen 2011). A mismatch of the study design applied in the research and the study design required for the answering of the

Footnote 6 continued
generally reveals limited use of social science research in policy-making” (Hird 2009). The source studies for these reviews tended to focus on instrumental use of research findings: precisely that form which seems to be presumed in the recommendation of clinical trials as an adequate foundation for prescriptive claims. Limiting interest in the effect of research on policy and practice are argued elsewhere to inevitably lead “to a dismal view of research impact, because it ignores the variety of ways in which research uptake and adoption can occur” (Cherney and McGee 2011). If knowledge use is expanded to include the more indirect forms of influence noted above, studies suggest a more hopeful (Cherney and McGee 2011; Paris 2011) but certainly less predictable picture.

In summary, there is limited consensus surrounding both an appropriate framework for understanding the use organizations make of research and the factors that shape that use. Even granting agreement on this point, studies that focus on the more predictable forms of use tend to produce rather dismal findings. As such, researchers do not have the ability at this time to predict the fate of their recommendations. Taking the next step, forecasting the effects of an organization’s conduct requires a causal understanding of what produces the desired effect and the ability to forecast that all analytically relevant conditions in the future will be arranged such that the desired effect can be reproduced. In the framework of social science research modeled on the natural sciences it is neither possible to establish causal relationships nor is it possible to make predictions of future social states with any degree of confidence.
CRQ invariably leads to wrong conclusions. While perhaps useful for descriptive studies, cross sectional research designs, for instance, lack a time component and therefore no definite claims regarding causes and effects can be made. Case study designs, again while appropriate in some applications, do not support inferences to larger populations.

5 What is the empirical basis of the findings of the report?

5.1 What order of data is required to answer the CRQ?

For the practical purpose of this protocol, we distinguish three basic orders of data. The first reflects the real world, the second experiences of that world and the third concerns what happens in the research context itself. This typology speaks to the question of how close to the real world a piece of research is trying to reach. Each of these is a legitimate domain for empirical research. Applied social science research tends to ask questions that stretch their reach the furthest: they often attempt claims with respect to future states in the world as it is in itself and as it is experienced outside of the research context.

Often in empirical work there is a gap between the order of data required to answer the CRQ and the order of the data generated by actual research. This gap must be closed in a way that is scientifically justified. Data presented as representing objects found outside of the research context itself are, in and of themselves, incomplete. Such data must be accompanied by a discussion of the scientifically justified inferential strategy by which the data generated in the research setting are connected to the time, place and objects of analytic interest. For example, data representing the experience of Afghan women with respect to justice requires to be complete, minimally, the narrative given by respondents accompanied by scientifically well-formed accounts of how the interview context shapes the narrative rendered, how the time lapse between the event narrated and the moment of research shaped the narratives given and all other variables that may have shaped the narrative as rendered in the research setting.\(^7\) Even with all of this, the report may only make claims with respect to how Afghan women reported experienced justice. Moving from these accounts of experience to what actually happened requires, yet another, set of scientifically justified inferences.

\(^7\) In preparing this essay we debated whether to include discussion of the role of other factors such as the researcher’s own cognition in shaping their hearing, understanding, memory and reporting of the narrative. We decided that this and others of its ilk were, for the limited purposes of this essay, a bridge too far.
One implication of recognizing that inferences beyond the research context require explicit and scientifically sound discussion is that questions regarding what people think, or what people perceive become very difficult to answer. Any research question that attempts to probe the mental state of respondents must rely either on the interpretations of an external party or the accounts of the respondents’ themselves. Third party interpretations of what an individual is thinking are not reliable. Answers respondents give to researchers are those that they provide to that researcher at that time in that context. Moving from these contextually conditioned responses to actual mental states, mental states as they would be espoused in a different context or conduct in another context, requires carefully and explicitly argued inference formed in a manner compatible with the paradigm in which the researcher is operating.8

5.2 What level(s) of data are required to answer the CRQ?

Applied social science research is often interested in objects that cannot answer questions. For example, research questions often take interest in objects like households, bureaucracies and communities. Applied research on these objects often engages the individuals whose collective behavior produces these objects of analytic interest. These individuals and the collectives they make up exist at different levels. It is not empirically defensible to, for example, talk to a male head of household and record the narrative he gives as that of a ‘household’ or to talk to a bureaucrat and make claims about a bureaucracy. Neither is it possible to infer findings at collective level to individuals. To mention just a well-known example, from the observation that European countries with high levels of protestant Christians have high levels of suicides per capita, it cannot be inferred that individuals with protestant faith are more likely to commit suicide (ecological fallacy). Data must be gathered at the level required to answer the CRQ, and data gathered at levels other than that of analytic interest are incomplete without explicit discussion of the inferences made to connect, for example, individuals and households.

5.3 What quality of data is required to answer the RQ?

Data quality may be degraded by any number of factors. The two main design determinants of the quality of obtained data are the sample and the instrument(s) used in the research. Lack of quality of one or the other destroys any possibility to draw any valid inferences based on the empirical parts of the research which of course renders the entire research project worthless. Any research project must produce transparent and convincing arguments that the sampling techniques are sound and that the instrument(s) of choice produce adequate measurements of the characteristic(s) of interest. If the design is adequate, the next concern is degradation of what was planned that occurs during implementation of the research plan. Tracking data quality loss during implementation requires careful documentation. All analytically relevant factors shaping data quality must be explicitly reported.

If the loss of accuracy associated with degraded data can be quantified, that risk must be carried through and be reflected in appropriately conservative confidence claims. If that loss of accuracy cannot be quantified, then the degraded data quality creates an uncertainty that must

---

8 Each research paradigm has its own notions regarding what constitutes legitimate claims. These apply both to findings and inferences based on those findings. If the researcher, for example, is conducting empirical research in a paradigm that supports validity and reliability then the inferential claims made must be based on valid and reliable empirical measures.
be appropriately noted. Discussion of data quality should be found in the methods and the limitations sections of a report and there is often some information in annexes. The methods section should specify and justify the design of both instruments and the sampling procedure, the limitations section should discuss how the findings are or are not robust with respect to threats encountered in implementation and the actual instruments, ideally accompanied by a discussion of how the instrument is contextually appropriate.

5.4 What analysis methods are required to answer the CRQ?

All data is subject to some form of analysis before it is presented to readers. At the most basic and unavoidable level, researchers structure what is reported and inevitably by doing so, introduce biases. Given that some form of analysis is inevitable, reports that make any suggestion that the ‘data speaks for itself’, that ‘themes emerged from the data’ or any other construction that is silent as to the role of the (tacit) theory informing the researcher in structuring what is reported are inadequate.

Assuming that the report has a theoretically informed and methodologically appropriate description of how data was analyzed, the next steps are to see if the data gathered supports the analysis methods used, if the analysis used was properly executed and appropriate, and if the results of that analysis are sufficient to answer the CRQ. Shortcomings or any mismatch between these renders the empirical research worthless. For example, the use of inferential statistics in non-random samples leads to invalid results, because of the non-representativeness of the sample with respect to the target population.

Unlike the previous steps, determining if the analysis methods are well suited to the task given can be a technical question best answered by people with that specific competence. In order to support this sort of review the report should contain a detailed discussion of the theory behind, the rationale for, the risks associated with and the content of the analysis methods used. This discussion is generally found in the analysis section of a report with, perhaps, some discussion in the limitations section.

5.5 Do the research results support the conclusions?

Conclusions emerge from the combination of appropriately specified and carefully gathered empirical data, the appropriate analysis of that data and transparent discussion of threats. If these are present the author can make well-argued inferences. Deficiency with regard to any one of these or reference to outside support removes the ground necessary to draw empirical conclusions. Research results are generally found in the findings and analysis sections of a report. These should be restated in conclusions. If there are differences between results and conclusions, the mechanisms by which inferences were made should be discussed in the analysis section. This discussion should place equal emphasis on empirical evidence found in favor of particular research hypotheses (verification) and on empirical evidence rejecting possible rival hypotheses (falsification). Silence with regard to rival hypotheses, particularly when the conclusions affirm notions held at the outset by the funder, should be regarded as deeply concerning. In all research the absence of evidence never implies evidence of absence. In applied research there, unfortunately, can be immediate benefit in not seeing certain things.10

---

9 In the natural sciences uncertainties, by default, make findings meaningless.

10 In our combined 35 years of experience the authors have never read a piece of directly commissioned policy relevant research that concluded that policy is irrelevant or the funder insignificant.
5.6 Do the conclusions generate an answer to the CRQ?

Although it is quite common for politicians to provide an answer to a different question than the one advanced, a researcher cannot follow their model. In scientific research, once you have a research question, you are stuck with it. If researchers accept terms of reference that contain an impossible CRQ, they are bound either to answer that question or to negotiate its modification. At the end of either path the conclusions must answer the CRQ. Whenever there is a mismatch between the conclusions drawn and the general research question, the researcher did not properly discharge their professional obligation to negotiate and then execute a solid piece of research. Conclusions are found in the executive summary and in the conclusions section of a report. The CRQ, and any modifications to that question, should be found in the introduction.

6 Closing remarks: applying the checklist in practice

In this essay we have outlined a (fast and simple) protocol that can be used by policymakers to decide if they should reject a report of empirical social science research. The first criterion tested if the report was structured around a researchable question, the second if the question was answerable, the third tested for the type of knowledge claim required and the fourth proposed a number of questions to test the empirical foundations of the report. Sequentially, the most efficient procedure to quickly assess a research report is by taking the following steps:

1. First check whether the research question is proper, that is aimed at gaining knowledge, not at changing reality (if not, dismiss report on the ground that it cannot involve research).
2. Then check whether it is potentially answerable or not (if not, dismiss report on the ground that it cannot be researched).
3. Then check whether or not there is a mismatch between the conclusion and the research question (if there is a mismatch, dismiss report on the ground that the research commission has not been fulfilled).
4. When there is no apparent mismatch between the conclusion and the CRQ, check if there is a mismatch between the conclusions and the empirical research findings (if so, dismiss report on the ground that the conclusions are not substantiated empirically).
5. When there is no apparent mismatch between the results and the conclusions, check for a mismatch between study design and the kind of knowledge required to answer the RQ (if there is a mismatch, dismiss report on the ground that research is badly designed).
6. If study design and kind of knowledge are compatible, check that the order and level of data used in the study match those required to formulate the conclusion (if there is a mismatch, dismiss report on the ground that it cannot substantiate claims empirically).
7. Check whether the data collecting process and its analysis is fully documented (if incomplete or unclear, dismiss report on the ground that its data cannot be trusted and/or its methods are not transparent).
8. If the report does not immediately fail on one of the counts 1–7 listed above, more careful study of the document is in order (still keeping in mind the checklist of Table 1).

It is our hope that in preparing this protocol those who pay for policy relevant social science research will be better able to specify terms of reference that are reasonable and critically
<table>
<thead>
<tr>
<th>Characteristic to be checked</th>
<th>Quality dimensions</th>
<th>Most likely location in the report</th>
</tr>
</thead>
<tbody>
<tr>
<td>1. Is it a research question?</td>
<td>knowing vs. prescribing</td>
<td>introduction, conclusion</td>
</tr>
<tr>
<td>2. Is the research question answerable?</td>
<td>suitability for empirical research</td>
<td>introduction, conclusion, executive summary</td>
</tr>
<tr>
<td>3. What kind of knowledge is needed?</td>
<td>data produced by design meets data required by question (descriptive, exploratory, confirmatory)</td>
<td>introduction, methods, conclusion</td>
</tr>
<tr>
<td>4. What order of data is required?</td>
<td>order of data secured (real world, experience, research context) is that required to answer question inferences justified</td>
<td>introduction, methods, analysis</td>
</tr>
<tr>
<td>5. What level of data is required?</td>
<td>level of research (sub-individual individual, collective) = level of claim inferences justified</td>
<td>introduction</td>
</tr>
<tr>
<td>6. What quality of data is required?</td>
<td>appropriate design accounts for field compromises</td>
<td>methods, limitations</td>
</tr>
<tr>
<td>7. What methods of analysis are required?</td>
<td>adequately discussed appropriate</td>
<td>methods, analysis, limitations</td>
</tr>
<tr>
<td>8. Do the research results support the conclusions?</td>
<td>sound inference accounts for threats</td>
<td>analysis, limitations</td>
</tr>
<tr>
<td>9. Do the conclusions provide an answer to the research question?</td>
<td>equivalence</td>
<td>introduction, conclusion</td>
</tr>
</tbody>
</table>

assess both the proposals and reports that flow therefrom. And of course, we hope to save policy makers the time they would otherwise spend reading ill formed “research” reports.

**Acknowledgments** We would like thank Prof. Dr. M. Paspalanova (Expert on Human Rights Indicators, OHCHR, Mexico), Prof. Dr. I. Tanasescu (Political Science Department, Free University of Brussels, Belgium) and Andy Tamás (Director, Tamás Consultants) for their helpful comments on an earlier draft of this article.

**References**


