A Critical Assessment of Quality and Validity of Composite Indicators of Innovation

Dániel Vértesy[1]
Joint Research Centre of the European Commission

DRAFT PAPER SUBMITTED TO THE OECD BLUE SKY III FORUM
ON SCIENCE AND INNOVATION INDICATORS

Do not cite without author's consent

29 Jul 2016

Abstract

Composite indicators are widely used to capture both the process and the outcomes of complex phenomena such as innovation. This paper aims to support this discussion by assessing the quality and validity of the most commonly used composite indicators of innovation. After discussing concepts related to the definition of innovation, the meaning of indicators and their use for innovation policy, we provide an overview of the practices of innovation indicators. We argue that a co-evolutionary framework is necessary to assess the quality and validity of composite indices of innovation, since the global economic environment, the policy goals, the availability of data and statistical methods keep changing. At the core of any assessment lies the quantification process, which is inherently dynamic and interactive. We conclude from an overview of assessment frameworks that assessing the relationship between component indicators and the conceptual framework needs to focus on both the development process and the results. This paper proposes that for the latter, instantaneous assessment is possibly by the analysis of the variance of composite scores explained by component indicators and comparison with the conceptual or policy aims of the indicator. We test these in the case of four country-level composite indicators of innovation (the Global Innovation Index, the Summary Innovation and the Innovation Output Index of the European Commission, and the Fraunhofer-consortium's Innovation Indicator). The study finds that the past decade has brought more incremental, rather than radical changes in composite indicators of innovation, and overall composite indicators essentially measure science and technology-based activities, while leave many conceptually important aspects are weaker drivers of country scores. The conclusions point in the direction of the need for innovation policies that are ready to embrace selectivity; for which the development of composite indicators offers useful inputs to highlight tradeoffs.

1 Introduction

Policy makers and business strategists usually monitor multiple indicators to compare and benchmark innovative capacities, activities and outcomes. While the cost of access to an ever-growing pool of statistical information is rapidly decreasing, the importance of selecting those relevant for making strategic decisions is becoming ever more crucial. This is especially true when quantitative evidence is a source of legitimation for policy interventions or business decisions. Given these trends, it is not

[1] email: daniel.vertesy@jrc.ec.europa.eu; address: Joint Research Centre of the European Commission, TP361, Via E. Fermi 2749 Ispra (VA), 21027, Italy

surprising that composite indicators are widely used for transforming multi-dimensional phenomena into a single (or a fewer, more manageable number of) dimension(s).

This is certainly the case with measuring "innovation". Annually published country rankings for the WIPO-INSEAD Global Innovation Index (GII) or the European Commission's Summary Innovation Index (SII) continue to attract broad public attention.[2] Over the last decade, the underlying concepts, selection of indicators, and modeling choices (weighting, aggregation methods) for these indices have been critically assessed (Grupp and Mogee, 2004; Schibany and Streicher, 2008; Grupp and Schubert, 2010; Gault, 2013, Hollanders and Foray, 2016 – among others) but also gradually refined (Sajeva et al, 2005; Hollanders and van Cruysen, 2008; Saisana and Filippas, 2013). Along the way, composites have exposed shortcomings in their component indicators, triggering changes in the data collection and dissemination process. While composite indicators are apparently here to stay, the very fact that there is an ongoing critical discussion keeps the "agora model" (Barré, 2001, 2010) of indicator development alive (Hardeman and Vertesy, 2016).

This paper aims to support this discussion by assessing the quality and validity of the most commonly used composite indicators of innovation. In our proposed framework, the *quality* of an indicator relates to its statistical properties (Saltelli, 2007; Saisana et al, 2005, 2011), while the *validity* of an index relates to the link between component indicators or aggregates and the conceptual framework (JRC-OECD, 2008; Saltelli et al, 2013). In the case of country-level innovation indicators, validity depends on how well an innovation index quantifies what is relevant for innovation policy. The paper concludes that both quality and validity are more normative concepts than what is usually acknowledged. This becomes evident from taking an evolutionary approach, recognizing the co-evolutionary dynamics of indicator development within changes in the global economic environment, the goals of science, technology and innovation policies, the availability of data as well as analytical and measurement tools. In this context, it is striking that the past decade has brought more incremental, rather than radical changes in composite indicators of innovation.

2 Measuring innovation with composite indicators

Before assessing composite indicators of innovation, we need to address some critical questions, such as (1) How to define innovation? (2) What is the meaning of indicators and how are they used (or misused) for (innovation) policy? (3) What is the connection between innovation policies and indicators? (4) What are the most widely used composite indicators of innovation? The aim of this section is to give an overview of some of the underlying concepts and practices.

2.1 Innovation: from concept to measurement

Our paper focuses on composite indicators that aim to compare countries in terms of innovation. The third wave of the *Oslo Manual* reflects the strongest common understanding of innovations: "the introduction of a new or significantly improved product (good or service), or process, a new marketing method, or a new organizational method in business practices, work-place organization or external relations" (OECD-Eurostat, 2005). While this definition is broad in the sense that it encompasses not only technological innovation, its focus is the firm that brings novelty to the (broadly understood) market. Aggregating the innovative activities of many firms to define and measure the innovative activities of geographical areas such as countries is not straightforward, and entails making many choices related to the weighting of different types of activities (i.e., Should innovations of all types, of all degree of novelty, in all economic activities, count equally? If not, what are the right weights to apply?).

[2] See i.e. "Global Innovation Rankings: The innovation game" *The Economist* 17 Sep 2015; or "Innovation in Europe Improves, But Only in a Handful of Countries" *The Wall Street Journal* 14 Jul 2016.

Moreover, scholarship on innovation showed that the innovation process occurs not only within firms, rather in a complex interaction that involves other firms as well as research organizations and universities, public funding agencies, regulating bodies, or users. In other words, if one aims to measure how countries innovate, the performance of national systems of innovation (Lundvall, 1992, 2007; Nelson, 1993; Freeman, 1995) need to be defined and quantified – challenges that multiply the already existing complexities of measuring the innovation performance of a set of firms.[3] Furthermore, comparability of countries is influenced by structural, scale and time factors that are central to innovation in a firm. Given the breadth of what national innovation systems encompass, when it comes to measurement, authors often focus on selected elements of it or for a certain purpose (i.e., technological capabilities or innovative capacities, scientific, technological or innovative interactions, the creation of new inventions, etc.). In sum, complex, often untraceable set of conceptual choices are inherently involved in the measurement of innovation at aggregate levels.

Driven by the definitions, multiple means are used to innovations: direct measures of innovations – following the object (the innovation itself) or subject (the innovator) approach and making use of surveys or expert opinions – exist alongside indirect measures of the innovative process. These include information on the innovation efforts (R&D or non-R&D innovation expenditures, personnel, etc.) and outcomes such as patents, or changes in the broader economic performance or structures (see Smith, 2005; Gault, 2013). Accordingly, statistical data stemming from the various ways to measure innovation is loaded with a similar depth of variety, as the reliability of data from innovation surveys as well as indirect measurements may be diverse. Nevertheless, a significant part of innovation survey data can be considered of sufficient quality for the purposes of robust econometric analyses (Mairesse and Mohnen, 2010).

The measurement of innovation system performance is a more complex task, which – depending on the purpose – may focus on the system functions (Hekkert et al, 2007), its ability to achieve certain outcomes, or directly capture the performance of its elements, such as actors, networks, and institutions. We conclude that beyond the identification of *what* is measured and *how* is measured, the outcomes of measurement are also influenced by *why* we intend to to define and measure innovation in a firm or in a country. Depending on whether the aim is to make an informed judgement to achieve high returns on investment, or to understand the dynamics of long-term economic development, technological change and catch-up (see Fagerberg and Srholec, 2008), innovation performance may turn out to differ significantly.

2.2 Indicators for policy, indicators as models

In the context of evidence-based policies, metrics are ever more frequently called upon to legitimate policy interventions (Wilsdon et al., 2015). The ability to measure innovation is perceived crucial for the design of effective and efficient policies (Grupp and Mogee, 2004), and the lack of measurement is often a way to hide weaknesses in the innovation system. Benchmarking and evaluation of distant, mostly similar entities – the key purposes of using indicator – is therefore highly informative. Nevertheless, especially because of the widespread political implications of “governance by indicators”, there is reason to exercise caution (Davis et al, 2012). In the domain of science, technology and innovation indicators, the cautionary remarks are also reflected in the scholarly literature both in terms of technical critiques (Grupp and Mogee, 2004; Grupp and Schubert, 2010) as well as normative critiques (related to development process or legitimacy, see Barré, 2010).

Debates around the development and use of indicators are centered on the understanding that quantification is always a conversion with involve a certain degree of simplification of reality, and the

[3] Not to mention that while national institutions matter, the geographical scope of interactions that influence innovation often encompasses many countries.

societal pressure calling for comparisons. Indicators, seen more optimistically, offer learning and self-discovery through the measurement, comparison and monitoring process, and may offer early warning signal, show trends and bottleneck, and identify strength and weaknesses. On the other hand, indicators are models, and thus selective. Normative choices (including the ability to manipulate by selection and weighting), made under uncertainty, are inherent in any kind of indicators (Saltelli et al, 2013) – but at the same time, it is important to recognize that indicators gain meaning only in a socio-political context. There is a two-way interaction between indicators and policies: indicator development is shaped by policy needs and the policy discourse (Godin, 2002), while in turn, indicators influence the policy discourse.

2.3 Innovation policy perspectives and the demand for indicators

The aspects of innovative performance governments are interested in monitoring and influencing are related to the innovation policy 'paradigm'. Is innovation policy conducted in isolation, or as part of a coordinated set of policies? What is the rationale for intervention?

If the rationale of innovation policy is limited to encouraging firms to engage in innovation and entrepreneurial activities, governments are concerned with providing framework conditions in which firms can operate, and recognizing and removing obstacles understood as market (or system) failures. This approach is likely associated with a 'neutral' understanding of innovation – the more innovation the better (i.e., larger share of firms reporting to innovate, larger spending on innovation, etc.) – regardless of the types of innovation, the size or sectoral belonging of firms that innovate, or how the outcomes relate to other policy goals, such as employment (if the kind of innovation pursued is not replacing jobs with machines), or redistribution of incomes. It is assumed that the outcomes of innovation and technological change will diffuse in the economy, spillover effects will lead to social gains. These implications are not the main concern of innovation policy understood in this narrow sense. On the other hand, however, scientific and technological activities, skills and capacity to innovate or engage in entrepreneurial activities – what may be a precursor to the innovation process –, is of interest.

In another understanding, innovation is perceived as a tool to achieve strategic aims. Paraphrasing Lundvall and Borrás (2005), such innovation policy is essentially about the promotion of production, diffusion and use of innovations in order to realize national or supra-national objectives. An objective may be economic development and catch-up, for which innovation serves to accelerate growth and increase competitiveness of industries. Other objectives could be the improvement of military capabilities, or, in contrast, finding new ways to tackle hitherto unresolvable societal problems (i.e., improvements in health, environment, or the survival of humanity). Tools to achieve such policies with fostering system transitions (Hekkert et al, 2007), or creating new markets (Mazzucato, 2016). The main distinction of this approach is that innovative activities of a certain character are preferred over another, depending on the aim.

Needless to add, there may be more approaches for innovation policy and these two approaches may be extremes. However, the key point is that they are rarely followed consistently in practice. All too often, an amalgam of the two emerges from the political process of harmonizing the strategies of different stakeholders (ministries, agencies, business associations, etc.). This is not without problems, as central questions such as the effectiveness of certain innovation policy measures depends on the perspective chosen. The different approaches may necessitate fundamentally different monitoring mechanisms.

How does this influence the demand for and the use of composite indicators? While the strength of scoreboards lies in allowing interested parties to monitor a variety of related processes in accordance with their preferred approaches and strategies, there are considerable overlaps as well as a few tradeoffs between processes that indicators measure. Carefully constructed composite indicators have the potential to summarize indicators without a significant loss of information, highlight an unobserved concept or normative choices and help analyze desired aspects of innovative performance (Saltelli, 2007; Ravallion, 2012). However, as statistics does not offer solutions for normative decisions, one cannot emphasize enough the need for clarity in the approaches and strategies for innovation policy even for the purposes of seemingly neutral practices such as monitoring. Without this, any indicator, composite indicators included, can be of limited support to policy.

2.4 Composite indicators of innovation

It is not the main purpose of this study to offer a complete survey or a taxonomy of composite indicators of innovation. Nevertheless, a brief overview of composite indicators of innovation is in place here to inform our selection of the indices we investigate closer in our study.

Composite indices of innovation come in different shapes and sizes. Some are side products of scoreboards or dashboards of indicators published by international organizations or think tanks, while in other cases, composite indicators are the "main products", which are accompanied by indicators that help understand country ranks.[4] It is noteworthy that many of the scoreboards are purposefully published without composite indices (see i.e. the OECD's Science, Technology and Innovation Scoreboard). Some are published by consultancy or media companies such as the Boston Consulting Group or Bloomberg[5], others by research think tanks or international organizations. The organization that publishes them also defines their impact – composite indices receive high visibility in the new media and are therefore strong advocacy tools. Some are published regularly (annually or biennially), while others only occasionally – for instance, the Economist Intelligence Unit's ranking of the world's most innovative countries was only published in 2007 and 2009. There are also overlaps with other indices. Apart from indices of innovation, indices of competitiveness (for instance the World Economic Forum's Global Competitiveness Report) or of research performance (even university rankings) typically build on a common pool of science, technology and innovation indicators, some even contain a pillar titled 'innovation'.

While the focus of our study is on country-level composite indicators, we note that innovation indices are also published at other levels. At the *regional level*, the European Commission's Regional Innovation Scoreboard and Index builds on the framework of the IUS/SII, or the Regional Competitiveness Index (see Annoni and Dijkstra, 2013) has an innovation pillar. At the *company level*, the Boston Consulting Group's "The Most Innovative Companies" reports annually publish league tables of leading innovators based on executive surveys. In a somewhat different context, composite indices are also used in micro-econometric studies to summarize different indicators of innovativeness at the firm level (for an example, see Coad and Rao, 2008).

Many of the composite indicators of innovation introduced in the early 2000s have been revised over time to accommodate methodological improvements, or developers' or publishers' changing priorities. For instance, the Summary Innovation Index was introduced as a tentative proposal in the 2001 European Innovation Scoreboard aggregating only the indicators in which a country performed strongest. In subsequent releases – also in line with the findings of robustness analyses – the index became more comprehensive both in terms of coverage of indicators, countries and years. Methodology changes may be particularly problematic as they make rankings according to different releases incomparable. Stronger indicators therefore pay attention to re-calculating past years' scores

[4] For a complete list, see Hollanders and Janz, 2013, Table 11.1 p.280-81

[5] See i.e. the Boston Consulting Group's "The Most Innovative Companies" [URL: https://media-publications.bcg.com/MIC/BCG-Most-Innovative-Companies-2015.pdf] "Bloomberg's 2015 ranking of the world's 50 most innovative countries" [URL: http://www.bloomberg.com/graphics/2015-innovative-countries/].

according to the latest methodology to allow intertemporal as well as cross-country comparison (i.e., the 2015 Summary Innovation Index publishes scores for eight years). This is crucial, as the ability to reflect trends is one of the central expectation from composite indices.

While many indicators – including those mentioned above – share a number of commonalities, the indicators are distinguished by a number of characteristics, such as who publishes them and how frequently, what concept or aspects of innovation they aim to measure, what policy goals they aim to support, their structure, their methodology of aggregation, their unit of analysis or geographical focus (Table 1). All these relate to choices developers make, which affect the quality and validity of the composite indices. In the next section, we propose a framework to assess them.

Table 1 Overview of regularly published, country-level composite indicators of innovation

Name (abbreviation)	**Global Innovation Index (GII)**	**Summary Innovation Index (SII) of the Innovation Union / European Innovation Scoreboards**	**Innovation Output Indicator (IOI)**	**Innovations-indikator (II)**
Published by (developed by)	INSEAD, WIPO, Cornell University	European Commission (UNU-MERIT)	European Commission	Fraunhofer ISI, ZEW, UNU-MERIT
Fist publication	2007	2001	2013	2005
Most recent	2015	2016	2016	2014
Revised in past?[a)]	Yes (2011)	Yes (...; 2011)	No	Yes (2011)
Frequency	Annual	Annual	Annual	Annual
Concept aimed to measure or benchmark	“the multi-dimensional facets of innovation”	The performance of innovation systems	“the outcomes of national innovation policies”	The performance of innovation systems
Policy goals aimed to support	long-term output growth, improved productivity, and job growth”	Industry (SMEs) and innovation policies	provide better jobs and make Europe more competitive	Growth and competitiveness
Structure (of most recent index)	79 variables aggregated step-wise in 21 sub-pillars, 7 pillars, 2 sub-indices and finally 1 index	25 variables aggregated directly in a single index; Separately, also aggregated into 8 dimension composites	4 components aggregated directly in a single index[b)]	38 variables aggregated (with overlap) into 5 sub-system scores and in an overall index
Aggregation method	Linear average with weights as scaling coefficients	Linear average with nominally equal weights	Linear average with effective equal weights as scaling coefficients	Linear average with nominally equal weights
Geographical focus	World (~141 countries)	EU Member States; EFTA countries; also selected OECD and BRICS countries (with a limited set of indicators)	EU Member States; EFTA countries; also benchmarked against selected OECD countries	35 Selected countries from Europe, OECD, BRICS, to benchmark Germany

Note: a) ‘Yes’ refers to major revision since first publication; b) The ‘competitiveness’ component is in fact an aggregate of two variables
Source: author’s summary based on multiple editions of the respective reports; and Hollanders and Janz (2013)

3 Methodology for the assessment of quality and validity of composite indices

In essence, this paper tries to answer the following two questions:

(1) How valid measures are the most commonly used composite indicators of innovation?
(2) How coherent are the most commonly used composite indicators of innovation statistically?

In an overview in the following sub-section, we discuss the key elements of what we consider as the most important assessment frameworks for indicators and composite indicators. Given the interaction between definition, measurement and (policy) use, it may be justified to look at multiple elements within a framework. We also aim to understand whether it is possible to establish a hierarchy, to see whether some of the elements are more important than others.

3.1 Key elements of assessment frameworks

The *OECD-JRC '10-steps' framework* (OECD-JRC, 2008) looks at composite indicators from the developers' perspective. While the development of a composite index is acknowledged to be iterative in nature, this "user's guide" tries to disentangle the key steps. From an evaluation perspective, the steps offer a useful entry point to assess the methodological soundness of an index, and the guide has in fact informed the statistical audits of composite indices (see i.e. Saisana and Saltelli, 2010, 2012).

Table 2 The OECD-JRC '10-steps' framework

Step 1. Developing a theoretical framework	Step 6. Weighting and aggregation
Step 2. Selecting variables	Step 7. Robustness and sensitivity tests
Step 3. Imputation of missing data	Step 8. Back to the details (de-constructing composites)
Step 4. Multivariate analysis	Step 9. Association with other variables
Step 5. Normalisation of data	Step 10. Presentation and dissemination

Source: OECD-JRC (2008)

In terms of the relative importance of the various steps, there is little guidance in the 10-step framework, apart from an emphasis on the importance of all steps as well as back-and-forth linkages in the development process. However, assessment is an inherent part of the framework under 'step 7'. Robustness and sensitivity analyses offer rich toolboxes to revealing the impact of uncertainty in modelling choices (selection of indicators, weighting, aggregation, etc.) on country rankings (Saisana et al, 2005, 2011). Considering the policy use of indicators, Saltelli et al (2013) highlight the importance of auditing beyond a technical process, but also considering the policy context in which it was developed.

Gisselquist (2014) proposed ten questions to guide both the development and evaluation of composite indicators. A rare feature of the framework is the distinction of the questions into six "critical" and four "less critical" ones (Table 3). For instance, while the definition of the concept, the operational definitions of the indicators to capture the concept are crucial, the ability to capture the full complexity of the concept is less crucial. Data reliability, validity and coverage is critical, as is the transparency and replicability of the computations or the discussion of sensitivity and robustness issues. However, the correctness of weighting is considered less critical, for the reason that theory gives little guidance on the relative importance of the various dimensions. While the article focuses on governance indicators, there no reason not to consider the framework applicable for other composite indices, such as innovation.

Table 3 The Gisselquist (2014) framework

Critical Questions:
1. What precisely does it aim to measure?
2. Content validity: does the operational definition capture the concept?
3. How good (reliable, valid and complete) are the data used?
4. Is the measure (including all of its sub-components) transparent and replicable?
5. How sensitive and robust is the measure to different data and design choices?
6. Does the measure allow the analyst to address key questions of interest?
Less Critical Questions:
7. Does the measure fully capture [the concept of interest] in all its complexity? [descriptive complexity]
8. Does the measure behave as theory predicts? [theoretical fit]
9. How precise are index values and are confidence intervals specified? [precision of estimates]
10. Is the weighting 'correct'?

Source: Gisselquist (2004)

The *United Nations Statistics Division's Generic National Quality Assurance Framework (NQAF)* provides quality assurance guidelines with regards to statistical output.[6] Despite their more general focus, the guidelines on managing statistical processes and outputs are applicable to any indicators, composites included. This framework covers not only a number of technical details, but importantly, highlights trade-offs (such as cost-effectiveness, or the burden of surveys on respondents). It also duly considers the multi-stakeholder context in which statistical products such as indicators are developed and used – for instance, *relevance*, is understood as the ability to reflect the degree to which the information meets the current and/or potential or emerging needs or requirements of clients, users, stakeholders, or the audience. It recognizes that relevance is subjective and depends upon the varying needs of users, which may be conflicting and need to be balanced. Regarding the quantification process, it considers *accuracy* in the technical sense (assessment of systematic and random errors) and *reliability* in the conceptual sense (consistency over time, and the ability to "measure the reality that they are designed to represent".

Table 4 Selected guidelines from The UNSD's National Quality Assurance Framework

managing statistical processes	**managing statistical outputs**
Assuring methodological soundness	Assuring relevance
Assuring cost-effectiveness	Assuring accuracy and reliability
Assuring soundness of implementation	Assuring timeliness and punctuality
Managing the respondent burden	Assuring accessibility and clarity
	Assuring coherence and comparability
	Managing metadata

Source: United Nations Statistics Division (see details in footnote 6)

In sum, at the heart of assessment, it appears from the three frameworks, is the translation mechanism between concepts and indicators (the quantification), making qualitative and quantitative assessments indistinguishable. Evidently, careful assessments should cover the *entire indicator development* process. However, there may be reason to find a tool for an "instantaneous" assessment of an

[6] See Guidelines for the Template for a Generic National Quality Assurance Framework (8 Feb 2012), [URL: http://unstats.un.org/unsd/dnss/docs-nqaf/GUIDELINES%208%20Feb%202012.pdf; retrieved: Jun 2016]

indicator, and our aim here is to identify whether it is sufficient to focus on certain elements to highlight key assessment of the quality and validity of an indicator. In other words, whether by looking at the *product*, rather than the entire development process, and assess its qualities. The ability to identify the qualities of a product is all the more important, given that indicators are developed in a changing environment – as elaborated in the following subsection.

3.2 Assessment of validity: the case for an evolutionary framework

Indicator development is a dynamic process, in which not only the conceptual framework influences the outcomes, but the results also reveal new features of the concept that was aimed to be measured in the first place. If this is valid for "single" indicators (Barré, 2001, 2010), it is certainly valid for composite indicators (beyond the iterative process of composite indicator development). "Goodhart's law" in practice also implies that due to strategic behavior of the assessed, indicators that become policy targets cease to be good indicators, and the search for replacement is a continuous process (Freeman and Soete, 2009).

These dynamics influence the validity of an index, which relates to the concept. There is the possibility that the concept remains stationary, but the indicator is "refined" along the development process. However, as suggested above, the concept may change as well. In fact, there is good reason to expect that the concept changes due to the co-evolution of a number of related elements.

Firstly, because business cycles in the world economy and developments such as the process of accelerated globalization (1990s) or global financial crises (2007/08) and jobless recovery define not only the real economic context in which innovation is carried out by firms, but also influence the policy context in which innovation policy interacts with other domains. For instance, the focus of economic policy that was concerned with productivity increase in order to increase wages and gaining competitive advantage in the 1990s and early 2000s shifted to "jobs and growth" in the aftermath of the global financial crises. While innovation policy was concerned in the 1990s with technological innovation and their diffusion to the rest of the economy, with increasing R&D spending, with human resources for tech-intensive activities, after 2007/08 it faced the new challenge to help "innovate out of the crisis". It is not surprising that policy focus changes, as stakeholders' values and interests are not fixed, but respond to broader socio-economic developments.

Second, there are parallel developments in the realm of statistics and indicators research. Although composite indicators have been around, the 2000s saw a rapid growth not only in their use by various organizations, but alongside, there were significant improvements in the development methodology (OECD-JRC, 2008), as well as in statistical methods of uncertainty and sensitivity analyses (Saltelli et al, 2008; Saisana et al, 2011; Paruolo et al, 2013).

Thirdly, the discipline of innovation studies has co-evolved with innovation policies (Mytelka and Smith, 2002). Considering the past two decades, the defining trend was rapid diffusion and dominance of the national innovation systems approach with its broad political appeal, although it appears to have peaked around 2011-13,[7] potentially giving room for a revival of linear models explaining the efficiency of innovation, or its impact on growth and jobs. While the latest changes in the discourse are yet to be seen, Gault (2010) highlights that the 'language' of discourse on innovation is also not static, and evolves through discussions.

While these co-evolutionary forces not necessarily make it easier to assess the validity of an indicator, these need to be taken into consideration. There are two direct implications. First, rather than absolute,

[7] The peak is apparent in bibliometric data (abstract, title or keywords) for published scientific articles; both in google scholar (decline after 2013) and scopus (flat since 2011).

validity is seen as specific to a context (geographical, historical, related to a certain scientific or technological paradigm), and reflects an agreement between developers and users, which may be bound to change (Boulanger, 2007; Turnhout et al, 2007). Second, it remains an empirical challenge to explore the actual co-evolution, for instance, to identify discontinuities in the various domains and how these relate to changes in (composite) indicator frameworks. A starting point is to understand the discontinuities in the indicator frameworks.

In what follows, we aim to provide a comparative assessment of composite indicators by measuring the variance between the composite index and the various underlying indicators. We argue that this is the key point in the assessment, as composites are effectively "defined" by the correlation structure between components. Our aim is to understand what indicators explain most or the least of the variance in composite scores, as – assuming that the selection of indicators was carried out carefully reflecting policy needs – this will reveal the effective importance of certain components. This is, in essence, the ultimate judgement of the operationalization process (selecting data and indicators; judging their accuracy and reliability; overcoming shortcomings in data; doing it in a transparent and reproducible way). In short, the assessment boils down to the question whether the indicator measures what the developers intend to measure. By doing so, we focus on internal validity, and not external validity of an index – that is, relating to an external comparator measure. We argue that while external validation is informative, it rests on the misleading assumption that objective measures exist (see discussion in section 2).

It is important to highlight that our analysis is a complementary tool to, and not a replacement for, a careful auditing of the development process. Rather, it aims to provide an instantaneous assessment for developers and users to the capacities and limitations of the index. This is an important feature to make a comparative assessment of multiple indicators aiming to measure the same phenomenon – as in the case of innovation, where multiple editions of multiple indicators exist. Assessing the evolution of importance in a certain component over time also offers an insight into the changes of its importance in multiple editions, in order to help uncover the co-evolutionary process discussed above. Correlation analysis is well in place in this regard, as we are less interested in exploring any direction of causality.

3.3 Selecting the cases for analysis

We select country-level composite indices of innovation, which fulfill a number of expectations. First, "innovation" should be part of the title of the index. Second, unless an indicator is supported by an elaborate description of methodology that describes in English at least the variables selected, the method of normalization, aggregation, we cannot make any assessment. Third, we expect the indicators to be published over multiple years, with the most recent edition (regardless of the actual year of most recent data, which differs in many cases) being not older than 2015. Multiple years also allows us to take into consideration the evolution of the indicator, and the nature of changes over the past years. Fourth, we expect the indicator to compare a larger set of countries.

Considering the overview of innovation scoreboards and surveys, this left us with four indices to study: the WIPO-INSEAD-Cornell Global Innovation Index, the European Commission's Summary Innovation Index and the Innovation Output Index, and the Innovationsindikator (see Table 1). Given the selection criteria, we left out regional indices as well as the Global Competitiveness Index from this study. This leaves us with some diversity of the indices: in terms of country coverage (one focuses on the world, two mostly on European countries, while the third focuses on Germany and benchmark countries); but also in terms of structure and number of indicators.

4 Comparison of most widely used composite indicators of innovation

In this empirical part, we focus the discussion on exposing the main purpose of an indicator and the assessment, by analyzing the correlation between composite scores and component indicators, of the actual importance of certain components for the four selected indices.

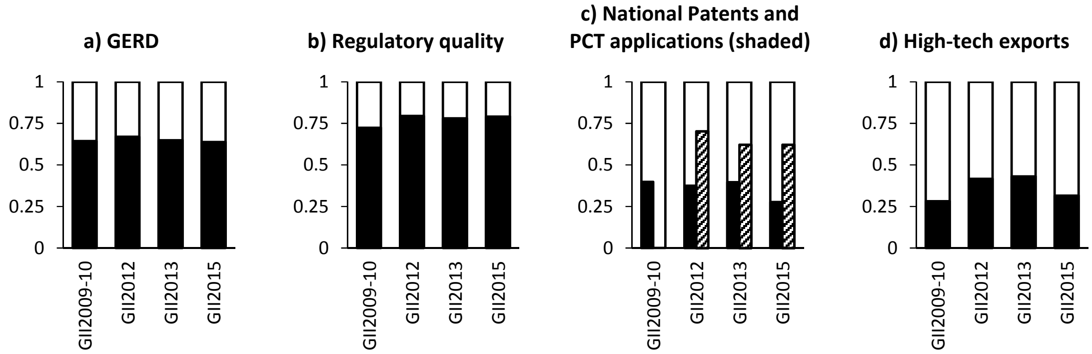
4.1 The Global Innovation Index

The Global Innovation Index, co-published by Cornell University, INSEAD, and the World Intellectual Property Organization (WIPO), intends to support policies "to promote long-term output growth, improved productivity, and job growth". The annual reports, alongside the composite index, have aimed for a global coverage: it has published around 80 indicators for some 141 countries. The GII composite indices are constructed in a multi-step aggregation. Individual indicators (selected on the basis of literature review, expert opinion, country coverage, and timeliness) are aggregated (using a weighted average) to obtain sub-pillar scores. In turn, a weighted average of these scores are the pillar scores. Five input pillars capture elements of the national economy that are understood as enablers of innovative activities: (1) Institutions, (2) Human capital and research, (3) Infrastructure, (4) Market sophistication, and (5) Business sophistication, while two output pillars capture what are considered as outputs of innovation: (6) Knowledge and technology outputs and (7) Creative outputs. The simple average of the input pillars are the input and output sub-indices; finally, the GII is the simple average of the Input Sub-Index and Output Sub-Index (see GII, 2015). The construction of the indicator takes into consideration the correlation structure of the underlying data, and applies weights as scaling coefficients (see Paruolo et al, 2013) for the sub-indices, since the 2012 edition. This introduces a robustness which is confirmed by the sensitivity analyses (see Saisana and Filippas, 2012)[8]. This builds on the assumption that sub-pillars are of key interest to for users of the index, rather than individual indicators. To balance this choice, detailed country scoreboards as well as composites of the various level of aggregation are published in the GII reports.

To understand, at the level of individual indicators, what aspects of the innovation system are the main driving factors behind GII scores, we looked at the proportion of variance explained by selected indicators. Curiously, we find that indicators of the Institutions pillar – regulatory quality (panel b in Figure 1) or the Rule of Law index (not shown here) are the strongest predictors of country performance, while overall R&D investment and PCT patent applications (panel a and shaded bars in panel c in Figure 1) are somewhat weaker, and patent applications in national patent offices or high-tech exports (panel c and d in Figure 1) show a variance of less than 0.5. Interestingly, R&D expenditure financed by business (not shown here) is typically accounting for about half of the variance in GII scores. As for the validity of the index, all this suggest that – notwithstanding the heterogeneity of the component indices – the GII boils down to the message that a government needs to strengthen first and foremost the institutional framework conditions (i.e., the rule of law) in order to improve its position in the ranking.

[8] PCA results confirm the presence of a single latent dimension in each of the seven pillars (one component with an eigenvalue greater than 1.0) that captures between 60% (pillar 4: Market sophistication) up to 84% (pillar 1: Institutions) of the total variance in the three underlying sub-pillars.

Figure 1 Variance explained by selected component indicators of multiple editions of the Global Innovation Index



Source: author's calculation based on multiple editions of the GII.
Note: The length of the bars indicate the proportion of variance explained by a certain component, ranging from none (0) to 100% (1).

4.2 The Summary Innovation Index

The Innovation Union Scoreboard and its predecessors, the European Innovation Scoreboard have been published regularly pursuant to the goals flagged out in the Lisbon European Council of 2000 to make the European Union the most competitive and dynamic knowledge-based economy in the world. The scoreboards help to monitor how Member States and a few selected international benchmark countries perform in a wide range of indicators related to innovation activities. It offers a comprehensive picture of the state of national innovation systems from different aspects, and thus reflects the understanding of innovation activities as a multi-dimensional phenomenon. The scoreboard has undergone a multiple major refinements since its introduction in 2001, and thus carries the outcomes of compromises between stakeholder needs and data availability (the history of changes can be traced in Sajeva et al, 2006; Hollanders and van Cruysen, 2008, and the various technical annexes). It is important to add though that the outcomes of the index also had an impact on the improvement of data coverage, including those coming from Community Innovation Surveys.

The SII aims to summarize the indicators of the Innovation Union / European Innovation Scoreboards in a single composite. In the most recent editions,[9] 25 indicators (min-max normalized, pooled data from multiple years, /8 in the 2015 edition for most indicators/) for 35 countries[10] are aggregated applying the arithmetic average with equal weights. Additionally, separate composite indicators are presented for each dimension. The technical annex to the IUS 2013 provides a description on the construction of this dataset (selection of indicators, definitions of numerators and denominators), outlier treatment, imputation, transformation, normalization and aggregation of data.

What makes the SII unique is the use of direct measures of innovation, alongside the indirect measures, as it contains a number of Community Innovation Survey variables. We notice however, that when it comes to aggregating the data in the SII composite, information contained in these components have relatively less influence on the scores. This is due to the correlation structure of the

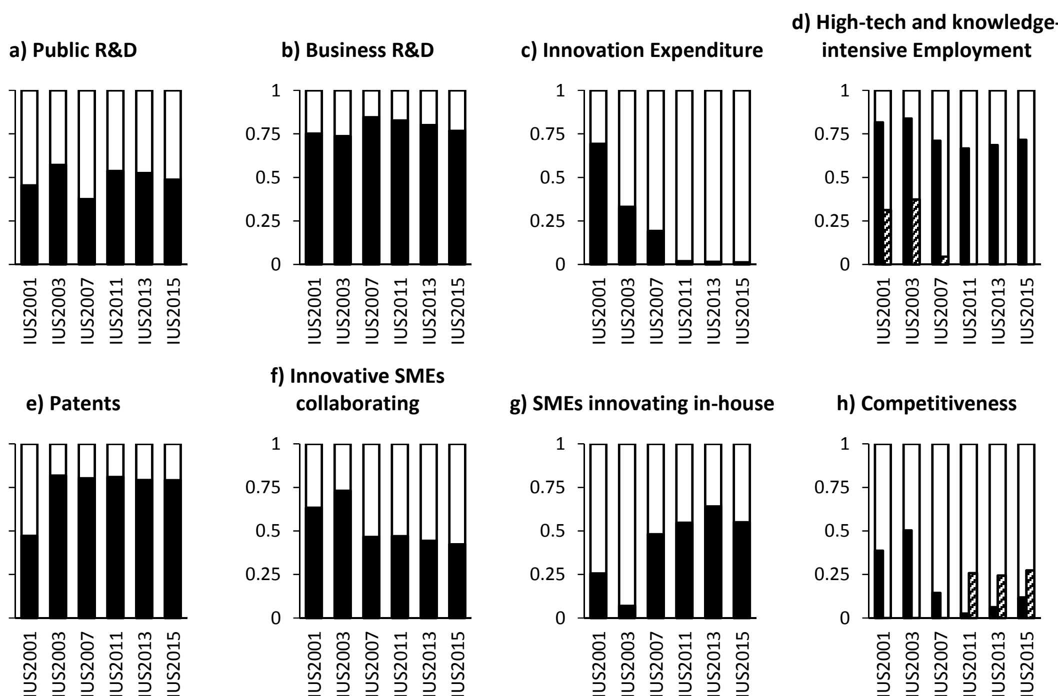
[9] The 2016 edition of the Index was published at the time when this report was being finalized, thus our focus is the 2015 edition.
[10] The analysis is focusing on the main benchmark indicator covering European countries (34 countries: the 28 EU member states, the EFTA countries, s well as TR, RS, MK and the EU itself as the 35th). The Scoreboard presents a subset of these indicators that are available for non-European (selected OECD and BRICS countries) to offer a global benchmark. The present analysis focuses on European countries only.

indicators. A correlation analysis of the 25 indicators showed a high collinearity between three indicator pairs (PCT patent applications and those in grand societal challenges; PCT patent applications and business R&D). Little or no correlation with the rest of the indicators for four of the indicators (youth with upper secondary level education; Non-R&D innovation expenditure; medium and high-tech product exports; and sales of new to market and new to firm innovations). These observations imply that aggregating such indicators results in a considerable loss of information in the summary index that was present in the underlying indicators. This is confirmed by the variance in the summary index explained by selected indicators (for the 2015 index, see first bar from the right in the panels of Figure 2). While business R&D expenditure (panel b) or PCT patent applications (panel e) explain most of the variance, Non-R&D innovation expenditure, for instance, explains hardly any.

Evidently, R&D- and patent-intensive, high-tech and knowledge-intensive industrial activities make countries score high on the Summary Innovation Index (Figure 2). The self-reported innovative activities (both innovating in-house as well as collaborating) of SMEs have a relatively lower, although non-negligible influence compared to indirect measures. It is interesting to note that the strongest 'predictors' of summary innovation scores – BERD, patents – have maintained their strength since at least the 2003 edition, despite the changes in methodology.

Alongside the SII, composite scores are also reported and discussed in the IUS/EIS reports for the eight dimensions. These unweighted averages have the potential to reveal differences in country performances in a more focused and conceptually clearer way. The indicators found to be "outliers" in terms of the correlation patterns are performing in a different way in these dimension composites. With the curious exception of the 'sales of new to market and new to firm innovations' variable, each of them influence at least modestly the overall composite scores.

Figure 2 Variance explained by selected component indicators of multiple editions of the Summary Innovation Index



Source: author's calculation based on the 2001, 2003 and 2007 editions of the European Innovation Scoreboard and the 2011, 2013 and 2015 editions of the Innovation Union Scoreboard.
Note: The length of the bars indicate the proportion of variance explained by a certain component, ranging from

none (0) to 100% (1). Shaded bars in panel d) refer to manufacturing, while in panel h) to services. Some of the indicators have been replaced or discontinued in given editions; we tried to use indicators that were available for most of the years, but this could not be carried out for all years for the following indicators: c) innovation expenditure: 2003: manufacturing only; 2011 and 2013: non-R&D innovation expenditures. e) the 2001 and 2003 editions use EPO high-tech patents per population, the 2007 edition EPO patents per population; the 2011 and 13 editions are represented by PCT patent applications.; f) and g) in the 2003 edition, manufacturing sector was considered only; h) 2001 and 2003 editions capture high-tech value added, while 2007 and later editions capture exports, also the formula has changed from contribution to the trade balance to share in total exports.

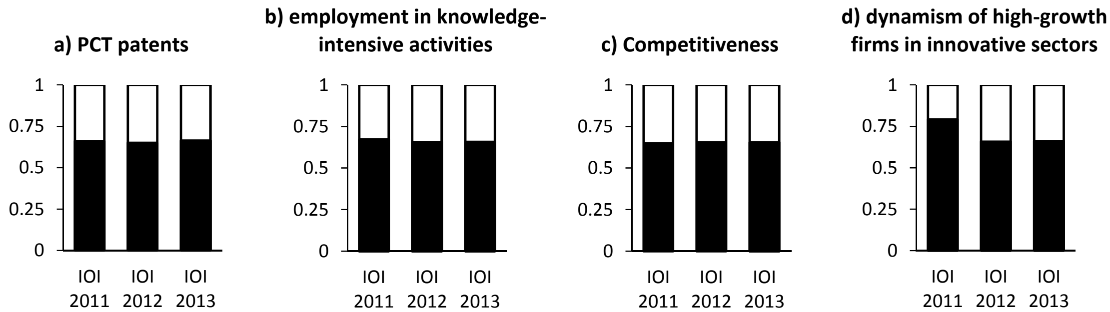
4.3 The Innovation Output Indicator

The Innovation Output Indicator (IOI) was introduced in 2013 by the European Commission with the aim to benchmark the outcomes of national innovation policies and to monitor the EU's performance against its main trading partners. Its main purpose has been to complement the R&D intensity indicator (a policy target criticized for being input-oriented) and focus on innovation output. It aims to support policy-makers by identifying bottlenecks preventing innovators from translating ideas into successful goods and services, with the overall goal to "provide better jobs and make Europe more competitive". IOI scores have been published in various reports of the European Commission (European Commission, 2013, 2014, 2016). It is a composite indicator of four components that cover technological innovation (measured by PCT patent applications), skills proxied by employment share in knowledge-intensive activities, the competitiveness of knowledge-intensive goods and services export, and employment dynamism of fast-growing enterprises in innovative sectors. The components are in fact selected variables of the European Innovation Scoreboard, however, in a different grouping.[11] The four components are aggregated using a weighted arithmetic average, where the weights (understood as scaling coefficients) are iterated in each edition of the indicator so that each component has an effectively equal contribution to the overall scores. Since the first publication of IOI2011 scores in 2013, refinements were introduced to accommodate changes in statistical definitions and slightly expand country coverage (Vertesy and Tarantola, 2014; Vertesy and Deiss, 2016).

All of the measures are indirect measures of innovation, and refer to a single latent dimension that explains 66% of the total variance observed in the data. An analysis of the correlation structure of the components shows that the real tradeoff is between the two trade variables within the competitiveness component: goods and services trade show a coefficient is 0.2, while all the other components have a modest to strong positive correlation – suggesting that the various goals, proxied by the indicators, do not fully overlap. By construction, the four components contribute equally to the IOI scores[12] (Figure 3), delivering the message that none of the aspects are more important than any of the other when it comes to setting policy targets. The IOI offers a rare case in which policy goals are made explicit in terms of outcome variables, it remains nevertheless an empirical challenge to assess to what extend do innovative activities influence performance in the various measures, and how much other factors, such as taxation, influence them.

[11] PCT patents are part of the firm activities group in the Scoreboard. Note also that the competitiveness component is an average of the goods and services trade variables.
[12] The higher contribution of the dynamism (d) scores in the IOI2011 are due to the imputation method.

Figure 3 Variance explained by selected component indicators of multiple editions of the Innovation Output Indicator



Source: author's calculation based on multiple editions of the IOI.
Note: The length of the bars indicate the proportion of variance explained by a certain component, ranging from none (0) to 100% (1). The competitiveness component (c) is an average of two indicators measuring the share of medium-high goods and knowledge-intensive services in the respective exports.

4.4 The 'Innovationsindikator' (II)

This indicator aims to benchmark the performance of the German innovation system against a selective set of European and global competitors (35 countries). The innovation system concept is reflected in the focus on five sub-systems (industry, education, state, public research and society), and 38 variables are aggregated (with overlap) into sub-system and overall scores, using arithmetic average with (nominally) equal weights.

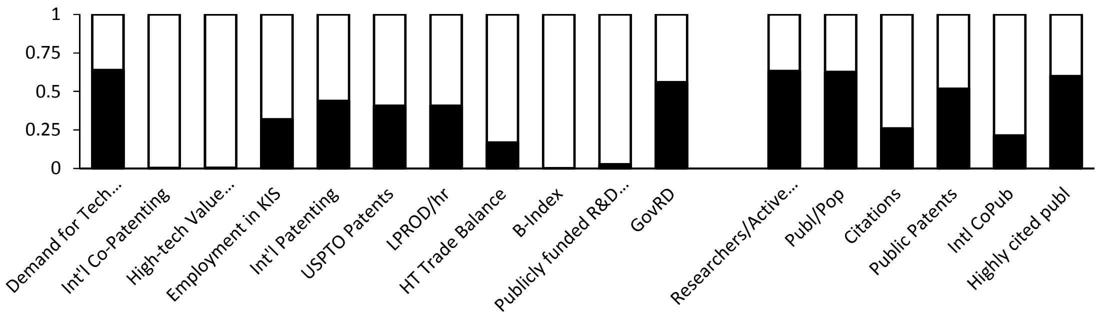
The report explicitly mentions that growth and competitiveness are the ultimate aims of innovation system performance. Hollanders and Janz (2013) pointed out that the selection of variables is closer to competitiveness than innovation (as none of the variables reflected information collected in innovation surveys). They found confirmation in the fact that country ranks of the Innovation Indicator correlated the strongest with those of the Global Competitiveness Indicator (of the World Economic Forum).

The II methodology puts a lot of emphasis on its careful conceptual reasoning and corresponding methodological approach to the selection of indicators; the developers selected the 38 variables from a pool of 100 using econometric regression models. While more transparency on this selection process would be welcome, it is interesting to find that some of the components are opinion-based indicators (although not from CIS data).

While the final II scores offer a balanced summary of the five subsystem scores (with above 0.94 Pearson coefficients for the industry and science dimensions, 0.89 for Education and State, and 0.78 for society), the same balanced structure cannot be seen at the level of individual components. In fact, the variables explaining relatively the highest share of variance in the final II scores are the survey response for demand for technological products, public R&D, researchers per active population or publications per population (Figure 4). Surprisingly, indicators such as the OECD B-index for tax-based funding of business R&D, publicly funded R&D in enterprises or high-tech value added have virtually no contribution to the final scores. Labor productivity (value added per hour worked) explains only about 41%, of the variance in II scores – an indicator that may be instrumental for reaching the main goal of innovation according to the framework, growth and competitiveness.[13]

[13] For the II, we note that we could not compute multiple years as done for the other cases, because only partial data was available online.

Figure 4 Variance explained by selected component indicators of the Innovationsindikator



Source: author's calculation based on the 2015 edition of the II.

5 Conclusions and policy recommendations

Assessing the main drivers of variance in the selected composite indices highlights interesting commonalities between the four composite indicators of innovation. Firstly, that overall they testify the diffusion of the systems of innovation approach. The complex, systemic interactions between the various components of inventions, research, technical change, learning and innovation (Soete et al, 2010; Freeman, 1987; Lundvall, 1992; Nelson, 1993; Edquist, 1997) warrants composite, rather than stand-alone indicators to support evidence-based policies. However, while hardly any of the indicators addresses explicitly choices between the different system components – on the contrary, by appearance, they argue that all elements are "weighted equally" –, a very clear choice emerges from looking at the main drivers of the scores. It can be argued that considering the multi-stakeholder involvement in all the indicators, the results reflect a shared intuition, at least when it comes to country rankings. Without a look at country rankings, and just by looking at relative importance of components, we found that this intuition is in the case of the SII, the IOI and II are centered on R&D-patenting- and knowledge-intensive activities. In the case of the GII, this intuition is about the importance of framework conditions for competitiveness and the rule of law.

The focus of our study was internal validity, but we can use the signals observed to assess the validity of indices in a broader context. Considering that composite indices are strong advocacy tools (Saltelli, 2007) and increase the public awareness of innovation, evidently, these indices appear to promote a technology-based view. This is, in many ways, at odds with the policy discourse of the 2000s as non-R&D-based aspects or entrepreneurship receives less attention, in the SII in particular. It is interesting to see that despite the multiple revisions, the introduction of new indicators, and even changes in policy discourse (the emergence of the jobless growth challenge) has brought little change to what the composite indicators of innovation actually capture. This is despite the recurrent expectations of policy makers to measure the success in producing a certain type of innovation for a certain strategy. In order to realize goals such as achieving competitiveness of selected segments of the economy, foster transition to more environmentally friendly growth trajectory, or create jobs and growth to speed up recovery from the financial crisis needs more selectivity in the definition of policy as well as in identifying monitoring tools.

We find that as innovation has become prominent on today's policy agenda in advanced economies, the advocacy function of innovation indices seems to have lost some relevance in comparison with their analytical function – that is, to assess the innovation system from different aspects based on more fine-grained policy needs. Paradoxically, the most commonly used composite indicators of innovation

(GII, SII) aggregate a large set of indicators that correspond to broad definitions of innovation and framework conditions. This comes at a cost of their quality and validity, and may lead to an increased use of scoreboards and dashboards, which often overwhelms policy makers with excess information. There appears to be a void for composite indices summarizing a more targeted, more limited set of indicators, which, if carefully constructed, efficiently reduce dimensionality and complexity. The aggregation of a more limited set of indicators (i.e. sub-indices) allows for a better articulation of normative choices involved in the selection of indicators. It also offers greater transparency concerning quality and validity assessment; two interlinked tools that are necessary ‘accompanying tools’ in order to strengthen to support the use of composite indices.

Our analyses revealed the presence of component indicators that explain little or none of the variance in overall composite scores. Such components are only seemingly part of the framework, misleading readers. We found differences between the various composite indices in terms of how they made explicit the presence of multiple, rather than one single latent dimension. These findings suggest that greater transparency may be necessary in terms of showing variance as discussed in section 4. This is crucial, as there may be a need to accommodate selectiveness: information relevant for only some of the stakeholders.

The diffusion of composite indicators measuring in effect a certain feature of innovation is also troublesome as it may further raise the expectation that innovation is a panacea. At best it nurtures false hopes; at worst policy makers may overlook unintended consequences of certain innovative activities they promote – if, indeed, they are ready to make choices for certain strategies (Mazzucato, 2016). We argue that it the root of the finding may very well be in the readiness of governments to set policy targets that address tradeoffs within innovation policy. Development and assessment of innovation indicators may very well support this process.

Acknowledgements: The author benefited hugely from interactions with colleagues Competence Centre on Composite Indicators and Scoreboards, Michaela Saisana, as well as Andrea Saltelli, Stefano Tarantola and Sjoerd Hardeman.

References

Annoni, P. and Dijsktra, L. 2013. *EU regional competitiveness index.* Luxembourg: Publ. Office of the European Union.

Barré, R., 2001. The Agora model of innovation systems: S&T indicators for a democratic knowledge society. *Research Evaluation*, 10(1), pp.13-18.

Barré, R. 2010. “Towards socially robust S&T indicators: indicators as debatable devices, enabling collective learning”. *Research Evaluation*, 19(3), pp.227-231.

Boulanger, P-M. 2007. “Political uses of social indicators: overview and application to sustainable development indicators”. *Int. J. Sustainable Development*, Vol. 10, p. 14-32.

Coad, A. and Rao, R., 2008. Innovation and firm growth in high-tech sectors: A quantile regression approach. *Research policy*, *37*(4), pp.633-648.

Davis, K.E., Kingsbury, B. and Merry, S.E. 2012. *Governance by indicators: global power through classification and rankings*. Oxford: OUP.

Edquist, C. 1997. *Systems of innovation: technologies, institutions and organizations*. London; Washington, D.C.: Pinter.

European Commission 2013. “Measuring innovation output: towards a new indicator” Communication COM (2013) 624final.

European Commission 2014. *Research and Innovation performance in the EU 2014*. (EUR 26334 EN) Luxembourg: Publications Office of the European Union.

European Commission 2016. *Science, research and innovation performance of the EU. A contribution to the open innovation, open science, open to the world agenda: 2016*. Luxembourg: Publications Office of the European Union.

Fagerberg, J. and Srholec, M. 2008. National innovation systems, capabilities and economic development. *Research policy*, *37*(9), pp.1417-1435.

Freeman, C. 1987. *Technology, policy, and economic performance: lessons from Japan*. London; New York: Pinter Publishers

Freeman, C., 1995. The ‘National System of Innovation’in historical perspective. *Cambridge Journal of economics*, 19(1), pp.5-24.

Freeman, C. and Soete, L. 2009. Developing science, technology and innovation indicators: What we can learn from the past. *Research policy*, 38(4), 583-589.

Gault, F. 2010. *Innovation strategies for a global economy*. Cheltenham: E. Elgar.

Gault, F. 2013. “Innovation indicators and measurement: an overview”, in: Gault, F. (ed.), *Handbook of Innovation Indicators and Measurement*, Edward Elgar, Cheltenham, UK and Northampton, MA, USA, pp. 3-40

Godin, B. 2004, “The obsession for competitiveness and its impact on statistics: the construction of high-technology indicators” *Research Policy* 33, Issue 8, October 2004, pp. 1217–1229

Grupp, H. and Mogee, M. E. 2004. “Indicators for national science and technology policy: how robust are composite indicators?” *Research Policy*, 33, Issue 9, November 2004, pp. 1373–1384

Grupp H. and T. Schubert. 2010002E “Review and new evidence on composite innovation indicators for evaluating national performance” *Research Policy* 39 (2010) 67–78

Hardeman, S. and Vertesy, D. 2016. “Evidence-based policy learning: the case of the Research Excellence Indicator” Paper presented at the *Workshop on Excellence Policies in Science*, Leiden, 2-3 Jun 2016.

Hekkert, M. P., Suurs, R. A., Negro, S. O., Kuhlmann, S., & Smits, R. E. 2007. Functions of innovation systems: A new approach for analysing technological change. *Technological forecasting and social change*, *74*(4), 413-432.

Hollanders, H. and Janz, N. 2013. “Scoreboards and indicator reports”, in Gault, F. (ed.), *Handbook of Innovation Indicators and Measurement*, Edward Elgar, Cheltenham, UK and Northampton, MA, USA, pp.279–97.

Hollanders, H., and van Cruysen, A. 2008. “Rethinking the European innovation scoreboard: A new methodology for 2008-2010.” *Inno-Metrics Publication*, European Commission.

JRC-OECD. 2008. *Handbook on constructing composite indicators: methodology and user guide*. Paris: OECD.

Lundvall, B-Å. 1992. *National Systems of Innovation: Towards a Theory of Innovation and Interactive Learning*. London: Pinter.

Lundvall, B-Å. and Borrás. 2005. “Science, technology, and innovation policy”. In: Fagerberg, J., Mowery, D. C., and Nelson, R. R. (eds.) *The Oxford handbook of innovation*. Oxford: OUP.

Lundvall, B-Å., 2007. National innovation systems—analytical concept and development tool. *Industry and innovation*, *14*(1), pp.95-119.

Mairesse, J. and Mohnen, P. Using innovation surveys for econometric analysis, in: Hall, B. H. and Rosenberg, N. (eds) *Handbook of the Economics of Innovation*, Amsterdam: Elsevier, 1130-1155.

Mazzucato, M., 2016. From market fixing to market-creating: a new framework for innovation policy. *Industry and Innovation*, *23*(2), pp.140-156.

Mytelka, L.K. and Smith, K., 2002. Policy learning and innovation theory: an interactive and co-evolving process. *Research policy*, *31*(8), pp.1467-1479.

Nelson, R.R. (ed.) 1993. *National innovation systems: a comparative analysis*. Oxford: OUP.

OECD and Eurostat. 2005. *Oslo Manual – Guidelines for Collecting and Interpreting Innovation Data*, Paris: OECD.

Paruolo, P., Saisana M. and Saltelli A. 2013. “Ratings and rankings: voodoo or science?” *Journal of the Royal Statistical Society: Series A*. 176(3), 609–634

Ravallion, M. 2012. “Mashup Indices of Development”. *The World Bank Research Observer*, 27(1), 1–32

Saisana M. and Saltelli A. 2010. *The Multidimensional Poverty Assessment Tool (MPAT): Robustness issues and Critical assessment*. Luxembourg: Publ. Office of the European Union.

Saisana M. and Saltelli A. 2012. *Corruption Perceptions Index 2012 - Statistical Assessment*. Luxembourg: Publ. Office of the European Union.

Saisana M., D’Hombres B. and Saltelli A. 2011. “Rickety Numbers: Volatility of university rankings and policy implications.” *Research Policy* 40, 165–177

Saisana M., Saltelli A. and Tarantola S. 2005. “Uncertainty and sensitivity analysis techniques as tools for the analysis and validation of composite indicators”, *Journal of the Royal Statistical Society A* 168(2), 307-323.

Sajeva, M., Gatelli, D., Tarantola, S. and Hollanders, H. 2006, “Methodology Report on the European Innovation Scoreboard 2005, European Commission, EUR 21883 EN

Saltelli, A. 2007. “Composite indicators between analysis and advocacy”. *Social Indicators Research*, 81(1), 65-77.

Saltelli, A., Ratto, M., Andres, T., Campolongo, F., Cariboni, J., Gatelli, D., Saisana, M. and Tarantola, S., 2008. *Global sensitivity analysis: the primer*. John Wiley & Sons.

Saltelli, A., Guimarães Pereira, Â., Van der Sluijs, J. P., & Funtowicz, S. 2013. “What do I make of your latinorum? Sensitivity auditing of mathematical modelling”. *International Journal of Foresight and Innovation Policy*, *9*(2-3-4), 213-234.

Schibany, A. and Streicher, G., 2008. The European innovation scoreboard: Drowning by numbers?. *Science and Public Policy*, *35*(10), pp.717-732.

Smith, K. 2005. “Measuring innovation”. In: Fagerberg, J., Mowery, D. C., and Nelson, R. R. (eds.) *The Oxford handbook of innovation*. Oxford: OUP, pp. 148-177.

Soete, L., Verspagen, B., ter Weel, B., 2010. Chapter 27 - Systems of Innovation, in: Hall, B.H., Rosenberg, N. (Eds.), *Handbook of the Economics of Innovation, Vol.2, Handbook of the Economics of Innovation*. North-Holland, pp. 1159–1180.

Turnhout, E., Hisschemöller, M., & Eijsackers, H. 2007. “Ecological indicators: between the two fires of science and policy”. *Ecological indicators*, 7(2007), 215-228.

Vertesy, D. and Tarantola, S. 2014. *The Innovation Output Indicator 2014. Methodology Report*. JRC Technical Reports, EUR 26936, Luxembourg: Publications Office of the European Union

Vertesy, D. and Deiss, R. 2016. *The Innovation Output Indicator 2014. Methodology Update*. JRC Technical Reports, EUR 27880, Luxembourg: Publications Office of the European Union.

APPENDIX

Table 5 Structure of the Summary Innovation Index 2015

	"Type"	"Dimension"	Indicator
SUMMARY INNOVATION INDEX (SII)	ENABLERS	Human resources	1.1.1 New doctorate graduates (ISCED 6) per 1000 population aged 25-34
			1.1.2 Percentage population aged 30-34 having completed tertiary education
			1.1.3 Percentage youth aged 20-24 having attained at least upper secondary level education
		Open, excellent and attractive research systems	1.2.1 International scientific co-publications per million population
			1.2.2 Scientific publications among the top 10% most cited publications worldwide as % of total scientific publications of the country
			1.2.3 Non-EU doctorate students as a % of all doctorate students
		Finance and support	1.3.1 R&D expenditure in the public sector as % of GDP
			1.3.2 Venture capital investment as % of GDP
	FIRM ACTIVITIES	Firm investments	2.1.1 R&D expenditure in the business sector as % of GDP
			2.1.2 Non-R&D innovation expenditures as % of turnover
		Linkages & entrepreneurship	2.2.1 SMEs innovating in-house as % of SMEs
			2.2.2 Innovative SMEs collaborating with others as % of SMEs
			2.2.3 Public-private co-publications per million population
		Intellectual Assets	2.3.1 PCT patent applications per billion GDP (in PPS€)
			2.3.2 PCT patent applications in societal challenges per billion GDP (in PPS€) (environment-related technologies; health)
			2.3.3 Community trademarks per billion GDP (in PPS€)
			2.3.4 Community designs per billion GDP (in PPS€)
	OUTPUTS	Innovators	3.1.1 SMEs introducing product or process innovations as % of SMEs
			3.1.2 SMEs introducing marketing/organisational innovations as % of SMEs
			3.1.3 Fast-growing firms in innovative industries
		Economic effects	3.2.1 Employment in knowledge-intensive activities (manufacturing and services) as % of total employment
			3.2.2 Medium and high-tech product exports as a % of total product exports
			3.2.3 Knowledge-intensive services exports as % total service exports
			3.2.4 Sales of new to market and new to firm innovations as % of turnover
			3.2.5 License and patent revenues from abroad as % of GDP