KU LEUVEN

KU Leuven, Faculty of Economics and Business
Dept. of Management, Strategy and Innovation (MSI)

Do long and short innovation survey forms yield comparable results?

Machteld Hoskens and Koenraad Debackere

MSI Discussion Paper No. 2407

Do long and short innovation survey forms yield comparable results?[1]

Machteld Hoskens[2] and Koenraad Debackere[3]

ECOOM, KU Leuven, Naamsestraat 61, bus 3551, 3000 Leuven, Belgium

August 2024

Abstract

Given the importance of innovation for economic growth, many countries conduct innovation surveys. International guidelines for such measurement have been established (OECD/Eurostat, 2018). The European Commission has made the measurement of innovation mandatory for EU member states. Many differences remain, however, between countries in the practical implementation of measuring innovation at the firm level, which complicates cross-country comparability. We conducted a randomized experiment in which we randomly assigned enterprises a long or a short form for measuring their innovation activities. We found clear differences between the two types of forms. We discuss implications of this work and put this in the broader perspective of other work done investigating questionnaire design issues in innovation surveys.

Keywords: questionnaire design, innovation survey, randomized experiment, questionnaire length, shortened survey form, nonresponse survey

[1] **Acknowledgements**: We thank Julie Delanote and Laura Verheyden for their help in conducting the survey and the experiment embedded in it. The financial support for this research was provided via the 5-year agreement with the Flemish government on the funding of the ECOOM activities.
The survey data were collected within the context of the European Statistical System and are subject to the statistical confidentiality constraints of Regulation (EC) No 223/2009.

[2] E-mail: machteld.hoskens@kuleuven.be, ORCID ID: 0000-0001-6813-3975.

[3] E-mail: koenraad.debackere@kuleuven.be, ORCID ID: 0000-0002-3411-0668.

1 Introduction

Since the seminal work by Schumpeter (1942), Solow (1956) and Arrow (1962) on the importance of technology and innovation for economic growth, the field of innovation research has made significant progress. Innovation has been studied in a macro- as well as a micro-economic setting, and in the context of different management and organizational structure theories. The ambition of the European Union to become the most innovative economy of the world, has - on a European level - set in motion a process of measurement, benchmarking and trend mapping, increasing the methodological needs for the valid and robust measurement of innovation activities, their antecedents, the supporting processes, and performances and outcomes across borders. To this effect, Eurostat - the European Statistical Bureau - has designed its own harmonizing process for statistical methods concerning Science and Technology variables, based on the OECD's pioneering work.

The valid measurement of various dimensions of innovation is considered by many countries to be of prime importance -- e.g., one of the flagship publications of the EU is its European Innovation Scoreboard, previously *Innovation Union Scoreboard*

(https://research-and-innovation.ec.europa.eu/statistics/performance-indicators/european-innovation-scoreboard_en); while OECD also publishes a Science, Technology and Industry Scoreboard (https://www.oecd.org/en/data/datasets/science-technology-and-innovation-scoreboard.html).

Both scoreboards include results obtained from surveys measuring innovation at the firm level. Those measurements are based on the Oslo Manual.

The Oslo manual developed by OECD offers guidelines for surveys on innovation. The first edition of the Oslo Manual (OECD, 1992), published in 1992, focused on technological product and process innovation. The second edition of the Oslo Manual (OECD/Eurostat, 1997) was published in 1997 as a joint effort between Eurostat and OECD. It built on experiences gained since the first edition had been published, and among other things, included measurement of innovation in services. The third edition

was published in 2005, again as a joint effort between Eurostat and OECD (OECD/Eurostat, 2005). It broadened the scope of innovation to also include non-technological innovations such as marketing and organizational innovation. A fourth edition was published by Eurostat and OECD in 2018 (OECD/Eurostat, 2018), as it was felt that after more than ten years the innovation landscape had sufficiently changed to warrant an update.

Going from one edition to the next, most attention went to clarifying and expanding the innovation concepts captured in the surveys, e.g. the concepts of product, process, organizational, and marketing innovation (cf. Arundel, O'Brien & Torugsa, 2013; Hoskens & Andries, 2015).

Over time, however, new questions were raised as to the comparability of results obtained using various survey formats. For example, a study done by OECD in 2012 using innovation survey metadata suggested some innovation survey design features negatively impacted the comparability of results between countries (OECD, 2012).

Delegates from the Netherlands reported at OECD and Eurostat meetings that they had encountered a significant increase in reported innovation rates when moving from a paper format to a web format for their 2011 innovation survey. In 2009 they surveyed the reference period 2006-2008 using a paper survey form (just like in previous waves) and obtained an overall innovation rate of 35% (CBS, 2010). In 2011 they surveyed the reference period 2008-2010, for the first time using a web survey form, and obtained an overall innovation rate of 48%, an increase of more than 10%. They noted in their technical report that this break in series might partially reflect a true increase in some innovating enterprises, but more important for explaining the break they thought was the change in survey format (CBS, 2012, p. 177).

Norway performed several randomized experiments in 2011 in which they compared a voluntary vs. mandatory survey and a stand-alone survey vs. one combined with the R&D survey (Wilhelmsen, 2012). For measuring R&D, the international guidelines of the Frascati Manual (OECD, 2002, 2015)

were followed. The design effects Norway investigated had a clear impact on reported innovation rates.

As will be explained in more detail below, the Flemish region (Flanders) in Belgium ran an experiment in 2015 in which responses to long vs. short forms measuring innovation were compared. Clear differences were found. Experiments performed by the World Bank Group also found strong differences between long vs. short questionnaires (Cirera & Muzi, 2016).

All these findings show that "how" innovation is measured at the firm level clearly does matter.

The attention paid to the survey methodology and the questionnaire design, has been rather limited in the first three editions of the Oslo Manual. For example, in the first edition of the Oslo Manual, only one page out of 62 is devoted to survey methods. The manual briefly considers questionnaire design issues. In the second and third edition of the Oslo Manual, the section on survey methods has expanded somewhat, but questionnaire design issues still only receive a few paragraphs of attention. For example, in the third edition, there is one chapter, or 13 pages out of 163, devoted to survey procedures. Most of this chapter discusses sampling issues, only two pages discuss questionnaire design issues, and only one paragraph in this section is devoted to issues requiring attention when running international innovation surveys (§ 455). In the fourth edition the chapter on business survey procedures has been expanded and runs 25 pages out of a total of 258, but the section on questionnaire design is still rather limited.

This can be contrasted with practices like the one applied in the European Social Survey (ESS; www.europeansocialsurvey.org). In this project a common questionnaire is developed with the explicit aim to be implemented in an international context. The team working on the ESS developed an extensive procedure to prevent potential translation problems. Likewise, they developed extensive guidelines to guard the quality of the items included in the common questionnaire. They use Split Ballot-Multi-Trait, Multi-Method (MTMM) design to study the impact of item wordings, item ordering etc.

The few empirical design studies done so far in the context of innovation measurement suggest innovation measurement as we have been doing it for years, is very vulnerable to questionnaire design effects. We thus see a clear need for more experimental studies documenting questionnaire design issues in measuring innovation at the firm level. Otherwise, we risk making invalid comparisons across countries, when we assume their measurements are comparable whereas in fact they are not. At best we might be limited to making comparisons within countries over years (at least if the questionnaire design has been sufficiently stable over time. Below, we describe more extensively a study that was done in Belgium comparing long and short forms for measuring innovation at the firm level.

In the European Union, regulation No 2019/2152 and its precursor regulation No 995/2012 require all EU countries to conduct a two-yearly innovation survey of the business enterprise sector. A harmonized survey form is developed under supervision of the European statistical office (Eurostat) and provided to all EU member states. Eurostat also provides methodological guidelines for conducting the survey. One of those guidelines reads that if a response rate below 70% is obtained, a nonresponse survey should be performed among a random subset of nonrespondents. The results of this nonresponse survey should then be used to adjust the weights applied to generalize the survey results to the target population, and to account for potential nonresponse bias. The methodological guidelines indicate that the nonresponse survey form used should be short, only containing the core innovation questions. (cf. the basic question approach, Kersten & Bethlehem, 1984)

In most EU member states the two-yearly innovation survey is mandatory, and hence by far most countries obtain response rates of 70% or higher (Eurostat, 2015). Only in four EU countries the innovation survey has a voluntary status. Three of these countries obtain response rates below 70%. Three of these countries have been performing a nonresponse survey and have been using its results to calculate nonresponse adjusted weights. Generally, they find that the innovation rates obtained when surveying a subsample of nonrespondents using the short form are higher than the innovation rates obtained with the regular long form (e.g., Peters & Rammer, 2013; OECD, 2012, p. 21, § 63-64). This is somewhat counterintuitive, as typically one finds that sampled units more familiar with or more

interested in the surveyed topic, are more inclined to respond (cf. Groves, Presser & Dipko, 2004; Groves, Couper, Presser, Singer, Tourangeau, Acosta & Nelson, 2006). For example, many sampled units in Belgium inform the surveying institute during the field phase that "the survey is not relevant for them, therefore they will not respond." When probed, for many of them the 'not relevant' meant that they had no meaningful innovation activities recently. Hence, when we see that non-innovators tend to consider the innovation survey as 'not relevant' and for that reason do not respond, we would expect to obtain a higher innovation rate from the regular, long form survey than from the short form nonresponse survey: among the remaining nonrespondents, we would hence expect to find more non-innovators. But this is not what is found.

A potential explanation for this counterintuitive finding is that the short form for the nonresponse survey is usually administered by phone, whereas the long form is administered either on paper or as a web survey. It is well known that response mode effects can occur in surveys, especially when comparing interviewer-administered surveys (such as phone interviews) and self-administered surveys (such as paper and web forms, e.g., de Leeuw, 2008). Social desirability and positive responses tend to be higher in interviewer-administered surveys. Hence, the higher innovation rates in the short form surveys that were performed by phone among nonrespondents might be response mode effects.

In Belgium, an attempt to administer the short form survey to a subset of nonrespondents by e-mail (a self-administered mode) in 2013 failed. The obtained response rate was too low to yield useful results. Generally, it is hard to get nonrespondents to respond, late in the field phase, when they already had had two paper reminders and, in some cases, also follow-up phone calls. Eliciting a late response by phone then appears to be better feasible than by e-mail or on paper.

Given the counterintuitive results obtained with the short form nonresponse survey, and the fact that other EU member states had reported significant method effects obtained with their innovation surveys at OECD and Eurostat STI meetings (CBS, 2012; Wilhelmsen, 2012), one of the three Belgian regions conducted a randomized experiment comparing long and short forms for surveying

innovation. (As Belgium has a highly regionalized structure, the production of official statistics is also partly decentralized and conducted at the regional level. The innovation survey is one of those regionalized surveys. The three regional entities producing those statistics do work closely together, though.)

In the section below we will first describe the randomized experiment that was performed in the Flemish region in Belgium (Flanders), comparing long and short forms for measuring innovation. Next, we will describe the results obtained, followed by a discussion of those results. We will also return to the method effects obtained in other EU countries in the discussion section.

2 Method

In the context of the 2014 Community Innovation Survey (CIS 2014), strata that were randomly sampled in Flanders were randomly split into two and were randomly assigned either a short form or a long form survey. Business enterprise strata that are generally completely enumerated for CIS (i.e., all enterprises within the stratum are surveyed rather than just a random subset of it), were not included in the randomized experiment. Such completely enumerated strata each consist of a smaller but influential group of enterprises, that are all included in the official survey, to ensure sufficient accuracy of results. For example, large enterprises and top R&D performing enterprises are all surveyed in the official CIS, given the high impact they have on overall results.

Randomly sampled strata in CIS 2014 in Flanders were a subset of small firms (10-49 employees): low-tech industry firms (LTI; NACE Rev. 2 codes 5-18, 31-32, or 35-39), medium low-tech industry firms (MLTI; NACE Rev. 2 codes 19, 23-25, or 33), and low-tech services firms (LTS; NACE Rev. 2 codes 46, 49-53, 58, 64-66, or 73). It should be noted that these small (medium) low-tech firms represent 60% of the surveyed population of enterprises in Flanders for CIS 2014. Hence, the trends observed in this subset have a significant impact on the overall results and on the concentration/dispersion of innovation activities in the region.

In total 2148 business enterprises were included in the randomized experiment. Table 1 shows their distribution over the various strata. Within each stratum, enterprises were randomly assigned either the short form or the long form innovation survey.

Table 1. Number of enterprises included in each of the strata of the randomized experiment

			Form type	
Stratum	Nace codes	Description	Long	Short
MLTI	23-25, 33	Minerals, metals, repair & installation	204	193
LTI	05-18, 31-32, 35-39	Mining, food, beverages, tobacco, textiles, paper, wood, furniture, other, electricity, gas, steam, water supply, sewerage & waste management	319	312
LTS (46)	46	Wholesale	334	325
LTS (oth)	49-53, 58, 64-66, 73	Transportation, storage, publishing, finance, advertising & marketing	246	215

The long form innovation survey was the official CIS 2014 form, i.e., the form used to calculate the official statistics that are reported to Eurostat and OECD. It was based on the harmonized form made available by Eurostat and included all its mandatory questions and a small number of its optional questions. It also included a small number of extra questions deemed useful for research, as well as a module on social innovation. Several optional questions of the Eurostat template (optional vs. mandatory status of questions was determined in EC regulation 995/2012) were not included, to limit the response burden for surveyed enterprises. Given the innovation survey is voluntary in Belgium, limiting response burden for respondents is a major concern. The long form includes definitions of the four types of innovation immediately preceding the questions relating to them.

The short form innovation survey corresponded to the short form that was going to be used for the nonresponse survey, to survey a subset of nonrespondents at the end of the field phase. It contained all core questions that Eurostat's methodological guidelines recommended to include in the CIS 2014 nonresponse survey (Eurostat, 2014), plus one extra question: whether R&D activities (if present) were performed continuously or occasionally. This extra question was included as it yields important information for the calculation of official R&D statistics. Each EU member state is required to report

innovation statistics two-yearly, but simultaneously with the reporting of the innovation statistics member states are also required to report three core R&D statistics. In the Flemish region of Belgium, those core R&D statistics are calculated based on responses obtained in the innovation survey. An English translation of the short form that was used can be obtained upon request. The short form contains one general definition of innovation, and then proceeds with slightly simplified questions on each of the four innovation types.

The randomized experiment was hence conducted within the context of the regular CIS 2014. The survey conducted was mixed mode: all sampled units were mailed a paper survey form. Enterprises could choose to either mail back their paper responses, or enter their responses online, in a web form. Sampled units were sent two paper reminders by regular mail. Sampled units sent the short form were sent a third reminder by e-mail. To some sampled units receiving the long form, follow-up phone calls were made to encourage them to respond. The survey was in the field from April through October 2015, six months.

3 Results

Figure 1 shows the response rates obtained for the four subgroups of the randomized experiment. We see that in all four subgroups the response rate obtained with the short form is higher than the one obtained with the long form. This is not so surprising, as the response burden is lower in the short form: four pages containing nine questions in the short form versus 20 pages containing 41 questions in the long form. The overall weighted response rate in the short form is 72% versus 54% in the long form. The differences in response rate are statistically significant, both within the strata as well as across the strata (chi^2 of 22,791, 17,041, 25,254, and 11,917 respectively for MLTI, LTI, LTS (46) and LTS (oth); and 74,566 overall, all $p < 0.001$).

Figure 1. Response rates by stratum and survey form

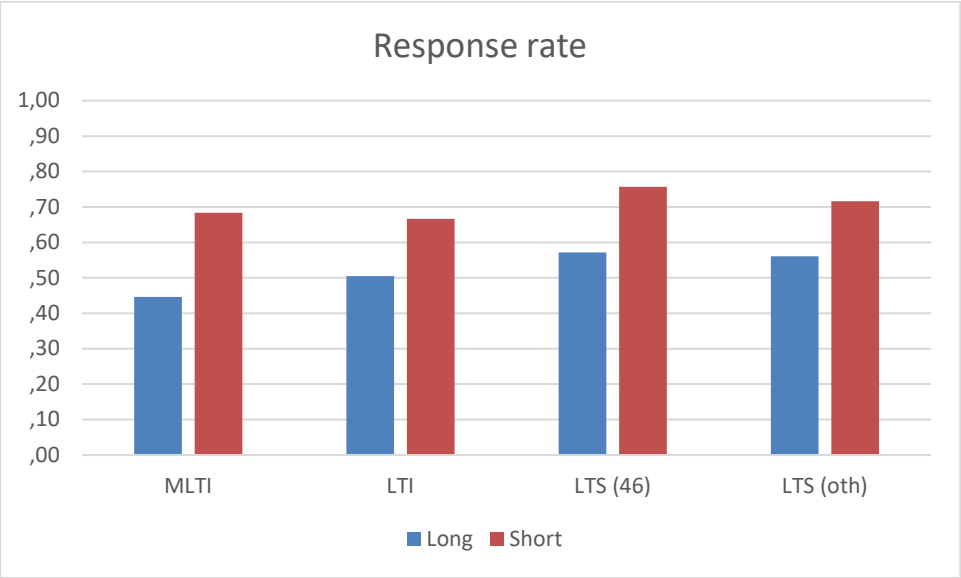


Figure 2 shows the innovation rates and accompanying confidence intervals[4] that were obtained for the four different types of innovation (product, process, organizational and marketing innovation) for the four subgroups of small firms. We see that for all subgroups and for all types, the innovation rates are higher for the short form as compared to the long form. For the subgroups low-tech industry (LTI) and wholesale (LTS (46)) the confidence intervals for the two types of forms do not overlap. For the subgroup of remaining low-tech services (LTS (oth)) the confidence intervals do not overlap for product innovation, but they do overlap for the three other innovation types. For the medium low-tech industry subgroup (MLTI), the confidence intervals for all four innovation types overlap, hence, none of these differences are statistically significant.

[4] Confidence intervals were obtained using the complex samples module in IBM SPSS, considering the sampling design.

Figure 2. Innovation rates by innovation type, stratum and survey form; 95% confidence intervals are indicated

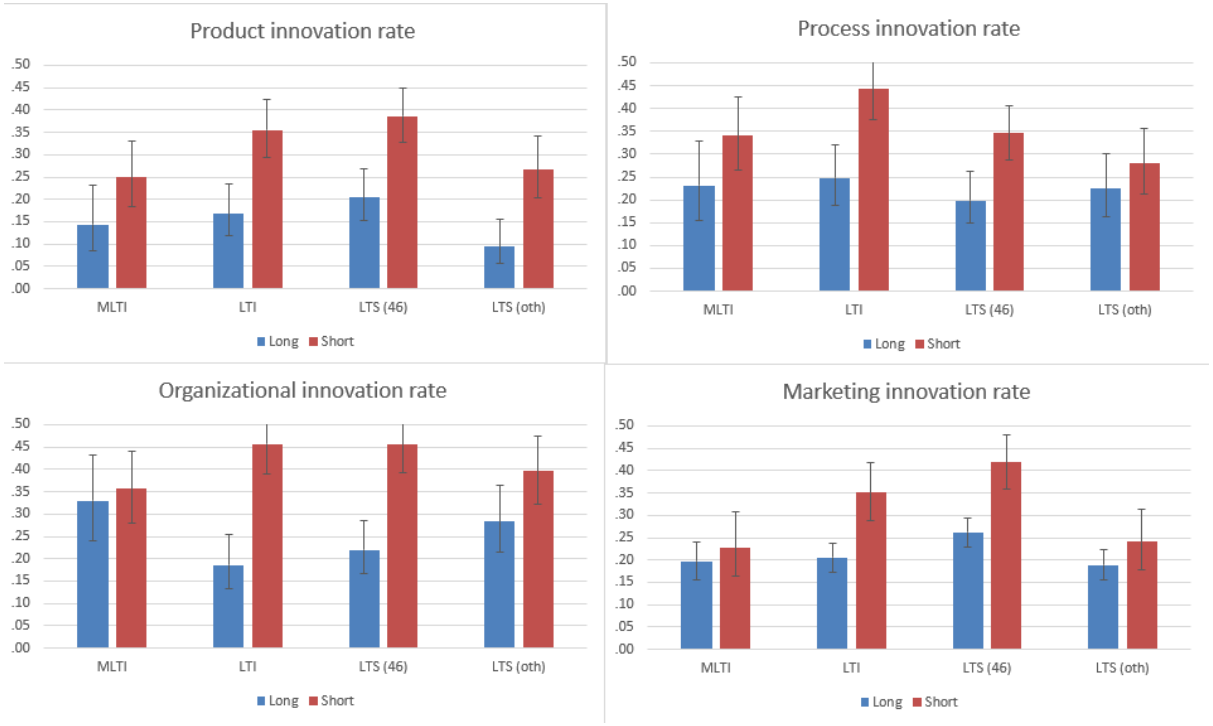


Table 2 shows the weighted innovation rates and 95% confidence intervals (95% CI) for the four types of innovation for the population of small (medium) low-tech enterprises included in the randomized experiment. We see that the confidence intervals for the innovation rates obtained with the long form and the short form do not overlap. For all four innovation types, innovation rates obtained with the short form are systematically higher. If such differences in innovation rates would be observed between countries or over time, policy makers would consider them to be highly significant. For product innovation, the innovation rate obtained with the short form is even almost double that of the long form. This is clear evidence of a method effect: as the only difference between the groups was their treatment, i.e., whether they were randomly assigned either the long or the short form, it is the mere difference in form length that explains the difference in innovation rates.

Table 2. Weighted innovation rates by survey form; 95% confidence intervals are indicated

Innovation type	Long (paper/web)		Short (paper/web)		Short (phone nrsp)	
	%	(95% CI)	%	(95% CI)	%	(95% CI)
Product	16%	(13% - 19%)	33%	(29% - 36%)	42%	(27% - 59%)
Process	22%	(19% - 26%)	35%	(31% - 38%)	34%	(22% - 49%)
Organizational	25%	(21% - 28%)	42%	(39% - 46%)	30%	(18% - 47%)
Marketing	22%	(19% - 26%)	33%	(29% - 36%)	29%	(16% - 45%)

Table 2 also shows the results of the nonresponse follow-up survey that was done by phone near the end of the field phase, in October (right-hand column). Here too, the innovation rates obtained with the short form are higher than those obtained with the long form, but for process, organizational and marketing innovation the confidence intervals overlap with those obtained for the long form. For product innovation the confidence intervals do not overlap. In the phone interviews respondents were probed and asked to give examples of innovations they introduced during the reference period, or were given examples of potential innovations, to ascertain that their responses were in line with the international guidelines on measuring innovation of the Oslo Manual (OECD, 2005).

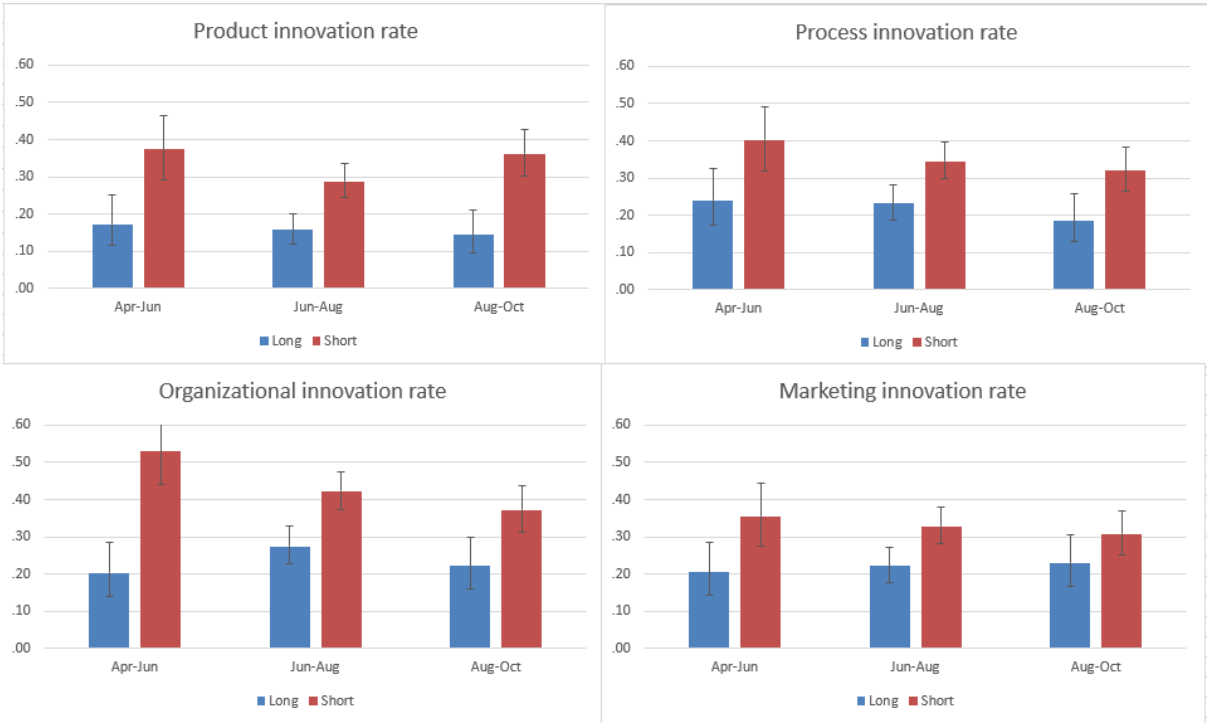
The difference between the results for the two short form conditions cannot merely be attributed to a measurement effect of the survey mode, i.e., a difference between the true value of a target variable and the value processed by a survey The paper and web responses for the short form in the experimental condition were obtained throughout the full field phase, whereas the phone responses for the short form nonresponse survey were only obtained near the end of the field phase, from converted nonrespondents. The comparison between both short form conditions contaminates a measurement error component of a mode effect (interviewer-assisted vs. self-administered) with a potential nonresponse bias effect. It would have been interesting to also randomize the survey mode (paper/web vs. phone interview) for the nonresponse survey. Converting nonrespondents late in the field phase into responding generally, however, is hard to achieve, and it was decided to spend the limited resources available to the standard, phone nonresponse survey.

Figure 3 shows the innovation rates and 95% confidence intervals for the two conditions in our randomized experiment, the long and the short form, spread over the field phase: (1) before the first reminder (April-June), (2) following the first reminder and before the second reminder (June-August), and (3) following the second reminder (August-October). The innovation rates over time are shown for the four types of innovation. A potential alternative explanation for the higher innovation rates traditionally obtained in the nonresponse survey, that many innovators tend to not respond in the regular phase because they see how much work a "yes, we innovate" response would require, is not confirmed, in this figure. In the long form, innovators are 'punished' for their "yes, we innovate" response by receiving many more detailed follow-up questions.

On the contrary, non-innovators get rewarded for their "no, we do not innovate" response by being allowed to skip many questions, significantly reducing their response burden. One alternative reasoning then goes that innovators are less inclined to respond to the long innovation survey form, as they realize how much work it would be, whereas non-innovators are more quickly won over to respond, as responding is not very demanding for them. We do not see any evidence supporting this reasoning in Figure 3, however. For the long form no clear trends over time can be discerned for the four innovation types. For the short form, innovation rates appear to slightly decrease over time for three of the four innovation types: process, organizational and marketing innovation, though the 95% confidence intervals for all three time periods overlap. The 95% confidence intervals also overlap for the three time periods for product innovation as measured using the short form, where the pattern looks slightly more curvilinear going from the beginning to the end of the field phase.

Figure 3 does not support the hypothesis that innovators tend to wait to respond (needing more reminders etc.) during the field phase, while non-innovators are thought to be more easily inclined to respond.

Figure 3. Innovation rates by innovation type and survey form, over time; 95% confidence intervals are indicated



In the short form survey there are no follow-up questions for the four types of innovation, hence the 'punishment' for innovators in terms of response burden is much less here than in the long form survey. Yet we do not see dramatically different response *patterns* over time for the short form as compared to those for the long form. Assuming respondents near the end of the field phase can be considered proxies for nonrespondents, we do not see any evidence that the near-the-end-of-the-field-phase respondents include more innovators than non-innovators.

4 Discussion

We performed a randomized experiment in which small (medium) low-tech business enterprises were randomly assigned either a long or a short innovation survey form. All other aspects of the survey process were kept equal between the two treatment conditions. The experiment was inspired by the fact that rather counterintuitive results had been obtained before, when using a short form for surveying nonrespondents in the context of the community innovation survey (CIS) in Europe: higher

innovation rates were obtained for the subset of surveyed nonrespondents than for the respondents in the regular, long form survey. Given that sampled units more familiar with or more interested in the surveyed topic are generally more inclined to respond, one would rather expect results in the opposite direction: non-innovators who are more inclined not to respond (as they deem the innovation survey 'not interesting/not relevant'), that are picked up with the short nonresponse survey form, resulting in lower innovation rates for the nonresponse survey.

The comparison of the results obtained with the regular long form survey and the short phone nonresponse survey innovation form, are contaminated by multiple factors, however. First, there is the comparison between early or earlier responders and late responders who originally had not responded but were 'converted' into responding. Second, there is a difference in response mode: the long form was self-administered, as respondents either responded on paper or in a web survey form, whereas the short form was interviewer-assisted and was administered in a phone interview. A third factor is the questionnaire length: the long form follows the Eurostat harmonized form and contains many follow-up questions on product and process innovation. It also contains extensive definitions of each of the four types of innovation (product, process, organizational and marketing innovation). The short innovation form is a simplified version of the long form. It contains one general definition of innovation and then four straightforward questions, one for each of the four types of innovation. There are no follow-up questions for any specific type of innovation, rather there are three follow-up questions that all respondents need to answer, on two of the most common innovation activities (R&D and acquisition of machinery).

In our randomized experiment, we isolated one of these three factors: questionnaire length.

When the only factor we manipulated was questionnaire length, we obtained considerable differences in innovation rate: higher innovation rates were obtained with the simplified short (self-administered) form than with the standard (self-administered), long form. This is a clear method effect: innovation measurements obtained with a short form versus with a long form are not equivalent.

A study by the World Bank (Cirera & Muzi, 2016) also found differences in innovation rates between short and long form innovation surveys when measuring innovation in developing countries. The short questionnaire was embedded in a general firm-level survey. The long form was a standalone innovation survey. The differences they found were in the same direction as those we found: innovation rates were higher in the short form than in the long form. The World Bank study included a within-enterprise design, however, i.e., the short innovation module embedded in the general enterprise survey was administered first, and a randomly selected subset of these enterprises were given the long form innovation survey later. Both forms were interviewer-administered, in face-to-face interviews. In the long form more probing was done, so the authors of the World Bank study consider the (lower) innovation rates obtained with the long form survey to be more trustworthy.

The third edition of the Oslo Manual (OECD/Eurostat, 2005) stated

> "For many small units and units in sectors with little innovation activity, the response burden for a full innovation questionnaire may be quite large relative to their innovation activity. Unit non-response rates may also be higher for these units. In such cases, shorter survey questionnaires that focus on a set of core questions can be useful. Short-form questionnaires can also be used in surveying units that have not reported innovation activity in previous innovation surveys. Conversely, for individual units in the above-mentioned groups (small units or less innovative sectors) which have previously reported substantial innovation activity, full questionnaires may be used." (pp. 124-125, § 456)

As the evidence obtained from the randomized experiment in Belgium and from the experiment conducted by the World Bank in developing countries indicates that short and long form innovation measurements cannot automatically be assumed to be equivalent, it was recommended that this paragraph be revised in the fourth edition of the Oslo Manual (OECD/Eurostat, 2018), and this has been done indeed.

One possible explanation for the differences found is that a sizable number of respondents have used satisficing strategies when responding to the surveys. Satisficing occurs in surveys when a respondent does not provide an optimal response to a question but provides a response that is 'good enough' instead, to reduce the cognitive burden required in responding to the survey (Krosnick, 1991, 1999). Hence, rather than reading the definition of innovation that is provided thoughtfully and carefully, and

then ponder evidence to decide whether any changes, improvements, etc., their enterprise made qualify as innovation, respondents might just browse the questions in the survey, and then, depending on the context, make snap judgments on what to respond.

As several scholars have described (e.g., Tourangeau, Rips & Rasinski, 2000), in a limited amount of time respondents ought to make sense of the questions asked, retrieve the relevant information, make a judgment, and report in accordance with the alternative provided in the survey. They might make spot judgments, mainly based on information that is accessible at that point in time. They might truncate the information search process as soon as they have collected enough information to formulate a judgment (Schwarz, 1999). By affecting the accessibility and salience of the information that respondents use to respond, prior items asked in the survey form may provide a framework to respond to later questions, generating context effects (Angelini, Bertoni & Corazzini, 2017).

When faced with the long form sampled units receive in the mail in the randomized experiment in Belgium, smaller firms might be somewhat intimidated with the 20-page booklet they see in front of them, and might reason: “we’re just small game, we’re not part of the pack of game changing market leaders, so we’ll just indicate we did not innovate during the period surveyed.” Larger firms might reason: “we’re market leaders, sure we innovated.” In both cases, a more generic judgment was made, without considering specific details of what constitutes innovation.

When faced with the short form sampled units in Belgium got by mail, the smaller firms that were included in the randomized experiment might have reasoned “sure we had novelties” without considering more details of those novelties in their judgment (the word innovation is only mentioned once in the short form, in an explanation box, that can easily be overlooked. Respondents oftentimes tend to overlook explanations. In Dutch the word ‘innovation’ also has more value judgment attached to it than the words ‘new’ and ‘significantly improved’ that were used in the actual wording of the questions. Many smaller firms are also not familiar with the word ‘innovation’.).

In the long form, respondents get 'punished' for a "yes, we innovated" response, as they are then steered towards several follow-up questions, that ask for more details of their innovations, including financial details for expense categories that they typically do not have in their accounting records (Galindo-Rueda & Van Cruysen, 2016). Non-innovators can skip all those follow-up questions, significantly reducing their workload. In the short form there are no such consequences: there are no follow-up questions attached to each of the four types of innovations included in the survey. The short form only includes two types of follow-up questions that are kept brief. They relate to two of the most common innovation activities (R&D and acquisition of machinery). All respondents to the short form survey need to respond to these follow-up questions, not just those who said 'yes' to one or more of the preceding innovation questions.

Method effects found in innovation measurement in other countries

Method effects have also been found in innovation surveys in other countries. Statistics Norway (Wilhelmsen, 2012) performed a randomized experiment amongst the strata that are generally randomly sampled in their innovation survey. They manipulated two factors: (1) whether the survey form sent to sampled units was voluntary or mandatory, and (2) whether the survey form was a combined R&D – innovation survey or a stand-alone innovation survey. Both factors had a clear impact on innovation rates: innovation rates in the voluntary survey were higher than those obtained in the mandatory form (e.g., respectively 45% and 38% for product and/or process innovation), and innovation rates obtained in the combined R&D – innovation survey form were lower than those in the stand-alone innovation survey (e.g., respectively 24% and 36% for product and/or process innovation). Selection bias might explain the impact of the voluntary vs. mandatory status of the survey: in the voluntary survey, enterprises more familiar with or more interested in innovation tend to self-select themselves into responding to the survey, whereas non-innovators are more inclined to reason "this survey is not relevant for us, so we won't respond", resulting in higher innovation rates when compared to the mandatory survey, where innovators and non-innovators both alike are obliged to respond.

In the combined R&D – innovation survey the R&D section preceded the section on innovation, which may have biased respondents to see innovation more in the light of R&D, that is, to set the bar higher for something to be considered an innovation and to use more stringent criteria for labeling something an innovation.

Statistics Netherlands went from a paper format for their innovation survey to a web format in 2011, and they found that this format change resulted in a break in series: their innovation rates jumped from 35% to 48% (CBS, 2012). Statistics Netherlands explained their findings by referring to satisficing: in the paper form respondents see their workload, they see that "yes, we innovate" responses are 'punished' by a series of follow-up questions, including questions asking for very detailed monetary aspects of innovation. "No, we don't innovate" responses are 'rewarded' by being allowed to skip entire sections of the survey form. Consequently, respondents might adjust their response behavior, and in many cases consider a "no, we did not innovate" a good enough response for the survey at hand. In the web format, Statistics Netherlands administered only one item per screen (cf. Couper, 2008, a 'paging design'), hence respondents were not aware of the punishments nor rewards that would follow a yes/no response. Statistics Netherlands reasoned that the web format they used, displaying one question per screen, reduced satisficing in respondents.

5 Conclusions

The method effect found in our study and in the study by the World Bank (long vs. short innovation forms), the method effects Statistics Netherlands and Statistics Norway found (paper vs. web survey format; voluntary vs. mandatory surveys; stand-alone vs. combined R&D and innovation survey) all indicate that innovation measurement in its current form is like attitude measurement, highly susceptible to method effects.

These studies confirm the conclusions drawn in an OECD study of metadata collected about innovation surveys that "some innovation survey design features have a significant and unintended impact on the comparability of reported innovation rates across countries" (OECD, 2012).

Method effects seriously compromise the international comparability of statistics derived from innovation surveys. Countries differ widely in the methods they use to measure innovation, and hence the innovation rates they obtain do not only reflect actual differences in innovation behavior between countries, but also include significant variation merely due to the measurement instrument that was used to measure it.

For example, one of the flagship publications of the European Union (EU), is the European Innovation Scoreboard. It includes several variables that are derived from the Community Innovation Survey. However, four of the 28 EU member states have a voluntary innovation survey, the remaining 24 member states have a mandatory survey. We know from the randomized experiment performed by Statistics Norway that these two types of measurement are not equivalent, that voluntary measurements tend to yield higher innovation rates than mandatory measurements. Three of those four member states with a voluntary innovation survey have been performing a nonresponse survey, to derive nonresponse adjusted weights. For this nonresponse survey they use a short form survey. We have demonstrated above, however, that results obtained with such short form innovation survey are not equivalent to results obtained with a long form innovation survey. The nonresponse adjustment obtained from the short form survey then adds even more noise to the measurement, obscuring the signal (information on true incidence of innovation) that we are really interested in. EU member states also differ in the mode used for their innovation survey: paper format, web format, phone interviews, face-to-face interviews (Eurostat, 2013). This variation in methodology used to measure innovation between countries increases the noise and the uncertainty in innovation measurement and reduces what signal we can capture from it.

The various studies demonstrating method effects in measuring innovation demonstrate once more that “the importance of good questionnaire design cannot be overestimated” (Statistics NZ, 2019).

6 Recommendations

Innovation is an important topic for policy, and innovation measurement is important as a foundation for this policy. We would therefore like to make some recommendations to improve innovation measurement.

The third edition of the Oslo Manual (OECD/Eurostat, 2005) mentions some standard questionnaire design recommendations such as pre-testing before fielding the questionnaire (§ 451, p.123), attention to be paid to translating questionnaires in the context of international testing (§ 455, p. 124). In the earlier mentioned European Social Survey (ESS) new modules are pre-tested in two countries prior to their inclusion in the main survey, to also test for usability of modules in a multinational context. Currently, in the European Union a harmonized survey form is used to measure innovation. New questions are cognitively tested in multiple countries before they are added to the harmonized form. To make a stronger case for usability of new questions – in terms of reliability and validity in a multinational context – we argue that preferably new questions should be pre-tested in at least two countries, and the results of these pre-tests carefully studied before inclusion in the harmonized innovation survey form. Moreover, translation issues should at least be investigated.

Besides the ESS there are other large-scale projects that run surveys and measurements in a multinational, multiregional and multicultural context (3MC surveys), e.g., Trends in International Mathematics and Science Study (TIMSS), Progress in International Reading Literacy Study (PIRLS), Programme for International Student Assessment (PISA). We would be wise to gain from the lessons learned there (see e.g., Harkness, Braun, Edwards, Johnson, Lyberg, Mohler, Pennell & Smith, 2010). For example, as mentioned above, the team working on the ESS developed an extensive procedure to prevent potential translation problems. Likewise, they have extensive guidelines they follow to guard

the quality of the items included in the common questionnaire. They use Split Ballot-Multi-Trait, Multi-Method (MTMM) design to study the impact of item wordings, item orders etc. We see a clear need for more experimental studies documenting questionnaire design issues in measuring innovation at the firm level. Such studies need to be carefully designed, as countries not only differ in how they design and conduct their surveys, but also in the innovation policies they implement.

In their overview of results obtained from cognitively testing innovation concepts, Galindo-Rueda and Van Cruysen (2016) make several recommendations that we think are worthwhile. One of these is:

"The Oslo Manual presents a rather marked distinction between functional and other forms of changes to products and processes. Future Oslo revision work should attempt to make more explicit what those criteria are and formulate them in survey environments. Specific questions asking firms to describe the various dimensions of novelty may assist in this process. The experience accumulated from this and other related projects on the analysis of design can help formulate these in a more concrete way." (§ 41, p. 13)

We agree with this recommendation. The way innovation is measured now in the European Community Innovation Survey (and in many other countries) is that first respondents are taught what innovation is (by means of the extensive definition given first), they then must indicate whether they had those particular types of innovation or not, followed by some follow-up questions. From the phone calls we get from sampled units, we know that even today, not all enterprises are familiar with the term innovation. Especially smaller enterprises tend to be less familiar with the term. From the cognitive interviews done in the context of the OECD study mentioned above (Galindo-Rueda & Van Cruysen, 2016), we know that especially respondents in large enterprises tend to have a higher threshold to consider something to be an innovation than respondents in other enterprises: respondents in large enterprises only considered new-to-market products as product innovations (new products that the enterprise launched on the market before its competitors did) and consider new-to-firm-only product innovations (products an enterprise introduced to the market for the first time, while competitors already offered this product on the market) not to be true innovations. The

Oslo Manual (OECD/Eurostat, 2018) and the Community Innovation Survey currently consider both types of products to be innovations.

One could draw a parallel with the measurement of life satisfaction here. Life satisfaction is an economically relevant measure of subjective wellbeing (Diener, 1984). The measurement of general life satisfaction, however, appears to be very volatile and susceptible to context effects. However, when the concept of life satisfaction is unpacked into its various constituents (satisfaction with one's health, satisfaction with one's relationships etc.) the measurement becomes more reliable (Angelini, Bertoni & Corazzini, 2017; cf. Tversky & Koehler, 1994).

Similarly, the unpacking of the various aspects of novelty that are included in innovation might help respondents pay attention to each these aspects separately, whereas otherwise they would make a global judgment, heavily influenced by the mood of the moment – and hence, very susceptible to context effects.

This way we might remove the wording 'significantly improved' from our innovation survey forms, as we might translate that more subjective judgment into a series of more objectively worded improvements or changes that enterprises may tick. Survey producers then later can decide on various cutoffs to be used for reporting.

On a more practical level also, the recommendation that the Eurostat methodological guidelines for CIS (Eurostat, 2023) currently includes, of performing a non-response survey in case the response rate is below 70%, and to use the results of this non-response survey for adjusting the grossing up weights, appears to increase rather than decrease the inaccuracy of innovation measurement. Therefore, it seems preferable to drop this recommendation or at least refine and then test it.

References

Angelini, V., Bertoni, M., & Corazzini, L. (2017). Unpacking the determinants of life satisfaction: a survey experiment. *Journal of the Royal Statistical Society. Statistics in Society Series A*, 180, 225-246. DOI: https://doi.org/10.1111/rssa.12190.

Arrow, K. (1962). Economic welfare and the allocation of resources or inventions. In R. Nelson (Ed.), *The Rate and Direction of Inventive Activity*. Princeton, NJ: Princeton University Press. Available at: http://www.nber.org/books/univ62-1 (accessed August 2024).

Arundel, A., O'Brien, K., & Torugsa, A. (2013). How firm managers understand innovation: implications for the design of innovation surveys. In F. Gault (Ed.), *Handbook of Innovation Indicators and Measurement* (pp. 88-108). Cheltenham: Edward Elgar. DOI: https://doi.org/10.4337/9780857933652.00012.

CBS (2010). *Kennis en economie 2009* [Knowledge and economy 2009]. Den Haag/Heerlen, Centraal Bureau voor de Statistiek. Available at: https://www.cbs.nl/nl-nl/publicatie/2010/25/kennis-en-economie-2009 (accessed August 2024).

CBS (2012). *ICT, kennis en economie 2012* [ICT, knowledge and economy 2012]. Den Haag/Heerlen, Centraal Bureau voor de Statistiek. Available at: https://www.cbs.nl/nl-nl/publicatie/2012/24/ict-kennis-en-economie-2012 (accessed August 2024).

Cirera, X., & Muzi, S. (2016). *Measuring firm-level innovation using short questionnaires. Evidence from an experiment* (Policy Research Working Paper 7696). Washington DC: World Bank Group. Available at: https://hdl.handle.net/10986/24531 (accessed August 2024).

Couper, M.P. (2008). *Designing effective web surveys*. New York: Cambridge University Press. DOI: https://doi.org/10.1017/CBO9780511499371.

De Leeuw, E. D. (2008). Choosing the method of data collection. In E.D. de Leeuw, J.J. Hox & D.A. Dillman (Eds.), *International handbook of survey methodology* (pp. 113-135). New York: Lawrence Erlbaum Associates. DOI: https://doi.org/10.4324/9780203843123.

Diener, E. (1984). Subjective well-being. *Psychological Bulletin*, 95, 542-575. DOI: https://doi.org/10.1037/0033-2909.95.3.542.

Eurostat (2013*). Synthesis Quality Report of CIS 2010* (Doc.Eurostat/G6/STI-ISS/2013/5). Available at: https://ec.europa.eu/eurostat/cache/metadata/en/inn_cis2_esms.htm (accessed August 2024).

Eurostat (2023). *Harmonized Data Collection CIS 2022. Methodological recommendations*.

Eurostat (2015*). Short synthesis of the CIS 2012 quality reports* (Doc.Eurostat/G4/STI/2015/10). Available at: https://ec.europa.eu/eurostat/cache/metadata/Annexes/inn_cis8_esms_an_6.pdf .

Galindo-Rueda, F., & Van Cruysen, A. (2016). *Testing innovation survey concepts, definitions and questions: Findings from cognitive interviews with business managers* (OECD Science, Technology and Innovation Technical paper). Paris: OECD.

Groves, R.M., Presser, S., & Dipko, S. (2004). The role of topic interest in survey participation decisions. *Public Opinion Quarterly*, 68, 2-31. DOI: https://doi.org/10.1093/poq/nfh002.

Groves, R.M., Couper, M.P., Presser, S., Singer, E., Tourangeau, R., Acosta G.P., & Nelson, L. (2006). Experiments in producing nonresponse bias. *Public Opinion Quarterly*, 70, 720-736. DOI: https://doi.org/10.1093/poq/nfl036.

Harkness, J.A., Braun, M., Edwards, B., Johnson, T.P., Lyberg, L., Mohler, P.P., Pennell, B.-E., & Smith, T.W. (Eds.). (2010). *Survey methods in multinational, multiregional and multicultural contexts*. Hoboken, NJ: Wiley. DOI: https://doi.org/10.1002/9780470609927.

Hoskens, M., & Andries P. (2015, December). *The distinction between product/process /organizational and marketing innovation: Does our respondents' understanding correspond to the Oslo view?* Paper presented at the OECD NESTI/EUROSTAT scoping workshop on the 3rd revision of the Oslo Manual, Oslo.

Kersten, H.M.P., & Bethlehem, J.G. (1984). Exploring and reducing the nonresponse by bias by asking the basic questions. *Statistical Journal of the U.N. Economic Commission for Europe*, 2, 369-380. DOI: https://doi.org/10.3233/SJU-1984-2404.

Krosnick, J. A. (1991). Response strategies for coping with the cognitive demands of attitude measures in surveys. *Applied Cognitive Psychology*, 5, 213-236. DOI: https://doi.org/10.1002/acp.2350050305.

Krosnick, J. A. (1999). Survey research. *Annual Review of Psychology*, 50, 537-567. DOI: https://doi.org/10.1146/annurev.psych.50.1.537.

OECD (1992). *OECD Proposed Guidelines for Collecting and Interpreting Technological Innovation Data – Oslo Manual*. Paris: OECD Publishing.

OECD (2002). *Frascati Manual 2002: Proposed Standard Practices for Surveys on Research and Experimental Development. The Measurement of Scientific and Technological Activities*. Paris: OECD Publishing. DOI: https://doi.org/10.1787/9789264199040-en.

OECD (2012). *The NESTI review of innovation survey design and implementation: Interim findings from a metadata collection and analysis* (DSTI/EAS/STP/NESTI(2012)5). OECD, Paris.

OECD (2015). *Frascati Manual 2015: Guidelines for Collecting and Reporting Data on Research and Experimental Development. The Measurement of Scientific, Technological and Innovation Activities*. Paris: OECD Publishing. DOI: https://doi.org/10.1787/9789264239012-en.

OECD/Eurostat (1997). *Proposed Guidelines for Collecting and Interpreting Technological Innovation Data – Oslo Manual.* Paris: OECD Publishing. DOI: https://doi.org/10.1787/9789264192263-en.

OECD/Eurostat (2005). *Guidelines for Collecting and Interpreting Innovation Data – Oslo Manual*. Third edition. Paris: OECD Publishing. DOI: https://doi.org/10.1787/9789264013100-en.

OECD/Eurostat (2018). *Oslo Manual 2018: Guidelines for Collecting, Reporting and Using Data on Innovation, 4th Edition*. Paris: OECD Publishing/Luxembourg: Eurostat. DOI: https://doi.org/10.1787/9789264304604-en.

Peters, B., & Rammer, C. (2013). Innovation panel surveys in Germany. In F. Gault (Ed.), *Handbook of Innovation Indicators and Measurement* (pp. 135-177). Cheltenham: Edward Elgar. DOI: https://doi.org/10.4337/9780857933652.00015.

Schwarz N. (1999). Self-reports: how the questions shape the answers. *American Psychologist*, 54, 93–105. DOI: https://psycnet.apa.org/doi/10.1037/0003-066X.54.2.93.

Schumpeter, J. (1942). *Capitalism, Socialism and Democracy*. New York: Harper & Brothers. Available at: https://periferiaactiva.wordpress.com/wp-content/uploads/2015/08/joseph-schumpeter-capitalism-socialism-and-democracy-2006.pdf (accessed August 2024).

Solow, R. M. (1956). A contribution to the theory of economic growth. *The Quarterly Journal of Economics*, 70, 65-94. DOI: https://doi.org/10.2307/1884513.

Statistics NZ (2019). *A guide to good survey design (5th ed.).* Available at: https://www.stats.govt.nz/assets/Uploads/Methods/A-guide-to-good-survey-design-fifth-edition/a-guide-to-good-survey-design-fifth-edition.pdf.

Tourangeau, R., Rips, L., & Rasinski, K. (2000). *The psychology of survey response*. Cambridge: Cambridge University Press. DOI: https://psycnet.apa.org/doi/10.1017/CBO9780511819322.

Tversky, A., & Koehler, D. J. (1994). Support theory: A non-extensional representation of subjective probability. *Psychological Review*, 101, 547–567. DOI: https://psycnet.apa.org/doi/10.1037/0033-295X.101.4.547.

Wilhelmsen, L. (2012). *A question of context. Assessing the impact of a separate innovation survey and of response rate on the measurement of innovation activity in Norway* (Document 51/2012). Statistics Norway, Oslo. Available at: https://www.ssb.no/en/teknologi-og-innovasjon/artikler-og-publikasjoner/a-question-of-context (accessed August 2024).

KU LEUVEN
Faculty of Economics and Business
Management, Strategy and Innovation (MSI)
Naamsestraat 69 bus 3535
3000 LEUVEN, Belgium
tel. + 32 16 32 67 00
msi@econ.kuleuven.be
https://feb.kuleuven.be/research/MSI/