

Development and validation of a questionnaire to measure research impact

Maite Solans-Domènech^{1,2,3,*}, Joan MV Pons^{1,2}, Paula Adam^{1,2}, Josep Grau⁴ and Marta Aymerich^{3,5}

¹Agency for Health Quality and Assessment of Catalonia (AQuAS), Roc Boronat 81-95, Barcelona, Catalonia 08020, Spain, ²Epidemiology and Public Health Network (CIBER ESP), Roc Boronat 81-95, Barcelona, Catalonia 08020, Spain, ³Faculty of Health Sciences, Universitat Oberta de Catalunya (UOC), Rambla del Poblenou, 156, Barcelona, Catalonia 08018, Spain, ⁴Research Planning Unit, Universitat Oberta de Catalunya (UOC), Rambla del Poblenou, 156, Barcelona, Catalonia 08018, Spain, and ⁵eHealth Center (eHC), Universitat Oberta de Catalunya (UOC), Av. Tibidabo, 39-43, Barcelona, Catalonia 08035, Spain

*Corresponding author. Email: mtsolans@gencat.cat

Abstract

Although questionnaires are widely used in research impact assessment, their metric properties are not well known. Our aim is to test the internal consistency and content validity of an instrument designed to measure the perceived impacts of a wide range of research projects. To do so, we designed a questionnaire to be completed by principal investigators in a variety of disciplines (arts and humanities, social sciences, health sciences, and information and communication technologies). The impacts perceived and their associated characteristics were also assessed. This easy-to-use questionnaire demonstrated good internal consistency and acceptable content validity. However, its metric properties were more powerful in areas such as knowledge production, capacity building and informing policy and practice, in which the researchers had a degree of control and influence. In general, the research projects represented an stimulus for the production of knowledge and the development of research skills. Behavioural aspects such as engagement with potential users or mission-oriented projects (targeted to practical applications) were associated with higher social benefits. Considering the difficulties in assessing a wide array of research topics, and potential differences in the understanding of the concept of ‘research impact’, an analysis of the context can help to focus on research needs. Analyzing the metric properties of questionnaires can open up new possibilities for validating instruments used to measure research impact. Further to the methodological utility of the current exercise, we see a practical applicability to specific contexts where multiple discipline research impact is required.

Key words: surveys and questionnaires; research impact assessment; validation study

Introduction

Over the past three decades, increasing attention has been paid to the social role and impact of research carried out at universities. National research evaluation systems, such as the UK’s Research Excellence Framework (REF) (Higher Education Funding Council of England et al. 2015) and the Excellence in Research for Australia (Australian Research Council 2016) are examples of assessment tools that address these concerns. These systems identify and define how research funding is allocated based on a number of dimensions

of the research process, including impact of research. (Berlemann and Haucap 2015).

Being explicit about the objective of the impact assessment is emphasized in the International School on Research Impact Assessment (ISRIA) statement (Adam et al. 2018) a 10-point guideline for an effective research impact assessment that includes four purposes: advocacy, analysis, allocation, and accountability. The last one emphasizes transparency, efficiency, value to the public and a return for the investment. With mounting concern about the

relevance of research outcomes, funding organizations are increasingly expecting researchers to demonstrate that investments result in tangible improvements for society (Hanney et al. 2004). This accountability is intended to ensure resources have been appropriately utilized and is strongly linked to the drive for value-for-money within health services and research (Panel on the return on investments in health research 2009). As policy-makers and society expect science to meet societal needs, scientists have to prioritize social impact, or risk losing public support (Poppy 2015).

To meet these expectations, the Universitat Oberta de Catalunya (UOC) has embraced a number of pioneering initiatives in its current Strategic Plan, which includes the promotion of Open Knowledge, a specific measure related to the social impact of research, (Universitat Oberta de Catalunya 2017) and the development of an institution wide action plan to incorporate it in research evaluation. The UOC is currently investigating how to implement the principals of the DORA Declaration in institutional evaluation processes, taking into account 'a broad range of impact measures including qualitative indicators of research impact, such as influence on policy and practice' ('San Francisco Declaration on Research Assessment (DORA)' n.d.). The UOC is also taking the lead in meeting the Sustainable Development Goals (SDG) of the UN 2030 Agenda, (Jørgensen and Claeys-Kulik 2018) having been selected by the International Association of Universities as one of the 16 university cluster leaders around the world to lead the SDGs ('IAU HESD Cluster | HESD - Higher Education for Sustainable Development portal' n.d.).

The term 'research impact' has many definitions. On a basic level, the 'academic impact' is understood as benefits for further research, while 'wider and societal impact' includes the outcomes that reach beyond academia. In our study we will include both categories and refer to 'research impact' as any type of output or outcome of research activities that can be considered a 'positive return or payback' for a wide range of beneficiaries, including people, organizations, communities, regions, or other entities. The pathways linking science, practice, and outcomes are multifaceted and complex (Molas-Gallart et al. 2016). Indeed, the path from new knowledge to its practical application is neither linear nor simple; the stages may vary considerably in terms of duration, and many impacts of research may not be easily measurable or attributable to a concrete result of research (Figure 1). This outputs and outcomes generated by research characteristics (inputs and processes) are context dependant (Pawson 2013). Therefore, a focus on process is fundamental to understanding the generation of impact.

Surveys are among the most widely used tools in research impact evaluation. Quantitative approaches as surveys are suggested for accountability purposes, as the most appropriate way that calls for transparency (Guthrie et al. 2013). They provide a broad overview of the status of a body of research and supply comparable, easy-to-analyze data referring to a range of researchers and/or grants. Standardization of the approach enhances this comparability and minimizes researcher bias and subjectivity, particularly in the case of web or postal surveys. Careful wording and question construction increases the reliability of resulting data (Guthrie et al. 2013). However, while ex-ante assessments instruments for research proposals have undergone significant study, (Fogelholm et al. 2012; Van den Broucke et al. 2012) the metric properties of research evaluation instruments have received little attention (Aymerich et al. 2012). 'Internal consistency' is generally considered evidence of internal structure, (Clark and Watson 1995) while the measurement of 'content validity' attempts to demonstrate that the elements of an

assessment instrument are relevant to and representative of the targeted construct for a particular assessment purpose (Nunnally and Bernstein 1994).

As the demand for monitoring research impact increases across the world, so does the need for research impact measures that demonstrate validity. Therefore, the aim of this study is to develop and test the internal consistency and the content validity of an instrument designed for accountability purposes to measure the perceived impacts of a wide range of competitively funded research projects, according to the perspectives of the principal investigators (PIs). The study will also focus on the perceived impacts and their characteristics.

Methods

A cross-sectional survey was used to assess the research undertaken at UOC. This research originates from four knowledge areas: arts and humanities, social sciences, health sciences, and information and communication technologies (ICT). Research topics include 'identity, culture, art and society'; 'technology and social action'; 'globalization, legal pluralism and human rights'; 'taxation, labour relations and social benefits'; 'internet, digital technologies and media'; 'management, systems and services in information and communications'; and 'eHealth'. UOC's Ethics Committee approved this study.

Study population

The study population included all PIs with at least one competitively funded project (either public or private) at local, regional, national, or international level completed by 2017 (n = 159).

The questionnaire

An on-line questionnaire was designed for completion by project PIs in order to retrospectively determine the impacts directly attributed to the projects. The questions were prepared based on the team's prior experience and questionnaires published in scientific literature. (Wooding et al. 2010; Hanney et al. 2013) The questionnaire was structured around the multidimensional categorization of impacts in the Payback Framework. (Hanney et al. 2017)

The Payback Framework has been extensively tested and used to analyze the impact of research in various disciplines. It has three elements: first, a logic model which identifies the multiple elements that form part of the research process and contribute to achieving impact; second, two 'interfaces', one referring to the project specification and selection, the other referring to the dissemination of research results; and third, a consideration of five impact categories: *knowledge production* (represented by scientific publications or dissemination to non-scientific audiences); *research capacity building* (research training, new collaborations, the securing of additional funding or improvement of infrastructures); *informing policy and product development* (research used to inform policymaking in a wide range of circumstances); *social benefits* (application of the research within the discipline and topic sector); and *broader economic benefits* (commercial exploitation or employment) (Hanney et al. 2013).

Our instrument included four sections. The first section recorded information on the PIs, including their sex, age, and the number of years they had been involved in research. The second focused on the nature of the project itself (or a body of work based on

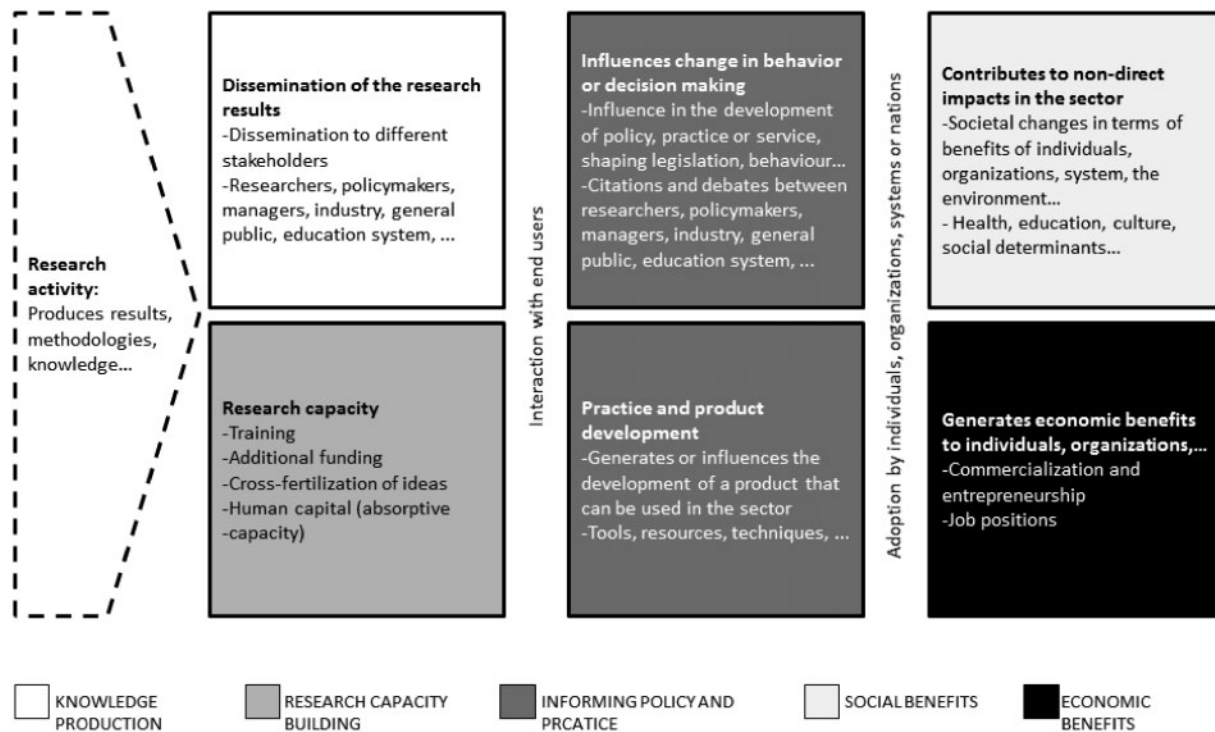


Figure 1. Effects of research impact.

continuation/research progression projects). PIs involved in more than one project (or a set of projects within the same body of work) were instructed to select one, in order to reduce the time needed to complete the survey and thereby to increase response rate. This section included the discipline, the main topic of research, the original research drivers, interaction with potential users of the research during the research processes, and funding bodies. The third section addressed the PIs’ perceptions of the impact of the research project, and was structured around the five impact categories of the aforementioned Payback Framework. The last section included general questions, one of which sought to capture other relevant impacts that might not fall within one of the previous five categories. The final question requested an evaluation (as a percentage) of the contribution/attribution of the research to the five impact categories. Respondents were required to decide the level of the contribution/attribution of the impacts according to three answer categories: limited, contribution from 1 to 30%; moderate, contribution from 40 to 60%; and significant, contribution from 70 to 100%.

Questionnaire items included questions with dichotomous answers (yes/no) and additional open box questions for a brief descriptions of the impacts perceived.

Prior to testing, we reviewed the abstracts of 72 REF2014 impact case studies (two per knowledge area). REF2014 (Higher Education Funding Council of England et al. 2015) is the first country-wide exercise to assess the impact of university research beyond academia and has a publicly available database of over 6,000 impact case studies, grouped in 34 subject-based units of assessment. Case studies were randomly selected and the impacts found in each mapped onto the most appropriate items and dimensions of the questionnaire. This review helped to reformulate and add questions, especially in the sections on *informing policy and practice* and *social benefits*.

Data collection

The questionnaire was sent to experts in various disciplines with a request for feedback on the relevance of each item to the questionnaire’s aim (impact assessment), which they rated on a 4-point scale (0 = ‘not relevant’, 1 = ‘slightly relevant’, 2 = ‘quite relevant’, 3 = ‘very relevant’) according to the definition of research impact included in our study (defined above). The experts were also asked to evaluate whether the items covered the important aspects or whether certain components were missing. They could also add comments on any item.

The PIs were contacted by email. They were informed of the objectives of the study and assured that the data would be treated confidentially. They received two reminders, also by email.

Analysis

A quality control exercise was performed prior to data analysis. The data were processed and the correct classification of the various impacts checked by comparing the yes/no responses with the information provided in the additional open box questions. No alterations were required after these comparisons. Questionnaire results provided a measure of the number of research projects contributing to a particular type of impact; therefore, to estimate each level of impact we calculated the frequency of its occurrence in relation to the number of projects. A Chi-squared test was used to test for group differences.

Internal consistency was assessed by focusing on the inter-item correlations within the questionnaire, indicating how well the items fitted together theoretically. This was performed using Cronbach’s alpha (α). An alpha between 0.70 and 0.95 was considered acceptable (Peterson 1994).

An expert opinion index was used to estimate content validity at the item level. This index was calculated by dividing the number of

experts providing a score of 2 or 3 by the total number of answers. Due to the diverse array of disciplines and topics under examination, values were calculated for all experts and for the experts of each discipline. These were considered acceptable if the level of endorsement was >0.5 .

All data were introduced into the statistical programme SPSS18, and the level of significance set at 0.05 for all tests.

Results

Sixty-eight PIs answered the questionnaire, a response rate of 42.8%. Respondents took an average of 26 minutes to complete the questionnaire. Table 1 shows the sample characteristics. Significant differences were found between the respondents and non-respondents for knowledge area ($p=0.014$) and age group

($p=0.047$). Arts and humanities investigators and PIs older than 50 years were more frequent among non-respondents. The proportion of women did not differ significantly between respondents and non-respondents ($p=0.083$).

Impact and its characteristics

An impact on *knowledge production* was observed in 97.1% of the projects, and an impact on *capacity building* in 95.6%. Lower figures were recorded for *informing policy and practice* (64.7%), and lower still for *economic benefits* (33.8%), and for *social benefits* (32.4%), although results were based on a formal evaluation in only 11, 8% of the cases included in *social benefits*. Estimations of the contribution of projects to the different impact levels were considered significant (between 70% and 100%) to *knowledge production*, moderate (between 40% and 60%), to *capacity building*, and

Table 1. Sample characteristics

		Respondents, n = 68 (42.8%)	Non-respondents, n = 91 (57.2%)
Knowledge area*	Social Sciences	42 (61.8)	47 (51.6)
	Information and Communication Technologies	14 (20.6)	16 (17.6)
	Health Sciences	7 (10.3)	7 (4.4)
	Arts and Humanities	5 (7.4)	24 (26.4)
Research subject ^a	Education	25 (36.8)	–
	Internet, digital technologies and the media	20 (29.4)	–
	Computation and artificial intelligence	15 (22.1)	–
	Health and sustainable lifestyles	14 (20.6)	–
	Art, culture and identity	11 (16.2)	–
	Society, social action and the environment	10 (14.7)	–
	Governance and social movements	7 (10.3)	–
	Management, systems and services in information and communications and innovation	4 (5.9)	–
	Globalization, legal pluralism and human rights	4 (5.9)	–
	Language, literature and cognition	4 (5.9)	–
	Tourism	2 (2.9)	–
	Others	8 (11.8)	–
	Original impetus for the project ^a	Scientific curiosity	32 (47.1)
The need to fill certain gaps in knowledge		38 (55.9)	–
Targeting to a practical application		39 (57.4)	–
Personal professional experience		23 (33.8)	–
Commissioned by third parties		2 (2.9)	–
Time elapse since the beginning of the project	Less than 4 years	17 (24.5)	–
	4–9 years	35 (50.7)	–
	More than 9 years	16 (23.9)	–
	Unknown	1 (1.4)	–
Interaction with end users ^a	Before the research process	24 (35.3)	–
	During the research process	48 (70.6)	–
	After the research process	41 (60.3)	–
	No interaction	7 (10.3)	–
PI's gender	Woman	34 (50.0)	33 (36.3)
	Man	34 (50.0)	58 (63.7)
PI's age (years)*	<30	3 (4.4)	2 (2.2)
	31–40	17 (25.0)	12 (13.2)
	41–50	38 (55.9)	45 (49.5)
	>50	10 (14.7)	28 (30.8)
	Unknown	–	4 (4.4)
PI's research experience	<5 years	5 (7.3)	–
	6–10 years	13 (19.1)	–
	>10 years	50 (73.5)	–

^aAnswers could include more than one response. PI: principal investigator.

* $p \leq 0.05$.

limited (1–30%) to *informing policy and practice, social benefits and economic benefits*. No additional impacts were reported.

Figure 2 shows the different impact categories and the distribution of impact subcategories. The size of the bars indicates the percentage of projects in which this specific impact occurred, according to the PIs.

Statistically significant differences were found according to the original impetus for the project: for projects intended to fill certain gaps in knowledge, the greatest impact was observed in *knowledge production* ($p=0.01$) and *capacity building* ($p=0.03$), while for projects targeting to a practical application, the greatest impact was observed in *informing policy and practice* ($p=0.05$) and in *social*

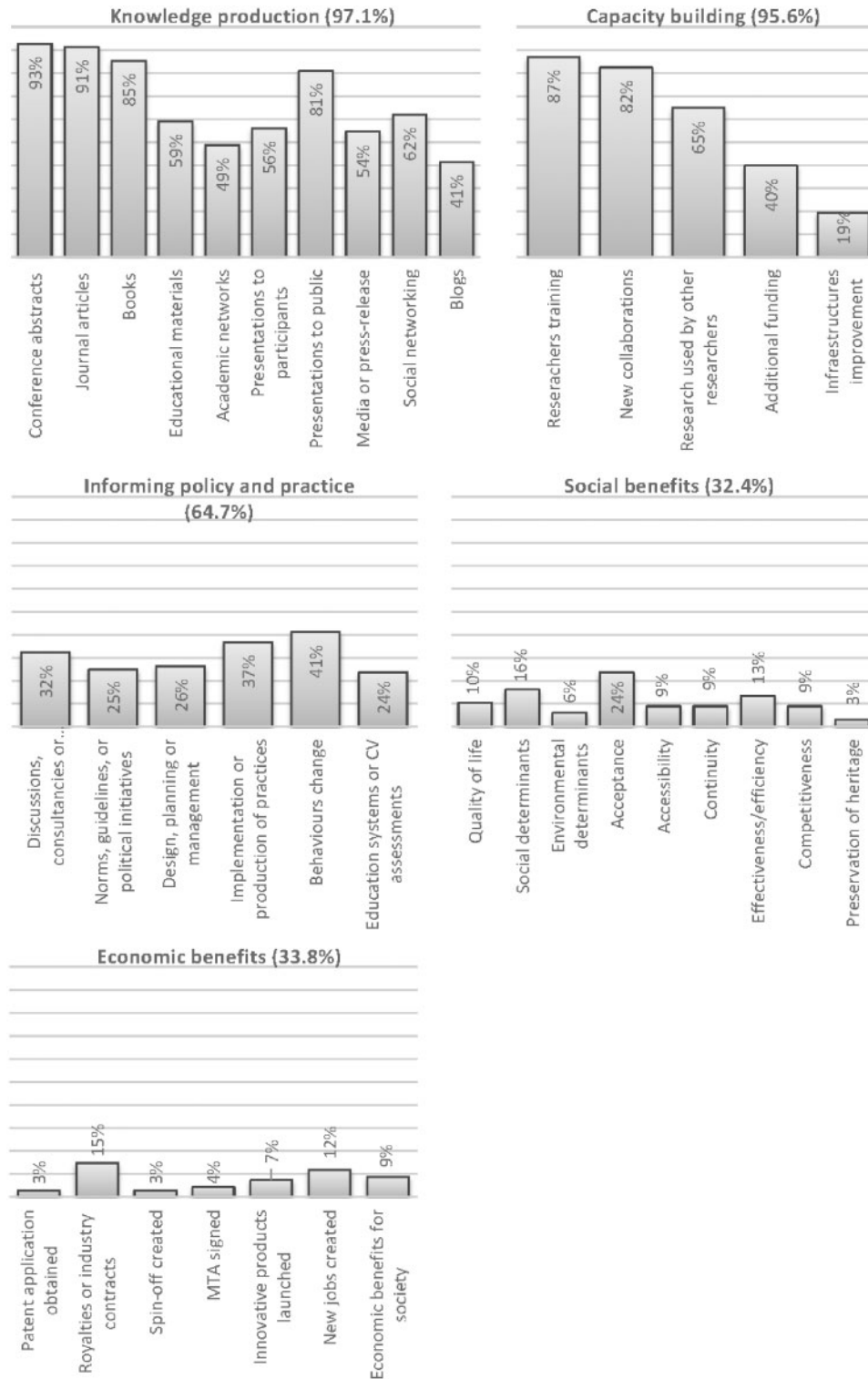


Figure 2. Achieved impact bars, according to level (n = 68).

Table 2. Internal consistency for each domain (impact level)

Domain	Cronbach's alpha
Knowledge Production	0.74
Capacity Building	0.74
Informing Policy and Practice	0.82
Social Benefits	0.89
Economic Benefits	0.47
Total domains	0.89

benefits ($p=0.01$). In general, projects that interacted with end users had more impact in the levels of *knowledge production* ($p=0.01$), *capacity building* ($p=0.03$), and *social benefits* ($p=0.05$). Projects that had begun over four years before the survey was completed was correlated with *knowledge production* ($p=0.04$), and PIs over 40 years of age and those with over 3 years research experience were correlated with more frequent impacts on *knowledge production* and *capacity building* ($p \leq 0.01$). No differences were found regarding the gender of PIs. The size of the differences can be found in the [Supplementary Table S1](#).

Internal consistency and content validity

The Cronbach's alpha score, which measures the internal consistency of the questions, was satisfactory ($\alpha=0.89$). [Table 2](#) shows its value in each domain (impact level). Internal consistency was satisfactory in all domains with the exception of *economic benefits*. However, the removal of any of the questions would have resulted in an equal or lower Cronbach's alpha.

Thirteen of the 17 experts contacted completed the content validity form and assessed whether the content of the questionnaire was appropriate and relevant to the purpose of the study. Seven were from social sciences and humanities, four from health sciences and two from ICT; 39% were women. All had longstanding experience as either researchers or research managers. The experts scored the 45 items according to their relevance and 76% of the ratings ($n=34$) had an index of 0.5 or greater. The results for each item are shown in [Table 3](#). In accordance with the expert review, an item relating to 'new academic networks' was added.

Ninety-one percent of the items in *knowledge production* were rated acceptable (expert opinion index ≥ 0.5), as were 89% of the items in *capacity building*, 83% of the items in *informing policy and practice*, and 63% of the items in *social benefits*. In contrast, only 43% of the items (three out of seven) in the *economic benefits* domain achieved an acceptable rating. Some items were of higher relevance in specific fields: for example, items relating to health and social determinants were considered acceptable by health experts; training for final undergraduate's projects was considered acceptable by ICT experts; influencing education systems and curricular assessments, was considered acceptable by social sciences and humanities, and ICT experts; and commercialization items were considered acceptable by health and ICT experts ([Table 3](#)).

Discussion

In this study, we tested the metric properties of a questionnaire designed to record the impact of university research originating from various disciplines. Tests of this kind, although rare in research impact assessment, are common in other study areas such as patient-reported outcome measures, education and psychology. The

questionnaire displayed good internal consistency and acceptable content validity in our context. Internal consistency for all items on the instrument was excellent demonstrating that they all measured the same construct. However, since 'impact' is a multidimensional concept and, by definition, Cronbach's alpha 'indicates the correlation among items that measure one single construct', ([Osburn 2000](#)) the internal consistency of each of the five domains required evaluation; this was found to be excellent in all cases except *economic benefits*. Low internal consistency in this domain may be related to the fact it contained relatively few items, and/or the fact that most of the researchers who answered the questionnaire worked in the social sciences and humanities, and therefore impacts relating to transfer, commercialization and innovation were less likely to occur. An alternative possibility is that the items are, in fact, measuring more than one construct.

There is a consensus in the literature that content validity is largely a matter of judgment, ([Mastaglia et al. 2003](#)) as content validity is not a property of the instrument, but of the instrument's interpretation. We therefore incorporated two distinct phases in our study. In the first phase of development, conceptualization was enhanced through the analysis and mapping of the impacts of the randomly selected REF case; in the second the relevance of the scale's content was evaluated through expert assessment. The expert assessment revealed that some items did not achieve acceptable content validity, especially in the domains of *social benefits* and *economic benefits*. However, it should be taken into account that while many of the items in the questionnaire were generic and thus relevant for all fields, a number were primarily specific to one field, and therefore, more relevant for experts in a particular field. Content validity was stronger in the domains 'closest' to the investigators. This may be due to the most frequently recognized impacts being both in areas where researchers have a degree of control and influence, ([Kalucy et al. 2009](#)) and those which have been 'traditionally' used to measure research. In other words, their understanding of the concept of impact in the *knowledge production*, *capacity building* and *informing policy and practice* domains: that is, those at the intermediate level (secondary outputs) display greater homogeneity ([Kalucy et al. 2009](#)).

Use of an online questionnaire in this research impact study provided data on a wide range of benefits deriving from UOC's funded projects at a particular moment and its results address a message of accountability. Questionnaires can provide insights into respondents' viewpoints and can systematically enhance accountability. Although assuming that PIs will provide truthful responses about the impact of their research is clearly a potential limitation, [Hanney et al \(2013\)](#) demonstrate that researchers do not routinely exaggerate the impacts of their research, at least in studies like this one, where there is no clear link between the replies given and future funding. International guidelines on research impact assessment studies recommend the use of a combination of methods to achieve comprehensive, robust results. ([Adam et al. 2018](#)) However, the primary focus of this study was the quality and value of the survey instrument itself, therefore the issue of triangulating the findings with other methods was not explored. The questionnaire could be applied in future studies to select projects that require a more in-depth and closer analysis, such as how an understanding of scientific processes works in this context. Previous attempts have been made to assess the impact of university's research in our context, but these have been restricted to the level of outputs (i.e. publications and patents), ([Associació Catalana d'Universitats Públiques \(ACUP\) 2017](#)) or inputs' level (i.e. contributions to Catalan GDP) ([Suriñach et al. 2017](#)).

Table 3. Content validity of items according to experts (n = 13)

Domain	Item	Number of answers	All experts (n = 13)	Social experts (n = 7)	Health experts (n = 4)	ICT experts (n = 2)
Knowledge Production	Presenting research findings in abstracts	13	0.77*	0.71*	0.75*	1.00*
	Presenting research findings in journal articles	13	0.85*	0.71*	1.00*	1.00*
	Presenting research findings in books or book chapters	13	0.69*	0.71*	0.50*	1.00*
	Presenting the research findings in educational materials	13	0.62*	0.57*	0.50*	1.00*
	Presentations of research findings to the public/patients/end-users	13	0.62*	0.43	1.00*	0.50*
	Presentations to the project volunteers	13	0.69*	0.57*	0.75*	1.00*
	Been mentioned by the media or the subject of a press release/conference	13	0.85*	0.71*	1.00*	1.00*
	Published through social networks	13	0.62*	0.43	1.00*	0.50*
	Published in influential blogging sites	12	0.67*	0.57*	1.00*	0.50*
	Concerts, recordings, or music hall presentations	9	0.11	0.25	0.00	0.00
Capacity Building	Training for PhD students	13	0.92*	0.86*	1.00*	1.00*
	Training for master's degree students	13	0.69*	0.57*	0.75*	1.00*
	Training for final undergraduate's projects	13	0.31	0.43	0.00	0.50*
	New collaborations at national level	11	0.91*	0.83*	1.00*	1.00*
	New collaborations at international level	12	0.92*	0.83*	1.00*	1.00*
	New academic networks	8	0.75*	0.75*	0.50*	1.00*
	Additional funding to create new research projects	13	0.77*	0.57*	1.00*	1.00*
	Additional funding for the research group	13	0.92*	0.86*	1.00*	1.00*
Informing Policy and Practice	Research or methods used by other researchers in subsequent research	9	0.78*	0.67*	0.75*	1.00*
	Contribution to the improvement of research infrastructures	13	0.62*	0.57*	0.50*	1.00*
	Informing as evidence in debates, discussions, or consultancies	13	0.62*	0.71*	0.50*	0.50*
	Informing as evidence in the formulation of norms, guidelines, political initiatives or recommendations by government bodies or other regulators	12	0.58*	0.67*	0.50*	0.50*
	Contribution in the design, planning and management of services and priorities	12	0.50*	0.50*	0.50*	0.50*
	In the implementation, adoption or production of practices within and beyond the professional world	12	0.58*	0.50*	0.50*	1.00*
	Influencing the behaviour of professionals or other people	13	0.54*	0.57*	0.50*	0.50*
Social Benefits	Influencing education systems and curricular assessments	10	0.40	0.50*	0.25	0.50*
	Improving health	11	0.27	0.20	0.50*	0.00
	Improving quality of life	11	0.55*	0.60*	0.50*	0.50*
	Improving social and cultural determinants	11	0.64*	0.80*	0.50*	0.50*
	Improving environmental determinants	11	0.36	0.40	0.50*	0.00
	Improving acceptability	11	0.73*	0.80*	0.75*	0.50*
	Improving accessibility*	11	0.55*	0.60*	0.50*	0.50*
	Improving continuity	9	0.56*	0.67*	0.50*	0.50*
	Improving effectiveness or efficiency	10	0.50*	0.50*	0.50*	0.50*
	Improving safety	10	0.20	0.25	0.25	0.00
Improving well-being and social benefits	11	0.64*	0.80*	0.50*	0.50*	
Improving heritage preservation	10	0.20	0.25	0.25	0.00	
Improving competitiveness and development of stimuli	10	0.50*	0.50*	0.25	1.00*	

(continued)

Table 3. Continued

Domain	Item	Number of answers	All experts (n = 13)	Social experts (n = 7)	Health experts (n = 4)	ICT experts (n = 2)
Economic Benefits	Patent application obtained	13	0.23	0.00	0.50*	0.50*
	Generating revenue from royalties, equities, industry contracts or any other compensation	13	0.23	0.14	0.50*	0.00
	Leading to the creation of a new business spin-off or start-up company	13	0.31	0.14	0.50*	0.50*
	Leading to Material Transfer Agreements	13	0.23	0.00	0.50*	0.50*
	Bringing innovations, products or devices to market	9	0.56*	0.33	0.50*	1.00*
	Creation of new jobs	13	0.62*	0.57*	0.75*	0.50*
	Bringing wider economic impacts	13	0.54*	0.57*	0.50*	0.50*

*Items rated greater than or equal to 0.5; ICT: information and communication technologies.

Evaluated as a whole, the research projects covered in this study was effective in the production of knowledge and the development of research skills in individuals and teams. This funded research has helped to generate new knowledge for other researchers and, to a lesser extent, for non-academic audiences. It has consolidated the position of UOC researchers (both experienced and novice) within national and international scientific communities, enabling them to develop and enhance ability to conduct quality research (Trostle 1992).

Assessing the possible wider benefits of the research process (in terms of *informing policy and practice*, *social benefits* and *economic benefits* for society) proved more problematic. The relatively short period that had elapsed since the projects finished might have limited the assessment of impact. There was a striking disparity, in our results, between the return on research measured in terms of scientific impact (*knowledge production* and *capacity building*), notably high and uniform, and the limited and uneven contribution to wider benefits. This disparity is not a local phenomenon, but a recurrent finding in contemporary biomedical research worldwide. The Retrosight study, (Wooding et al. 2014), which analyzed cardiovascular and stroke research in the United Kingdom found no correlation between knowledge production and the broader social impact of research. Behavioural aspects such as researcher engagement with potential users of the research or mission-oriented projects (targeted to practical applications) were associated with higher social benefits. This might be interpreted as strategic thinking on the part of researchers, in the sense that they consider the potential ‘mechanisms’ that might enhance the impact of their work. These results do not appear to be exceptional, since the final impact of research is influenced by the extent to which the knowledge obtained is made available to those in a position to use it.

Although the response rate was lower than expected, 43% is within the normal range for on-line surveys. (Shih and Xitao 2008) In addition, arts and humanities researchers were underrepresented among PIs, but not between experts for considering content validity. One possible reason for this is that investigators are not fully aware of the influence of their research; another is the belief that research impact assessment studies are unable to provide valuable data about how arts and humanities research generates value. (Molas-Gallart 2015) Arts and humanities is a discipline where in some cases the final objective of the research is not a practical application, but rather to change behaviours or people perspectives, which are therefore, more difficult to measure. According to Ochsner et al. (2012)

there is a missing link between indicators and humanities scholars’ notions of quality. However, questionnaires have been used to successfully measure the impact of arts and humanities research, including in an approach adapted from the Payback Framework (Levitt et al. 2010), and research impact analyses such as REF2014 (Higher Education Funding Council of England et al. 2015) and the special issue of Arts and Humanities in Higher Education on the public value of arts and humanities research (Benneworth 2015) have demonstrated that research in these disciplines may have many implications for society. Research results provide guidance and expertise and can be easily transferred to public debates, policies and institutional learning.

Weiss describes the rationale and conceptualization of assessment activities relating to the social impact of research as an open challenge (Weiss 2007). As well as the well-known practice of attributing impact to a sole research project and the time-lag between the start of a research project and the attainment of a specific impact, in this study we also had the challenge to assess the impact of research from a diverse variety of topics and disciplines. Research impact studies are prevalent in disciplines such as health sciences (Hanney et al. 2017) and agricultural research (Weißhuhn et al. 2018) but less common in the social sciences and humanities, despite the REF2014 results revealing a wide array of impacts associated with various disciplines. (Higher Education Funding Council of England et al. 2015) Our challenge was to analyze projects from highly diverse disciplines—social sciences, humanities, health sciences, and ICTs—and assess their varied impacts on society. We have attempted to develop a flexible and adaptable approach to assessing research impacts by utilizing a diverse amalgamation of indicators, including impact subcategories. However, due to ‘cultural’ differences between disciplines, we cannot guarantee that PIs from different knowledge areas have a homogeneous understanding of ‘research impact’: indeed the diversity of respondents when assessing the relevance of questionnaire items suggests otherwise. For this reason, a context analysis in which research is carried out and assessed, as described in the literature (Adam et al. 2018) may help to decide which questionnaire items or domains should be included or removed in future studies.

To conclude, this study demonstrates that the easy-to-use questionnaire developed here is capable of measuring a wide range of research impact benefits and provides good internal consistency. Analyzing the metric properties of instruments used to measure research impact and establishing their validity will significantly

contribute to research impact assessment and stimulate and extend reflection on the definition of research impact. Therefore, this questionnaire can be a powerful instrument to measure research impact when considered in context. The power of this instrument will be significantly improved when combined with other methodologies.

What is already known about this topic

Surveys are widely used in research impact evaluation. They provide a broad overview of the state of a body of research, and supply comparable, easily analyzable data referring to a range of researchers and/or grants. The standardization of the approach enhances this comparability.

What this study adds

To our knowledge, the metric properties of impact assessment questionnaires have not been studied to date. The analysis of these properties can determine the internal consistency and content validity of these instruments and the extent to which they measure what they are intended to measure.

Supplementary data

Supplementary data is available at *Research Evaluation Journal* online.

Acknowledgements

We thank the UOC principal investigators for providing us with their responses.

Funding

This project did not receive any specific grants from funding agencies in the public, commercial, or not-for-profit sectors.

Transparency

The lead authors (the manuscript's guarantors) affirm that the manuscript is an honest, accurate, and transparent account of the study being reported; that no important aspects of the study have been omitted; and that any discrepancies from the study as planned have been explained.

References

Adam, P. et al. (2018) 'ISRIA Statement: Ten-Point Guidelines for an Effective Process of Research Impact Assessment', *Health Research Policy and Systems*, 16/1, DOI: 10.1186/s12961-018-0281-5.

Associació Catalana d'Universitats Públiques (ACUP) (2017) *Research and Innovation Indicators of Catalan Public Universities*. Report 2016.

Australian Research Council (2016) *State of Australian University Research 2015-2016: Volume 1 ERA National Report*. Canberra: Commonwealth of Australia.

Aymerich, M. et al. (2012) 'Measuring the Payback of Research Activities: A Feasible Ex-Post Evaluation Methodology in Epidemiology and Public Health', *Social Science and Medicine*, 75/3: 505–10.

Benneworth, P. (2015) 'Putting Impact into Context: The Janus Face of the Public Value of Arts and Humanities Research', *Arts and Humanities in Higher Education*, 14/1: 3–8.

Berlemann, M. and Haucap, J. (2015) 'Which Factors Drive the Decision to Opt out of Individual Research Rankings? An Empirical Study of Academic Resistance to Change', *Research Policy*, 44/5: 1108–15.

Clark, L. A. and Watson, D. (1995) 'Constructing Validity: Basic Issues in Objective Scale Development', *Psychological Assessment*, 7/3: 309–19.

Fogelholm, M. et al. (2012) 'Panel Discussion Does Not Improve Reliability of Peer Review for Medical Research Grant Proposals', *Journal of Clinical Epidemiology*, 65/1: 47–52.

Guthrie, S. et al. (2013) *Measuring Research: A Guide to Research Evaluation Frameworks and Tools*. RAND Corporation.

Hanney, S. et al. (2004) 'Proposed Methods for Reviewing the Outcomes of Health Research: The Impact of Funding by the UK's "Arthritis Research Campaign"', *Health Research Policy and Systems*, 2/1: 4.

Hanney, S. et al. (2013) 'Conducting Retrospective Impact Analysis to Inform a Medical Research Charity's Funding Strategies: The Case of Asthma UK', *Allergy, Asthma, and Clinical Immunology: Official Journal of the Canadian Society of Allergy and Clinical Immunology*, 9/1: 17.

Hanney, S. et al. (2017) 'The Impact on Healthcare, Policy and Practice from 36 Multi-Project Research Programmes: Findings from Two Reviews', *Health Res Policy Syst*, 15/1: 26.

Higher Education Funding Council of England, et al. (2015) *The Nature, Scale and Beneficiaries of Research Impact: An Initial Analysis of Research Excellence Framework (REF) 2014 Impact Case Studies*. London: HEFCE.

'IAU HESD Cluster | HESD - Higher Education for Sustainable Development portal' (n.d.) <<http://iau-hesd.net/en/contenu/4648-iau-hesd-cluster.html>> accessed 17 Dec 2018.

Jørgensen, T. E. and Claeys-Kulik, A.-L. (2018) *Universities' Strategies and Approaches Towards Diversity, Equity and Inclusion. Examples from Across Europe*. Brussels: European University Association.

Kalucy, E. C. et al. (2009) 'The Feasibility of Determining the Impact of Primary Health Care Research Projects Using the Payback Framework', *Health Research Policy and Systems*, 7: 11.

Levitt, R., Celia, C. and Diepeveen, S. (2010) *Assessing the Impact of Arts and Humanities Research at the University of Cambridge*. Technical Report. RAND Corporation, 104.

Mastaglia, B., Toye, C. and Kristjanson, L. J. (2003) 'Ensuring Content Validity in Instrument Development: Challenges and Innovative Approaches', *Contemporary Nurse*, 14/3: 281–91.

Molas-Gallart, J. (2015) 'Research Evaluation and the Assessment of Public Value', *Arts and Humanities in Higher Education*, 14/1: 111–26.

Molas-Gallart, J. et al. (2016) 'Towards an Alternative Framework for the Evaluation of Translational Research Initiatives', *Research Evaluation*, 25/3: 235–43.

Nunnally, J. C. and Bernstein, I. H. (1994) *Psychometric Theory*. New York: McGraw-Hill.

Ochsner, M., Hug, S. E. and Daniel, H.-D. (2012) 'Indicators for Research Quality in the Humanities: Opportunities and Limitations', *Bibliometrie - Praxis und Forschung*, 1/4: 1–17. DOI: 10.5283/bpf.157.

Osburn, H. G. (2000) 'Coefficient Alpha and Related Internal Consistency Reliability Coefficients', *Psychological Methods*, 5/3: 343–55.

Panel on the return on investments in health research. (2009) *Making an Impact: A Preferred Framework and Indicators to Measure Returns on Investment in Health Research*. Ottawa, ON (Canada): Canadian Academy of Health Science (CAHS), Ed.

Pawson, R. (2013) *The Science of Evaluation: A Realist Manifesto*. London: Sage Publications Ltd. <http://dx.doi.org/10.4135/9781473913820>

Peterson, R. A. (1994) 'A Meta-Analysis of Cronbach's Coefficient Alpha', *Journal of Consumer Research*, 21/2: 381.

Poppy, G. (2015) 'Science Must Prepare for Impact', *Nature*, 526/7571: 7.

'San Francisco Declaration on Research Assessment (DORA)'. (n.d.) <<https://sfedora.org/>> accessed 17 Dec 2018.

Shih, T. H. and Xitao, F. (2008) 'Comparing Response Rates from Web and Mail Surveys: A Meta-Analysis', *Field Methods*, 20/3: 249–71.

Suriñach, J. et al. (2017) *Socio-Economic Impacts of Catalan Public Universities and Research, Development and Innovation in Catalonia*. Barcelona: Catalan Association of Public Universities (ACUP).

Trostle, J. (1992) 'Research Capacity Building in International Health: Definitions, Evaluations and Strategies for Success', *Social Science & Medicine*, 35/11: 1321–4.

- Universitat Oberta de Catalunya (2017) *Strategic Plan Stage II 2017-2020*. Barcelona: UOC.
- Van den Broucke, S., Dargent, G. and Pletschette, M. (2012) 'Development and Assessment of Criteria to Select Projects for Funding in the EU Health Programme', *The European Journal of Public Health*, 22/4: 598–601.
- Weiss, A. P. (2007) 'Reviews and Overviews Measuring the Impact of Medical Research: Moving from Outputs to Outcomes', *Psychiatry: Interpersonal and Biological Processes*, 164/February: 206–14.
- Weißhuhn, P., Helming, K. and Ferretti, J. (2018) 'Research Impact Assessment in Agriculture—A Review of Approaches and Impact Areas', *Research Evaluation*, 27/1: 36–42.
- Wooding, S. et al. (2010) *Mapping the Impact: Exploring the Payback of Arthritis Research*. Santa Monica, CA: RAND Corporation.
- Wooding, S. et al. (2014) 'Understanding Factors Associated with the Translation of Cardiovascular Research: A Multinational Case Study Approach', *Implementation Science*, 9/1: 47.