Final draft

The dilemmas and uncertainties in assessing the societal impact of research

Juha-Pekka Lauronen*

Faculty of Social Sciences, University of Helsinki, Helsinki 00014, Finland

*Corresponding author. Phone: +358415049642; Email: juha-pekka.lauronen@helsinki.fi

Abstract

In the 2000s, many EU countries have established research impact assessment practices as part of the scrutiny of the third mission of universities. The Finnish research evaluation system has widely adopted a societal impact criterion. In this study, the question is what guiding principles does impact assessment rely on. This point of view is based on the experiences and opinions of experts interviewed for the study. Four implicit principles of impact assessment which guide assessment goals and practical implementations were found. The guiding principles have several social and methodological dilemmas due to liminal interpretations between social interests and academic endeavor. This study recommends that the evaluation scholars should consider the actual social purpose vis-à-vis consistent methodological approaches to assessment before applying "all-round" solutions to scholarly fields.

Key words: Societal impact; Research assessment; Evaluation methods; Liminality; Uncertainty

1. Introduction

During the past 20 years in European research policy, there have been attempts to demonstrate the societal impact of research as a means of representing legitimatization of the third mission of universities that is, interacting with society at large. The pursuit culminated in the European Union's (e.g. 2017) policy to develop models and methods for socio-economic impact assessment and to maximize the impact of mission-oriented research. Societal impact has become a great challenge for research evaluation scholars. In the development and implementation of impact assessment, there have been contesting tendencies. Academia has preferred more technical refinements of the assessment methods in consideration of the diversity of disciplines (Martin 2011), but the government's bureaucratic research evaluation systems have emphasized control and cost-efficiency as much as broader indicators of research excellence (Hicks 2012; Molas-Gallart 2012).

Impact assessment has gone along a hard road from heterogeneous quantifications, qualitative approaches and conceptual impreciseness towards more sophisticated and formalized methods of case studies and interaction processes (Bornmann 2013). The theoretical frameworks and methodical approaches have become more elegant, also considering the characteristics of the social sciences and humanities (e.g. Miettinen et al. 2015; Muhonen et al. 2018). However, at the same time, impact assessment has become a formal practice following normalized rules aided by manual guidelines (cf. Martin 2011). Social science and humanities scholars have tried to make sense of what the impact endeavor is all about and what the explicable agenda behind it might be (e.g. Benneworth et al. 2016).

In Finland, the discussion on the societal impact of research has been ongoing since the early 2000s, showing signs of institutional standardizations. However, the dust seems not to have settled yet, as the academic criticism, disputes, redefinitions and administrative fatigue continue. Regardless of a profound scholarly development in assessment methods and diverse assessment designs at all levels of the Finnish research evaluation system, the perceptions of societal impact have remained vague and the assessment practices unstructured (e.g. FINEEC 2019). The motive in this paper is to understand why there is such dissatisfaction with impact assessment and its incompleteness. The sketchy research impact policies and deliberate debate provides an inspiring context, which contributes to the understanding of how research policy preferences encounter in research assessment and how assessors adjust their interpretations between intellectual endeavor and sociopolitical interests.

Instead of considering suitable assessment methods, this study reports on an investigation of the meaning of impact assessment and specific methods when attempting to verify the impact of scholarly endeavor. This article focuses on the social and methodological ground of the disputes and problems of impact assessment by investigating diverse assessment preferences through experts' experiences and interpretations in a social setting. The question investigated in this study through the empirical material used has two parts. First, the study is concerned with how the experts explain the grounds for the diverse purposes of impact assessment and how they justify and criticize methodical choices to support these purposes. Second, the study asks what social and methodological preferences guide these

choices. Do they form guiding principles? Through these questions the aim of the study is to provide interpretation of the characteristics of impact assessment in Finland.

The interpretation is based on the presumption that impact assessment has implicit principles comprised of diverse social and methodical preferences. In this study it is argued that impact assessment methods (such as case studies) find it difficult to address sociopolitical aims and the characteristics of diverse disciplines at the same time because of a multitude of liminal interpretations. The discussion on impact has been concentrated on technical justifications for assessments but has little considered the relationship between the technical methods, policy, administrative frameworks and academic practices. The paper continues with sections on theory framework (2), policy background (3), data and methods (4), analysis (5-6) and conclusive discussion (7).

2. The social and methodological tensions of impact assessment

2.1. Impact representing contextualized knowledge

As the sciences have begun to lose ground in holding the monopoly of legitimate knowledge in public debates on the multiplicity of rationalities and uncertainty, extra-scientific justifications have increased dramatically (Beck et al. 2003). Science is needed to explain complex problems but may offer only a diversity of explanations. David H. Guston (2000) points out how the boundary between politics and science has become blurred, as multiple stakeholders and agencies involve producing knowledge. In the old social contract for science, academia's self-regulation of research assured research integrity and productivity. The new forms of assurance formalize the self-regulative aspects of integrity and productivity. They transfer collaboration to intermediary institutions between academia and politics. Academia is supposed to actively seek partnerships and social opportunities rather than waiting for knowledge entering society.

However, according to Helga Nowotny et al. (2001, 46-47) these social demands for accountability have taken a form of outcome orientation and measurable objective indicators to realize social control of research endeavor. Academic institutions attempt to internalize the social control into self-control by pushing the limits of traditional research evaluation. Academia has found it difficult to create corresponding criteria for quality judgements of social collaboration and impact. Reconciling intermediate organizations such as research councils has led to difficulties in setting priorities for basic research, as they attempt to apply additional social criteria for quality. Interdisciplinary justification of knowledge in peer review is one way to respond to the demands for social accountability (Huutoniemi 2012, 24-25).

Research evaluation has had a major transition of focus from pure scientific quality to applications and societal utility. At the same time, sciences have become more globalized, standardizing local research objects and rationalities (Drori et al. 2003; Mosbah-Natanson & Gingras 2014). The change from the knowledge production principles of Mode 1 to

Mode 2[1] is argued to be one reason for the transitions and growing wave of research evaluation (Nowotny et al. 2001). As research funding is more tangled up with the needs of other societal sectors, research is more accountable to stakeholders.

Some scholars have argued how the inadequate indicators of research quality need to be improved because of this transition (Ernø-Kjølhede & Hansson 2013). They think that research evaluation should show how the sciences enter society. Research impact assessment is a way to respond to the social demands of accountability to ensure contextualization of knowledge and local collaboration. Impact assessment is a way to encounter the uncertainties of the sciences and politics (Dahler-Larsen 2011, 94; Eräsaari 2009; Weingart 1999), but it is also a tool to control university research through pre-established priorities responding to the very same uncertainties (Whitley 2014).

Furthermore, impact assessment designs seem to suffer from the same uncertainties, which they encounter through traditional tripod expertise by established scholarly practice (e.g. theory building and methods), institutionalized rules (e.g. evaluation offices in universities) and profession (people trained for assessments) (Eräsaari 2009). Formalization of assessment methods, such as through a case study, means standardization of data, codification of assessment questions a priori and reduction of interpretations and explanations (Firestone & Herriot 1983).

Since the additional criteria for research quality represent simplistic pictures of the research process and are outcome oriented, they can easily turn into the tyranny of transparency (Strathern 2000). Such audit mechanisms would counter the actual contextualization of social knowledge created in free interaction with society (Nowotny et al. 2001, 115). These remarks leave uneasy questions about whether there are adequate indicators of growing uncertainty of science by impact assessment and whether impact assessment can be the right tool to respond to these uncertainties.

2.2. Strategies to verify impacts

Societal impact is usually understood as a concept which broadens the scope of research evaluation to the third mission of universities (Bornmann 2013). There are three distinct strategies for verifying impact and these strategies accomplish the approaches taken by different impact models. The different approaches get tangled with each other in practical assessments, but they have a verification logic of their own. One of the main strategies to define societal impact flexibly is through multiple social and epistemic systems and pathways such as the economy or the environment (e.g. Miettinen et al. 2015; Muhonen et al. 2018). The second way is to define a general phenomenon for practical indicators (Bastow et al. 2014, 53). It underlines that further contemplations on causalities are infeasible.

The third approach focuses solely on the interactions between researchers and stakeholders having a logic of a proxy measure which can easily be incorporated into the idea of societal

---

[1] Mode 1 and 2 refer to the structure of the science system, in which knowledge is produced. Mode 2 transfers knowledge production from traditional disciplines to plural and interactive system focusing on profiling research problems in an interdisciplinary manner.

sectors. Productive interaction is the best example of this approach (e.g. Spaapen & Drooge 2011). It is more a comprehensive framework than an assessment tool. Jack Spaapen and Leonie van Drooge (2011) assume that interactions form categories which can be used to create a more robust set of measurement instruments. The same idea has been introduced into the social sciences (e.g. Esko et al. 2012). Reetta Muhonen et al. (2018) have taken the step to combine the interaction model with pathway logic in respect of the impact of the social sciences and humanities. It brings interactions closer to the first strategy of impact. Basically, the models of the impact of the social sciences are not significantly different from their general development. They are mere modifications having the same verification logic.

Penfield et al. (2014) list four data sources for providing evidence of research impact: metrics, narratives, surveys/testimonies and external citations beyond academia. All these methods have their strengths and shortcomings. Claire Donovan (2008) argues that there have been three phases in societal impact evaluation: technometric, sociometric and the case study approach. Only the latest, the case study, combines several methods and verification strategies into one assessment model, which can be used flexibly in a range of contexts. The case study approach implicitly represents the development towards more pluralistic assessment methods based on constructive thinking (cf. Aledo-Tur & Domínguez-Gomez 2016). It is also often participatory and negotiating, in the sense that the academic community may take part in the design of the evaluation and interpretation of knowledge.

The case study approach has not been able to solve the tension between metric-based outcomes and processual narratives. Altmetrics has begun to take its place in impact assessment (e.g. Bornmann 2017). Altmetrics usually means a mixture of quantitative methods to assess societal impact retrospectively. Most commonly, the metrics focus on citation data on social media (e.g. tweets), online publications (e.g. Wikipedia) and policy documents, but also on peer-reviewed meta-data. However, altmetrics pushes impact assessment towards the one-size-fits-all type of model causing the meaning of disciplinarity, locality and specificity of impact to fade, which the policy-makers seem to prefer. Moreover, in practice, case studies often utilize academic bibliometrics instead of social metrics, making it more difficult to understand the verification logic of impact. The pursuit of normalizing impact criteria and evaluation mechanics also encounter academic peer-review, which is often used as expertise in case studies and other forms of panel assessment (Derrick & Samuel 2017; Derrick 2018, 11).

Theories of impact tend to emphasize linear feedback systems in which research outputs together with stakeholder interaction and outcomes enable impact (e.g. Penfield et al. 2014). The mechanical feedback approach fits poorly in many fields, especially in the social sciences and humanities. Not only do they enter society through research outputs and outcomes, they also define the social problems under public discussion and the proper approaches to discuss these social problems. This double hermeneutic perspective cannot be reduced to simple feedback, pathways nor interactions. Impact as measurable or documented sociopolitical utility becomes a less relevant question than reconsidering the purpose and design of impact assessment.

2.3. Liminality in research impact assessment

As the interest in this study is expert justifications for diverse ways to assess research impact, there is an investigation of these cross-cutting preferences in the perspective of liminality. The concept of liminality derives from an anthropological background in which it has described transitional states of rites. From this tradition, it has transferred to sociological and political studies. Liminal sociopolitical situations have ambivalence of meanings that strive them for new interpretations (Giesen 2015).

Peter Dahler-Larsen (2011, 16) have seen evaluation as liminal sociopolitical practice that creates extra space for interpretations and unusual perspectives apart from everyday institutional practices. The extra space between practices is artificial in the sense that evaluation is purposely created for additional interpretations. The artificial nature makes evaluation procedures controversial because they provide optional ways and methods to distance oneself from everyday experience leading to ambiguous identities and definitions. There is uncertainty and potentiality to create new ways to interpret practices. The controversy of evaluation considers, particularly, diverse interpretations of assessed practices and the ways to assess these practices. Furthermore, there is the uncertainty between official rhetoric of evaluation and actual implementation procedures.

This study concerns liminality through intersecting science political, administrative and academic interpretations of impact assessment. Impact assessment is considered to be an optional tool developed to communicate between scholarly endeavor and politics. Understanding impact assessment has social and methodological liminalities because various preferences and justifications of impact create intermediate interpretations between several fields of practices: the evaluation profession, academia, politics and administrative rationality. Impact assessments attempt to merge these interpretations in a practice of assessment. Liminality has also an effect on how scholars make academic judgments on research between traditional criteria for quality and new criteria for social accountability (Watermeyer & Chubb 2018). Pertaining to the theoretical discussions on research evaluation, there are four possible tensions in impact assessment: local/contextual vis-à-vis global knowledge; social control vis-à-vis academic autonomy; social research vis-à-vis impact assessment methods; and ideals of constructionist assessment approaches vis-à-vis handcrafted guidelines. The concept of liminality has been employed in this study to understand impact assessment preferences and their dilemmas through these tensions. The empiric analysis attempts to deepen the understanding of liminality in assessment in the sense identified by Dahler-Larsen (2011).

3. Impact assessment in Finnish research policy and evaluation

3.1. Enhancing rationality and legitimation

This overview outlines the policy lines and assessment practices in Finland. The overview is based on selected policy documents (see policy references) of the main research policy and evaluation institutions: The Ministry of Education and Culture (MEC), the Academy of Finland (AF) (a governmental funding body for scientific research in Finland) and the Finnish Higher Education Evaluation Centre (FINEEC). They describe the purpose of impact assessment through jargon of science policy, which is mentioned below.

The European Commission has framed the understanding of societal impact by emphasizing governmental objectives of the knowledge-based economy and strategies to tackle complex social problems: grand challenges (EU 2010; 2012; 2014; 2017). In Finland, the government's impact policies have meant shaping the research infrastructure and funding for the purpose of societal collaboration and politicization of research agendas in relation to national socio-economic aims (e.g. Finnish Government 2013).

According to the Ministry of Education and Culture (MEC 2015b) and the Academy of Finland (2016), establishing research profiles for impact is an incremental part of the strategic thinking of the universities' higher education policy. They suppose that profiles will increase research impact throughout. The Ministry emphasizes university profiles and their tasks in the sense of concentrating on strong research fields which focus on common problems, regional strengths and closing units that have weak academic and social performance. Interaction and collaboration with multiple stakeholders is an essential part of well-established research profiles for impact (Ranki 2015). However, the Ministry prefers collaboration between universities over competition (MEC 2015a).

The Ministry's research report (MEC 2015b) emphasizes the strategic management of impact. Not only supporting ready strategies, but also predicting future strategies has become the meaning of research impact assessment (e.g. Ranki 2015). Universities may be proactive in creating infrastructure and conditions leading to an anticipated change in an uncertain society (Heikkilä & Jokinen 2015). This proactive approach takes a step forward from only measuring outcomes towards monitoring processes. Impact assessment is supposed to provide information on what works and what does not. Impact assessment provides knowledge on knowledge production itself. This means constant monitoring of progress and verification of the "right" courses of action in research and university administration.

Impact also means general "excellence" in the policy discourse of the Ministry and the Academy of Finland emphasizing the legitimation purpose of the concept (e.g. MEC 2015a). Societal impact has a multitude of meanings to support, justify and create strategies along the general legitimation purpose of university funding (e.g. Aarrevaara et al. 2015). Thus, impact assessment has a two-fold purpose: enhancing rationality and enhancing legitimation (cf. Ahonen 2015).

3.2. Self-organized impact assessments

Perhaps the most rigorous system of impact assessment is the Research Excellence Framework (REF) in the United Kingdom which has a clear emphasis on societal impact (REF 2018). In addition, the Netherlands has incorporated a criterion for valorization for their universities in the Standard Evaluation Protocol (SEP 2016). Like Sweden and Norway, Finland has not established clear practices and standards for societal impact, although it has been interested in formalizing impact assessment (e.g. Swedish Research Council 2015). In similar fashion to Sweden and Norway, Finnish assessors have preferred to use independent university assessments and evaluations of research councils for advisory purposes (see e.g. The Research Council of Norway 2018).

Finland has incorporated impact assessment at different levels of the research evaluation system based on a dual funding system (block grants from the government and competitive funding from public and private institutions). FINEEC has set guidelines for universities on impact and its quality judgements (FINEEC 2019). FINEEC has listed criteria for advanced societal impact emphasizing regional collaboration and strategic research development. FINEEC's guidelines suggest that the universities would set clear goals for impact and encourages them to develop monitoring mechanisms themselves. FINEEC admits the challenges of the impact assessment but has no suggestions in its directive to tackle these challenges by assessment designs.

The Ministry of Education and Culture has employed an application of impact evaluation in a form of university strategies in its funding scheme (MEC 2015a, 34-36). Strategic development of the universities is based on agreed outcomes and goals, which they report to the Ministry. There are no specific indicators for this, and such a goal can be local science education development, for example.

The universities have an obligation to undertake independent research impact assessments fitting into their own periodic strategic planning every six years and to use them when having an audit for FINEEC and the Ministry (FINEEC 2017). The universities are given free rein to design the assessments, but they must follow a few basic guidelines such as describing how societal interaction supports their strategy. Evaluators have often implemented the case study approach with self-assessment reports and panel assessments (e.g. Saari & Moilanen 2012). Evaluators have usually constructed the assessments as faculty level self-assessments providing detailed information. The purpose of the assessment exercises is developmental and to give directions to strategic planning linked to the goals agreed with the Ministry.

Most of the funding institutions have included impact criteria in their funding programs prospectively and many of them evaluate the impact of their research programs and projects retrospectively by utilizing diverse quantitative and qualitative methods (e.g. Hjelt et al. 2009). For example, the Strategic Research Council allocates funds purely for research relevant to the government's goals. These programs are planned, monitored and evaluated tightly according to the impact it generates (Mickwitz & Maijala 2015). The impact of the programs is based on interaction and co-creation with stakeholders.

To sum up, the institutional criteria for impact assessment are imprecise, as much as the guidelines and practices are diffuse. Although FINEEC has established basic guidelines, the universities do not have common standards on how these methods should be used or according to what design. On the contrary, there is considerable incoherence concerning what impact is and how it should be verified. In this respect, the Finnish impact assessment of universities is unstructured compared to the English and Dutch evaluation models. The impact assessment practices in the research evaluation system are internalized by academia and research funding institutions emphasizing self-control of these actors. Regardless of the diffuse standards, impact is required at all levels of research evaluation: university, faculty and individual researchers.

4. The data and method

Interview material finds answers to how specific experts justify impact assessment through their personal experiences and professional perceptions: what preferences guide their overall perceptions and what problems do they expect to encounter? The empirical material includes 14 semi-structured interviews (13h 8 min; 94 pp.) combining face to face, telephone and email communication. The interviewees were selected due to their expert position, professional status and representative academic field. The interviews were in either Finnish or English in 2016. They were recorded and transcribed for further analysis. Lastly, the Finnish quotes were translated into English.

The experts represent different points of view of research evaluation. They have backgrounds either in social research, evaluation studies or implementation of research evaluation. The interviewees were research policy intermediaries, evaluation scholars, and academic professors. They were selected in respect of their experience in research evaluation, evaluation methods and/or profound understanding of social research methodology. Two of them came from outside Finland (one evaluation scholar and one policy expert) for a wider cross-national perspective of European policy on research impact. They also discussed Finnish policy and assessment. This paper is the first part of a research project focusing on research impact assessment.

By using an organic interview sample and the snowball method to gather the data, the aim of the study was to find an overall view in the discussion on the purposes and methods answering to those purposes. By interviewing informants representing a specific perspective and institutional position, it was possible to find diverse arguments over impact assessment and its design. Policy documents can show similar traits on assessment methods, but they tend to be formal and conceal any practical problems experienced. Selection of many of the informants was based on a recommendation from other professionals or an informant. The informants can be considered to be reliable sources of information in their field and knowledge at the time of their selection.

The informants were firstly asked to describe their personal experience and interpretation of specific methods and designs of impact assessment, and secondly, they were encouraged to explain the grounds for their claims and critics. The interviews were structured according to key topics (e.g. methods, expertise and problems) derived from the literature and policy. The informants represented their academic and professional expertise in research evaluation and in the social sciences. The analysis considered impact assessment from a broader perspective than from the informants' individual academic discipline and profession, since interdisciplinary and inter-sectorial characteristics of impact assessment. The interviewees were categorized into three areas of expertise and were also used as narrative codes: policy intermediaries, evaluation scholars and academic professors. This coding helped to locate the position and expertise the informants have in relation to impact assessment, regardless of many of them having overlapping experience of these positions. The validity of the informants' answers was interpreted against the policy and research evaluation setting.

The analysis was undertaken by using the Atlas.ti 8 program. The analysis was conducted in two parts to answer the research questions. The first part was discussion about the aims

of impact assessment. The second part described how assessment methods support the aims. The analysis was conducted thematically by the following steps (e.g. Attride-Striling 2001). First, the interviews were coded descriptively to categorize similar issues in the data, for example, about quantitative metrics. In the second phase, these codes were re-examined on a more conceptual level in relation to issues and problems rising in the literature, for example the technical characteristics of altmetrics, their justifications and critics and their social purpose (e.g. altmetrics utilization, purpose and credibility). Thirdly, the conceptual codes were transformed into general subthemes (e.g. social control/internal control) by utilizing the analysis framework on possible social and methodic preferences and their liminalities. Finally, the subthemes were conceptualized into main themes (guiding principles) of impact assessment by examining how the driving preferences and their dilemmas (e.g. external transparency/internal capture) form thematically meaningful entities and respond to each other. The analysis used the interpretative/constructive approach. The analysis presented an interpretation of possible assessment principles based on diverse preferences. These preferences can vary contextually.

5. The main tasks of impact assessment

The experts implicitly revealed four guiding principles of impact assessment: stimulation of impact thinking, interaction loop, integrating by formalized methods and framing impact. The first and second are discussed in this section. These two principles echo the rationality and legitimation aims of impact assessment in the Finnish research policy and evaluation. In a sense, they are the purpose of impact assessment reconciling with the sociopolitical expectations of impact assessment. The interviewees discussed these principles from different stand points, having different modes of justifications. The similarities and differences in the arguments occurred regardless of the category the interviewee was in.

5.1. Stimulation of impact thinking

The interviewees brought out the importance of encouraging researchers to think about the impact of their research, which would eventually help them to state the impact and normalize societal impact as a part of research activity. One can call it the art of the **stimulation of impact thinking**. It underlines societal impact as a normalized routine of research and makes it a discursive preoccupation. Impact is not understood as merely an additional quality criterion (cf. Nowotny et al. 2001, 46), but a way to comprehend scholarly work throughout. In the following quotation, one professor described this preoccupation by how impact assessment forces universities to think about their impact to make it transparent for stakeholders, even though researchers generate impact regardless of incentives.

> The biggest effect is that faculties must think about impact and what makes the impact visible. It is so built-in to a researcher's daily life. They show up on TV and write in newspapers. They have always done impact, but now they must make it visible. Allocation is not the big deal, but that researchers must think about impact. (Professor 1)

In practice, the incorporation of impact thinking is implemented through incentives, for example in funding programs which guide researchers to frame their activities in relation to social questions or the work of other researchers. Another informant employed the idea of self-organizing and bottom-up management to impact assessment as an incentive. This idea is closely connected to the previous one of nudging researchers to think about impact but having a responsibility to indicate this thinking.

> In bottom-up funding, we encourage researchers to recognize wider connections and why they are doing their small piece of work in relation to others and should they do it a bit differently when it is in relation to the bigger picture of research. This is how we could get a bit more relevance. I see this in the point of view of self-organizing, so that we could increase incentives to think about the bigger picture and have a little more direction for research. (Policy intermediary 3)

Impact thinking becomes more structured and guided when it is applied to funding programs which are explicitly oriented to social problems. A third informant explicitly connected self-organizing to tailor-made structures and incentives to help researchers to understand their interaction with other stakeholders. These ready frameworks are usually intrinsically in the funding formats aiming for societal collaboration. Policy intermediaries' arguments underline pre-established standards or frameworks to help researchers to find the self-motivated direction for impact. These ready frameworks are persuasive.

> In our mission-oriented projects, we utilized a bottom-up style approach successfully in basic research applications so that we asked the researchers how they could utilize their research societally. This was an extra incentive for the researchers to think how to valorize their own research. (Policy intermediary 5)

On the one hand, this principle represents lightly managed societal activities without imposing the high pressure of performance. On the one other hand, there is a presumption that researchers need to make formal statements to make impact visible. The purpose of this preoccupation is to monitor and verify societal activities by indirect or direct incentive mechanisms in public funding by increasing normative pressure of having impact. The policy intermediaries agreed that societal impact is stimulated through impact frameworks such as pathways and even encouraged through specific funding and assessment devices. However, some of the academic informants explicitly pointed out the harmful side of the priority control mechanism behind the bottom up and self-organizing orientation of impact assessment emphasizing short sighted goals:

> If impact assessment is done poorly, it will lead to a decline in theoretical social research rather fast. […] Imposing restrictions on the direction of research easily defeats the purpose of the desired research impact. It would be good if these practical utility indicators could also favor abstract thinking, because after all, there is an

interest in having universities practice abstract thinking, by which we can organize society in a way that could not be imagined in laymen's terms. (Professor 5)

"Abstract thinking" refers to social research, which distances itself from tailor-made directions by indicators. Abstract thinking does not clearly oppose impact thinking but seems to require autonomy from pursuing dictated directions of political agendas. The informants thought that the downside of impact thinking is that the research evaluation system does not recognize the burden on researchers to report and verify impact continuously in their limited working time, although it emphases self-organized preoccupation of researchers. They understood social activity indicators as external control disturbing scholarly endeavors. From the researchers' point of view, impact assessment seems to suffer from the apparent need to measure fetishes, having precise information on phenomena that indicators cannot show. The constructive aspects of the assessment incentives are related to researchers' responsibility to be more open and to communicate their research.

Hence, the informants struggled between **external social control, internal self-control and undirected orientation** when finding an interpretation of impact thinking. The informants preferred impact thinking that is naturalized in scholarly work, but at the same time, they experienced impact assessment as external social control, which nullifies their scholarly endeavors. The policy intermediaries attempted to tame this threat by emphasizing self-organizing of basic research. This preference indicates internal self-control by established frameworks and priorities as signposts of accountability (cf. Nowotny et al. 2001, 46-47). The liminal interpretations of impact thinking creates this dilemma. Incentives by assessment take impact thinking closer to the problems of transparency (Strathern 2000). Preoccupation with impact relies on understanding societal impacts as a natural part of conventional research practice, although through limiting audit mechanism. There is also a conflict between academic quality and societal impact judgements, as a criterion for impact is not part of a traditional review process (cf. Watermeyer & Chubb 2018).

5.2. Interaction loop: monitoring, verifying and learning from impact

The experts believed that interaction within academia and between researchers and public and private stakeholders is necessary in respect of generating an understanding of possible impact through a dialogue with societal actors. Thus, in this study, this principle is understood as an **interaction loop**, which is located on two levels in the discussion. First, interaction is a guiding tool helping to produce impact by specific funding criteria, as in the following quotation about policy intermediary highlights.

Impact can be seen in the first round of strategic funding, in the sense that some of the money is designated for interaction with the end-users, meaning €5 million in three years. We did not have that sort of money before. […] In this kind of interaction, knowledge can be utilized immediately and if it exists already, the stakeholders will notify the researcher. (Policy intermediary 5)

In this sort of interaction framework, the research process is co-created with the stakeholders and research participants. The funders have pre-established priorities, a research process guided by steering groups and immediate measured interaction. Along the internal control of research process, interaction also has a second higher level of control. One informant described interaction as a meta-goal of research, which can be promoted through research assessment. As a meta-goal, it includes values of simplified communication with policy-makers and other stakeholders, transparency and openness.

> It is the interaction between researchers and other social worlds: the benefit of impact indicators as evaluation is their role as 'boundary objects', which enables the collaboration and critical dialogue between the worlds. And the interaction between the assessor and the principal helps us to decide what kind of indicators would be the most feasible for this purpose. (Scholar 1)

Interaction as a meta-goal can be directly linked to co-creation, as funding programs may assess their research projects and collect data from the programs. The scholar understood impact assessment as a communication tool between social worlds. In this sense, impact forms standardized tools for interaction. One informant brings out how knowledge on impact mechanisms can help to enforce impact itself. Impact assessment produces information to orient research proactively in societally-meaningful directions and becomes more of a learning process between the academy, societal sectors and policy-makers.

> It is a two-way interaction process that we illuminate through the kind of pathways research can connect to society. It is important for researchers to understand impact transfers to societal process. Case studies could make these processes visible: show the various mechanisms and support those mechanisms further. (Scholar 4)

The mechanism means a loop in which the assessment information guides the planning of the future research priorities by showing what kind of collaboration works. The informants tend to think that it is possible to establish a sort of interaction loop based on this mechanism. However, there is tension in the pursuit of dialogue and transparency of research illustrated in following quotes. This tension considers explicitly the insecurity of academic self-control, which the overly emphasized transparency may hamper. The openness and visibility turn to terror of transparency not only by an outcome-oriented approach of assessment, but also by pre-established and controlled interaction processes (cf. Strathern 2000). As one professor pointed out, intensified interaction with policy-makers may also hamper the intrinsic motives of scholarly endeavor.

> Yes, impact assessment increases interaction. But the downside of this kind of openness and interaction is the danger of blind spots. This is an argument of the social scientist: when we have discussions in any groups, we form a common perception of the directions we want to strive for. And that direction is not necessarily the one where we ought to go with respect of new knowledge. (Professor 3)

The professor hinted that transparency may have a restricting effect on scholarly endeavors, stressing worries related to social control of the research process and priorities. Another informant pointed out that impact assessment devices are supposed to illuminate research investigation as a communication tool but not interfere with the academic internal control mechanisms:

> They must somehow know, but they mustn't interfere with how research is funded, because they don't necessarily see what is really needed, for example in methodology or in history studies. Those needs enable something more tangible later. If having more openness, citizens could discuss more research and somehow participate in it. (Policy intermediary 3)

Preoccupation with interaction loops seems to be justified by democratic openness and transparency. The ideal of transparency justifies monitoring mechanisms orienting researchers' impact thinking. Impact assessment is expected to produce simple and efficient information and dialogue with policy-makers. Monitoring interactions may enhance the transparency of academic research in the sense that policy-makers and the public are aware of what kind of research is funded and what "public good" it may provide. Although the informants thought that impact assessment might enhance the rationale of proactive respondence to social change by meta-evaluations and feedbacks, they tended to give assessments mere legitimizing role in research policy.

This idealized assessment mechanism has tension between **transparency** and **internal capture** of research agendas in interpretations of communication via impact assessment. Transparency limits the discursive space of the social sciences and humanities in interaction processes. Stakeholders tend to use already-naturalized concepts and language, which may hamper the development of new knowledge by narrowing down perspectives. The double hermeneutic, self-reflective and emancipatory nature of scholarly endeavor is jeopardized in intense monitoring processes. The degree of **concealment** of scholarly practice remains an unsolved problem.

6. Integration of liminal assessment preferences

The second part of the analysis illuminates how the informants justify the most commonly-used assessment methods. These justifications show diverse, disputed and incoherent arguments, which form guiding principles of their own vis-à-vis the previously described aims of impact assessment. Methodical choices attempt to tackle tensions between ideals of methodological credibility and simply trimmed legitimation and formal rationality of impact.

6.1. Quantitative metrics showing realistic impact?

The informants had two-fold arguments about impact assessment methods, which consider basic forms of altmetrics, narrative reporting and case studies combining miscellaneous data often assessed by panel experts. They pursue both **credibility** and **functionality** of

assessment in the sense of concordance with policy-makers and stimulating effect of impact assessment on researchers. According to the informants' arguments, the credibility of these assessment methods is constantly on the line. However, credibility itself does not bother all the informants, but how the methods handle the purpose of impact assessment is important.

According to one informant, altmetrics of online occurrence try to demonstrate how popular a researcher is on internet platforms, as in the extract below. In this sense, popularity is an incentive for researchers, and it means social presence and fame, which is measured through social media citations such as Facebook shares. It is thought that this indicator of social presence will help to guide research socially and extend researchers' academic profile, which is traditionally valued through publications.

> I think that this is also an interest of researchers because now their merit system is based solely on publications and citations. With altmetrics, we could influence the merit system and researchers' profiles. On the other hand, the funders' interest is also to see what the broader impact on society is and who benefits from research. Regarding societal impact, we definitely need many indicators for different fields. (Scholar 2)

The informants disputed whether altmetrics have a real (positive) effect on researchers' motivation and ability to generate more impact. Some informants thought that altmetrics contradict the purpose of impact assessment because standardized incentives may cause unwanted behavior (e.g. Campbell's law[2]). In general, they saw indicators as a means to stimulate researchers to be socially active according to a preset framework, though none of the interviewees claimed that researchers would act outside the interests of society.

On one hand, many of the interviewees saw that quantitative data are reliable for realistically interpreting impact regardless of the problems that erode their credibility as a method to show broader societal impact, which the policy-makers so desire to have. One of the policy intermediaries complained that it is difficult to show the causality and logic between impact and individual social media events or other forms of interaction, as the following extract shows:

> If we measure, we ought to have some feature that is measured. We can't measure a researcher's impact because it is not a feature of a project, but it is a consequence in the sense that end-users are having activity. Research impact is on the end-user's activity. This is my opinion, perhaps more than others think. (Policy intermediary 3)

This point of view emphasizes innovative research ideas rather than measurable feedback from end-users (cf. Penfield et al. 2014). The informants had difficulties seeing what quantitative social indicators mean regarding academic research: what is the information

---

[2] The more any quantitative social indicator is used for social decision-making, the more it will be subject to corruption pressures and the more apt it will be to distort and corrupt the social processes it is intended to monitor (Campbell, 1979, p. 85).

they provide? They stressed that academic fields need diverse and plural indicators because societal impact is a complex phenomenon and because differentiated academic fields have divergent perspectives on reality. The plurality means a multitude of possible interpretations of impact. Along with the problems of causality, one of the interviewees expressed her concerns on how quantitative data communicate with other stakeholders, as the following extract shows.

> No, indicators do not increase transparency. If measuring such dull metrics, no one cannot even understand them. […] Problem is that they can dictate research directions. […] Where can one find reasonable meters because they tend to become a normative guide?

The question around altmetrics and other social meters remain dubious. The inconclusive arguments, experiences and feelings on social indicators seem to emphasize the standardized guiding effects of indicators, for better or for worse. The informants seemed to be cautious about the normative guiding effect of indicators and about the overemphasis of researchers' social accountability. They had insecurities about which indicators would work in the end. Indicators seemed to meet the purpose of measuring popularity and building a merit system but it was difficult to interpret impact in society and enhance communication with policy-makers. The informants' notions suggested that indicators may provide pretentious information on social presence and fame but be lacking in liminal communication between scholarly work and politics.

6.2. Qualitative narratives supporting the complexity of impact

Many of the informants believe that narrative reporting by researchers can avoid mechanistic interpretation problems of indicators. The informants tend to lean on the general fashion of narratives referring to the Research Excellence Framework (REF). Narrative reporting is not only more encouraging because researchers have more discursive space to describe impacts in a specific framework but also more communicating, as one policy intermediary describes in the following quote:

> The narrative method can describe the contribution of the research in language that is also comprehensible outside the academic community, for example among politicians. Moreover, narratives enable interweaving multiple studies together, in which case one can get rid of the problem of attribution of impacts. We live in a time of stories and a good narrative can be more convincing for decision-makers and the public. (Policy intermediary 6)

A few of the informants pointed out that the narrative method seems to suffer from a lack of standardization leading to several problems. According to them, narrative reports are not free of specific problems of interpretation, design and manipulation. Narratives also need to be constructed for the purpose and goals of assessment to increase the credibility of the communication. Another policy intermediary pointed out these structural issues below:

Narratives must be structured well. And they must be considered throughout, whether they are connected to research allocation, or whether they are understood as an attempt to legitimate the wider contribution of the research community. […] Impact assessment and allocation may limit free research activity of the network. Yes, we need these narratives, but the question is how we use them and how those are connected to the system. (Policy intermediary 3)

According to the policy intermediary, the narrative method encounters attribution and causality problems in ways other than those encountered by quantitative indices. Narrative reports show how complex collaboration can be. According to this logic, narrations should be limited and formalized to have meaningful interpretations and messages for stakeholders. Narrative reporting, therefore, is an attempt to formalize the form of communication about research impact between scholars and policy-makers. Policy-makers need simplified information, but researchers must also see this communication as meaningful.

The interviewees also considered a case study to be a feasible compromise which can lean on the diverse and contextual quantitative indices alongside narrative reporting. However, many of the informants pointed out that a case study tends to be too laborious compared with the benefits it provides. The case study is not free of methodic ambivalences. Below, a professor pointed out how a case study can grasp the complexity of a research process, and hence satisfy the scholarly demands for credible assessment, but at the same time, he was doubtful about the validity of the narrative method in a case study:

If they are not only statistical but are comprehensive, I can sympathize with these case models. No doubt they are good methods alongside indicators. Assessment must be multi-dimensional: quantitative and qualitative, complex. […] On the other hand, we have evidence on how easily we can be misled by these case narrations: one can use just the right choices of words in a case study. (Professor 3)

In conclusion, the main methodical concern seems to be that societal impact is disturbingly complex in respect of reliability and interpretation of research impact. As the interviewees pointed out, there are contradictory tendencies regarding the impact assessment methods. The methods are supposed to guide scholarly work and build a quality system but not standardize social collaboration according predictable behavior or social control. They should also increase transparency and communicate research impact to policy-makers and other stakeholders of research in a simple but truthful manner, but follow methodical rigorism and ideals. They should also be administratively reasonable regarding cost-efficiency. Thus, they functionalize impact assessment.

Methodical choices support not only the credibility of impact to safeguard research practices from overly-simplistic assumptions about research outcomes and interaction, but also help to orient researchers into social interaction with stakeholders and standardize the communication and interaction forms. One can describe this operation as **integrating** the

functions and credibility of impact assessment **by formalized methods**. Formalized methods help to constitute a liminal review between causal and constructed interpretations of assessment outcomes, and between standardized and creative forms of social interaction. However, the ideals of methodical rigorism often result in unsatisfying compromises between political, administrative and disputed scholarly preferences.

6.3. Meta-expertise across discipline boundaries

This section describes the final guiding principle manifested by the informants: **framing impact** through meta-expertise. This principle is characterized by the importance of having experts interpret data on impact comprehensively in any assessment. The assessment data should be transmuted in the sense that assessors make it understandable. The interviewees emphasized comprehensive (non-linear and qualitative) understanding of impact through peer review panels, as one evaluation expert underlined:

> I would say that impact assessment is expert work. It is the job of an expert to interpret. And it is exactly an expert group which knows the research system, interprets information and indicators. And in a manner of speaking, that is their expertise: how well they can interpret impact. There cannot be any mechanical indicators. This fact leads us to expertise, which is a kind of human capital. (Scholar 3)

This approach highlights expert-driven peer review against pre-determined quantitative evaluation mechanics (Derrick 2018). Impact outcomes are generated in a process of expert deliberation representing a more constructive approach to impact assessment. However, the informants perceived that the success of the expert interpretation depended heavily on the competence of the panel, their skills and interaction, but also on their creative skill and capacity to compensate for the information gaps during the assessment process. There always seems to be a risk of impact being attached as a superficial label after trivial skimming. The interviewees had no clear idea how to compose an ideal panel for an impact assessment, but surprisingly they thought that external representatives are needed. One informant would solve this bias with clear standards:

> But a panel is a panel. No matter how well you compose it, biases are inevitable. The key is to select experienced people who understand academic work but also understand how society functions. You need to develop understandable criteria on what impact is, such as having some reliable benchmarks. (Professor 4)

An opposing approach to standards emphasizes the dialectical relationship between diverse experts. The informants emphasized academic experts who have understanding and personal experience in societal sectors and their functions, but also the opportunity to have academically-oriented civil servants or other external experts. Possibly there should be interaction between academic and societal experts, but the main argument is a

comprehensive understanding and the interaction between academic experts who can positively influence each other, as one professor describes in the following quote:

> […] Panels function well when they are based on dialogue over the boundaries of expertise and when the panelists are not defending their narrow field of expertise. In layman's terms, we may speak of open-mindedness. And I don't mean polite open-mindedness in the sense that panelists just pretend to be curious and are able to argue over boundaries. I would say, one of the main skills in this kind of impact assessment is that one can think metaphorically about familiar concepts. (Professor 3)

According to the interviewees, a panel should follow what is here described as "boundary expertise". Boundary experts are scholars who are close to the scholarly discipline operating in the similar field and having a skill to openly review similar but disciplinary unfamiliar social phenomena in the sense of meta-expertise. The reviewers ought to be interdisciplinary virtuosos. However, this meta-expertise is based on a disciplinary foundation, as a research manager pointed out:

> There is always a danger that one has insufficient understanding of the nature of specific disciplines. But if disciplines are defined too narrowly, they cannot be compared to each other. Then there would not be a reasonable way to aim resources in the direction of the highest utility. (Scholar 4)

According to these interviewees, interdisciplinarity has a major role in impact assessment and breaks down old disciplinary boundaries. The purpose of such meta-expertise seems also to be monitoring differentiated disciplines regarding wide research policy priorities. Meta-experts ought to be "watchdogs" of their fellow academics, as an informant hinted in the following quote:

> My suggestion is to use experts from neighboring fields as much as possible, in the sense that they wouldn't be too close [to the assessed field]. The best possible assessor is the one who can understand enough but can take a couple of steps back and see the big picture. If we want to have an assessor who understands the discipline assessed as well as possible, they are only able to see whether we are doing well enough. (Policy intermediary 1)

The informants underlined impact assessment expertise, which considers mainly how the peer review panel can coordinate ideally between measurement-based outcomes and personal interpretations (cf. Derrick 2018, 11). Panel review has liminality between the level of expertise (**disciplinary**, **interdisciplinary** and **inter-sectorial** expertise) and between forms of expert valuation (**standards** versus individual **experience**). Ambiguously, the informants emphasized both personal experience and strict standards, which could self-regulate the understanding of impact through orientation. Interestingly, three of the informants highlighted interdisciplinary accountability in impact assessment. In this sense, interdisciplinarity is a form of an internal self-control for epistemic

judgements on social impact (Huutoniemi 2012, 24-25). However, the informants also supported external representatives on panels, but they did not have a coherent idea how this would work out. It seems that interdisciplinarity as a form of academic self-control is easier to comprehend than external social control indicating lack of assessment culture for such conceptualization (cf. Derrick & Samuel 2017). Ostensibly, academic differentiation seems not to fit in the frame of impact expertise emphasizing multidisciplinary aims of knowledge, but, yet, disciplinary knowledge and understanding had their place in the informants' views.

7. Discussion: Observations on the tragic uncertainty of impact assessment

7.1. Problems with the rationality of impact assessment

Impact assessment in Finnish science policy serves as a conduit for national policy strategies (cf. Whitley 2014). The concept of impact has a circular reasoning of predicting strategic choices and legitimizing science by demonstrating its success. Institutional impact assessment is extensively internalized and self-organized having a high degree of devolution within a limited and guided understanding of its interpretations (cf. Hicks 2012).

The interviews revealed four guiding principles of impact assessment: **stimulation of impact thinking**, **interaction loop**, **integrating by formalized models** and **framing impact** (Table 1.). The first two are the actual purposes of assessment and the latter two balance diverse preferences on what assessment should represent and how it should be done. Orienting research impact encounters liminal interpretations between external social control, internal self-control and undirected orientation, as well as liminal judgements between societal impact and academic quality. The informants argued that ideally impact thinking would be normalized in daily research work. However, they found it difficult to adopt impact judgements into scholarly endeavor through assessment performance. Academic quality and impact seem to have inner contradictions because impact assessment methods may contradict academic purposes.

Table 1. Guiding principles of societal impact assessment and their dilemmas

Guiding principles	Dilemmas	Examples of methods
Stimulation of impact thinking	External social control / internal self-control / undirected orientation Academic quality / societal impact	Altmetrics as an internalized quality system and as external disturbing metrics
Interaction loop	External transparency / internal capture / conceal of research	Narrative report as a communication tool of interactions and as a directive device of naturalized social agendas
Integrating by formalized methods	Functionality / credibility	Case study as a functional tool for standardized information of impact and as a complex social research method
Framing impact	Disciplinary / interdisciplinary / inter-sectorial expertise Standardization / deliberation	Peer review as expertise in disciplinary knowledge and as meta-expertise in monitoring disciplines

The informants tended to think of impact as an additional criterion which can disturb academic work. Academic quality is often considered to be primary to impact (e.g. de Jong et al. 2015). Impact represents local objects of study in the sense of contextualized knowledge (Nowotny et al. 2001). However, the sciences have become increasingly globalized and institutionalized, standardizing local research agendas and perspectives on problems (Drori et al. 2003; Mosbah-Natanson & Gingras 2014). National socio-economic impact is over-represented in research policy, although global scholarly rationalities are defining these priorities. Realistic assessments should consider all levels of research impact (local, national and global), as local problems are often related to global rationalities.

As the rationality of impact assessment responds to demands of social accountability, it tends to emphasize research performance and prospective outcomes at the cost of processes and prospects (Hicks 2012). In this sense, the pursuit of transparency turns to internal capture of research agendas by pre-established agendas and vocabulary. Monitoring research processes not only become outcome-oriented, but also begin to determine the researcher's perspective on outcomes (cf. Strathern 2000). The sophisticated control mechanisms of impact assessment have the risk of trivializing problems.

The uncertainty of impact assessment is perhaps in the rationality of impact itself. It is pursued to solve social uncertainties through strategic research and research impact to have its own dilemmas. The guiding principles find ways to respond to the political rationality of impact: orienting towards strategic research priorities (stimulation of impact thinking), guiding and monitoring research processes and priorities (interaction loop), having tools to communicate them convincingly (formalized methods) and having clear interpretations for further strategies (framing impact).

7.2. Idealized assessment

The informants' methodological justifications and critics are diffuse, hinting that there is no clear picture of good guidelines. This remark supports a notion that their understanding of impact frameworks is largely based on examples of previous field trials and errors. Formalization of impact assessment seem not to build methodic soundness. Instead the methods seem to be a collection of possible choices responding to the tasks of impact assessment. They also seem to be compromises between credibility, political expectations and institutional readiness. In this sense, the methodical choices and how they are represented in assessments are mere ideals of well-thought verification logic turned into formal solutions.

The major benefit and problem of altmetrics is how they can create common standards of desirable social outcomes at the international level. However, at the same time, these indicators de-contextualize impact assessment in relation to local activities (cf. Feretti et al. 2018). Rather, they create a global system of merit parallel to bibliometrics (cf. Mosbah-Natanson & Gingras 2014) indicating a dim reflection of public discourse. Qualitative narrations enable contextual interpretations and messages to policy-makers, but lack generalizability of mechanisms. The case study approach cannot solve this only by mixing up methods, as it should also have explicit methodological justification for how it interprets assessment outcomes. Framing the assessment outcomes through expertise partly answers the problem of methodical choice but faces other dilemmas between standards, deliberation and level of expertise (cf. Derrick & Samuel 2017). Peer review of impact seems to have a risk of high academic interdisciplinary orientation instead of inter-sectorial interests in assessment.

In conclusion, the liminality of impact assessment between academia and politics and between politics, evaluation studies and academia confuses the design of a balanced assessment model. Designing impact assessment is sociopolitical. Knowledge production practices are not only conditional by knowledge assessment practices (Feretti et al. 2018). Assessment practices are also conditioned by liminalities, in which the evaluators, evaluation scholars and academia transgress not only new interpretations of research but also how these interpretations should be done (cf. Dahler-Larsen 2011, 16). However, these interpretations do not necessary address the social opportunities and characteristics of academic fields nor political interests in them.

7.3. Final remarks

This study was undertaken to understand the dissatisfaction and incompleteness of impact assessment by investigating social and methodological tensions through multiple expertise. The interviews provided a diverse source of material to draw a view about liminal interpretations of impact assessment. Liminality can help us to understand the guiding principles and problems common to divergent assessment designs and practices but it lacks details of the processes. It would be essential to know the processes and grounds for noteworthy decisions in research evaluation and what those decisions mean to divergent disciplines. The connection and process between science policy, evaluation studies and evaluation management should be studied critically in the future to gain an understanding of how assessment models are formed in this interaction and in a national context vis-à-vis international isomorphism.

The study presumed that formalized impact assessment encounters a multitude of liminalities. The study found four frameworks having several uncertainties. Assessors attempt to solve these uncertainties by formalized methods, but they cannot be treated as purely logical choices due to liminalities. The assessment models are often a result of previous mediocre solutions, such as linear outcomes, causing problems in redefining impact assessment mechanics. There are no simple answers for dealing with these uncertainties, as idiosyncratic assessment designs find unique ways. Yet, they can be facilitated by corresponding assessment to the social opportunities of academic disciplines.

The dilemma of social control can be addressed by moderating the assessment culture and concentrating on collaborative activities by bringing parties together and introducing innovative research findings (cf. Guston 2000). Impact assessment should correspond to contextual problems through global academic endeavor. Quantitative indicators should answer questions about contextualized research interaction and knowledge, for example, by focusing on specific data sources such as policy documents. Assessments should leave enough space for negotiation of diverse understandings of impact and new interpretations of problems, for example in narrative reporting. Avoiding "all-purpose" designs and opening the intentions of specific assessments could deal with standardization and credibility issues: expressing explicitly to whom and what purpose the assessment information is produced. Space for free collaborations, interpretations and open intentions could help to frame correct expertise.

References

Ahonen, P. (2015) 'Aspects of the Institutionalization of Evaluation in Finland: Basic, Agency, Process and Change '. *Evaluation* 21/3: 308-24.

Aledo-Tur, A., and Domínguez-Gómez, J. A. (2017) 'Social Impact Assessment (SIA) from a multidimensional paradigmatic perspective: Challenges and opportunities', *Journal of Environmental Management,* 195/1, 56-61.

Attride-Stirling, J. (2001) 'Thematic networks: an analytic tool for qualitative research. *Qualitative Research* 1(3): 385-405.

Bastow, S., Dunleavy, P., and Tinkler, J. (2014) *The impact of the social sciences*. Los Angeles: Sage.

Beck, U., Bonss, W., and Lau, C. (2003) 'The Theory of Reflexive Modernization: Problematic, Hypotheses and Research Programme', Theory, Culture & Society, 20/2: 1–30.

Benneworth, P., Gulbrandsen, M., and Hazelkorn, E. (2016) *The Impact and Future of Arts and Humanities Research*. London: Palgrave Mac Millan.

Bornmann, L. (2013) 'What is societal impact of research and how can it be assessed? a literature survey', *Journal of the American Society for Information Science and Technology,* 64/2: 217-33.

Bornmann, L. (2017) Validity of altmetrics data for measuring societal impact:A study using data from Altmetric and F1000Prime', *Journal of Informetrics* 8: 935–95.

Campbell, D. T. (1979) 'Assessing the impact of planned social change', Evaluation and Program Planning 2/1: 67-90. 10.1016/0149-7189(79)90048-X

Dahler-Larsen, Peter (2011) *The evaluation society*. Stanford, California: Stanford Univ. Press.

de Jong, S., Smit, J., and van Drooge L. (2016) 'Scientists' response to societal impact policies: A policy paradox', *Science and Public Policy,* 43/1: 102-14.

Derrick, G. E., and Samuel, G.S. (2017) 'The future of societal impact assessment using peer review: pre-evaluation training, consensus building and inter-reviewer reliability', *Palgrave communications,* 70:1-10.

Derrick, G. E. (2018) *The Evaluators' Eye: Impact Assessment and Academic Peer Review*. Cham, Switzerland: Palgrave Macmillan.

Donovan, C. (2008) 'The Australian Research Quality Framework: A live experiment in capturing the social, economic, environmental, and cultural returns of publicly funded research', *New Directions for Evaluation,* 118: 47-60.

Drori, G. S., Meyer, J. W., Ramirez, F. O., and Schofer, E. (2003) Science in the Modern World Polity: Institutionalization and Globalization. California: Standford University Press.

Ernø-Kjølhede, E., and Hansson, F. (2011) 'Measuring research performance during a changing relationship between science and society', *Research Evaluation,* 20/2: 131-43.

Eräsaari, R. (2009) 'Open-context expertise'. In: Peters, M. A., Besley A.C., Olssen, M., Maurer, S., and Weber S. (eds.) *Governmentality studies in education*, pp. 55-76, Rotterdam: Sense Publishers.

Esko, T., Tuunainen, J., and Miettinen, R. (2012)' Social impact and forms of interaction between university research and society in humanities and social sciences', *International Journal of Contemporary Sociology* 49/1: 17-46.

Feretti, F., Pereira Â., Vértesy D., and Hardeman, S. (2018) 'Research excellence indicators: time to reimagine the 'making of'?' *Science and Public Policy*, 45(5): 731–41

Firestone, W., and Herriot, R. E. (1983) 'The Formalization of Qualitative Research: An Adaptation of "Soft" Science to the Policy Word, *Evaluation Review*, 7/4: 437-66.

Giesen, B. (2018) 'Inbetweenness and Ambivalence'. In Horvath A., Thomassen B., and Wydra H. (eds.) *Breaking Boundaries: Varieties of Liminality*, pp. 61-71, New York: Berghahn Books.

Guston, G.D. (2000) *Between Politics and Science: Assuring the Integrity and Productivity of Research*. Cambridge: Cambridge University Press.

Hicks, D. (2012) 'Performance-based university research funding systems', *Research Policy,* 41/2: 251-61.

Huutoniemi, K. (2012) *Interdisciplinary Accountability in the Evaluation of Research: Proposals Prospects for academic quality control across disciplinary boundaries*. Academic dissertation. Helsinki: University of Helsinki.

Martin, B.R. (2011) 'The Research Excellence Framework and the 'impact agenda': are we creating a Frankenstein monster?' *Research Evaluation*, 20/3: 247-54.

Miettinen, R., Tuunainen, J., and Esko, T. (2015) 'Epistemological, artefactual and interactional-institutional foundations of social impact of academic research', *Minerva,* 53/3: 257-77.

Molas-Gallart, J. (2012) 'Research Governance and the Role of Evaluation: A Comparative Study'. *American Journal of Evaluation* 33/4: 583-98.

Mosbah-Natanson, S., and Gingras, Y. (2014) 'The globalization of social sciences? Evidence from a quantitative analysis of 30 years of production, collaboration and citations in the social sciences (1980–2009)'. Current Sociology 62/5: 626-46.

Muhonen, R., Benneworth, P. and Olmos J. (2018) *From productive interactions to impact pathways: Understanding the key dimensions in developing SSH research societal impact*. Cheps working paper 02. <https://research.utwente.nl/en/publications/from-productive-interactions-to-impact-pathways-understanding-the> last accessed 28 December 2018.

Nowotny, H., Gibbons, M., and Scott, P. (2001), *Re-thinking science.* Cambridge: Polity.

Penfield, T., Baker, M. J., Scoble, R., and Wykes, M. C. (2014) 'Assessment, evaluations, and definitions of research impact: A review', *Research Evaluation,* 23/1: 21-32.

Spaapen, J. B. and van Drooge, L. (2011) 'Introducing 'productive interactions' in social impact assessment', *Research Evaluation,* 20/3: 211-18.

Strathern, M. (2000) 'The Tyranny of Transparency'. *British Educational Research Journal* 26/3: 309-22.

Watermeyer, R. and Chubb, J. (2018) 'Evaluating 'impact' in the UK's Research Excellence Framework (REF): liminality, looseness and new modalities of scholarly distinction. *Studies in Higher Education,* 44/9: 1554-1566.

Weingart, P. (1999) 'Scientific expertise and political accountability: paradoxes of science in politics', *Science and Public Policy,* 26/3: 151-61.

Whitley, R. (2014) 'How do Institutional Changes Affect Scientific Innovations? The Effects of Shifts in Authority Relationships, Protected Space, and Flexibility'. In: Whitley, Richard, and Jochen Gläser (eds.) *Organizational Transformation and Scientific Change: The Impact of Institutional Restructuring on Universities and Intellectual Innovation.* Research in the sociology of organizations 42, pp. 367-407. Bingley: Emerald Group Publishing Limited.

Policy references

Aarrevaara, T., Puukka, J., Ritsilä, J., and Wikström, J. (2015) 'Korkeakoulujen yhteiskunnallinen vaikuttavuus – vaikuttavuuden kanavat' [Societal impact of Higher Education Institutes – The Pathways of Impact]. In: The Ministry of Education and Culture (MEC). 2015a. *Vastuullinen ja vaikuttava: Tulokulmia korkeakoulujen yhteiskunnalliseen vaikuttavuuteen.* [Accountable and impactful: Point of views into societal impact of higher education institutions], pp. 78-98. <http://www.minedu.fi/OPM/julkaisut> last accessed 12 March 2017.

European Commission (EU) (2017) LAB – FAB – APP — *Investing in the European future we want. Report of the independent High Level Group on maximising the impact of EU Research & Innovation Programmes.* EU evaluations. <http://ec.europa.eu/research/evaluations/pdf/archive/other_reports_studies_and_document s/hlg_2017_report.pdf> last accessed 12 January 2018.

European Commission (EU) (2014*) Horizon 2020 in brief: The EU Framework Programme for Research and Innovation.* doi:10.2777/3719

European Commission (EU) (2012) *The Grand Challenge: The design and societal impact of Horizon 2020. Directorate-General for Research and Innovation.* <https://ec.europa.eu/programmes/horizon2020/en/news/grand-challenge-design-and-societal-impact-horizon-2020> last accessed 10 October 2018.

European Commission (EU) (2010) *Assessing Europe's University-Based Research: Expert Group on Assessment of University-Based Research.* Belgium: EU publications. <https://ec.europa.eu/research/science-society/document_library/pdf_06/assessing-europe-university-based-research_en.pdf> last accessed 28 May 2017.

The Finnish Education Evaluation Centre (FINEEC) (2019) *Laatu hallussa: Yhteenveto korkeakoulujen aditoinneista 2012-2018* [Quality in control: A summary of higher

education institutions' audits 2012-2018]. <https://karvi.fi/en/publication/laatu-hallussa/> last accessed 10 May 2019.

The Finnish Education Evaluation Centre (FINEEC) (2017) *Audit Manual for Higher Education Institutions 2018-2014*. <https://karvi.fi/en/higher-education/audits-higher-education-institutions-2018-2024/> Last accessed 10 May 2019.

Finnish Government (2013) *Valtioneuvoston periaatepäätös valtion tutkimuslaitosten ja tutkimusrahoituksen kokonaisuudistukseksi* [Government's decisions on the comprehensive reform for state's research institutes and research funding]. < https://vnk.fi/tula> accessed 10 May 2018

Heikkilä, K., and Jokinen, L (2015) Korkeakoulujen yhteiskunnallinen vaikuttavuus [Societal Impact of Higher Education Institutes]. In: The Ministry of Education and Culture (ed.) *Vastuullinen ja vaikuttava: Tulokulmia korkeakoulujen yhteiskunnalliseen vaikuttavuuteen*. [Accountable and impactful: Point of views into societal impact of higher education institutions], pp 32-46. <http://www.minedu.fi/OPM/julkaisut> last accessed 28 March 2017.

Hjelt, M., Ahonen P.P., and Pessala P. (2009) *Impact Evaluations of Finnish Programmes for Centres of Excellence in Research 2000-2005 and 2002-2007*. Publication of Academy of Finland 2, Gaia Consulting Ltd. <http://www.aka.fi/globalassets/awanhat/documents/tiedostot/julkaisut/2_09-coe-in-research.pdf> Last accessed 12 January 2018.

Mickwitz P., and Maijala, R. (2015) 'Strateginen tutkimus ja Strategisen tutkimuksen neuvosto' [Strategic research and Strategic Research Council]. *Tieteessä tapahtuu* 6. <https://journal.fi/tt/article/view/53324> Last accessed 10 May 2019.

Ranki, S. (2015) 'Korkeakoulun vaikuttavuus strategisen johtamisen näkökulmasta' [Higher Education Impact in the Perspective of Strategic Management]. In The Ministry of Education and Culture (ed.) *Vastuullinen ja vaikuttava: Tulokulmia korkeakoulujen yhteiskunnalliseen vaikuttavuuteen*. [Accountable and impactful: Point of views into societal impact of higher education institutions], pp. 236-58. <http://www.minedu.fi/OPM/julkaisut> last accessed March 2017.

REF (2018) *REF 2021: Draft Guidance on submissions*. <https://www.ref.ac.uk/publications/draft-guidance-on-submissions-201801/ > last accessed 10 December 2018.

Saari, S., and Moilanen, A. (eds.) (2012) *International Evaluation of Research and Doctoral Training at the University of Helsinki*. University of Helsinki. <http://www.helsinki.fi/julkaisut/aineisto/hallinnon_julkaisuja_81_2012.pdf> Last accessed 28 March 2017.

SEP (2016) *Standard Evaluation Protocol 2015 – 2021: Protocol for Research Assessments in the Netherlands*. <https://www.knaw.nl/nl/actueel/publicaties/standard-evaluation-protocol-2015-2021> last accessed 10 December 2018.

Swedish Research council (2015) *Research quality evaluation in Sweden – FoKus: Report of a government commission regarding a model for resource allocation to universities and*

university colleges involving peer review of the quality and relevance of research.
<https://www.vr.se/english/analysis-and-assignments/we-analyse-and-evaluate/all-publications/publications/2015-06-25-research-quality-evaluation-in-sweden---fokus.html>
last accessed 12 December 2018.

The Academy of Finland (the AF) (2016) Tieteen tila 2016 [The State of Scientific Research 2016]. <http://www.aka.fi/globalassets/30tiedepoliittinen-toiminta/tieteentila/aka_tieteen_tila_yksi.pdf> last accessed 11 March 2017.

The Ministry of Education and Culture (MEC) (2015a) *Ehdotus yliopistojen rahoitusmalliksi 2017 alkaen* [Proposal for the funding model of universities as of 2017]. Report and Appendices.
<julkaisut.valtioneuvosto.fi/bitstream/handle/10024/75157/tr19.pdf> last accessed 11 March 2017.

The Ministry of Education and Culture (MEC). (2015b). *Vastuullinen ja vaikuttava: Tulokulmia korkeakoulujen yhteiskunnalliseen vaikuttavuuteen*. [Accountable and impactful: Point of views into societal impact of higher education institutions]. <http://www.minedu.fi/OPM/julkaisut> last accessed 11 March 2017.

The Research Council of Norway (2018) *Evaluation of the Social Sciences in Norway: Report from the Principal Evaluation Committee*.
<https://www.forskningsradet.no/en/Article/Evaluation_of_social_science_research_in_Norway/1254020218541> last accessed 12 December 2018.