JRC SCIENTIFIC AND POLICY REPORTS

Composite Indicators of Research Excellence

WP3 Deliverable,
Project ERA_MONITORING

Daniel Vertesy, Stefano Tarantola

2012

Report EUR 25488 EN

European Commission
Joint Research Centre
Institute for the Protection and Security of the Citizen

Contact information
Forename Surname
Address: Joint Research Centre, Via Enrico Fermi 2749, TP 361, 21027 Ispra (VA), Italy
E-mail: stefano.tarantola@jrc.ec.europa.eu
Tel.: +39 0332 78 9928
Fax: +39 0332 78 5733

http://ipsc.jrc.ec.europa.eu/
http://www.jrc.ec.europa.eu/

JRC72592

EUR 25488 EN

ISBN 978-92-79-26260-9

ISSN 1831-9424

doi:10.2788/45492

Luxembourg: Publications Office of the European Union, 2012

Printed in Italy

EXECUTIVE SUMMARY

This report on Research Excellence is the deliverable of the third work package (WP3) of the feasibility study 'ERA MONITORING', financed by DG RTD. The objective of the work package was to explore the possibility to develop a composite indicator of research excellence in Europe, in coherence with the orientations of the EU 2020 strategy and the Innovation Union initiative.

The study built on the theoretical framework proposed by the 2011 report of the Expert Group on the Measurement of Innovation '*Indicators of Research Excellence'*, co- authored by Rémi Barré (CNAM, France), Hugo Hollanders (UNU-MERIT, The Netherlands) and Ammon Salter (Imperial College, UK), from now on identified as the Expert Group report. The report identified a scoreboard of indicators to describe research excellence in the context of a research and innovation system. The proposed indicators characterize knowledge production as well as the institutional arrangements and interactions through which research activities take place, and refer to [basic] research actors, industrial innovation actors and societal and political actors.

We considered the quality profile of 22 indicators for 41 countries and performed multivariate analyses. We concluded that 16 indicators can be used for aggregation into a composite indicator, some only after the treatment of outliers and imputation of missing values. We also found that the indicators could be computed for all EU Member States, most EFTA countries, Candidate countries and the main international competitors of the European Union (United States, Japan and China). However, given that some indicators are meaningless for non-ERA countries, we concluded that modified setups would be required for global country comparisons.

We proposed three alternative conceptual frameworks of research excellence with different underlying indicator structures, and tested their statistical coherence. In the first theoretical framework, we aimed to follow as closely as possible the Expert Group recommendation of 6 dimensions. In the second framework, we tried to consider only two dimensions (basic and applied science), also based on the Expert Group report. The third framework was derived from the data and three dimensions were identified directly from a principal component analysis.

First, we tested the 6-dimensional conceptual framework originally proposed by the expert group (Framework 1). The multivariate statistical analysis suggested that the framework only partially holds. We decided to remove some indicators from each dimension so that this better expresses a distinct aspect of research excellence. We identified four dimensions (given the no data were available on indicators of societal relevance), which consisted of a total of 13 indicators. In sum, we found that:

dimension 1a would obtain greater statistical coherence if it were narrowed down to include only five indicators (*Publications* - 1a1, *Share of highly cited publications* -

1a2, *Share of 250 top scientific Universities* - 1a4, *Share of ERC grants* - 1a5 and *Specialisation in Grand Societal Challenges publications* - 1a9, as defined in section 0);

dimension 1b is a statistically coherent pillar composed by 4 indicators, as long as indicator *Share of international co-publications* - 1b1 is replaced by another (*International Collaboration Index* - 1b1_collind);

dimension 2a needs to be reduced to three indicators (*Patent applications by the public sector* - 2a1, *Patent applications by industry* 2a4 and Patenting in *Key Enabling Technologies* - 2a8) for a statistically coherent pillar;

dimension 2b could only be represented by indicator *Public – private co-publications* - 2b1, given the negative correlation between the only two indicators available.

We found that the dimensions were populated by different numbers of indicators; each pillar captured one single latent dimension that explained at least 55% of variance in the data. In addition, the factors loadings for pillar 2a were rather unbalanced.

Second, we tested the two-dimensional framework distinguishing basic and applied research (Framework 2). The multivariate analysis showed that ‘basic research’ can only be captured in three distinct, conceptually heterogeneous, dimensions, while the ‘applied research’ pillar is relatively sound, and consists of three indicators (2a1, 2a4 and 2b1). In summary, this second framework was found weak because of the existence of three “basic research” pillars, for which the statistical profile was not matched with any theoretical underpinning.

Thirdly, aiming to achieve a framework which is both statistically sound and rich in indicators, we analysed by principal components the whole set of 22 indicators starting from Framework 1 and testing alternative specifications. In this third attempt we considered a modified indicator (2b3_gdp) to measure financial flows from business to public research. In Framework 3, we found three coherent and statistically sound pillars:

1. excellence of public research (6 indicators);
2. *interactions, collaborations (4 indicators)*
3. excellence in industrial actors (3 indicators)

The framework was found to accommodate 13 indicators (similarly to Framework 1), with each pillar capturing one single latent dimension, which explains at least 63 % of variance in the data. The factors loadings were overall rather balanced.

Composite indicators were computed for each of the three frameworks, using geometric aggregation across pillars. Each pillar was an arithmetic average of its indicators normalized between 0 and 100, using the min-max method and taking into consideration the two years simultaneously. Initially, the geometric aggregation was computed using equal weights, which were adjusted in light of the global sensitivity analysis results, carried out at the pillar level. This improved the balance between the pillars in all three cases.

For **Framework 1** the composite indicator shows a clear North-West vs South-East divide with centers of excellence in the Nordic and North-Sea countries. Almost all countries have improved their excellence between 2005 and 2009. The considerable spread of scores at the pillar level provides useful insights to research excellence performance: high and low scores are detected for pillars 1b and 2b, respectively.

For **Framework 2**, the pillars of the resulting composite indicator do not contribute equally to the overall composite indicator. A re-adjustment of weights could improve the framework imbalances, but leaves the conceptual difficulties of the theoretical framework unresolved.

According to **Framework 3**, three countries are clearly distinguished as the leaders in research excellence: Switzerland, Israel and the Netherlands, all with a score of 70 or above. They are followed by countries of North Western Europe with scores exceeding the EU27 score, and Southern and Eastern member states and Associate Countries below the EU27 score. Of the countries with the largest population, United Kingdom and Germany are neck and neck, ahead of France and Italy. Three countries trail the list with single-digit scores: Romania, Lithuania and Turkey.

The scores resulting from the frameworks, with adjusted weights, are shown in Figure 1 for 2009. Frameworks 1 and 3 are closely correlated (Pearson correlation coefficient = 0.97), while Framework 2 scores show greater variance.

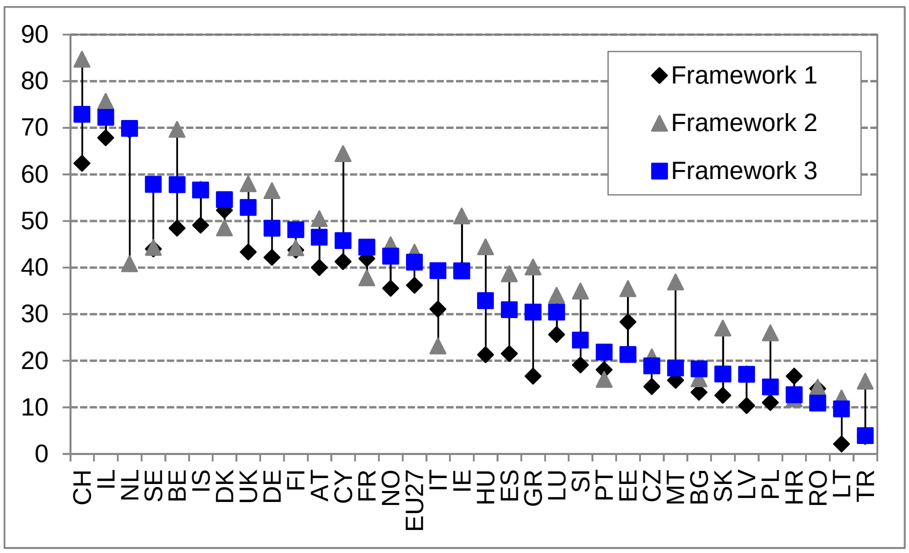


Figure 1 Research excellence composite scores, comparison of 3 conceptual frameworks (adjusted weights, 2009)

We tested the composite indicator of Framework 3 against other established composite indicators of the research and innovation system. We found that the research excellence composite scores correlate strongly with the Innovation Union Scoreboard 2011's Summary Innovation Index. Comparing the scores with GERD figures, we also found that composite score changes over time of around 20-30% are associated with R&D investments of at least 1.5% of GDP.

Our final recommendation is to consider Framework 3 as the basis for the development of the fully-fledged composite indicator on research excellence. The pillars in this framework explain a greater share of variance in the data, and each variable contributes to pillar scores in a more balanced way than in Framework 1. At the same time, the contribution of each pillar to the composite index is intrinsically more balanced.

Introduction and Objectives

The EU2020 strategy contains a blueprint for transforming Europe into an 'Innovation Union' by 2020. The Innovation Union flagship initiative (COM (2010) 546 final, October 6th, 2010) commits the EU to boosting investment in research and making Europe an attractive place to develop innovative products. Consequently, national governments will have to reform their innovation systems to boost cooperation between industry and universities, ensure a modernization of framework conditions for enterprises, and a number of other measures to enhance cross-border cooperation and to embrace joint programming. All these innovation aspects need to be carefully monitored by policy-makers in the European institutions and Member states.

This feasibility study, entitled 'ERA monitoring', focuses on monitoring the progress of Europe towards the completion of the European Research Area (ERA), towards the structural change of national and supra-national innovation systems and towards the modernization of higher-education institutions.

The project addresses the feasibility to develop three conceptual frameworks (organised in three work packages – WPs) and the potential to further aggregate the underlying components into composite indicators to measure:

> WP 1:
> progress in the construction and integration of a European Research Area (ERA), to monitor the overall performance of the Science and Technology system.
>
> WP 2:
> structural change, to monitor the increase towards a more knowledge-intensive economy in Europe coherently with the orientations of the EU 2020 strategy and the Innovation Union initiative.
>
> WP 3:
> research excellence in Europe, meaning the effects of European and National policies on the modernization of research institutions, the vitality of the research environment and the quality of research outputs in both basic and applied research.

The present deliverable represents the outcome of WP 3 of the project.

The objective of this work-package is to test the feasibility to develop a composite indicator to measure the excellence of the research systems of all EU Member States, most EFTA countries, Candidate countries and the main international competitors of the European Union

(United States, Japan and China). The aim of the study is to propose alternative conceptual frameworks of research excellence and to test their statistical coherence in order to identify feasible composite indicators.

In this WP – as well as in WP1 and WP2, the steps mentioned in the OECD/JRC Handbook[1] have been followed:

step1. Development of theoretical frameworks for the measurement of research excellence. The proposed frameworks were derived by the report entitled *Indicators of Research Excellence*, co- authored by Rémi Barré, Hugo Hollanders and Ammon Salter of the Expert Group (from now on identified as the Expert Group) on the Measurement of Innovation, finalized on 8 Oct 2011.

The first theoretical framework, consisting of 6 dimensions, was proposed by the Expert Group. The relevant data sources were collected at country level, and for as many years as possible, for all EU27 Member States, most EFTA countries, Candidate countries, and the main international competitors of the European Union (United States, Japan and China). Another framework consisting of two dimensions (basic and applied science) was extracted from the Expert Group report and tested. Finally, an alternative framework, in which the indicators were combined according to three dimensions deriving from a principal component analysis, was proposed and validated.

step2. Multivariate statistical tools have been used to assess the suitability of the data set and to ease the understanding of the implications of the methodological choices, e.g., weighting and aggregation, during the construction phase of the composite indicator. Statistical analysis has been used for imputing missing data, detecting outliers, and to suggest suitable transformations of indicators due to skewness or kurtosis. Principal components analysis has been used to verify whether the structure of the underlying data is consistent with the proposed conceptual framework and therefore is appropriate to describe the phenomenon. Note that principal component analysis has not been used as a weighting method.

step3. Construction of composite indicators. The composite indicator has been calculated by considering geometric aggregations of the pillars using both equal and adjusted weights. The composite scores for each pillar have been calculated by taking the arithmetic averages of the underlying indicators.

step4. Sensitivity analysis was conducted to appreciate the relative importance of the pillars on the overall composite. The results show how balanced the composite structure is

[1] Nardo M., Saisana M., Saltelli A., Tarantola S., Hoffman A., Giovannini E., (2008) *Handbook on constructing composite indicators: methodology and user guide*. OECD publishing. http://www.oecdbookshop.org/oecd/display.asp?CID=&LANG=en&SF1=DI&ST1=5KZN79PVDJ5J

in its components and suggest the adoption of adjusted weights in case the degree of balance has to be increased.

Theoretical Framework proposed by the Expert Group

The main aim of the Expert Group Report was "to recommend a short list of indicators and a method of assessment to "*describe the progress to excellence of European research*". In an initial step, the Report interprets the measurement of excellence at the country level. According to the Report, three stages of extension translate excellent pieces of research (validated by the peer-review process) to research excellence measured at the country level. First, having a "sufficient number of scientific articles and research projects that are considered excellent" define an excellent researcher. Second, research units (laboratories, university departments and even research institutes or universities) are considered as excellent by "a sufficient proportion of [excellent] researchers". Third, regions and countries are excellent if "a sufficient proportion of their research institutes and universities are considered excellent".

Why and how does research excellence matter? The main reason, as stated in the Report, for governments to be concerned about research excellence and "a 'better functioning' national research and innovation system" is the need to maximize efficiency when allocating resources to research organizations through various schemes. Three "contexts of engagement" are identified:

- *research* ("engagement with actors in the academic context with issues of scientific relevance and scientific quality"),
- *innovation* ("engagement with actors in a socio-economic context with issues of socio-economic relevance, in particular related to firms", or how efficiently to convert R&D investment into value) and
- *society* ("engagement with the citizens - the general public – stakeholders and concerned groups as well as the political and public authorities, through issues of public policies, quality of life, sustainability with attention to issues of risks and ethics in the political context defining political relevance").

In each of these three contexts the *production of knowledge* as well as the *institutional arrangements and interaction schemes* in which knowledge activities take place matter for excellence. The reason for focusing not only on the end result of knowledge production but also on the mechanisms through which knowledge is produced is the time lag between the activity and the impact.

In addition to measuring the existence and size of excellent research activities, further measures of excellence include: *impact*, *openness* and *attractiveness* of the research and innovation system. Impact is more closely associated with knowledge production, while openness and attractiveness are more closely associated with institutional arrangements and interaction schemes.

The Expert Group Report concludes that the measurement of research excellence should consider six types of activities, or dimensions, which are presented in a matrix in Table 1 (taken over from the Report). The issues at stake are both the excellence of each dimension as well as “the balance of the portfolio”.

As words of caution, the Report points out that excellence should not be seen as an absolute term, and that one need to appreciate the various dimensions and their relative weights in context. It also points out potential problems with taking into consideration sectoral and disciplinary diversity and life-cycle dynamics of research actors.

Table 1 Research Excellence - Dimensions of Analysis

Issues	**Impact**	**Openness - attractiveness**
Context of engagement	**Knowledge production**	**Institutional arrangements, interaction schemes**
Research actors	1a Production of generic knowledge Scientific publications	1b Coordination, networks Collaborative schemes, infrastructures & instruments open to the scientific community
Industrial innovation actors	2a Knowledge and expertise orientated towards industrial innovation	2b Industrial and professional partnerships and collaborative schemes (in particular in the context of clusters) Public research – industry linkages, consulting
Societal and political actors	3a Knowledge and expertise orientated towards societal concerns and policy regulations S&T diffusion and culture	3b Civil society and public policies partnerships and collaborative schemes

Source: Expert Group Report Table 1, p.7)

Potential indicators and data treatment

Table 1 reports the indicators proposed by the Expert Group which are robust and currently available (Table 8 of the EG Report). As it can be noted, no indicators are currently available for the societal dimensions (3a and 3b)

Table 1: available and reliable indicators proposed by the Expert Group for the assessment of research excellence

1a. Engagement with research actors - Knowledge production / impact	
1a1	**Publications per 1000 researchers in public research**
1a2	**% of Highly cited publications / % publications**
1a3	**Average of relative citations (ARC)**
1a5	**% (EU) ERC and/or Marie-Curie grantees / % (EU) public RD spending (HERD + GOVERD)**
1a9	**Specialisation in publications in the fields of the Grand societal challenges**
1a10	**Specialisation in publications in the fields of the Key enabling technologies (KETs)**
1b. Engagement with research actors - Institutional arrangements, interaction schemes / openness – attractiveness	
1b2	**Collaboration index (with emerging countries)**
1b5	**% Foreigners in doctoral programmes**
1b7	**% (EU) Coordination position in FP projects/ % (EU) participation in FP projects**
2a. Engagement with industrial innovation actors – Knowledge production / impact	
2a1	**Patent applications by HEIs+PROs (per 1000 researchers)**
2a7	**Specialisation in patenting in the fields of Grand social challenges**
2a8	**Specialisation in patenting in the Key enabling technologies (KETs)**
2b. Engagement with industrial innovation actors – Institutional arrangements, interaction schemes / openness - attractiveness	
2b1	**Public – private co-publications per million population**
2b3	**% (national) HERD+GOVERD financed by business / RD financed by business**
2b4	**% (national) of industry funded HEIs + PROs budget**
3a. Engagement with societal and political actors - Knowledge production / impact	
3b. Engagement with societal and political actors - Institutional arrangements, interaction schemes / openness – attractiveness	

Table 7 of the Expert Group report provides additional available indicators with some problems of "reliability, robustness and comparability". We could collect reliable data for 7 indicators. These are listed in Table 2 using the language of the Expert Group report

Table 2: Additional available indicators with some problems of reliability, robustness and comparability proposed by the Expert Group

1a4	% (EU) 250 top scientific universities / % (EU) public RD spending (HERD + GOVERD)
1a6	% (EU) ERC and/or Marie-Curies grantees / % (EU) of [HE researchers + government RD personnel] in FTE
1a7	% women among researchers
1a8	% (EU) Scientific prizes / % (EU) HEI+PRO spending
1b1	International collaborations index
2a3	% (national) innovative firms that use HEIs+PROs as a source for their knowledge for innovation (CIS)
2a4	Patent applications by industry (per 1000 researchers and/or relative to BERD)

List of indicators used

We have collected data for the indicators provided in Tables 1 and 2 and for some additional indicators which are not in these lists. The 22 indicators used in this WP are defined below, some of which may not coincide with the indicators proposed by the Expert Group, because of data availability.

Note that indicators 1a5, 1a6, 1b7 and 2a3 defined below were not meaningful beyond the ERA (e.g., research institutes in the US or Japan are not receiving ERC grants). These indicators could therefore not be considered for any global comparison.

1a1: Publications per 1000 researchers

Definition: Total number of publications by country divided by 1000 researchers.

In the EG Report, the denominator is defined as 1000 researchers in "public research". However, since the set of publications in the numerator covers all publications, including those with authors in the private sector, the denominator was adjusted to similarly cover all sectors for consistency reasons.

Sources and notes:

Numerator: Science Metrix (Scopus) data on the total number of publications by country (full counting method, based on Scopus 2011 bibliometric data); due to incomplete coverage for 2009, 2008 figures shown as latest year available.

Denominator: Eurostat, OECD and UNESCO data on R&D personnel and researchers by sectors of performance. Occupation-based definition applied, figures in full-time equivalents (FTEs). Data extrapolated for missing years for GR, FI, HR, US and interpolated for CH; 2005 and 2008 values used.

1a2: Highly cited publications as a share of total publications

Definition: Total number of publications within the top 10% most cited publications as a share of all publications by country.

Sources and notes: Science Metrix (based on Scopus data) computation, using 'Relative Citation' indices (see explanation for indicator 1a3 below); full counting method applied (co-authored publications shown for each author's country). Since a 4-year citation window was applied (reference year + 3 years), 2007 was the latest reference year available.

1a3: Average of relative citations (ARC)

Definition: 'Relative citation' indices are computed by dividing the citation count of a publication by the average citation count of all publications of the corresponding document type (reviews benchmarked against reviews, articles against articles) published in the same year in the same scientific subfield. The ARC is obtained by computing the average of such type- and field-normalized citation scores. A value above 1 indicates that an entity is cited more frequently than the world average.

Sources and notes: Science Metrix (based on Scopus data) computation; a citation window includes the three years following the year of publication (i.e. 2005 scores cover the period of 2005-2008), however, 2007 scores refer only to a window of 2007-2009 due to the incompleteness of reference year 2010 at the time of the analysis.

1a4: Share in Top 250 scientific universities / public R&D spending (HERD+GOVERD)

Definition: Number of universities in a country included in the list of the world's top 250 universities based on scientific impact, divided by public R&D spending of the higher education and government research institutes.

Sources and notes:

Numerator: The Leiden Ranking (CWTS, Leiden University) ranking for 2008 and 2010 based on CWTS computations of size-independent and field-normalized average impact. In order to measure global excellence and to be able to benchmark against non-European countries, world top 250 ranking was used.

Denominator: Eurostat and OECD data on Total intramural R&D expenditure (GERD) by sectors of performance and source of funds; HERD+GOVERD calculated using constant 2000 PPPs. (IL 2009 values extrapolated.) Previous year's figures were used (i.e. 2007 and 2009) in the numerator.

Note that the indicator does not take into account the position in the top 250, which was not seen as a problem as it would refer to less than the top 10% universities (a recent EUMIDA study[2] identified some 2500 higher education institutes in Europe alone).

1a5: Amount of ERC grants received by country / public RD spending (HERD + GOVERD)

Definition: The total amount of European Research Council (ERC) grants by country of host organization spread over duration of project divided by HERD+GOVERD.

Sources and notes:

Numerator: DG-RTD data on ERC grants (retrieved: 18 May 2011). The total amount of ERC grant funding received by the country of the host organization, spread equally over a project's years of duration. Includes both starting grants and advanced grants; first year available is 2008. (I.e., 1 million EUR grant responding to a call in 2007 for a project starting in 2008 and ending in 2011 will be counted as 250 thousand EUR for the years 2008, 2009, 2010 and 2011.)

Denominator: Eurostat and OECD Total intramural R&D expenditure (GERD) by sectors of performance and source of funds data used to computed HERD+GOVERD, with current prices in millions of EUR PPPs.

Note that the indicator is only applicable for European Research Area countries.

1a6: Amount of ERC starting grants researchers by country / number of researchers in institutes of higher education and public research institutes in FTE

Definition: The total number of ERC starting grant receivers by country of principle investigator divided by the total number of researchers in HEI and PROs, in full-time equivalents (FTEs). The indicator aims at measuring the excellence of young researchers in a country in receiving ERC funding.

Sources and notes:

---

[2] European Commission (2010) *Feasibility Study for Creating a European University Data Collection, Final Study Report*, DG-RTD. (URL: ec.europa.eu/research/era/docs/en/eumida-final-report.pdf)

Numerator: DG-RTD data on ERC grants (retrieved: 18 May 2011). The total number of ERC grants (projects) awarded to a country of the principle investigator by year of grant call. (ERC Starting grants are aimed at offering young researchers with 2-12 years of experience opportunities to develop independent careers.) First year of grant calls was 2007; subsequently grants were advertised from 2009 onwards. 2009 and 2010 figures averaged in order to avoid fluctuations.
Denominator: Eurostat Total R&D personnel by sectors of performance, occupation and sex data; occupation-based definition; sum of full-time equivalents computed for Higher education and public research organizations.
Note that the indicator is only applicable for European Research Area countries.

1a7: Share of women among researchers

Definition: The number of women intramural researchers divided by the total number of intramural researchers, computed in terms of headcount.
Sources and notes: UNESCO, OECD, Eurostat data on R&D personnel by sector of employment and sex; occupation-based definition, including all sectors of employment. Data of US differs in definition, and refers to "all persons with bachelor's or higher degrees in science or engineering (S&E)" as published by the National Science Foundation.

1a8: Number of highly valued scientific prizes by HEI and PRO R&D spending

Definition: The number of Nobel prizes in natural sciences and economics plus Field Medals by country divided by HEI and PRO R&D spending.
Sources and notes:
Numerator: The total number of Nobel prizes in chemistry, medicine, physics and economics, and Fields Medal (in mathematics) awarded to researchers by country of affiliation within the 5 most recent years (i.e., 2005 values refer to 2001-2005).
Note that the distribution of these prizes is highly skewed, '0' values found for over 77% of the countries within our scope; and highly concentrated to one country (US), which accounts for 61% and 45% of all prizes in 2005 and 2010 respectively.
Denominator: HERD+GOVERD (same as 1a4).

1a9: Specialisation in publications in the fields associated with Grand Societal Challenges

Definition: A specialization index = number of a country's publications within FP7 thematic priorities classified as Grand Societal Challenges (GSC) divided by the total number of publications in a country, over the share of GSC publications in the world.
Sources and notes:
Numerator: Science Metrix (based on Scopus data) data; the total number of publications by country published in journals classified in any of the following FP7 thematic priorities 1-Health, 2a-Food, Agriculture and Fisheries, 5-Energy or 6-Environment (incl. Climate Change), over the total number of publications by country (fractional counting method to avoid duplications).
Denominator: The share of GSC publications in the world.

1a10: Specialisation in publications in the fields associated with Key Enabling Technologies

Definition: A specialization index = number of a country's publications within FP7 thematic priorities classified as Key Enabling Technologies (KET) divided by the total number of publications in a country, over the share of KET publications in the world.

Sources and notes:

Numerator: Science Metrix (based on Scopus data) data; the total number of publications by country published in journals classified in any of the following FP7 thematic priorities 2b-Biotechnology, 3-Information and Communication Technologies, 4a-Nanosciences and Nanotechnologies, 4b-Materials (excl. nanotechnologies) or 4c-New Production Technologies, over the total number of publications by country (fractional counting method to avoid duplications).

Denominator: The share of KET publications in the world.

1b1_collind: International collaboration index

Definition: The ratio between the predicted number of international co-publications and the observed number of co-publications by country.

Sources and notes:

Science Metrix calculations, based on the overall number of publications in Scopus; this is a scale-adjusted indicator of collaborations, computed by adjusting for the power-law relationship between the number of publications and the number of co-publications. A value above 1 means that a country produces more publications in collaboration with at least another country than expected based on the size of its scientific production. The collaboration index (CI) was computed by log transforming the number of publications and number of co-publications, and performing a log-log linear regression to estimate the constants (*a* and *k*) of the following equation:

Expp (M) = a * (M^k), *where Expp: expected number of co-publications of a country; M: the observed number of publications by a country*[3]

As this indicator was only computed by Science Metrix for the entire period 2000-2009, a non-scale-adjusted share of collaborations index was also considered (see below 1b1).

1b1: Share of international co-publications to total number of publications

Definition: The ratio of international co-publications to the total number of publications by country.

Sources and notes:

Numerator: Science Metrix (based on Scopus data) data on the number of co-publications from a country in which the co-authors are from at least two different countries (full counting, all fields)

Denominator: Science Metrix (based on Scopus data) data on the total number of publications from a country (full counting, all fields).

Note that this indicator is not scale adjusted (c.f. indicator 1b1_collind).

[3] See detailed description on p.54 of Science Metrix Suite of Methods – Methods Associated to Report 2.3.1 to European Commission

1b2: Share of international collaborations with non-EU partners in total publications

Definition: The ratio of international co-publications with at least one partner from a non-EU member state to the total number of publications by country.

Sources and notes:

Numerator: Science Metrix (based on Scopus data) data on the number of co-publications from a country in which one of the co-authors is from a non-EU country (other collaborator may or may not be from an EU country; fractional counting, all fields).

Denominator: Science Metrix (based on Scopus data) data on the total number of publications from a country (full counting, all fields).

Note that this indicator is not scale adjusted (c.f. indicator 1b1). The indicator is not, as originally intended, a measure of co-publication with emerging countries since co-publications with (among others) US, JP, CH, NO, IL could not be excluded.

For non-EU member states all international collaborations are counted. The difference with 1b1 is that this indicator uses fractional counting.

1b5: Share of foreigners in doctoral programmes

Definition: Share of foreigners from other EU or non-EU countries in doctoral programmes within the total number of doctoral candidates in a country.

Sources and notes: Eurostat Education Statistics and MORE Survey (Table D of MOB-ST4: 'Number and share of doctoral candidates (ISCED 6) with the citizenship of another EU27 member state in the reporting country in the EU27' and 'Number of doctoral candidates (ISCED 6) with another non-EU27 citizenship per country in the EU27'). Data for DE taken from DESTATIS, Statistik der Studenten, Studierende: Deutschland, Semester, Nationalität, Geschlecht, Angestrebte Abschlussprüfung.

(For the US National Science Foundation (NSF) 'S&E and non-S&E foreign students enrolled in U.S. higher education institutions, by academic level: 2006–09' table provides the closest matching data, however, since it does not cover only science and engineering students, it was not considered.)

Note that data coverage is low, no data is available for GR, IE, LU or NL, or any non-EU country except for the US.

1b7: Share of coordination position in FP projects / participation share in FP projects

Definition: A specialization index = the share of coordination position by a country to the total number of coordination positions of an FP6 or FP7 project, divided by a country's share of participation to the total number of FP6 and FP7 projects.

Sources and notes:

Special tabulations from CORDIS E-Corda (retrieved 12 Sep 2011); data for 2006 includes all FP6 signed grant agreements by country over the period 2002 to 2006; data for 2011 includes the same for FP7 over the period 2007-2011.

Numerator: Number of FP projects coordinated by an entity from a given country divided by the total number of FP projects.

Denominator: Number of FP project participation by an entity from a given country divided by the total number of FP project participation of the country.

Note that the indicator is only applicable for European Research Area countries.

2a1: Patent applications by HEIs and PROs per 1000 researchers

Definition: Patent applications to the EPO by HEIs and PROs divided by the number of researchers in HEIs and PROs in head count.

Sources and notes:

Numerator: Eurostat 'Patent applications to the EPO by priority year at the national level by institutional sector', filed by HEIs and PROs, fractional counting. Aggregate of preceding 3 years counted to avoid fluctuation bias, values below 5 were not considered in order to avoid small number bias. (In view of this rule, the values for BG, CY, LT, LU, LV, HR, MK, IS, LI for all years and EE, RO, SK for 2005 were imputed.)

Denominator: Eurostat 'Total R&D personnel by sectors of performance, occupation and sex' and OECD 'Main Science and Technology Indicators'; researchers in HEIs and PROs, headcount. For US, missing headcount values were estimated by applying US to UK non-business sector FTE ratios, and extrapolating for other years using FTE series. CN data from 'China Statistical Yearbook 2010: Education, Science and Technology', Tables 20-41 and 20-42, definitions may differ. No data available for BR, IL and IN.

2a3: Share of innovative firms that use HEIs and PROs as a main source for their knowledge for innovation

Definition: The share of firms that indicated universities and public research organizations as a highly important source of information for innovation.

Sources and notes: Eurostat Community Innovation Survey (CIS) data for 2004, 2006 and 2008, covering enterprises with technological innovation (product, process, ongoing or abandoned), regardless organizational or marketing innovation, all activities. Data for time point 2005 refers to 2004 for all countries except AT, MT, PT, SI and TR, which refer to 2006. EU-27 value is an unweighted average.

Note that the indicator is only applicable for European Research Area countries.

2a4: Patent applications by industry relative to BERD

Definition: Patent applications to the EPO by business enterprises divided by business expenditure on R&D (BERD).

Sources and notes:

Numerator: Eurostat 'Patent applications to the EPO by priority year at the national level by institutional sector', filed by business enterprises. Aggregate of preceding 3 years counted to avoid fluctuation bias (i.e., 2002, 2003 and 2004 for first time point; 2005, 2006 and 2007 for second time point). 2008 values not considered, decline assumed due to incomplete data.

Denominator: Eurostat Total intramural R&D expenditure funded by business, millions of PPPs at constant 2000 prices. If data was missing for a year, average of neighboring years used (ie: DK, GR, LU, MT, NL, SE, IS, NO, CH) Extrapolations: GR for 2006-2008; IT for 2000-2004.

Note: alternative definition (using 1000 researchers as denominator) not used due to lower data coverage.

2a7: Specialization in patenting in the fields of Grand Societal Challenges (GSCs)

Definition: A specialization index = number of patent applications filed under patent cooperation treaty (PCT) in the fields associated with GSCs divided by the total number of PCT patents, over the share of GSCs within all PCT patents in the world.

Sources and notes:

Numerator: OECD Patent Statistics data; the total number of PCT patents in B. Energy generation from renewable and non-fossil sources; D. Technologies specific to climate change mitigation; F. Emissions abatement and fuel efficiency in transportation; G. Energy efficiency in buildings and lighting over the total number of PCT patents of a country.

Denominator: The share of GSC patents in the world.

For both numerator and denominator, the sum of 3 preceding years used (i.e. 2003-04-05 for 2005 figures) to avoid fluctuation bias; in the numerator, sums less than 5 not considered to avoid small numbers bias (this affected BG, EE, LT, LV, MT, RO, SK and MK for all years, and CY for 2005). PCT patents were used to allow for an unbiased global comparison.

2a8: Specialization in patenting in the fields of Key Enabling Technologies (KETs)

Definition: A specialization index = number of patent applications filed under patent cooperation treaty (PCT) in the fields associated with KETs divided by the total number of PCT patents, over the share of KETs within all PCT patents in the world.

Sources and notes:

Numerator: OECD Patent Statistics data; the total number of KETs patents in Biotechnologies, Nanotechnologies and ICTs over the total number of PCT patents of a country.

Denominator: The share of KET patents in the world.

For both numerator and denominator, the sum of 3 preceding years used (i.e. 2003-04-05 for 2005 figures) to avoid fluctuation bias; in the numerator, sums less than 5 not considered to avoid small numbers bias (this affected CY, MK and LI for all years). PCT patents were used to allow for an unbiased global comparison.

(NOTE: ICTs added in line with KETs definition and for the sake of consistency vis-à-vis publications specialization indicator 1a10, although it was not included in EG Report.)

2b1: Collaborations with industry

Definition: Public-private co-publications per million population.

Sources and notes:

Data source: CWTS (Thomson Reuters database); Computed by CWTS (Leiden University, http://www.cwts.nl) for the Innovation Union Scoreboard 2011 (see indicator 2.2.3)

Numerator: Number of public-private co-authored publications (includes all research-related papers (document types: ‘research articles’, ‘research reviews’, ‘notes’ and ‘letters’) published in the Thomson Reuters Web of Science database. These co-publications have been allocated to one or more countries according to the geographical location of the business enterprise (or enterprises) that are listed in the authors affiliate address(es); as a result the geographical location of the public sector research partner(s) in those addresses is not relevant. Each co-publication is counted as one publication for each country, irrespective of the number of co-authors and (parent) organisations listed in the author affiliate

address(es). The definition of the “private” sector excludes the private medical and health sector.
Denominator: Total population as defined in the European System of Accounts (ESA 1995). Data are two-year averages.

2b3: HERD and GOVERD financed by business / R&D financed by business

Definition: R&D performed by the higher education and government sectors and funded by business, divided by R&D financed by business.
Sources and notes: Eurostat ‘Total intramural R&D expenditure (GERD) by sectors of performance and source of funds’ (retrieved 25 Jan 2012).
2008 values used for CH, JP, KR, CN, US; others for 2009. No data available for GR after 2006.

2b3_gdp: HERD and GOVERD financed by business / GDP

Definition: R&D performed by the higher education and government sectors and funded by business, divided by GDP at purchasing power standard
Sources and notes: Eurostat ‘Total intramural R&D expenditure (GERD) by sectors of performance and source of funds’ (retrieved 25 Jan 2012); Eurostat ‘GDP and main components’; OECD GDP data. 2008 R&D and GDP figures were used for CH, JP, KR, CN, US; others for 2009. No R&D data was available for GR after 2006.
The denominator for 2b3 was modified after having found negative correlations between the original 2b3 and the majority of the excellence indicators. Normalizing by GDP offers a more balanced indicator of financial flows from private to public research then normalizing by BERD, which rewards countries with relatively low levels of business R&D expenditure.
Since this indicator was only introduced at a late stage of the study, it was only considered for the alternative framework 3.

Descriptive statistics

Table 2 provides summary information for the 22 indicators defined above. The information includes main sources, percentage of missing values (after the first level of imputation, see next sub-section), minimum, maximum, mean values and standard deviations across countries and for two reference time points.

Table 2 Descriptive Statistics of the Research Excellence Dataset (considering 41 countries)

Name of Indicator	1a1		1a2		1a3		1a4	
Description	Publications / 1000 researchers		Highly cited publications / total publications		Average of relative citations (ARC)		% Top 250 universities / (HERD+GOVERD)	
Main Sources	Science Mtrx (Scopus) / Eurostat		Science Metrix (Scopus)		Science Metrix (Scopus)		CWTS / Eurostat, OECD	
Reference years	2005	2008	2005	2007	2005	2007	2008	2010
% of missing values	2%	5%	0%	0%	0%	0%	5%	5%
Min	72.9	73.8	4.3	4.3	0.53	0.53	0.00	0.00
Max	879.9	1034.5	17.7	18.2	1.62	1.64	4.58	4.27
Mean	368.0	406.5	10.2	10.8	1.07	1.10	0.90	0.81
Standard Deviation	176.3	201.3	3.8	3.9	0.29	0.30	1.19	1.05
Name of Indicator	1a5		1a6		1a7		1a8	
Description	Tot ERC grants value / (HERD + GOVERD)		Nr. ERC starting grants / HEI+PRO researcher FTE		% of women among researchers (HC)		Scientific awards / (GOVERD + HERD)	
Main Sources	DG-RTD, Eurostat		DG-RTD, Eurostat		Eurostat, OECD, UNESCO		Nobel, Fields / Eurostat	
Reference years	2008	2010	2007	2010	2005	2008	2005	2010
% of missing values	17%	17%	24%	24%	7%	7%	0%	0%
Min	0.00	0.00	0.00	0.00	0.12	0.13	0.0	0.0
Max	3.89	14.3 (7.2)*	5.3 (1.60)*	1.6 (1.03)*	0.52	0.52	41.0	32.0
Mean	0.81	2.97	0.32	0.34	0.34	0.34	1.6	1.7
Standard Deviation	1.00	2.55	0.36	0.27	0.09	0.09	6.5	5.6
Name of Indicator	1a9		1a10		1b1		1b1 Coll Ind	
Description	Specialisation in GSC publications		Specialisation in KET publications		% international co-publications / total publ.		International collaboration index	
Main Sources	Science Metrix (Scopus)		Science Metrix (Scopus)		Science Metrix (Scopus)		Science Metrix (Scopus)	
Reference years	2005	2009	2005	2009	2005	2009	2009	2009
% of missing values	0%	0%	0%	0%	0%	0%	0%	0%
Min	0.35	0.41	0.40	0.52	0.13	0.14	0.43	0.43
Max	1.53	1.34	2.24	1.99	0.72	0.74	1.51	1.51
Mean	0.98	1.00	1.03	1.02	0.41	0.43	1.00	1.00
Standard Deviation	0.26	0.24	0.44	0.39	0.14	0.14	0.26	0.26
Name of Indicator	1b2		1b5		1b7		2a1	
Description	% int'l collaborations with non-EU partners		% of foreigners in doctoral programmes		% coord,pos. in FP projects / % particip.		HEI+PRO Patent appl./ 1000 researchers HC	
Main Sources	Science Metrix (Scopus)		Eurostat, MORE		DG-RTD, CORDIS		Eurostat	
Reference years	2005	2009	2005	2007	2006	2011	2002	2005
% of missing values	0%	0%	39%	39%	7%	7%	34%	27%
Min	0.10	0.11	0.00	0.00	0.2	0.1	0.1	0.1
Max	0.40	0.46	0.41	0.51	1.4	2.0	15.4	20.3
Mean	0.22	0.24	0.13	0.14	0.8	0.8	3.9	4.3
Standard Deviation	0.07	0.07	0.11	0.13	0.3	0.4	4.3	4.9
Name of Indicator	2a3		2a4		2a7		2a8	
Description	firms with HEIs and PROs as main source of innov.		Patent applications by industry / BERD		Specialization in GSC patents (PCT)		Specialization in KET patents (PCT)	
Main Sources	Eurostat CIS		Eurostat		OECD		OECD	
Reference years	2004	2008	2005	2008	2005	2009	2005	2008
% of missing values	34%	44%	7%	7%	20%	17%	10%	7%
Min	0.03	0.03	0.03	0.03	0.30	0.24	0.28	0.28
Max	0.12	0.14	0.87	0.76	2.17	2.47	1.37	1.69
Mean	0.06	0.06	0.31	0.29	1.08	1.16	0.79	0.82
Standard Deviation	0.02	0.03	0.23	0.20	0.40	0.46	0.28	0.29
Name of Indicator	2b1		2b3		2b3_gdp			
Description	Co-publications with industry / mln Pop		HERD+GOVERD financed by business / BERD		HERD+GOVERD financed by business / GDP			
Main Sources	CWTS (IUS)		Eurostat		Eurostat			
Reference years	2005	2008	2005	2009	2005	2009		
% of missing values	20%	20%	7%	10%	12%	12%		
Min	0.9	1.2	0.2	0.5	0.000	0.002		
Max	183.0	198.5	29.1	19.5	0.177	0.151		
Mean	38.9	45.1	7.9	7.2	0.038	0.041		
Standard Deviation	45.0	51.1	7.6	5.9	0.036	0.036		

* value after outlier treatment

Missing data

Missing data was imputed at two levels. **At the first level**, when figures were missing in the annual time series of the data that was collected for numerators and denominators of all the variables considered, two methods were applied: linear interpolation (missing data between two known years) and linear trend extrapolation (several missing values after or before a given year). Imputation at this first level was only used if a linear trend could be identified, with no breaks or fluctuations prevalent in the data source. Such a quality-improvement-at-the-source exercise was crucial in order to minimize the number of missing values for ratio-type indicators where a numerator or denominator was not available for one of the selected benchmark years. (A typical example where imputation in the original data was useful were variables using head count of the number of researchers.) As a result, there was still a considerable share of missing values, as shown in Table 2.

Variables most affected were 1a6 (ERC starting grant receivers), 1b5 (foreigners in doctoral programmes) and 2a3 (CIS innovators that value public sources) where in general non-EU countries were missing, and variable 2a1 (public researcher patent applicants) where data coverage for small countries was unreliable and thus missing. Countries with the highest share of missing values in the total set of indicators were non-EU countries, including Israel, Brazil and India (with 42, 42 and 49% missing values) and China, South Korea and Japan (all with 28%). Missing data for EU countries was significantly lower, with an average of 6% in all indicators and no missing data was observed for variables 1a1, 1a2, 1a4, 1a7, 1a8, 1a9, 1a10, 1b1 (1b1_collind), 1b2, 1b7, 2a4, and 2b1. The indicators most affected were 2a1 (33% and 22% in the two time points respectively), 2a7 (30% and 26%) 2a3 (11% and 22%), 1b5 (15% both time points).

In this study, our aim was to impute missing data at the next level in a way that it does not influence the final composite scores. So, using the normalized dataset, missing data was imputed **at the second level** with the average of the various indicators for the same country within the relevant pillar. In this way, the values imputed for a country could differ depending on the conceptual framework applied, as different conceptual frameworks are made by different pillars. We also note that this “no imputation” method could overestimate the performance of countries with missing values, and another unintended consequence is that it may encourage countries not to report low values.

Outlier treatment, normalization

Excessive skewness and kurtosis was observed for ERC grant-related indicators 1a5 (skewness: 2.4, kurtosis: 6.7) and 1a6 (skewness: 6.0, kurtosis: 43.3). Applying an upper control boundary which was computed by adding 1.5 times the interquartile range to the upper quartile value, the 2010 (referring to time point 2009) values for indicator 1a5 for CH, IL, CY, NL and UK were winsorized. In the case of indicator 1a6, the 2007 (used for first time point) value for CY was replaced by the original 2010 value (used for second time point), and the 2010 value was replaced by the upper control limit value obtained in the same way as described for indicator 1a5.

In the case of indicator 1a8 (Nobel prizes and Field medals obtained) we observed such a highly skewed and peaked (kurtosis 30.2) distribution that could not be treated by any kind of method without a significant loss of information. We therefore chose to report the values in Figure 2 but not to include it when computing the composite indicators.

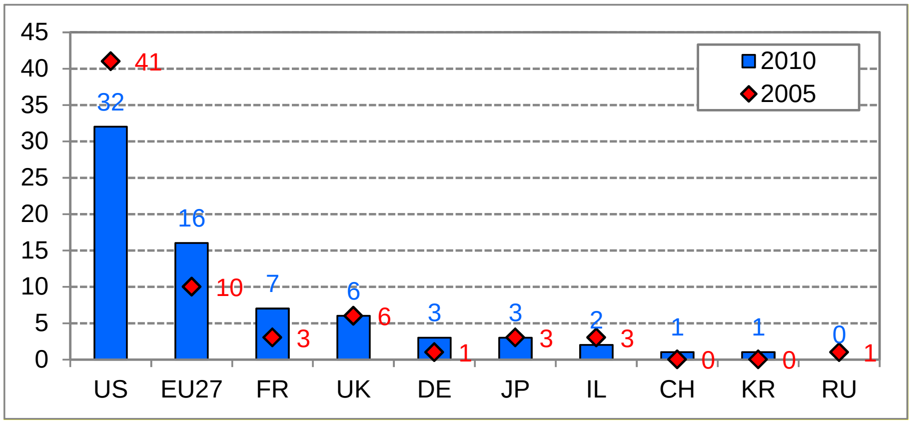


Figure 2 Nobel Prize and Field Medal winners by country of affiliation; sum of current and 4 preceding years; 0s not shown for remaining 33 countries

As a matter of standard practice, the dataset was normalized by applying the 'min-max' transformation method.

Conclusion:

Indicator 1a8 cannot be used as input for the composite indicator because of highly polarized distribution; the use of the other indicators depends on the conceptual framework applied, which is a result of the multivariate statistical analysis (see next section). Missing data are present for many indicators, especially in the case of non-European countries. These are imputed by averages of normalized indicators within each pillar, in order not to influence the composite scores.

Multivariate analysis and three theoretical frameworks

Multivariate analysis was carried out as a tool to verify the internal statistical consistency of the data within each pillar of the conceptual framework. Ideally, a composite indicator is structured in a way that each pillar describes a single latent component. This requires positive and high level of correlation within each pillar. There are two ways to test this: using a correlation table including all indicators, and by conducting classical Principal Component Analysis (PCA) as a method of dimensionality reduction. Based on the PCA results, relevant dimensions can be accepted if they adhere to the following criteria: (a) have an eigenvalue above 1 (Kaiser criterion); (b) account for at least 10% of total variance; (c) cumulatively contribute to more than 60% of total variance (OECD-JRC 2008).

For the multivariate analysis, we used the set of 21 indicators described in section 3. The pairwise correlation coefficients before second level imputation are reported in Table 3. The correlation matrix is split in four groups (1a, 1b, 2a and 2b), according to the conceptual framework proposed by the EG.

Table 3 Correlation matrix, 21 indicators, all time points combined

	1a1	1a2	1a3	1a4	1a5	1a6	1a7	1a9	1a10	1b1	1b1collin	1b2	1b5	1b7	2a1	2a3	2a4	2a7	2a8	2b1	2b3
1a1	1.00																				
1a2	0.46	1.00																			
1a3	0.28	0.59	1.00																		
1a4	0.53	0.66	0.30	1.00																	
1a5	0.55	0.55	0.21	0.47	1.00																
1a6	0.50	0.30	-0.05	0.29	0.46	1.00															
1a7	-0.14	-0.48	-0.27	-0.35	-0.30	-0.36	1.00														
1a9	0.32	0.56	0.37	0.45	0.21	-0.14	-0.32	1.00													
1a10	-0.34	-0.48	-0.40	-0.34	-0.17	0.14	0.09	-0.80	1.00												
1b1	0.26	0.52	0.50	0.11	0.29	0.28	-0.04	0.09	-0.17	1.00											
1b1collind	0.39	0.73	0.51	0.48	0.48	0.42	-0.21	0.27	-0.29	0.81	1.00										
1b2	0.31	0.69	0.52	0.33	0.57	0.33	-0.20	0.28	-0.21	0.77	0.79	1.00									
1b5	0.21	0.69	0.29	0.67	0.50	0.21	-0.49	0.54	-0.38	0.06	0.69	0.54	1.00								
1b7	0.42	0.64	0.38	0.43	0.49	0.27	-0.51	0.51	-0.39	0.26	0.50	0.50	0.84	1.00							
2a1	0.39	0.63	0.21	0.52	0.44	0.63	-0.51	0.34	-0.26	0.34	0.47	0.46	0.60	0.42	1.00						
2a3	0.05	-0.03	0.11	-0.14	0.19	-0.07	-0.13	0.15	0.06	-0.01	0.00	0.00	0.08	0.13	-0.24	1.00					
2a4	0.55	0.64	0.35	0.51	0.45	0.49	-0.49	0.34	-0.34	0.43	0.59	0.51	0.35	0.42	0.58	-0.16	1.00				
2a7	0.17	0.16	0.30	0.05	0.11	-0.16	0.34	0.16	-0.20	0.32	0.15	0.37	-0.03	0.05	0.11	0.04	-0.11	1.00			
2a8	-0.17	0.40	0.17	0.13	0.33	0.22	-0.30	-0.07	-0.04	0.08	0.16	0.29	0.16	0.25	0.27	-0.13	0.11	-0.05	1.00		
2b1	0.34	0.78	0.42	0.62	0.44	0.06	-0.35	0.54	-0.42	0.41	0.63	0.67	0.51	0.33	0.35	-0.02	0.42	0.22	0.33	1.00	
2b3	-0.16	-0.56	-0.34	-0.27	-0.25	-0.23	0.61	-0.25	0.19	-0.23	-0.34	-0.30	-0.44	-0.25	-0.25	0.18	-0.41	-0.09	-0.34	-0.36	1.00

The correlation matrix indicated the following key messages:

- The variable with the highest number of positive correlations is *1a2* (highly cited publications), which is very much in line with the concept of research excellence.
- Variables *1a7* (share of women among researchers), *1a10* (specialization in KETs publications) and *2b3* (public R&D finance by business) have mostly significant negative correlations with many of the other indicators, and only correlate positively with one another. This suggests that variables *1a7*, *1a10* and *2b3* will most likely have to be excluded from the aggregation. However, they provide useful information on their own.
- Variables that show little if any correlation with any of the others include CIS variable *2a3*, and GSC and KETs patent specialization indicators *2a7* and *2a8*.

- We included two indicators of international collaboration: *1b1*, the share of international co-publications, and *1b1_collind*, computed as an index correcting for scale effects. The correlation matrix shows that *1b1_collind* fits better in the composite framework than *1b1*.
- The dimensions proposed by the Expert Group report appear to be problematic for groups 2a and 2b, as they lack positive and significant correlations. This is especially the case with respect to indicators *2b1* and *2b3*. Instead, we observe that indicator *2b1* may be more associated with the *1a* and *1b* group. An exception is the pair of indicators *2a1* and *2a4*, on patent applications.

As a next step, we conducted PCA with various specifications in order to test the statistical coherence of the framework proposed by the Expert Group. The PCA has also allowed us to test alternative frameworks. In summary, three different frameworks have been tested:

I. The framework proposed by the Expert Group, using dimensions 1a, 1b, 2a and 2b;
II. A 2-dimensional framework distinguishing indicators associated with basic and applied research;
III. A modified 3-dimensional framework obtained by successive re-grouping of indicators, according to the results of the PCA.

Framework according to the Expert Group proposal (Framework 1)

The Framework proposed by the Expert Group considered 3x2 dimensions of research excellence:

1a. Engagement with research actors – knowledge production
1b. Engagement with research actors – institutional arrangements, interaction schemes
2a. Engagement with industrial innovation actors – knowledge production
2b. Engagement with industrial innovation actors – institutional arrangements, interaction schemes
3a. Engagement with societal and political actors – knowledge production
3b. Engagement with societal and political actors – institutional arrangements, interaction schemes

In fact, only 3 dimensions out of the 6 could be analysed in its merits. Data availability precluded us from considering the last two dimensions (3a and 3b) and, due to the correlation structure, dimension 2b in fact became a single-indicator pillar, consisting only of indicator *2b1*.

Dimension 1a

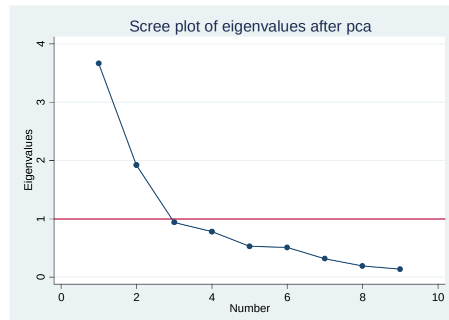
We first conducted a Principal Component Analysis (PCA) using all indicators proposed by the expert group for pillar 1a. (Note that in addition to the indicators deemed robust and reliable, for the sake of the test we also included other indicators with available data, namely *1a4*, *1a6*, *1a7* and *1a8*).

As shown by the PCA results in Table 4, pillar 1a is in fact composed of at least two latent dimensions (with eigenvalues greater than 1). The first one shows positive significant correlations with indicators *1a1, 1a2, 1a3, 1a4, 1a5, 1a6* and *1a9*; and another one correlating with *1a10*, but also with *1a6*, *1a1* and *1a5*. Negative correlations were also observed in case of both components. Indicator *1a7* (as expected) does not correlate with any of these two components, and calls for a separate dimension on its own.

Table 4 Results of the Principal Component Analysis: pillar 1a of the original framework

Component	Eigenvalue	Diff.	Prop.	Cumm.
Comp1	3.66	1.74	40.7%	40.7%
Comp2	1.92	0.98	21.4%	62.1%
Comp3	0.94	0.16	10.5%	72.5%
Comp4	0.78	0.25	8.7%	81.2%
Comp5	0.53	0.02	5.9%	87.1%
Comp6	0.51	0.20	5.7%	92.8%
Comp7	0.32	0.12	3.5%	96.3%
Comp8	0.19	0.06	2.2%	98.5%
Comp9	0.14		1.5%	100.0%

Nr. Obs. = 62



Variable	Comp1	Comp2	Comp3
_1a1	0.294	0.380	0.384
_1a2	0.451	-0.028	-0.207
_1a3	0.253	-0.229	-0.550
_1a4	0.413	0.080	0.182
_1a5	0.327	0.327	0.135
_1a6	0.216	0.530	-0.125
_1a7	-0.337	0.009	0.504
_1a9	0.329	-0.445	0.319
_1a10	-0.318	0.458	-0.290

In an attempt to adjust pillar 1a in order to obtain greater statistical coherence (only 1 latent component with an eigenvalue above 1) but keep as many of the proposed indicators as possible, we found that this pillar holds if it only consists of indicators *1a1*, *1a2*, *1a4*, *1a5* and *1a9* (see detailed results in Table 5). We take note that this specification only explains 57% of the variance of 5 indicators, just falling short of the 60% benchmark; to correct for that, indicator *1a9* could be removed and pillar 1a would have to be limited to 4 indicators: *1a1*, *1a2*, *1a4* and *1a5*, in which way the variance explained increases to 64.5%. Nevertheless, in this framework our main goal was to use as many as possible of the indicators proposed by the expert group, also allowing for a degree of flexibility.

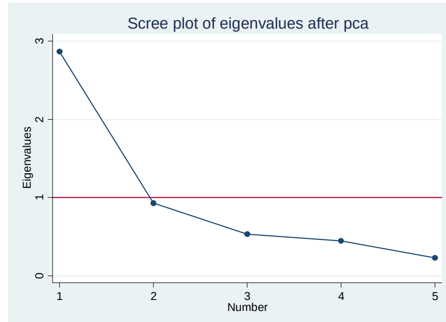
We also note the surprizing behaviour of the indicator 1a3 (ARC), which would have been expected, by definition, to best describe research excellence. However, in the multivariate context proposed by the Expert Group, we found that 1a3 has no significant loading in any of the first three components. For this reason, we did not consider it for the aggregation in this framework.

Table 5 Results of the Principal Component Analysis: pillar 1a of the adjusted framework

Component	Eigenvalue	Diff.	Prop.	Cumulative
Comp1	2.86	1.93	57.3%	57.3%
Comp2	0.93	0.40	18.6%	75.9%
Comp3	0.53	0.08	10.6%	86.5%
Comp4	0.45	0.22	8.9%	95.4%
Comp5	0.23		4.6%	100.0%

Nr. Obs. = 66

Variable	Comp1	Comp2
_1a1	0.422	-0.500
_1a2	0.500	0.245
_1a4	0.498	0.062
_1a5	0.435	-0.453
_1a9	0.367	0.693



Conclusion 1:
The principal component analysis carried out on the 41-country, 2-year research excellence dataset indicates that pillar 1a would obtain greater statistical coherence if it were narrowed down from nine to include the following five indicators: 1a1, 1a2, 1a4, 1a5 and 1a9. Indicators 1a7 and 1a10 were found not to be measuring the same latent phenomena as the rest of the indicators.

Dimension 1b

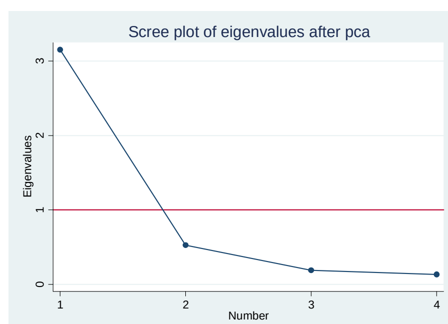
The PCA carried out using all available indicators of pillar 1b revealed the existence of only a single latent dimension, as long as indicator 1b1_collind is used in place of 1b1. The results, shown in Table 6, indicate a strong framework in the case of this dimension, with 1 clear latent component which explains a high proportion of variance (79%). However, if international collaborations are measured not by the scale-adjusted collaboration index (*1b1_collind*) but by the share of international collaborations within total publications (*1b1*), the framework becomes weaker and reveals the existence of another latent dimension.

Table 6 Results of the Principal Component Analysis: pillar 1b, original framework

Component	Eigenvalue	Difference	Proportion	Cumulative
Comp1	3.15	2.63	78.8%	78.8%
Comp2	0.52	0.34	13.1%	91.9%
Comp3	0.19	0.05	4.7%	96.6%
Comp4	0.13		3.4%	100.0%

Nr. Obs. = 48

Variable	Comp1	Comp2
_1b1_collind	0.523	0.221
_1b2	0.464	0.735
_1b5	0.507	-0.411
_1b7	0.504	-0.491



Conclusion 2:
The principal component analysis carried out on the 41-country, 2-year research excellence dataset indicates that dimension 1b built with four indicators is a statistically coherent pillar, as long as indicator 1b1_collind replaces indicator 1b1.

Dimension 2a

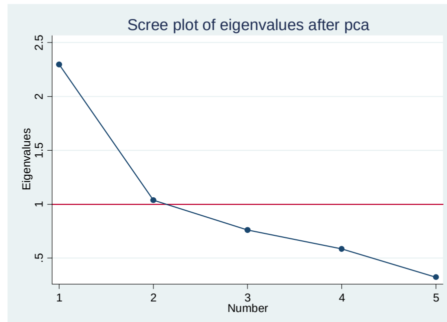
The PCA results for the indicators in group 2A show the presence of at least two latent components (Table 7). Indicators 2a1, 2a4 and 2a8 are correlating with component 1, and 2a7 associates with component 2. Indicator 2a3 (as already seen in the correlation matrix in Table 3), does not fit with any of these two components and requires a third component. It is therefore necessary to reduce the framework to a narrower set of indicators.

Table 7 Results of the Principal Component Analysis: pillar 2a, original framework

Component	Eigenvalue	Diff.	Prop.	Cumul.
Comp1	2.30	1.26	45.9%	45.9%
Comp2	1.04	0.27	20.7%	66.6%
Comp3	0.76	0.18	15.2%	81.9%
Comp4	0.58	0.26	11.7%	93.5%
Comp5	0.32		6.5%	100.0%

Nr.Obs.=32

Variable	Comp1	Comp2	Comp3
_2a1	0.523	0.214	0.444
_2a3	-0.337	-0.607	0.572
_2a4	0.556	0.112	-0.022
_2a7	-0.324	0.663	0.586
_2a8	0.447	-0.366	0.363



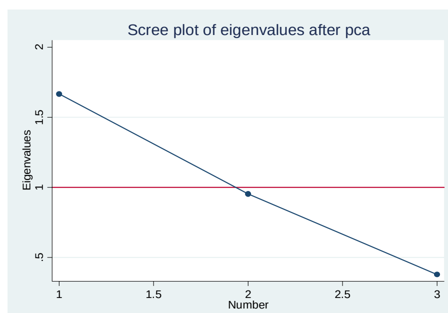
In a reduced specification (Table 8), we excluded the GSC patent specialization indicator 2a7 (second component) and the innovation survey-based indicator 2a3 (third component). For the three remaining indicators (*2a1*, *2a4* and *2a8*), the PCA showed a more coherent framework, with only one component explaining 56% variance. Adding the second component (2a7) could increase variance to 87%. However, this additional component has significantly low correlation with the indicators of the first component.

Table 8 Results of the Principal Component Analysis: pillar 2a, adjusted framework

Component	Eigenvalue	Diff.	Prop.	Cumul.
Comp1	1.67	0.71	55.5%	55.5%
Comp2	0.95	0.57	31.8%	87.3%
Comp3	0.38		12.7%	100.0%

Nr. Obs. = 57

Variable	Comp1	Comp2
_2a1	0.693	-0.059
_2a4	0.635	-0.421
_2a8	0.341	0.905



Conclusion 3:
The principal component analysis carried out on the 41-country, 2-year research excellence dataset indicates that dimension 2a needs to be reduced to three indicators, 2a1, 2a4 and 2a8 for a statistically coherent pillar of a composite indicator.

Dimension 2b

Given that data was only available for 2 indicators of the 2b group (2b1 and 2b3), and these two indicators were negatively correlating with each other (the coefficient was -0.36), a PCA could not be applied in this case. It is therefore a matter of arbitrary choice between the two indicators, which one to select as the single indicator representing the institutional arrangements and interaction schemes related to applied research. We decided to take only indicator 2b1 (public-private co-publications) which correlates positively with most of the indicators of the other pillars; selecting the indicator 2b3 would have shown negative correlation with the other pillars, hence the existence of a trade-off between pillars, a not suggested practice.

Conclusion 4:
Due to the negative correlation between the only two indicators available for dimension 2b, this pillar could only be represented by one indicator. Only indicator *2b1* correlates positively with the indicators in the other components.

A 2-pillar framework: basic research vs. applied research (Framework 2)

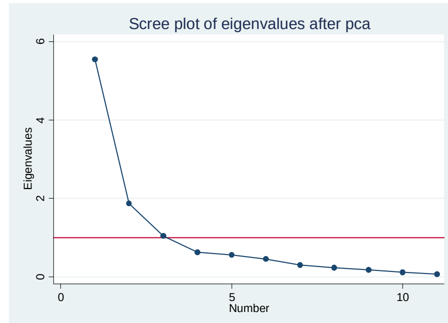
The conceptual framework proposed by the Expert Group is built on 6 pillars (1a, 1b, 2a, 2b, 3a, 3b) and, as the multivariate statistical analysis has shown, the first four pillars can well be represented by 13 indicators. We now try to simplify it into a 3-dimensional structure composed by Basic Research, Applied research and Societal relevance, associated with indicators 1a and 1b, 2a and 2b, and 3a and 3b, respectively.
We decided to test how the data structure could support a distinction between basic research and applied research into two pillars. There was no available data on indicators of societal relevance.

Basic Research (1a and 1b combined)

Seeking statistical support for combining all ‘basic research’ indicators in one pillar, we carried out a principal component analysis on all 1A and 1B indicators (with the exception of 1a7 and 1a10 that were negatively correlated with the rest). As could be expected from the similar analysis done for the Expert Group framework, basic research was found to include in fact three components (Table 9).

Table 9 PCA Results, 'Basic Research' (1A+1B) Indicators

Component	Eigenvalues	Diff.	Prop.	Cumulative
Comp1	5.54	3.67	50.4%	50.4%
Comp2	1.87	0.83	17.0%	67.4%
Comp3	1.04	0.41	9.5%	76.9%
Comp4	0.63	0.07	5.7%	82.6%
Comp5	0.56	0.10	5.1%	87.7%
Comp6	0.45	0.15	4.1%	91.8%
Comp7	0.30	0.07	2.8%	94.5%
Comp8	0.24	0.06	2.1%	96.7%
Comp9	0.18	0.06	1.6%	98.3%
Comp10	0.12	0.05	1.1%	99.4%
Comp11	0.07		0.6%	100.0%
Nr.Obs.=44				



Variable	Comp1	Comp2	Comp3
_1a1	0.175	0.502	0.398
_1a2	0.364	-0.158	0.076
_1a3	0.128	-0.223	0.787
_1a4	0.331	-0.253	-0.002
_1a5	0.286	0.307	0.012
_1a6	0.230	0.502	0.083
_1a9	0.246	-0.434	0.205
_1b1_collind	0.369	0.030	-0.295
_1b2	0.368	0.185	-0.077
_1b5	0.348	-0.177	-0.218
_1b7	0.352	-0.090	-0.163

A first group of indicators (1a2, 1a4, 1a9, 1b1_collind, 1b2, 1b5 and 1b7) correlated with the first component that explained 50% of data variance. These indicators measure many facets of excellence in basic research, from the production of highly cited publications, to the existence of top research universities, to specialization in grand societal challenges, to collaboration, to the attraction of foreign doctoral candidates, to coordinating research projects. In sum, the group combines both the production and the institutional dimension.

However, additional dimensions feature on their own, which are more difficult to explain conceptually. The second latent component is correlated with indicators 1a1, 1a5 and 1a6, explaining 17% of variance, while indicator 1a3 represents a third dimension on its own, adding a further 10% of variance explained (resulting in 77% all together). The indicators measuring researcher productivity (1a1) and specialization in attracting ERC grants (1a5 and 1a6) should, in principle, be associated with the first component that captured many dimensions of excellence.

In a validation round, PCA conducted on indicators identified by the first component (1a2, 1a4, 1a9, 1b1_collind, 1b2, 1b5, 1b7) confirmed the presence of a single latent dimension, explaining 70% of total variance, with relatively balanced loadings observed for the indicators. Similarly, PCA was conducted on indicators 1a1, 1a5 and 1a6, confirming the existence of a single latent component, accounting for 66% of variance with a well-balanced loading structure.[4]

[4] Alternatively, indicator 1a1 could join the first group, but as a result, the variance explained by the single component decreases to 63% and 1a1's 0.19 loading is much lower than the loading of the others. In this case, the second component would be composed of only 2 indicators, and the explained variance would increase to 73%.

In sum, we find that a ‘basic research’ pillar requires at least three sub-dimensions. In the absence of a theoretical underpinning for such sub-dimensions, it is not advisable to aggregate the three “sub-pillars” into a ‘basic research’ pillar.

Conclusion 5:
The multivariate analysis showed that there was no single ‘basic research’ pillar, but three distinct, albeit conceptually heterogeneous dimensions.

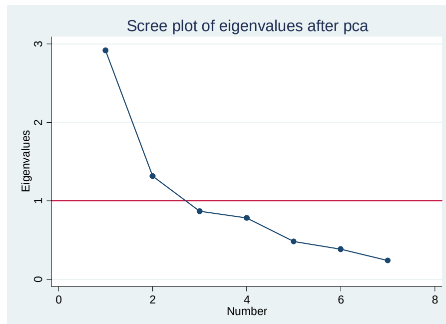
Applied Research (2a and 2b combined)

In a similar fashion, a PCA was conducted for all the indicators of the 2a and 2b group. The results, shown in Table 10, display 2 latent components.

Table 10 PCA Results, ‘Applied Research’ (2a+2b) Indicators

Component	Eigenvalues	Diff.	Prop.	Cumm.
Comp1	2.92	1.60	41.7%	41.7%
Comp2	1.32	0.45	18.8%	60.5%
Comp3	0.87	0.09	12.4%	72.9%
Comp4	0.78	0.30	11.2%	84.1%
Comp5	0.48	0.10	6.9%	91.0%
Comp6	0.39	0.14	5.5%	96.5%
Comp7	0.24		3.5%	100.0%
Nr.Obs.=32				

Variable	Comp1	Comp2	Comp3
_2a1	0.424	0.049	0.449
_2a3	-0.295	0.503	-0.346
_2a4	0.471	-0.003	0.360
_2a7	-0.246	-0.540	-0.042
_2a8	0.426	0.214	-0.482
_2b1	0.462	0.116	-0.352
_2b3	-0.237	0.628	0.437



In order to keep just one latent component, we removed indicators 2a3, 2a7 and 2b3 (as they are anti-correlated with the rest) from the framework. Repeating the PCA on the remaining indicators showed one principal component with an eigenvalue greater than 1, however, the total variance explained was 54%. We then removed indicator 2a8 (the indicator with the lowest loading value of 0.34 vs. 0.50-0.58 found for the others), and obtained a final robust framework with 3 variables: 2a1, 2a4 and 2b1. The variance explained thus increased to 66%, and the loadings for the indicators were rather balanced (0.53-0.63).
Conceptually, the pillar captures both the process and the results of applied research: both public-private co-publications as well as patenting by public and private sectors.

Conclusion 6:
According to the multivariate analysis, a pillar of ‘applied research’ can be statistically supported as long as it consists of three indicators: 2a1, 2a4 and 2b1.

Conclusion 7:
The conceptual framework is weak because the “Applied research” pillar has been placed side by side to three “Basic research pillars”, for which there is no clear theoretical underpinning

A three-dimensional framework (Framework 3)

In Framework 1 the latent components explain relatively low levels of variance in the data, which can only be improved by removing a considerable amount of underlying indicators. Framework 2 is hard to justify conceptually. With a view to achieving a framework which is both statistically sound and rich in indicators, we decided to analyse by principal components the 22 indicators starting from the original structure (Framework 1) and testing alternative specifications.

In this framework we considered a modified indicator (2b3_gdp) to measure financial flows from business to public research. In 2b3_gdp the amount of R&D financed by business and performed by PROs and HEIs is divided by GDP (at PPPs) instead of the total amount of business R&D. This new definition is the same as the one used in the European Commission/DG-RTD’s Innovation Union Competitiveness Report 2011 (see Fig.II.2.2, p.I-203). While indicator 2b3 could not be used for composite indicators as it would have only added noise without affecting the final scores, the modified indicator 2b3_gdp could fit well in the framework. The updated correlation matrix (Table 11) confirms that the change in the denominator matters: while the GDP-normalized indicator 2b3_gdp is completely unrelated to the “old” BERD-normalized indicator 2b3, it is correlated with key indicators such as 1a2 (highly cited publications’ share) or, as expected, 2b1 (public-private co-publications).

Table 11 Correlation Matrix for Updated Indicator 2b3

	1a1	1a2	1a3	1a4	1a5	1a6	1a7	1a9	1a10	1b1	1b1_collin	1b2	1b5	1b7	2a1	2a3	2a4	2a7	2a8	2b1	2b3	2b3_gdp
2b3_gdp	0.18	0.55	0.32	0.38	0.31	0.08	-0.23	0.50	-0.32	0.24	0.40	0.53	0.40	0.37	0.47	0.11	0.38	0.05	0.25	0.67	0.00	1.00

What follows below is the discussion of the results of an iterative process in which we identified an optimal alternative framework (Framework 3) for a composite indicator to measure research excellence. The three pillars that were identified are described below, in terms of the variables included together with a possible explanation of their meaning.

Pillar 1 “excellence of public research”

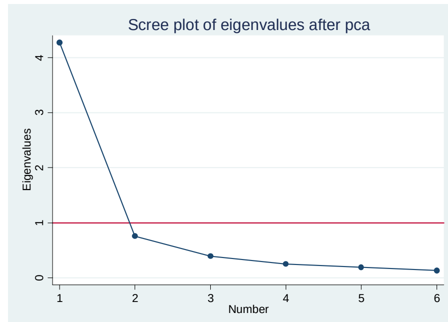
The pillar measures the excellence of research actors. It is a multi-faceted pillar which was created by taking the ‘core’ indicators of the 1a group as identified by the Expert Group and adding correlating indicators from the other groups, while keeping conceptual integrity in mind. The resulting pillar combines highly cited publications (1a2); excellence in research carried out at top universities and PROs (1a4); the excellence of research system in terms of its the ability to attract foreign doctoral students (1b5); specialization in publications in the fields of grand societal challenges (1a9); coordination position in FP projects (1b7) and public applied research excellence measured through PCT patent applications (2a1). While it is a heterogeneous pillar, a common feature in the indicators used is that they measure excellence in research mostly carried out by the public sector. Table 12 presents the results of the PCA.

Conclusion 8:
The framework well accommodates the 6 variables, and the pillar captures one single latent dimension, which explains 67% of variance in the data. The factor loadings are well balanced.

Table 12 PCA results, Pillar 1 of alternative framework

Component	Eigenvalues	Diff.	Prop.	Cumul.
Comp1	4.02	3.29	67.0%	67.0%
Comp2	0.72	0.09	12.1%	79.0%
Comp3	0.63	0.33	10.5%	89.6%
Comp4	0.30	0.10	5.0%	94.5%
Comp5	0.20	0.07	3.3%	97.8%
Comp6	0.13		2.2%	100.0%
Nr. Obs. = 35				

Variable	Comp1	Comp2
_1a2	0.437	-0.167
_1a4	0.432	-0.367
_1a9	0.380	-0.611
_1b5	0.417	0.503
_1b7	0.411	0.343
_2a1	0.369	0.305



Pillar 2 "interactions, collaborations"

The second pillar was constructed around variables that measure interactions and collaborations, and were not included in the first pillar. Interactions between researchers are captured in two forms: directly, through international co-publications (1b1_collind and 1b2), and indirectly, through successfully receiving ERC grants (1a5 and 1a6). Co-publications refer to collaboration in general as well as with non-European co-authors. In our interpretation, ERC grants indicators very much refer to interaction between supranational government and research actors, as well as between research partners across countries. We noted that a few other indicators of collaboration could not fit in this pillar, but are better associated with other pillars (i.e., coordination position in FP projects, indicator 1b7, is associated to pillar 1; public-private co-publications, indicator 2b1, is associated with pillar 3). The results of the validating PCA for pillar 2 are shown in Table 13.

Conclusion 9:
PCA confirms the presence of a single principal component, capturing 63% of variance in the data, with relatively balanced component loadings, although the loading value for 1a6 is weaker than for the other three.

Table 13 PCA results, Pillar 2 of alternative framework

Component	Eigenvalues	Diff.	Prop.	Cumul.
Comp1	2.52	1.77	63.1%	63.1%
Comp2	0.76	0.24	18.9%	82.1%
Comp3	0.52	0.33	13.0%	95.1%
Comp4	0.20		4.9%	100.0%
Nr.Obs. = 62				

Variable	Comp1	Comp2	Comp3	Comp4

_1a5	0.492	0.222	-0.810	-0.229
_1a6	0.413	0.791	0.416	0.176
_1b1	0.539	-0.355	0.412	-0.643
_1b2	0.545	-0.447	0.008	0.709

Pillar 3 "Excellence in industrial research"

The third pillar combines variables on industrial contributions to research excellence from three aspects: EPO patent applications by industry (2a4), co-publications between industry and academia (2b1) and business financing research carried out at institutes of higher education and public research organizations (2b4_gdp).

The PCA confirmed that financial flows could be captured in the industrial research excellence pillar as long as the indicator 2b4 is normalized by GDP as opposed to BERD.

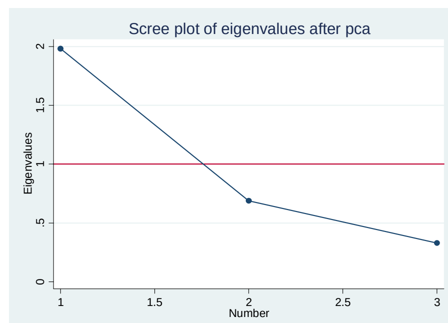
Conclusion 10:
The pillar was found to be statistically sound with the three indicators; the single latent component explains 66% of variance in the data and factor loadings are relatively balanced (Table 14).

Table 14 PCA results, Pillar 3 of alternative framework

Component	Eigenvalue	Diff.	Prop.	Cumul.
Comp1	1.98	1.29	66.1%	66.1%
Comp2	0.69	0.36	22.9%	89.0%
Comp3	0.33		11.0%	100.0%

Nr. Obs. = 66

Variable	Comp1	Comp2
_2a4	0.493	0.866
_2b1	0.623	-0.284
_2b3_gdp	0.607	-0.412



Summary of Multivariate analysis

The multivariate analysis of the data led us to conclude that the framework proposed by the expert group based on theoretical considerations only partially holds in light of the empirical data collected for 41 countries. With some modifications of the original framework we could achieve three statistically coherent frameworks.

The first one was based on the expert group's proposal, but excluded indicators that do not correlate positively with the others. The second one distinguished basic research and applied research (still in principle based on the export group report). The third framework was influenced by the original structure, but reiterated that in a way to optimize statistical coherence. The allocation of indicators within the three proposed frameworks is presented in Table 15.

With Framework 3, we found that a few indicators could be interpreted in different ways and thus could be associated with pillars of seemingly different meaning. For instance, the share of foreigners in doctoral programmes can at the same time be an indicator of cross-border interaction as well as an indicator of the presence of excellent research actors and infrastructures. This, in our reading, offered a degree of flexibility in the construction of pillars of the research excellence composite indicator.

Table 15 Overview of the three Research Excellence frameworks considered

Framework:	I. Expert Group proposal				II. Basic vs. Applied Research				III. Alternative		
Indicator (Description)*	EG-1A	EG-1B	EG-2A	EG-2B	BR-P1	BR-P2	BR-P3	AR	P1	P2	P3
_1a1 (Publications)	x					x			–	–	–
_1a2 (% highly-cited p.)	x				x				x		
_1a3 (ARC)	–	–	–	–			x		–	–	–
_1a4 (top universities)	x				x				x		
_1a5ʷ (ERC grants)	x					x				x	
_1a6ʷ (ERC starting gr.)	–	–	–	–		x				x	
_1a7 (% women res.)	–	–	–	–	–	–	–	–	–	–	–
_1a9 (GSC publ. spec.)	x	–	–	–	x				x		
_1a10 (KET publ. spec.)	–	–	–	–	–	–	–	–	–	–	–
_1b1 (% int'l co-publ.)	–	–	–	–	–	–	–	–	–	–	–
_1b1_collind (coll. index)		x			x					x	
_1b2 (non-EU co-publ.)		x			x					x	
_1b5 (% foreigner PhDs)		x			x				x		
_1b7ʷ (FP coord. pos.)		x			x				x		
_2a1 (HEI&PRO patents)			x					x	x		
_2a3 (publ inn. source)	–	–	–		–	–	–	–	–	–	–
_2a4 (IND patent app)			x					x			x
_2a7 (GSC pat. spec)	–	–	–	–	–	–	–	–	–	–	–
_2a8 (KET pat. spec)			x		–	–	–	–	–	–	–
_2b1 (co-publ w industry)	–	–	–	X				x			x
_2b3 (BES funded publ R&D)	–	–	–	–	–	–	–	–	–	–	–
_2b3_gdp (2b3 norm'd by GDP)	–	–	–	–	–	–	–	–			x

Notes: * for a full description of indicators, please refer to the section on indicator definition above; indicator 1a8 was already excluded in the univariate analysis; A dash (–) indicates that the variable is available but not used in any of the pillars of the given frameworks; ᵂ indicates indicator not appropriate for comparison with world; indicator 2b3_gdp was only introduced for the alternative pillar.

We further see that four indicators (1a3 – average relative citations; 1a7 – women among researchers; 1a10 – specialization in KETs publications and 2b3 – HERD+GOVERD financed by business) do not fit in any of the former groups. Indicator 1b1 was also not used in order to avoid confusion with the scale-normalized '1b1_collind'.

Computation of composite indicators and results

Composite indicator according to the Expert Group Proposal

First, we computed the composite indicators using the structure proposed by the Export Group (and reduced to the 13 indicators for which data was available and which statistically made sense to aggregate). We used an equal weight geometric aggregation of the four pillars (1a, 1b, 2a and 2b), each of which was calculated as the arithmetic average of the normalized indicators it is composed of (as defined by Table 15). Taking the arithmetic mean within pillars is justified if the component loadings identified by the PCA are relatively balanced – which is the case (although the least balanced pillar was 2a). At the overall composite level, the advantage of the geometric average over the arithmetic mean is that countries cannot completely compensate a weaker performing pillar with an above-average performing one. This motivates them to increase performance in their weakest pillar. Note that the fourth pillar only consists of one single indicator (2b1), due to limited data availability and the correlation structure of the data. The standard min-max normalization we used to rescale each variable to the interval of 0 to 1 necessarily results in a value of 0 for the worst performing country in one year. For this country, the composite indicator, which is the geometric average of all pillars, will in turn be 0. This country was Bulgaria in 2005. In order to avoid having a score of 0, it could be possible to define the normalization interval in the range 0.1-0.9; the country rankings would minimally be affected in this case.

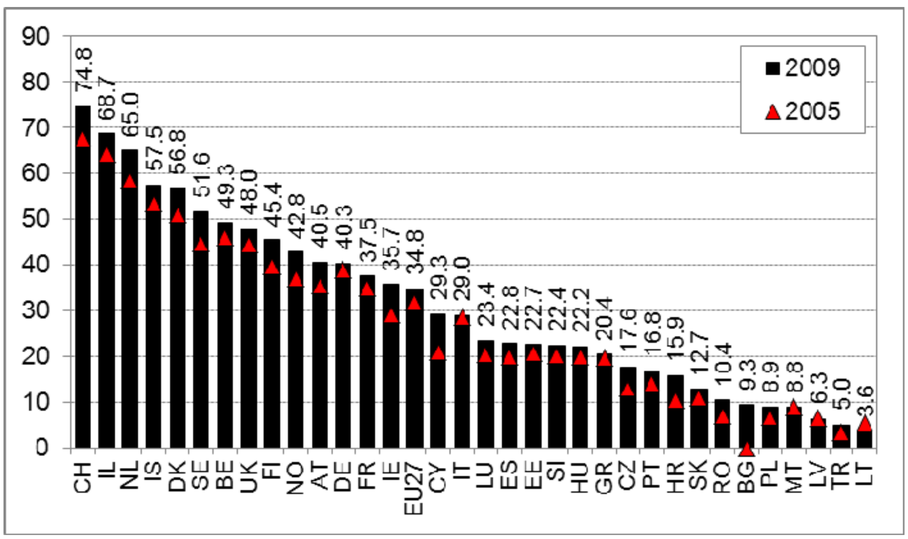


Figure 3 Research Excellence Composite Scores based on the Expert Group proposed framework: ERA countries (geometric aggregation, equal weights)

The composite scores according to the framework proposed by the Expert Group are presented in Figure 3. We notice that Switzerland, Israel and the Netherlands are placed in the top ranks, with scores above 65 on a scale of 0 to 100. The graph also shows a clear North-West versus South-East divide, with many of the Member States from the former Eastern bloc as well as Mediterranean countries (including Portugal, Greece, Spain and Italy) below the EU27 average, and centers of excellence in Nordic and North Sea countries. Of the EU members with the largest population count, we find the UK outperforming Germany, France and Italy.

The composite indicators were computed for two time points**. Time point 2009 in fact refers to the most recent year for which data was available (typically between 2007 and 2009, see**

Table 2 for the year actually used for the various indicators). The indicators were also computed for time point 2005 in order to present trends over time. This analysis based on the framework proposed by the Expert Group shows that, with the exception of Lithuania, Latvia and Malta, all countries have increased their performance in excellence over time. The most significant increases were registered in the cases of Turkey, Croatia, Romania, Cyprus, the Czech Republic and Poland. Positive, but slow changes were found for Italy, Germany and Greece.

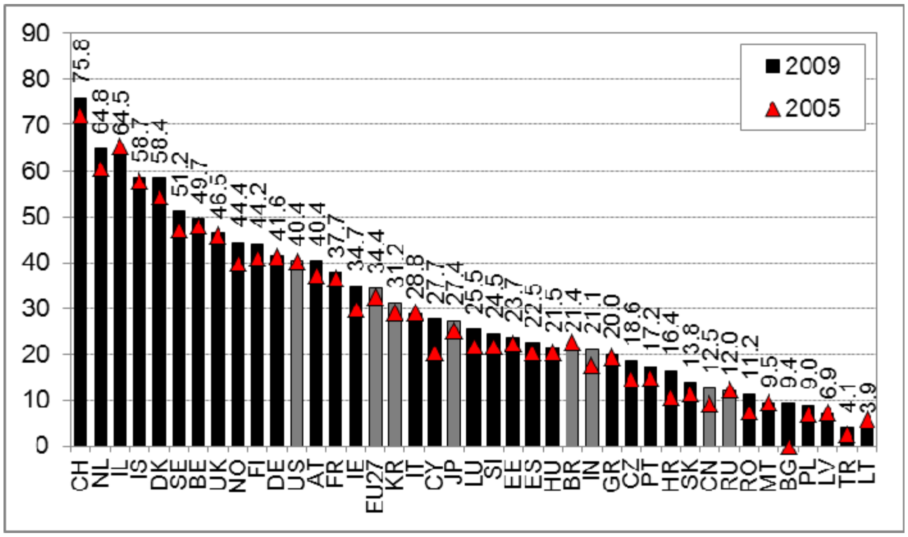


Figure 4 Research Excellence Composite Scores based on the Expert Group proposed framework: global comparison (geometric aggregation, equal weights)

As we noted earlier, certain indicators were found inappropriate for a **global comparison**. Therefore, in order to benchmark the countries against global competitors, indicators 1a5 and 1b7 were removed from the list. The results are presented in Figure 4. When comparing EU27 and individual countries with the US, we found (somewhat surprisingly) that the "American challenge" so often cited as a main driver of institutional reforms and increases in research efforts in Europe, is not apparent within these results. The US outperforms the EU27 average, however, it lags behind Switzerland, the Nordic and North Sea countries as well as Germany. This could be explained by the fact that many indicators of pillar 1b included in this framework measure international collaboration (1b1_collind, 1b2), in which indicators the US performs below-average given the fact that domestic collaborations are not captured. We also remind the reader of the towering US leadership in top scientific prizes (Nobel prizes and Fields medals), that were captured in indicator 1a8 but could not be implemented in the composite framework due to the lack of data variance.

We found Korea and Japan following closely the EU27, with similar scores as Italy. Two large emerging countries, Brazil and India are in the ballpark of Greece, Hungary or Spain, whereas China and Russia trail far behind, with scores similar to Romania and Slovakia.

If we plot the pillar scores together with the composite scores, we can better identify the main drivers behind countries' performance. This is shown in Figure 5 for year 2009 for both the ERA countries as well as for the global comparison.

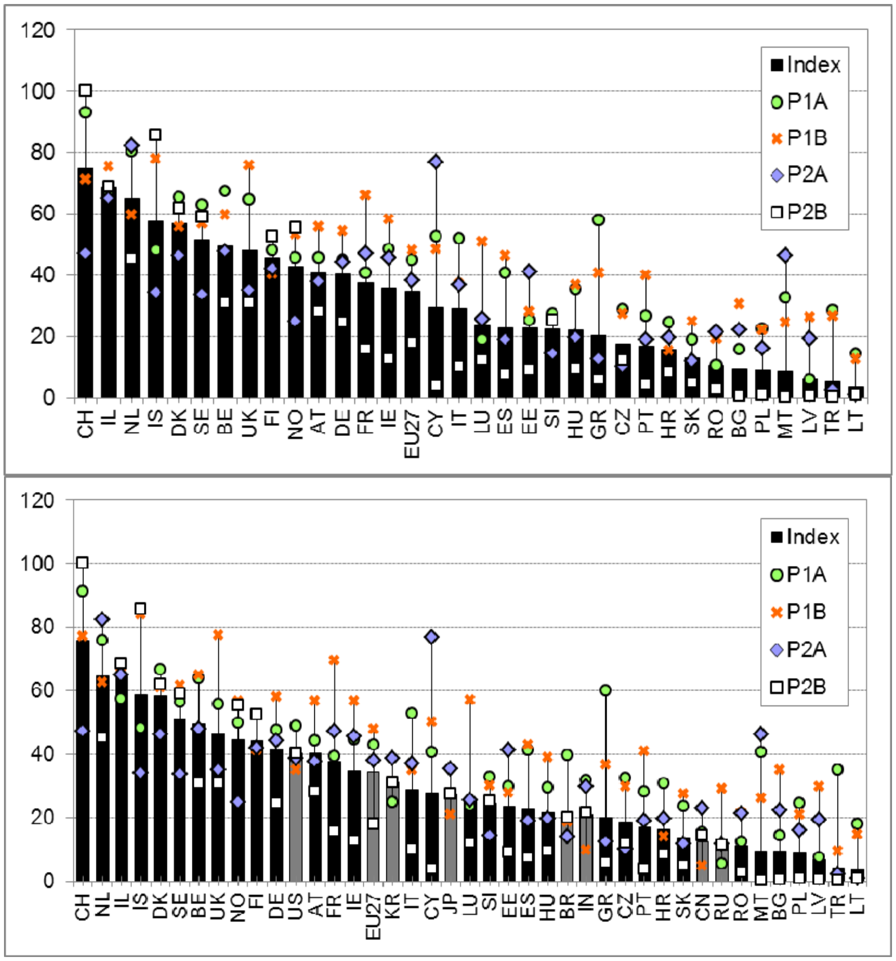


Figure 5 Research Excellence Composite and Pillar Scores, 2009, based on the Expert Group proposed framework (geometric aggregation, equal weights)

In general, we see a rather large difference between the various pillar scores of the same country. For instance, Switzerland, which is the leader based on the composite score, is a towering leader in terms of the 2B and 1A pillars, but is in the mid-ranks based on its 2A pillar score. For most EU member states except for the Nordic countries, we find that pillar 2B holds the scores back. A few countries have particularly large distribution of scores, such as Cyprus, Greece or Malta, while the scores of a few (Israel, Denmark, Finland or Slovenia) are less spread out.

This indicates that should any of the pillars receive more weight than the other, the ranking can change. So far, we have used equal weights. But if we look at the correlation results between the various pillar scores and the overall composite, we find rather different values: the coefficients for pillars 1A and 2B are close to 0.9, for pillar 1B 0.8, but for 2A only 0.66. By **readjusting the weights** with the aim to achieve similar correlation scores for all pillars, the ranking will indeed change considerably. The pillars 1A, 1B, 2A and 2B were given the following weights, respectively: 0.1, 0.22, 0.52, 0.16. As a result, country ranks that increased by at least 10 positions included that of Malta and Cyprus at both time points. Conversely, the following countries fell back at least 10 positions in the rank: the Czech Republic, Sweden, and Croatia in 2005 and Norway at both time points.

We conducted a global **sensitivity analysis** to appreciate the relative importance of the domains for the overall composite in terms of the so-called first order sensitivity indices (S_i).[5] The value of these indices may vary within the range of 0.0 to 1.0 and indicate how important each domain is in terms of driving variability on the overall composite indicator.

In our model we assumed that all indicators (and thus pillars) are of the same importance for the composite on research excellence. The sensitivity analysis confirmed that the equal weights specification is in fact resulting in an unbalanced structure, as Pillar 2A has a lower effective contribution to the overall composite, both in the ERA-specific and in the global benchmark frameworks (see black bars in Figure 6). With pillar weights adjusted as described above, the sensitivity analysis confirms a relatively balanced structure (see gray bars in Figure 6).

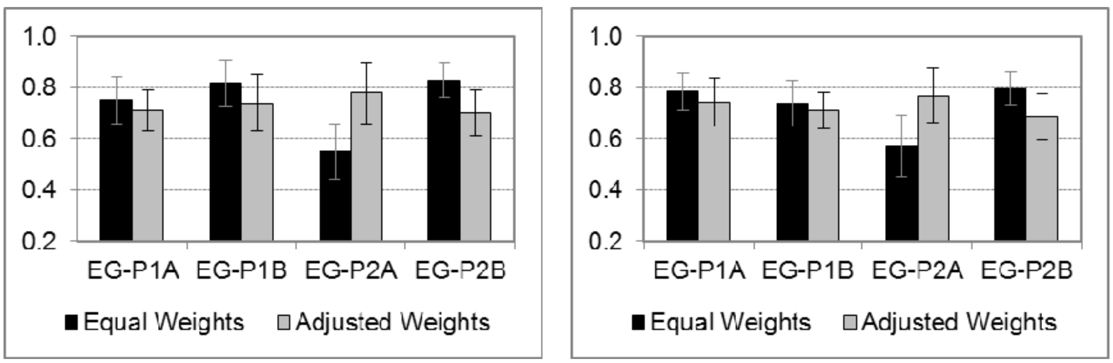


Figure 6 Sensitivity Analysis Results for Framework 1
(LEFT: ERA countries; RIGHT: Global benchmark)

Conclusion:

The composite indicator shows a clear North-West vs South-East divide with centers of excellence in the Nordic and North-Sea countries. Almost all countries have improved their excellence between 2005 and 2009. The considerable spread of scores at the pillar level provides useful insights to research excellence performance: high and low scores are detected for pillars 1b and 2b, respectively.

[5] The sensitivity indices of the first order $S_i=V[E(Y|X_i)]/V(Y)$ were computed from the $E(Y|X_i)$ curve, obtained by kernel regression of the original data points after Gasser et al. (1991). S_i has been computed by weighted averaging of the regression curve. Instead of a single estimate based on the 34 or 41 points available we have boot-strapped the points for the computation of $E(Y|X_i)$ using as many replicas as the sample size and computed S_i mean and standard deviation. The resulting average S_i can be taken as a robust measure of importance.

A two-dimensional framework (framework 2)

As we learned from the multivariate analysis (Section 4), the structure of the composite combines 3 basic and one applied research pillar. Therefore, here we aggregate the four pillars up to a composite indicator and analyze the resulting scores

We first construct the composite indicator from the min-max normalized data for the two time points (2005 and 2009), using equal weights, **for the ERA countries**. The results are shown in the upper part of Figure 7.

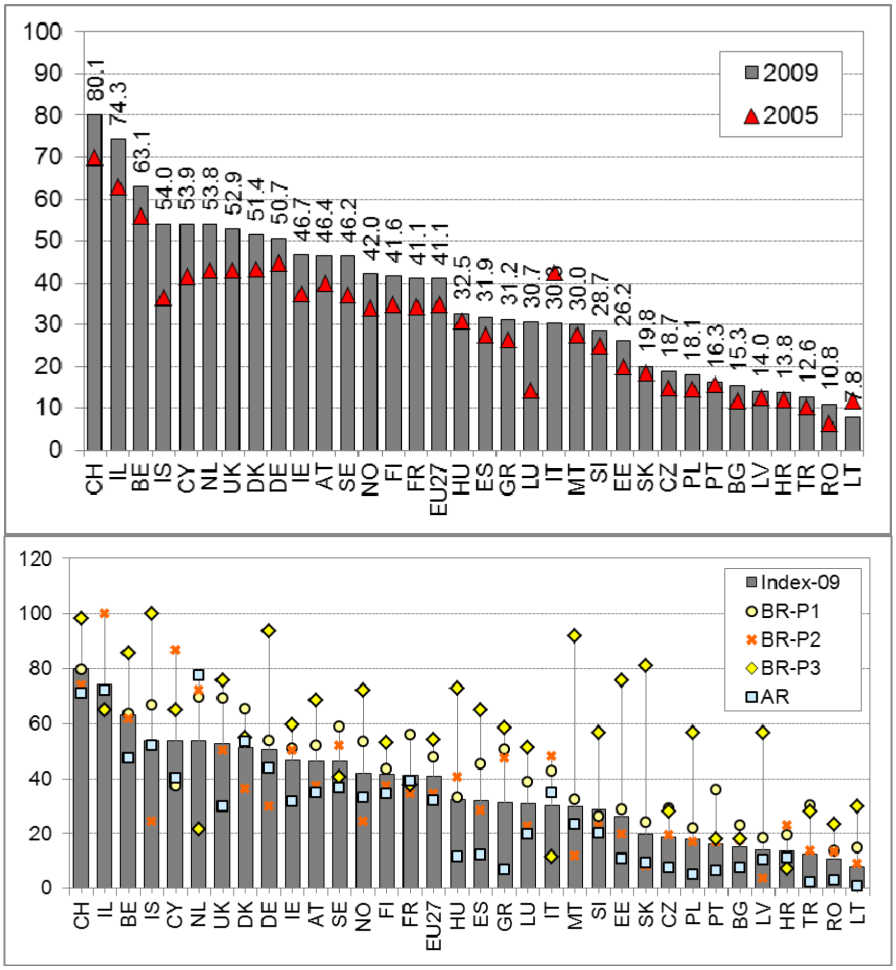


Figure 7 Research Excellence Composite and Pillar Scores based on the 'Basic vs. Applied research' framework, for ERA countries (geometric aggregation, equal weights)

Similar to the framework proposed by the Expert Group, Switzerland and Israel lead the ranks and Nordic and North Sea countries occupy the prominent positions. A big change with respect to the framework proposed by the Expert Group is the promotion of Cyprus to the 5th position compared to a below average position. The UK is still the leader among the larger Member States but, in this framework, Germany follows more closely and France is lagging further behind, at the level of the EU27 average. A considerable reshuffling of relative positions can also be observed in the lower ranks where, for instance Malta performs better in this setting, but Romania relatively worse.

Considering change over time, almost all of the countries have increased their composite score, with the clear exception of Italy and Lithuania. The main reason behind the decline of Italy's score is the decline in the Basic Research Pillar 3 (BR-P3), caused by the drop in the average of relative citations indicator 1a3, as well as a modest decline in the Applied Research (AR) pillar. The AR pillar and more specifically, business patenting indicator 2a4, are behind Lithuania's decline. The relatively biggest gains were found for Luxemburg, Iceland and Romania. In the case of Luxemburg, this is explained by the rapid growth in Basic Research Pillar 2 (BR-P2), which even offset some decline in pillars BR-P1 and BR-P3. The growth in the overall composite score of Iceland and Romania were driven by Basic Research Pillars 2 and 3.

A look at pillar level scores for year 2009 (lower part of Figure 7) helps us to understand country performance in more detail. Clearly, Basic Research Pillar 1 (BR-P1) and Pillar 3 scores (BR-P3 – in fact, indicator 1a3) pull scores up, while the Applied Research pillar (AR) works mostly as a drag (some noticeable exceptions include the Netherlands and Italy, with above-average Applied Research Scores). The second Basic Research pillar (BR-P2) scores are more distributed across countries.

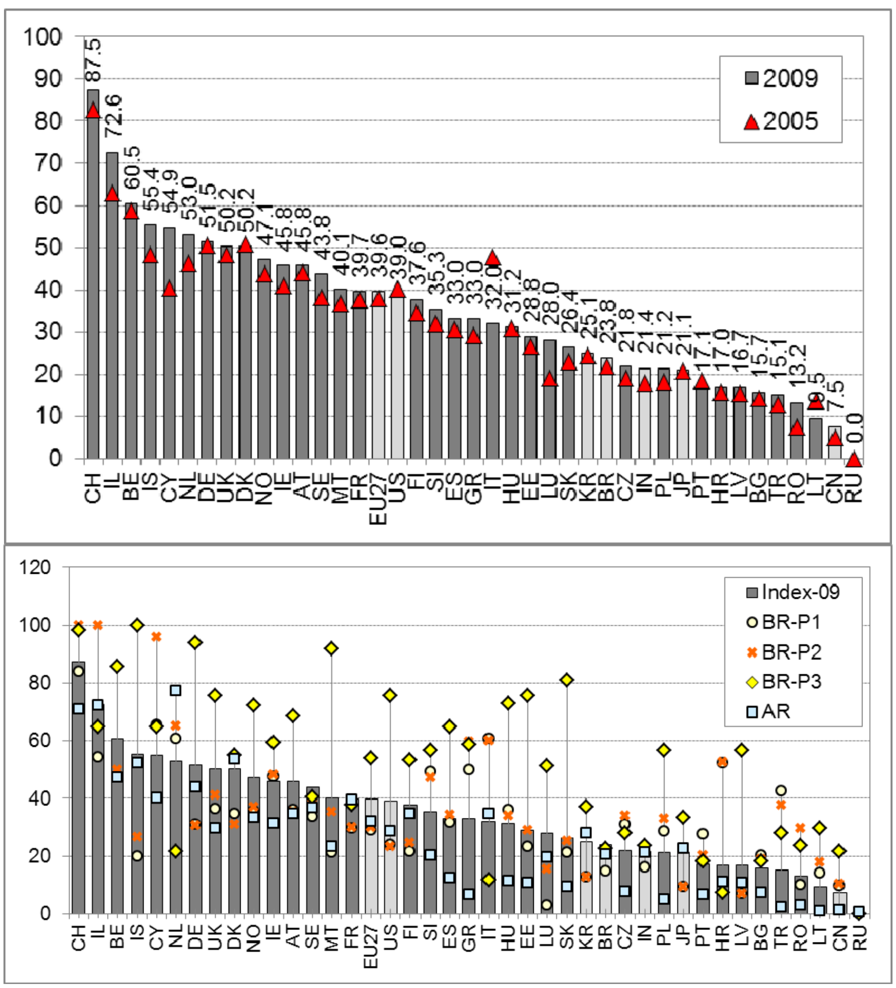


Figure 8 Global Benchmark of Research Excellence Composite Scores based on 'Basic vs. Applied research' framework (geometric aggregation, equal weights)

The **international comparison** composite scores are shown in the upper chart of Figure 8 for time points 2005 and 2009. This specification follows a slightly modified structure in which Basic Research Pillar 1 excludes indicator 1b7, and Basic Research Pillar 2 excludes indicators 1a5 and 1a6, as these indicators are only meaningful for ERA countries. According to this framework, in 2009, the EU27 and the US were neck and neck, followed in a distance by the cluster including South Korea, Brazil, India and Japan, and finally, China and Russia occupying the last two positions. Considering pillar scores, the US outperforms the EU in Basic Research P3 (in fact, indicator 1a3, average of relative citations), but lags behind in the other two basic research pillars and in applied research. China and other international competitors perform weaker in all dimensions, Russia scoring the lowest in all pillars.

In the composite indicator analyzed so far the four pillars were aggregated using equal weights (0.25 for each pillar). However, the four pillars were clearly not balanced: the correlation coefficients between the composite indicator and BR-P1 or AR pillars were 0.89, while 0.8 for BR-P2 and only 0.66 for BR-P3. This imbalance could be corrected by adjusting the weights for each pillar. Initial efforts to balance the structure suggested that nearly doubling the weight for BR-P3 (to 0.48) and slightly increasing for BR-P2 (to 0.32), while decreasing significantly for the BR-P1 and AR pillars (to 0.06 and 0.14, respectively), resulted in correlation coefficients close to 0.8 for all pillars. The improved distribution is shown in Figure 9 – for better readability, only time point 2009 is shown and countries are only labeled for BR-P3 pillar, which is the most affected by the weight adjustments. (Note that the weight adjustments were made by a rough iteration process and not a proper optimization process, but it does show the main directions of change.)

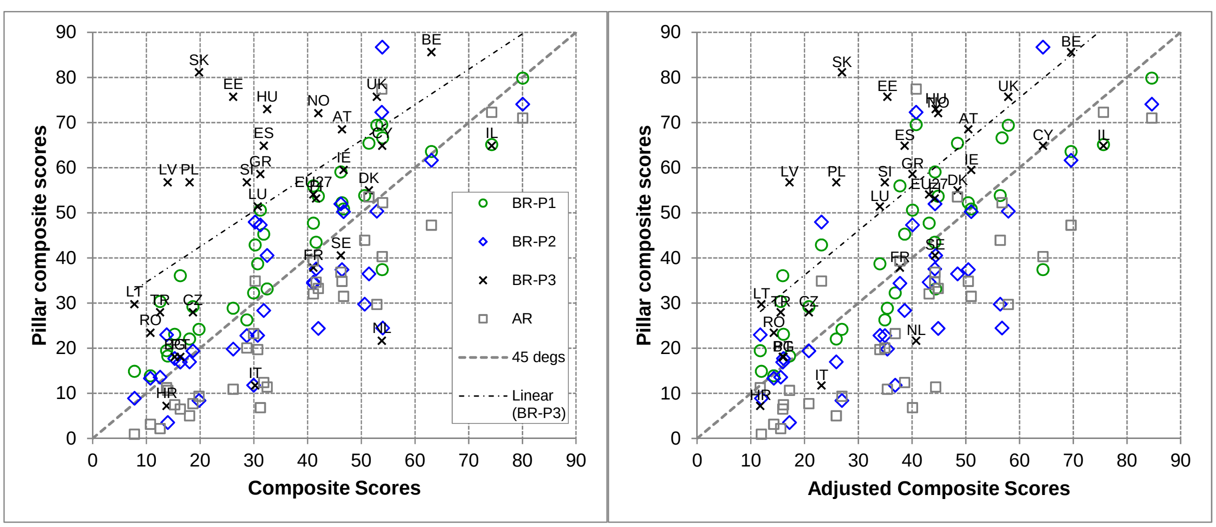


Figure 9 Plotting Pillar Scores against Composites Scores, 2009
(LEFT: Equal weights; RIGHT: adjusted framework)

Many countries changed their ranks by a few positions as a result of the weight adjustments (Figure 10). Amidst a general improvement of nominal scores, Hungary and Estonia were the biggest relative winners (by 5 and 3 positions respectively), as a result of giving less weight to applied research and more weight to the 3rd basic research pillar. Conversely, the Netherlands, Italy and France fell back in the ranks (by 10, 5 and 4 positions, respectively) due to the same forces. Similarly adjusting the framework for a global comparison, we found little rank changes for the main competitors; Europe and United States were still neck and neck, but with the US outranking EU27.

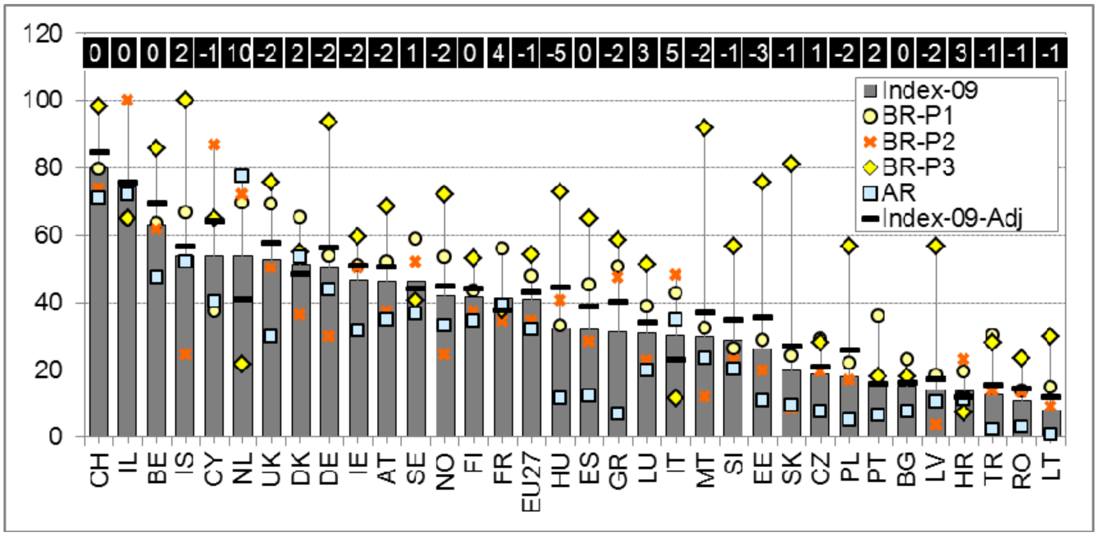


Figure 10 Score and rank changes after weight adjustment, 2009, basic v. applied research framework

Note: black marks indicate new composite index scores; numbers in black box show rank changes (a negative value refers to improvement).

A global **sensitivity analysis** shows that in fact the re-adjustment of weights is necessary in order to balance out the too high contribution of AR and BR- P1, but imbalances persist after weights are adjusted (Figure 11), because the weights adjustments were made by a trial-and-error iterative process and not through a proper optimization.

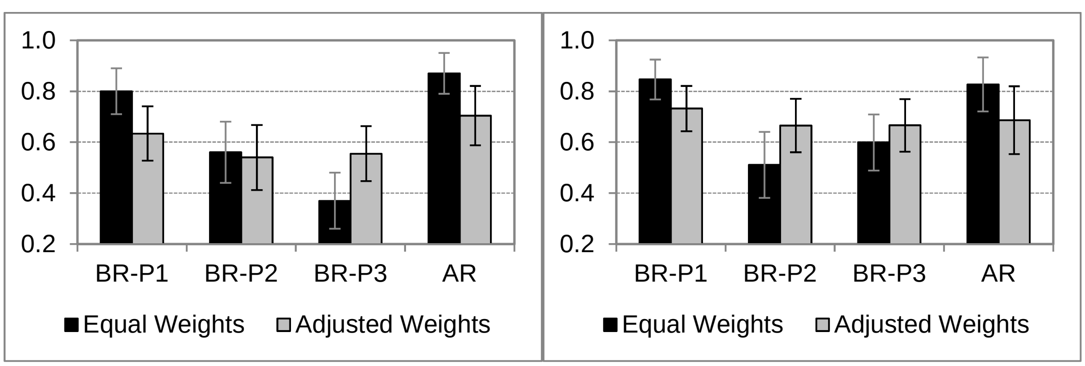


Figure 11 Sensitivity Analysis Results for Framework 2
(LEFT: ERA countries; RIGHT: Global benchmark)

Conclusion:
The resulting composite indicator is quite unbalanced in its pillars: the Basic Research (BR-P1) and the Applied Research (AR) pillars are responsible for disproportionately larger variance in data than the other two. Moreover, the framework is conceptually weak because of the 3 Basic research pillars which were warranted by the multivariate analysis. A re-adjustment of weights can improve the framework imbalances, but not the conceptual difficulties.

A three-dimensional framework (framework 3)

According to Framework 3, the research excellence composite indicator was computed as the geometric average of 3 pillars: “P1 - genuine excellence of research actors”, “P2 - interactions, collaborations” and “P3 - excellence in industrial research”. Each pillar is an arithmetic average of its indicators normalized between 0 and 100 using the min-max method and taking into consideration the two years simultaneously – as had been done for the previous two frameworks. A major difference with respect to the first two frameworks is the use of a modified indicator to measure financial flows from business to public research. Instead of dividing the amount of R&D [financed by business and performed by public research organizations and higher education institutes] by the total amount of business R&D as was done in the case of indicator 2b3, this time we used GDP (at purchasing power parities) in the denominator. While the “old” 2b3 could not be used for composite indicators as it would have only added noise but would not have affected the final scores, this “new” indicator could fit in the framework.

This new indicator *2b3_gdp* showed positive, significant correlation with the other indicators of research excellence, thus could be included in the framework. (For an overview of the 13 indicators used in this framework, see Table 15). Initially, only ERA countries were considered and all three pillars were aggregated with equal weights.

The resulting composite scores are shown for time points 2005 and 2009 in the upper panel of Figure 12. Three countries are clearly distinguished as the leaders in research excellence: Switzerland, Israel and the Netherlands, all with a score of 70 or above. They are followed by countries of North Western Europe with scores exceeding the EU27 score, and Southern and Eastern member states and Associate Countries below the EU27 score. Of the countries with the largest population, the United Kingdom and Germany are neck and neck, ahead of France and Italy. Three countries trail the list with single-digit scores: Romania, Lithuania and Turkey.

The lower panel of Figure 12 shows pillar scores against the composite scores, all for time point 2009. The general pattern is a relatively narrow variance in scores in the upper ranks, except for Belgium and the UK, with increasing variance in the lower-middle and lower ranks. Pillar 1 scores are, in most of the cases, greater than the composite (which is an equally weighted geometric average of all pillar scores) as well as Pillar 3 scores. This is especially true for the UK, Greece or Turkey. However, some countries stand out with an opposite trend, i.e. Germany, Finland, Cyprus or Latvia, where Pillar 1 is performing relatively worse than the other pillars. For Pillar 3, there is, conversely, a general pattern in which this pillar performs relatively worse than the other two. The UK, Ireland, Greece, Portugal or the Czech Republic are good examples for this, while the Netherlands, Germany, Finland or Slovenia go against the trend. In other words, two models are apparent: in some countries, research excellence is driven by research carried out in universities and public research organizations, while industry is the main driver of excellence in others. Compare the UK with Germany: both countries have an average score of around 49, but it is driven by Pillar 1 and 2 in the case of the UK, and by Pillar 3 in Germany.

Pillar 2 (collaborations) scores can in most cases improve the average composite scores, especially for smaller countries.

We also note the unique case of Turkey. The country shows average performance in Pillar 1, but extremely low score for Pillar 2. This is mostly the outcome of very low scores on the ERC-related indicators 1a5 and 1a6, and rather low performance in the other two co-publication indicators used in Pillar 2.

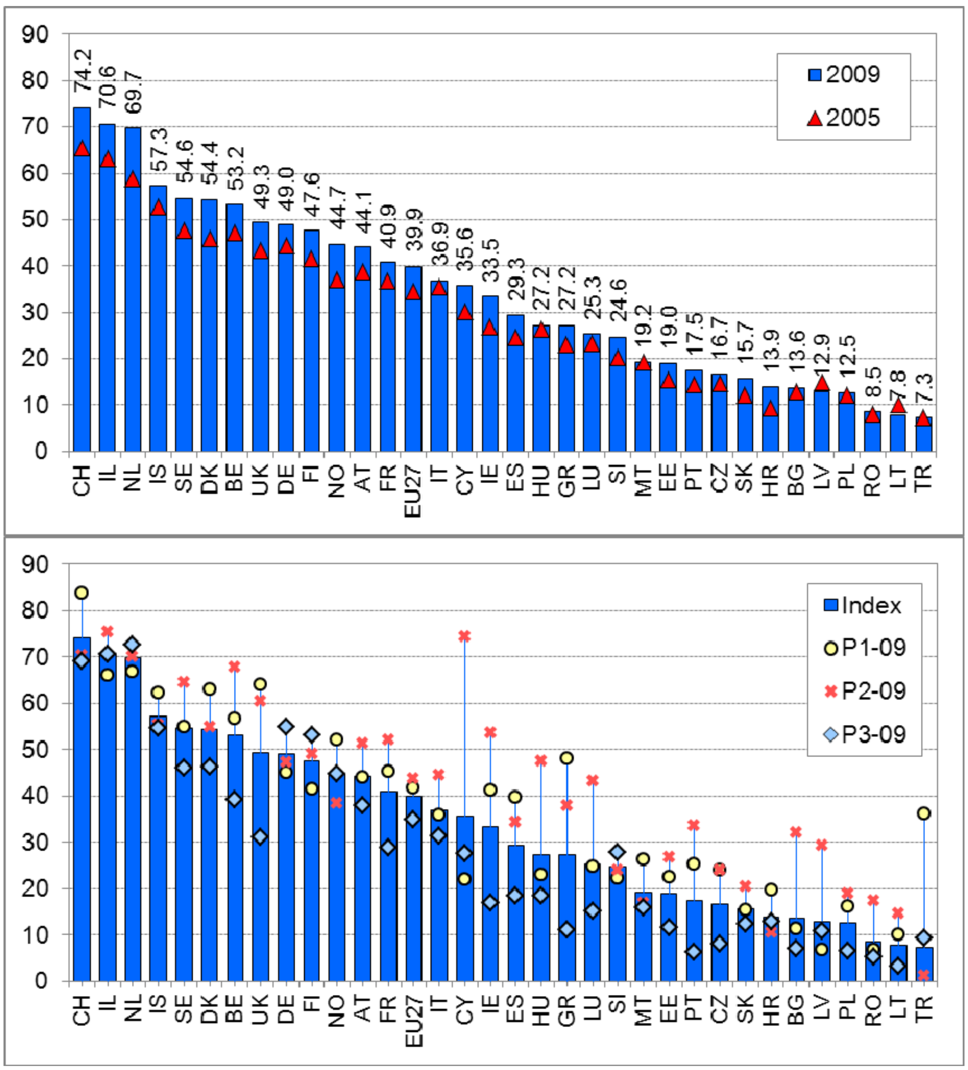


Figure 12 Research Excellence Composite Scores based on the Alternative Framework, for ERA Countries (geometric aggregation, equal weights)

Composite scores were also computed to enable comparison of ERA with non-ERA countries. As in the case of the other frameworks, the comparison excluded three variables that were not applicable or severely biased against non-ERA countries. Hence indicator 1b7 was excluded from Pillar 1, and 1a5 and 1a6 were excluded from Pillar 2 of the global benchmark indicator, while Pillar 3 remained unchanged. As a result, we obtained a slightly modified ranking due to the differences in pillar scores, but could include 7 additional countries in the list.

The composite scores for this international comparison are shown for both time points in the upper panel of Figure 13. Interestingly, we observe a clear leadership of EU-27 over not only BRIC countries, but also main international competitors, such as the US and Japan. Brazil and South Korea are also in the same ballpark as the latter two, with scores similar to Slovenia, Hungary or Greece. India, Russia and China are scored within the lowest quartile.

The lower panel of Figure 13 reveals which pillars were responsible for the low country scores. At the same time, the opposite is true for large countries, especially non-European ones. In the case of the US, we see that in P1 ("genuine excellence"), the US outperforms Europe and would rank within the top quartile. However, its low scores are due to its weaker performance in indicators of the P2 and P3 pillars. Even if excellent research output is produced through collaboration between [world leader] universities across the US, this does not increase P2 scores. The use of EPO instead of USPTO patents in the case of indicator 2a4 is unfavourable towards the US; normalizing the level of public R&D financed by business with GDP in the case of the modified indicator 2b3_gdp similarly pulls back the scores of P3, which cannot be compensated by public-private co-publications in 2b1 due to the absence of data for countries beyond Europe. To a lesser extent, the same patterns can be observed for Brazil, India and China as well. For these reasons, it may be most informative to revert to P1 only, instead of the full composite, when making global comparisons.

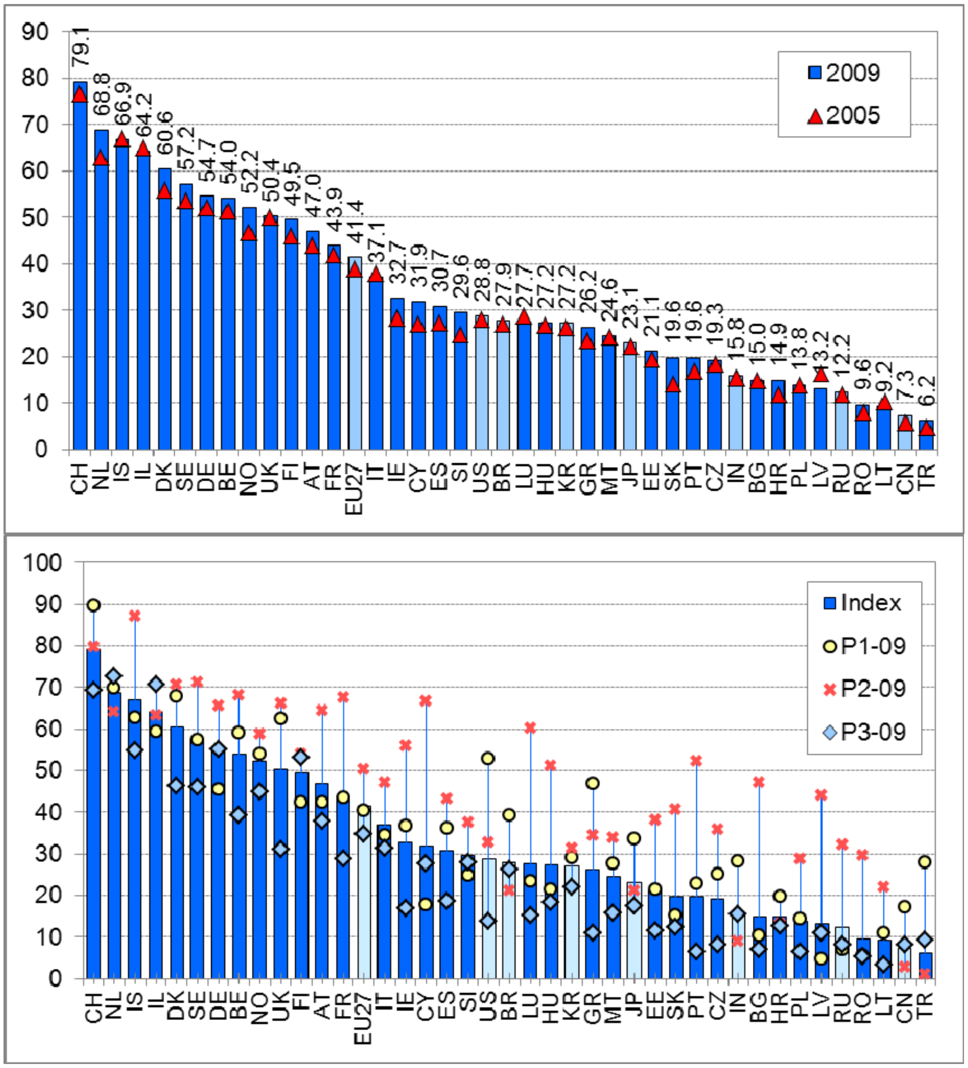


Figure 13 Research Excellence Composite Scores based on the Alternative Framework, for Global Benchmark (geometric aggregation, equal weights); 2005 and 2009 (UPPER); Pillar scores (LOWER)

Adjusting the weights

A global sensitivity analysis conducted on the pillars and composite scores revealed that the three pillars do not equally contribute to the final composite score, neither when it was computed for the ERA countries, nor for the global benchmark (black bars of Figure 14). Although all sensitivity scores were relatively strong (above 0.7, meaning a consistent contribution of the pillar to the composite), the variance of pillars 1 and 3 influences the variance of the composite scores more than that of pillar 2. In order to balance the sensitivity scores, we tested the effect of adjusted weights. This was done by increasing the weights for pillar 2, while decreasing them for the other two pillars. For the ERA countries, the adjusted weights were 0.22, 0.56 and 0.22; for the global benchmark 0.27, 0.49 and 0.24 for pillars 1, 2 and 3 respectively. (These should be seen as initial tests, not a full optimization exercise).

As the gray bars of Figure 14 indicate, the resulting sensitivity scores were more balanced. A full optimization of the weights shall be the object of the full-fledged composite indicator.

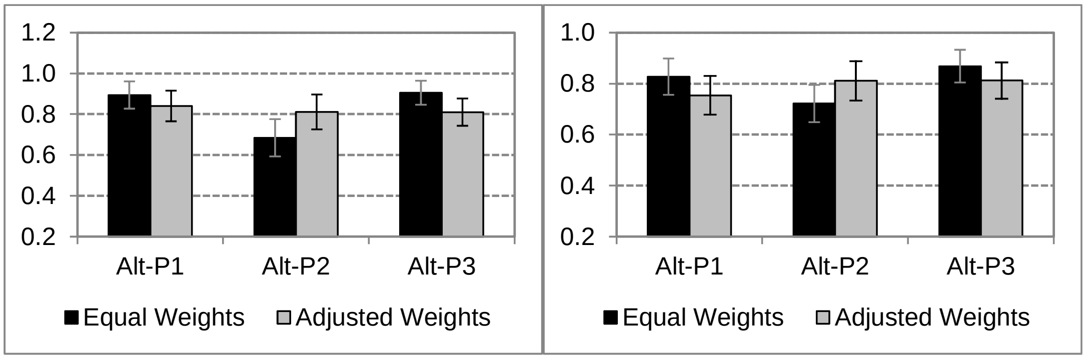


Figure 14 Sensitivity Analysis Results for Framework 3
(LEFT: ERA countries; RIGHT: Global benchmark)

Figure 15 shows how the composite scores changed after the adjustment of the weights as described above, and how did that affect country rankings. Rank changes are shown in blue rectangles at the top of the charts: a negative sign indicates advancement in rank, a positive sign indicates the opposite. For instance, among the ERA countries (upper panel) Cyprus advanced four positions, while Iceland fell back two positions compared to the scores based on equal weights for all pillars.

Countries which have greater variance in pillar scores should in principle be most affected by adjusted weights, as can be seen in the case of Cyprus, Ireland, Portugal, Bulgaria or Latvia. However, absolute changes do not necessarily translate into rank position changes – for instance, the near 7 point increase in the Irish score left the country's relative rank unchanged, while the nearly tied positions of Norway and Croatia were much more affected by only a marginal change of scores.

In the international comparison, the adjustment of weights has a positive effect only on EU27 and Russia, while it leaves the scores of the US, Japan, Brazil, India, South Korea and China unchanged. However, their ranks deteriorate because of the increase in scores in neighboring ERA countries.

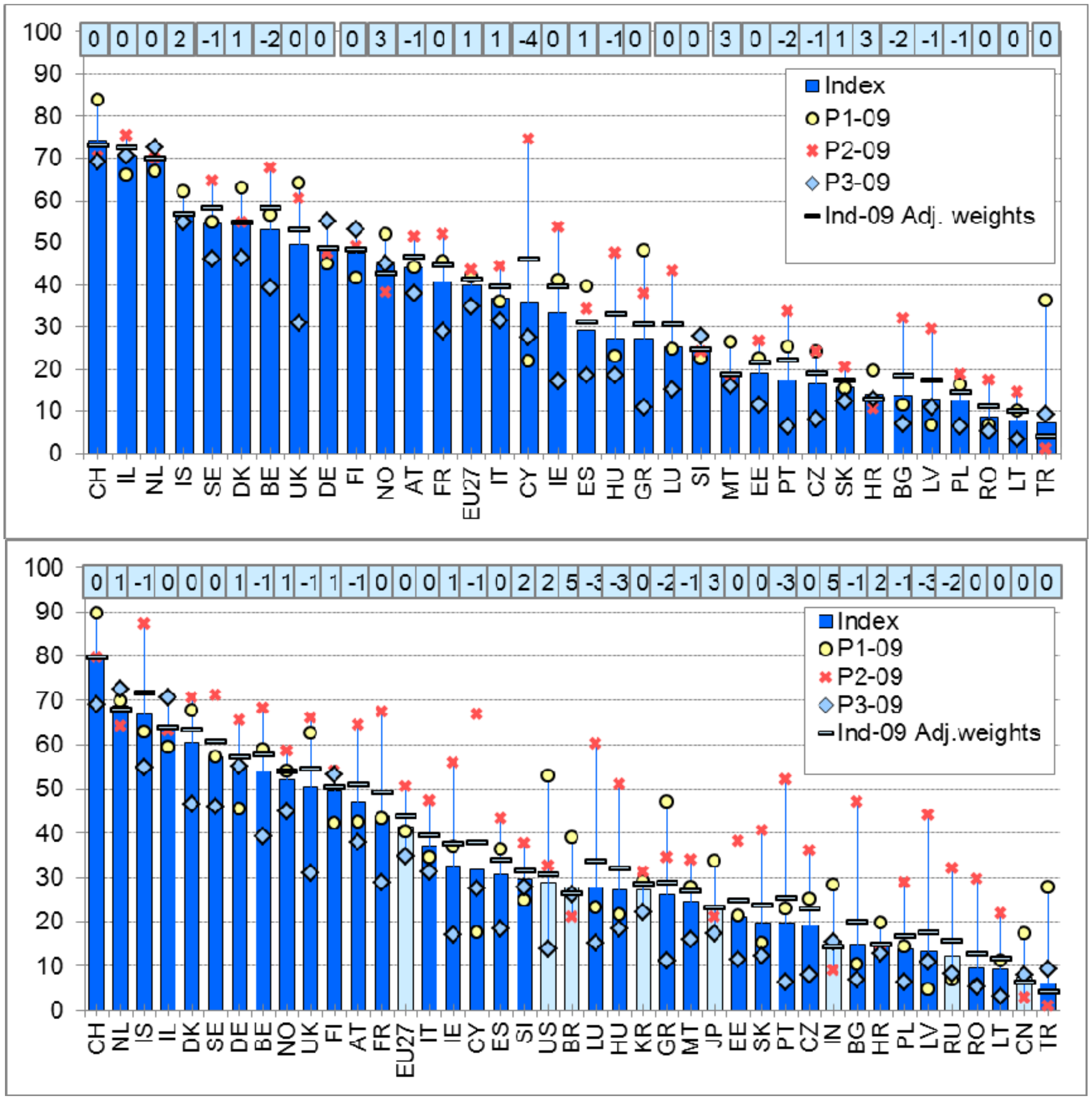


Figure 15 Composite Score Changes after adjustment of weights, alternative framework, 2009 (UPPER: ERA countries; LOWER: Global benchmark)

Figure 16 shows the map of composite scores for the ERA countries, computed using adjusted weights and the pillars of the alternative framework. A 'core-periphery' pattern is evident: scores decrease by moving from North West to South East Europe, although some centers of excellence are found outside the "core" i.e. in Israel. But evidently, EU member states from the former Eastern block as well as Turkey and Croatia need to do the most in order to catch up in research excellence.

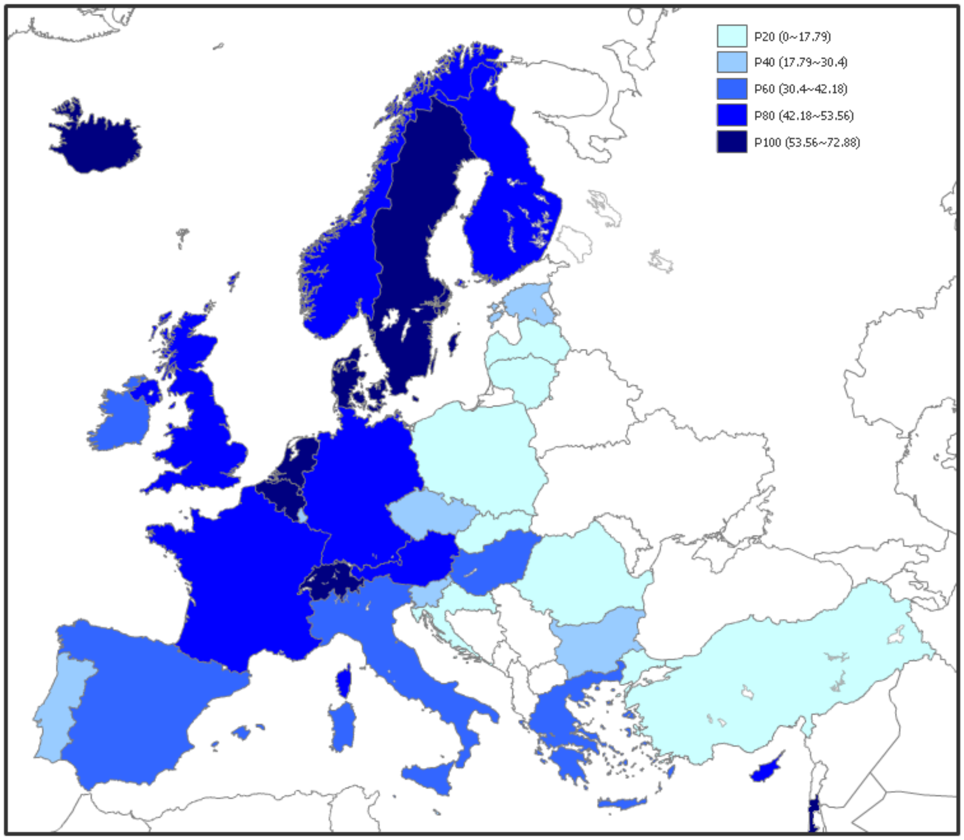


Figure 16 Map of research excellence scores in Europe, (Framework 3, adjusted weights, 2009)

Change over time

The score changes between time points 2005 and 2009 are plotted on the map in Figure 17. We notice some resemblence with the map showing the scores of 2009 (Figure 16). Most of the countries that performed worst in 2009 were also the ones that showed the least (or even negative) relative change, a group which includes Latvia, Lithuania, Poland, Romania and Turkey. These Eastern countries, together with other below-average countries (i.e. Bulgaria or the Czech Republic), need to show more dynamism in order to “catch up” with the leaders. Croatia, Slovenia and Estonia have relatively increased their performance over time, but this dynamism needs to be maintained in order to further increase their scores that were below-average even in 2009. At the upper end of the scale, the Netherlands and Denmark have been among the leaders in 2009 composite scores and also showed strong growth over the period examined.

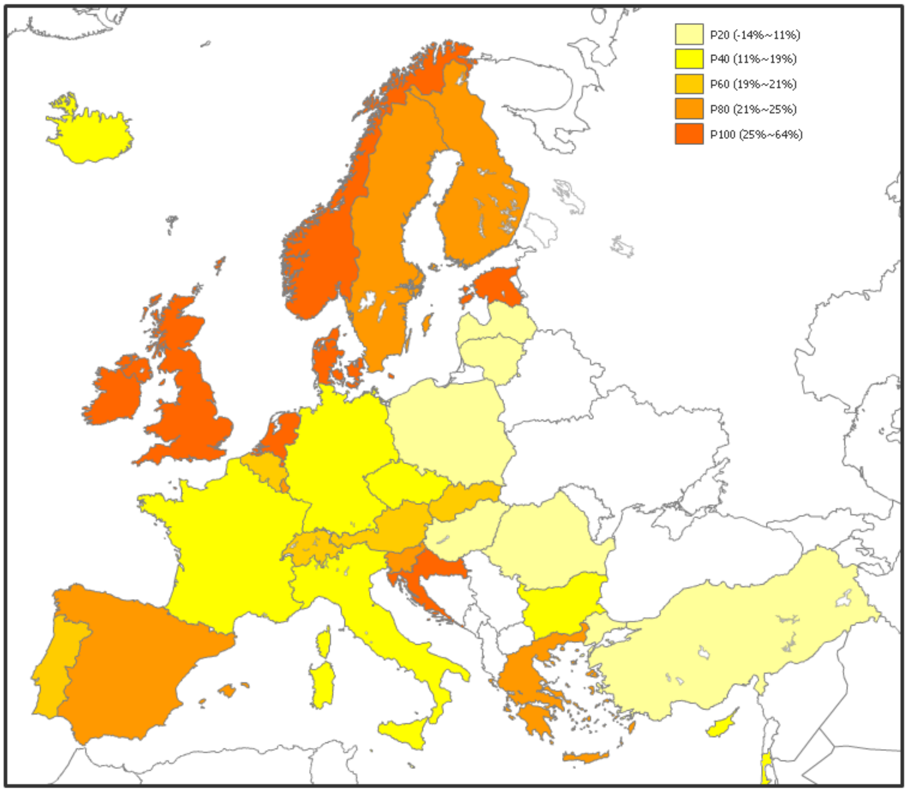


Figure 17 Map of change in research excellence scores, 2005 to 2009 (in %, ERA countries, Framework 3, adjusted weights)

Dynamics can also be presented by comparing performance in 2009 and progress from 2005 to 2009 in a single graph. The quadrangles in Figure 18 highlight that most of the countries with scores below the EU27 have also grown slower over time than the EU27 (the laggards group), some even declined (Latvia, Lithuania, Turkey, Romania). At the same time, only very few (most prominently Estonia) have been catching up with the EU27 (Note that the EU27 scores represent a weighted average of the Member States). Most of the countries with an above-average performance in 2009 are declining relative to the EU27 (or 'losing momentum', i.e. Germany, France, Israel or Austria), while the Netherlands, the United Kingdom, Denmark and Norway are forging ahead.

Figure 19 is based on the global benchmark (and since the framework uses a more limited set of indicators, the 2009 scores as well as the rates of change differ from the ones shown in Figure 18). We see that China is catching up with the EU27 at high rates (although it still remains at the end of the list), Japan and the US neither converge, nor fall behind the EU27, while all other countries (Russia, Korea and India) are falling further behind. Brazil's score has even declined over time.

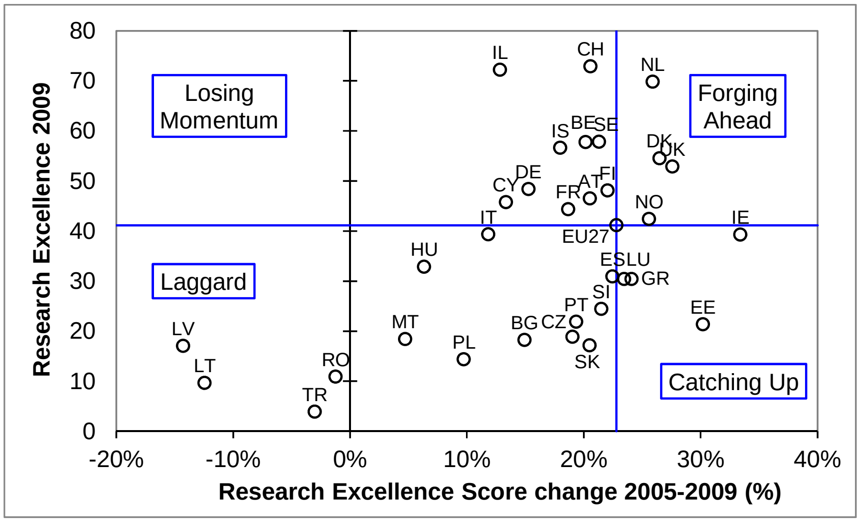


Figure 18 Performance vs. Progress: Research Excellence score change 2005-09 vs. 2009 scores, ERA countries (Framework 3, adjusted weights)

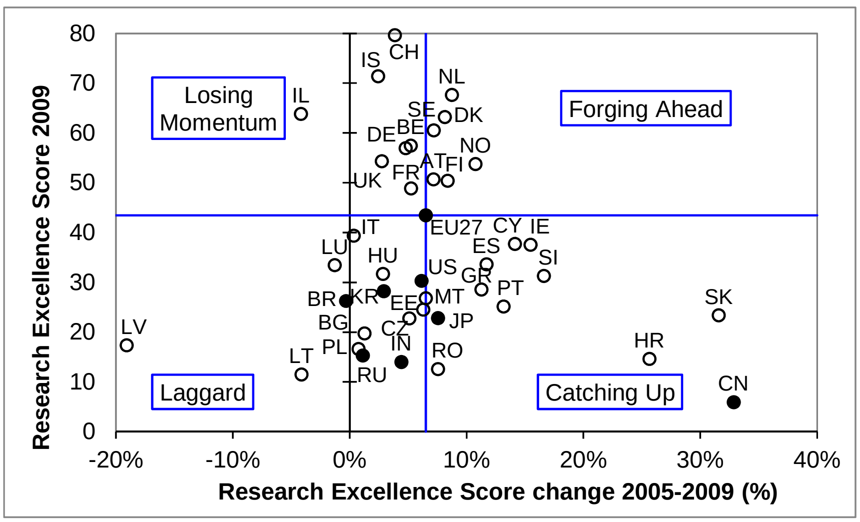


Figure 19 Performance vs. Progress: Research Excellence score change 2005-09 vs. 2009 scores, global benchmark (Framework 3, adjusted weights)

Pillar-by-pillar score changes are plotted against the change in the total composite score from 2005 to 2009 in Figure 20-Figure 22. Pillar 1 shows a slight growth for most of the countries, with Malta, Slovenia and Cyprus achieving relatively the biggest gains (Figure 20). The few countries that experienced decreasing P1 scores were Bulgaria, Luxemburg, Latvia and Romania. The reason behind this decline varied from country to country – in the case of Luxemburg, for instance, it was due to indicators 1a9, 1b5 and 1b7. The chart also shows that two countries were outliers: Turkey with a relatively stronger, and Cyprus, with a relatively weaker P1 score compared to their composite scores.

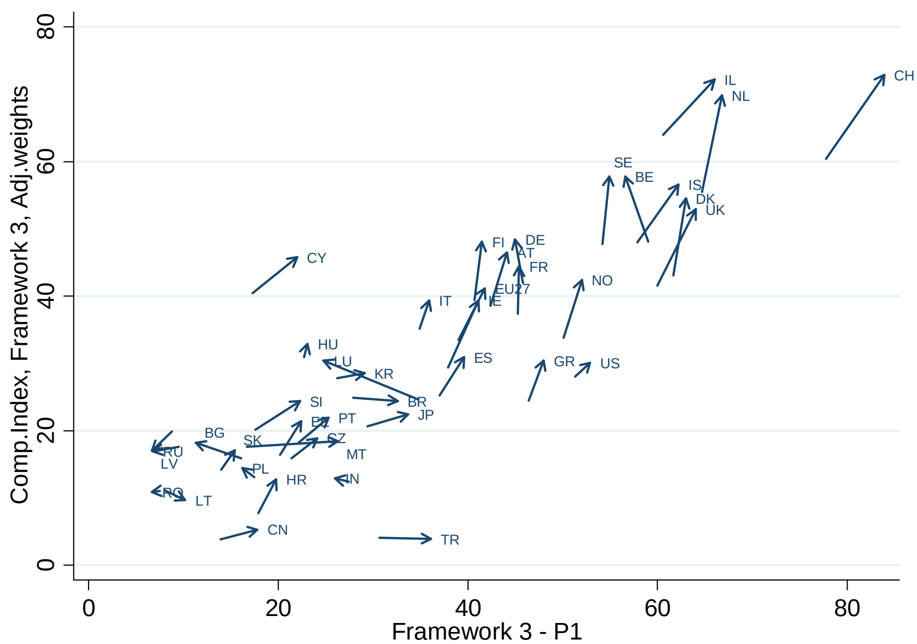


Figure 20 Changes in Pillar 1 and Composite Scores from 2005 to 2009, Framework 3

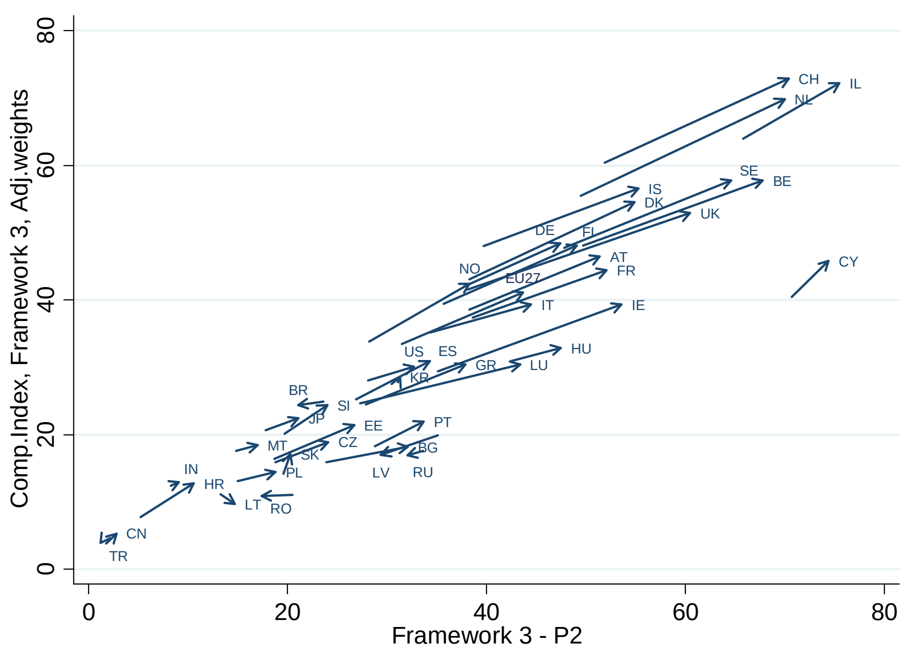


Figure 21 Changes in Pillar 2 and Composite Scores from 2005 to 2009, Framework 3

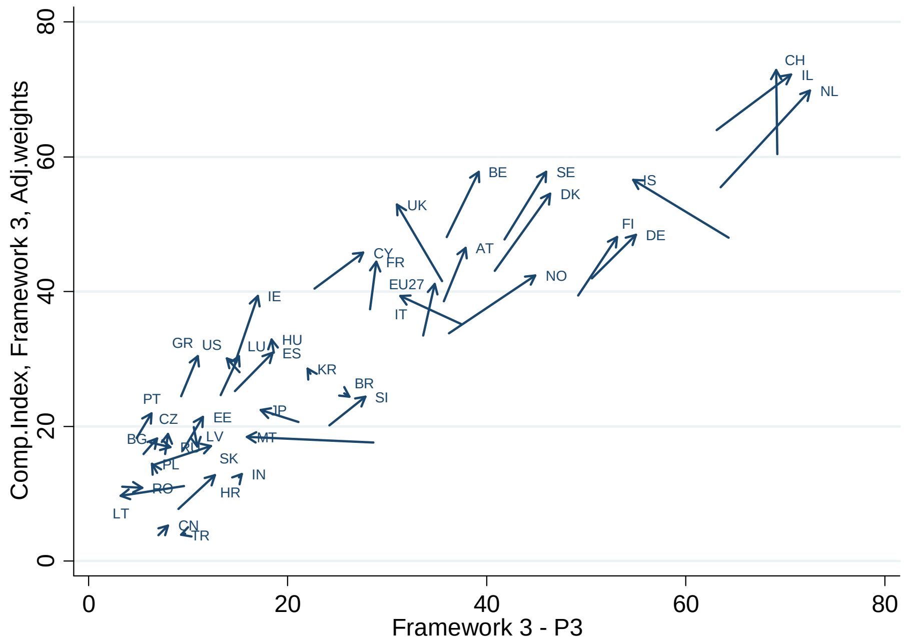


Figure 22 Changes in Pillar 3 and Composite Scores from 2005 to 2009, Framework 3

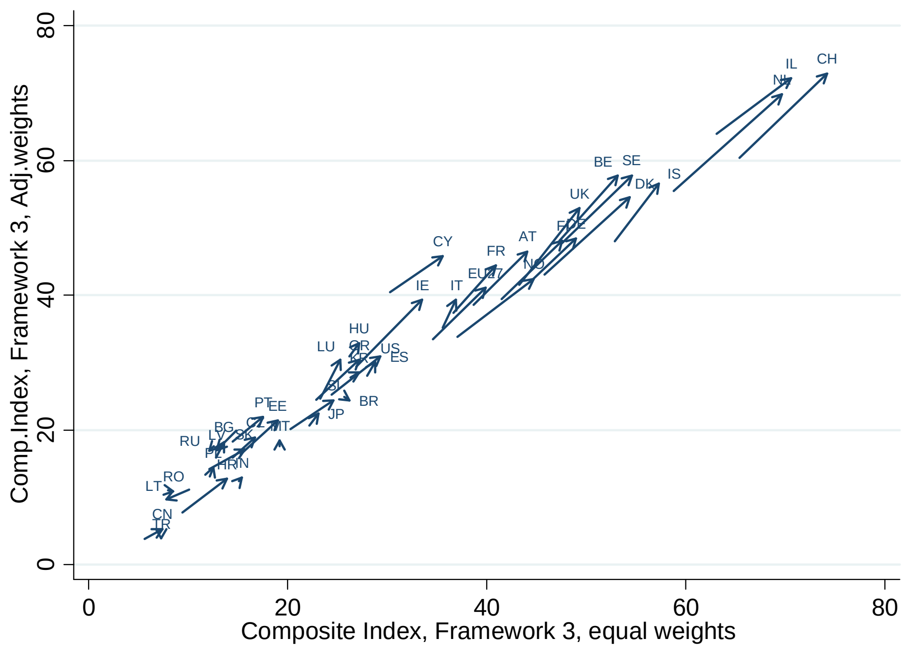


Figure 23 Composite Scores changes with equal and adjusted weights, from 2005 to 2009, Framework 3

The changes of Pillar 2 scores indicate a pattern of increasing international collaboration (Figure 21). The increase especially affected the countries that already had well-established collaborations. There were a few cases that did not follow the trend: Latvia, Romania and Turkey experienced a slight decline in their Pillar 2 scores, while Cyprus was an outlier with

relatively greater score for Pillar 2 than for the composite index. This can also be seen in Figure 23, where the equal and adjusted weight frameworks are compared over time: Cyprus benefits from increased weights given to Pillar 2.
Pillar 3 showed mixed dynamics over time (Figure 22). Countries with low scores (below around 20) were stagnating, while others experienced growth – with some exceptions including the UK, Italy and Iceland.

Comparison against other indicators

We used the weight-adjusted Framework 3 composite scores to test how it relates to two selected indicators of the research and innovation system. This has to be seen as an initial, exploratory attempt to identify a few, meaningful correlation patterns, not an in-depth analysis of causal relationships.

First of all, it is interesting to contrast composite scores of research excellence with a highly visible, crude measure of research excellence: the number of Nobel Prize and Field Medal winners per population (same as the numerator of indicator 1a8, but normalized by total population and standardized between 0-100). As apparent in Figure 24, they measure different aspects of research excellence. The US and Israel (below the 45 degrees line) score much better in terms of hosting prize-winner scientists than the other countries. With similar levels of prize winners per population, the UK and France significantly outperform the US.

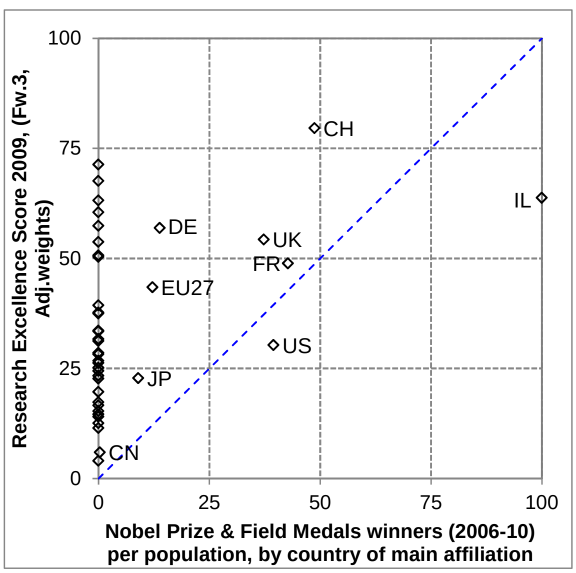


Figure 24 The crude vs. the composite measure of research excellence: Nobel Prize and Field Medal winners (2006-2010) vs. Research excellence scores (2009)

In Figure 25 we plotted gross R&D investments (GERD) as a share of GDP against composite score *changes* between 2005 and 2009. The blue dotted lines in the graph indicate the averages of GERD as well as composite score changes. We found a steady, positive growth rate of around 20-30% for the composite index associated with GERD levels of at least around 1.5% of GDP. If GERD is below this level, research excellence scores vary more, suggesting that other factors may influence them more. (Croatia was an outlier.)

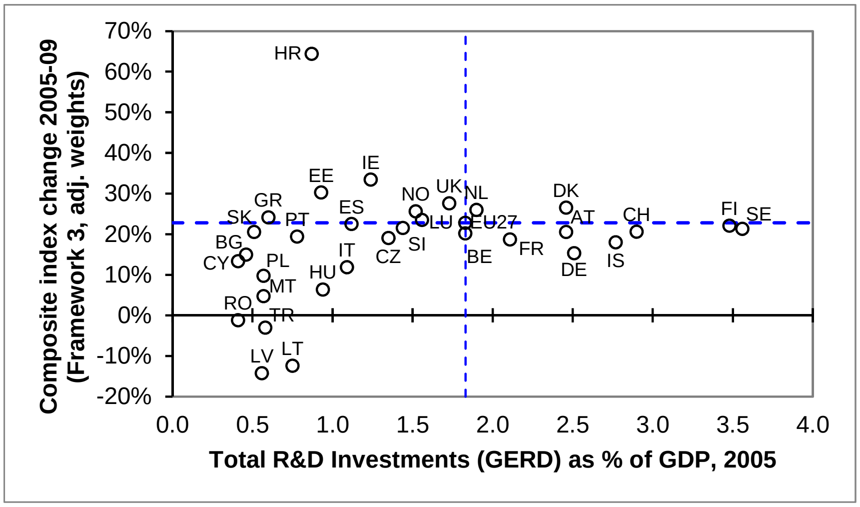


Figure 25 Composite Scores changes 2005-09 (Framework 3, adjusted weights) compared with Gross R&D Investments 2005

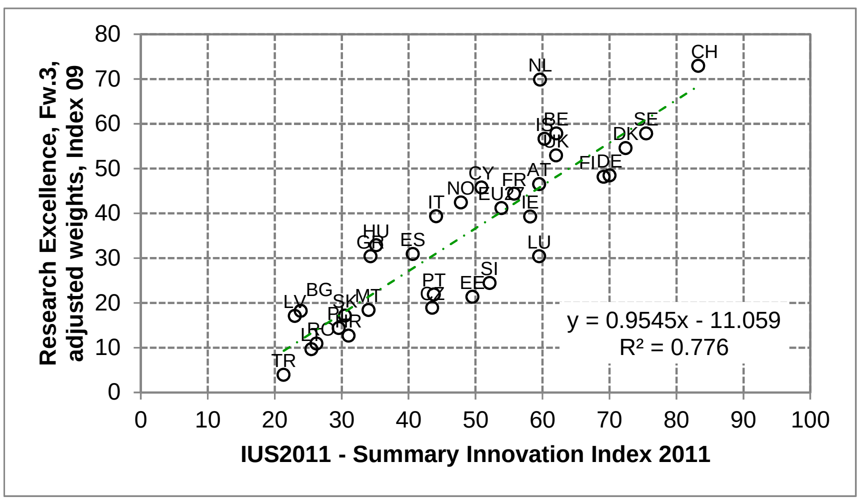


Figure 26 Composite Scores 2009 (Framework 3, adjusted weights, ERA countries) compared with Summary Innovation Index of the Innovation Union Scoreboard of 2011

Figure 26 compares the Summary Innovation Index (SII) of the 2011 Innovation Union Scoreboard[6] with the research excellence composite (framework 3, adjusted weights) score of 2009. (The years used may nominally differ, but they were both computed from the most recently available data. In any case, the Pearson correlation coefficients remain stable (0.87-0.88), regardless of using the 2011, the 2010 or the 2009 edition of the SII for comparison.) We do find that the SII does well capture research excellence. Indeed, one of its sub-pillars ("Open, excellent, attractive research systems") is built from three of the indicators used also

[6] Innovation Union Scoreboard 2011 [http://ec.europa.eu/enterprise/policies/innovation/files/ius-2011_en.pdf]

for the excellence composites. The graph shows a few outlier countries: for instance, the Netherlands is relatively stronger in research excellence, for instance Luxemburg, Slovenia and Estonia are stronger in the more comprehensive index of innovation. On this note we point out a peculiarity, that in these cases, the relatively better performance in the SII is not matched with a relatively better performance of these countries in Pillar 2, or excellence of business research. This also shows that excellence in business research does not directly translate to innovation performance. Comparing for instance the Netherlands with Switzerland both with similar scores, the Netherlands clearly needs to do more to reap the results of its excellent research performance through innovations.

Another possible use of the research excellence score is to compare with levels of GDP per capita. In Figure 27, 2009 GDP per capita levels (at purchasing power parity) are standardized between 0 and 100 on the horizontal axis. Without presuming any causal relationships between the two indicators, the graph shows that at a similar GDP per capita level the Netherlands and Denmark outperform Ireland and Austria, or that Israel (with even lower GDP per capita levels) outperforms the EU27 in research excellence – implying the presence of a more efficient research system in those countries. In the future, it may be useful to conduct similar comparisons with other indicators on productivity, the labour market and institutional conditions or product market regulations.

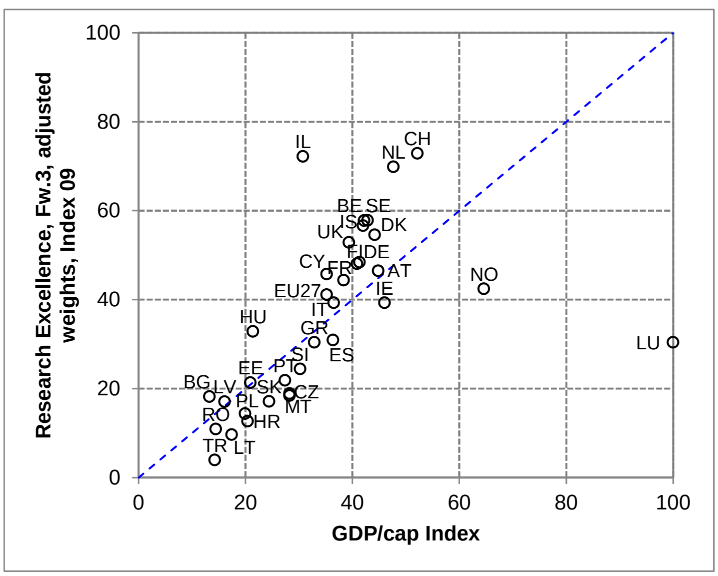


Figure 27 Research Excellence Scores (2009, Fw.3, Adj.weights) vs. GDP Index (2009 GDP per capita PPP levels standardized between 0-100)

Conclusions

The objective of the present deliverable has been to test the feasibility to develop a composite indicator to measure the excellence of the research systems of all EU Member States, most EFTA countries, Candidate countries and the main international competitors of the European Union (United States, Japan and China). We have proposed three alternative conceptual frameworks of research excellence and we have tested their statistical coherence in order to identify feasible composite indicators.

The first theoretical framework, consisting of 6 dimensions, was proposed by the expert group on measuring innovation through the report *Indicators of Research Excellence*. Another framework consisting of two dimensions (basic and applied science) was extracted from the expert group report and tested. Finally, a third framework, in which the indicators were combined according to three dimensions deriving from a principal component analysis, was proposed and validated.

Starting from the 6-dimensional conceptual framework originally proposed by the expert group, the multivariate statistical analysis led us to conclude that the framework proposed by the expert group based on theoretical considerations only partially holds in light of the empirical data collected for 41 countries. We could identify four dimensions composed by an overall set of 13 indicators (there was no available data on indicators of societal relevance, i.e. pillars 3a and 3b):

1. dimension 1a would obtain greater statistical coherence if it were narrowed down to include the following five indicators: 1a1, 1a2, 1a4, 1a5 and 1a9;
2. dimension 1b is a statistically coherent pillar, as long as indicator 1b1_collind replaces 1b1 (composed by 4 indicators);
3. dimension 2a needs to be reduced to three indicators, 2a1, 2a4 and 2a8 for a statistically coherent pillar;
4. due to the negative correlation between the only two indicators available for pillar 2b, this dimension could only be represented by indicator 2b1.

The dimensions were quite heterogeneous in terms of number of indicators; each pillar captures one single latent dimension, which explains at least 55 % of variance in the data. The factors loadings for pillar 3 are rather unbalanced.

We then tested how the data structure could support a two-dimensional framework distinguished between basic and applied research. The multivariate analysis showed that there was no single ‘basic research’ pillar but three separate, albeit conceptually heterogeneous, dimensions; further, a pillar of ‘applied research’ could be statistically supported as long as it consists of three indicators (2a1, 2a4 and 2b1). In summary, this second framework is weak because the “applied research” pillar is placed side by side to three “basic research” pillars, for which there is no clear theoretical underpinning.

With a view to achieving a framework which is both statistically sound and rich in indicators, we analysed by principal components the whole set of 22 indicators starting from the original structure (Framework 1) and testing alternative specifications. Here we considered a modified indicator (2b3_gdp) to measure financial flows from business to public research. We found three coherent and statistically sound pillars:

4. excellence of public research (6 indicators);
5. *interactions, collaborations (4 indicators)*
6. excellence in industrial actors (3 indicators)

The framework well accommodates 13 indicators (as in Framework 1) and each pillar captures one single latent dimension, which explains at least 63 % of variance in the data. The factors loadings are overall rather balanced.

Composite indicators were computed for each of the three frameworks, using geometric aggregation across pillars. The geometric aggregation adopts both equal and adjusted weights in order to make pillars more equally balanced. Each pillar is an arithmetic average of its indicators normalized between 0 and 100 using the min-max method and taking into consideration the two years simultaneously.

For Framework 1 the composite indicator shows a clear North-West vs South-East divide with centres of excellence in the Nordic and North-Sea countries. Almost all countries have improved their excellence between 2005 and 2009. The considerable spread of scores at the pillar level provides useful insights to research excellence performance: high and low scores are detected for pillars 1b and 2b, respectively.

For Framework 2, the pillars of the resulting composite indicator do not contribute equally to the overall composite indicator. A re-adjustment of weights could improve the framework imbalances, but leaves the conceptual difficulties of the theoretical framework unresolved.

According to Framework 3 three countries are clearly distinguished as the leaders in research excellence: Switzerland, Israel and the Netherlands, all with a score of 70 or above. They are followed by countries of North Western Europe with scores exceeding the EU27 score, and Southern and Eastern member states and Associate Countries below the EU27 score. Of the countries with the largest population, the United Kingdom and Germany are neck and neck, ahead of France and Italy. Three countries trail the list with single-digit scores: Romania, Lithuania and Turkey.

Figure 28 compares the composite scores for year 2009 resulting from the three frameworks, using adjusted weights for all. Frameworks 1 and 3 strongly correlate (0.97 Pearson correlation coefficient), and are providing a similar message. Apart from Estonia, Romania and Croatia, the scores for Framework 1 are smaller than for Framework 3. Framework 2 scores are less correlated with the scores of Frameworks 1 and 3.

These results suggest the use of Framework 3 as the basis for the development of the fully-fledged composite indicator on research excellence. The pillars in this framework explain a greater share of variance in the data, and each variable contributes to pillar scores in a more balanced way than in Framework 1. At the same time, the contribution of each pillar to the composite index is intrinsically more balanced.

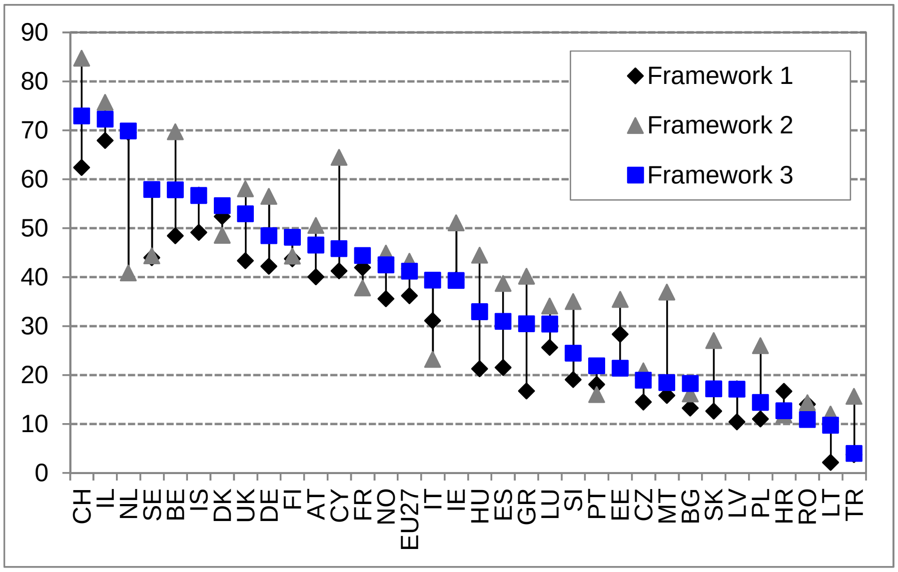


Figure 28 Comparison of the 3 proposed frameworks with equal weights and framework 3 with adjusted weights, ERA countries, 2009 – rank order by Framework 3

Comments from Experts

Rémi Barré

(Professor, CNAM University, Paris)

This report is a methodologically rigorous in its building of a sound composite indicator of research excellence in Europe. It is to be praised for building upon an explicit theoretical framework published in the 2011 report of the Expert group on the measurement of innovation "Indicators of research Excellence" and for going beyond that work on two counts:

- the reconfiguration of the set of indicators in terms of a selection of these indicators and of regrouping them into 'pillars' in order to get a coherent base for the computation of the composite indicator,
- the actual computation of the indicators and of the composite for 41 countries at two dates, allowing for empirical results of high relevance.

The depth and rigour of the assessment of the data sources is to be noted, as well as the quality of the handling of the internal coherence issue of a composite indicator - through a correlation matrix and a principal components analysis.

The development and assessment of three alternative frameworks, the sensitivity analysis and the explicitation of the adjustments made for the indicators – provide a high level of credibility and usefulness for this report.

My regret – which is not related to the authors' work – is that the lack of indicators of 'engagement of S&T with society' has logically led to the disappearance of this issue as a 'pillar'.

This - rare – fully fledged exercise in the development of a composite indicator shows in a most demonstrative way both the appeal and feasibility of such an indicator and also that there is no analytically ‘best’ composite indicator: the discussions on the relative weights of the pillars as well as those related to the computation formulae are eloquent in this respect.

At the very end, the ultimate shape chosen for the composite indicator, is a ‘political’ decision, to be made in all transparency – and this work can be a cornerstone in this respect. This is not a limitation of the indicators relevance and importance – but the recognition that they are an expression of the articulation between science and society.

Robert Tijssen

(Professor of Science and Innovation Studies, CWTS, Leiden University, The Netherlands)

I've read your report with great interest. I found it quite sophisticated and helpful on deciding how to design policy-relevant and statistically robust composite indicators. Your model 3 is definitely the best option.

There's just one small comment: I would replace the term 'research excellence' by something more fitting. That's not easy because the selected list of indicators is so divers. How about 'knowledge creation and utilization'?

Acknowlegements

We thank Matthieu Delescluse for his in-depth comments on the definition of indicators and interpretation of results which were incorporated in the present version of the report. We also thank Andrea Saltelli for providing technical assistance with the sensitivity analysis and Michela Nardo and Michaela Saisana for their review of a preliminary version of the report.

Abbreviations:

ARC:	Average of Relative Citations
BERD:	Business Expenditure on R&D
BRICS:	Brazil, Russia, India, China, South-Africa,
Candidate countries:	Croatia, Macedonia, Turkey,
EFTA:	European Free Trade Association composed of Switzerland, Iceland, Liechtenstein, Norway,
EPO:	European Patent Office
ERA countries:	Austria, Belgium, Bulgaria, Cyprus, Czech Republic, Germany, Denmark, Estonia, Greece, Spain, Finland, France, Hungary, Ireland, Italy, Lithuania, Luxembourg, Latvia, Malta, Netherlands, Poland, Portugal, Romania, Sweden, Slovenia, Slovakia, United Kingdom, Croatia, Macedonia, Turkey, Switzerland, Iceland, Liechtenstein, Norway, Israel;
ERC:	European Research Council;
Eurostat:	Statistical office of the European Union;
FP:	Framework Programme
GERD:	Gross Expenditure on Research and Development;
GOVERD:	Government (Public) expenditure on R&D;
GSCs:	Grand Societal Challenges: Health, Energy, Environment (including Climate Change), Food, Agriculture, Fisheries;
HEI:	Higher Education Institute;
HERD:	Higher Education sector's expenditure on R&D;
KETs:	Key Enabling Technologies: Biotechnology, ICT, Nanotechnologies, New Production Technologies;
OECD:	Organisation for Economic Co-operation and Development;
PCA:	Principal Component Analysis
PCT:	Patent Cooperation Treaty
PRO:	Public Research Institute
R&D:	Research and Development
UNESCO:	United Nations Educational, Scientific and Cultural Organization;
USPTO:	United States Patent and Trademark Office

European Commission
EUR 25488 EN– Joint Research Centre – Institute for the Protection and Security of the Citizen

Title: **Composite Indicators of Research Excellence**

Authors: Daniel Vertesy, Stefano Tarantola

Luxembourg: Publications Office of the European Union

2012 – 62 pp. – 21.0 x 29.7 cm

EUR – Scientific and Technical Research series – ISSN 1831-9424 (online), ISSN 1018-5593 (print)

ISBN 978-92-79-26260-9

doi:10.2788/45492

Abstract

This report on Research Excellence is the deliverable of the third work package (WP3) of the feasibility study 'ERAMONITORING', financed by DG RTD. The objective of the work package was to explore the possibility to develop a composite indicator of research excellence in Europe, in coherence with the orientations of the EU 2020 strategy and the Innovation Union initiative. The study built on the theoretical framework proposed by the 2011 report of the Expert Group on the Measurement of Innovation 'Indicators of Research Excellence', co- authored by Rémi Barré (CNAM, France), Hugo Hollanders (UNU-MERIT, The Netherlands) and Ammon Salter (Imperial College, UK). We proposed three alternative conceptual frameworks of research excellence with different underlying indicator structures, and tested their statistical coherence. In the first theoretical framework, we aimed to follow as closely as possible the Expert Group recommendation of 6 dimensions. In the second framework, we tried to consider only two dimensions (basic and applied science), also based on the Expert Group report. The third framework was derived from the data and three dimensions were identified directly from a principal component analysis.

LB-NA-25488-EN-N

As the Commission's in-house science service, the Joint Research Centre's mission is to provide EU policies with independent, evidence-based scientific and technical support throughout the whole policy cycle.

Working in close cooperation with policy Directorates-General, the JRC addresses key societal challenges while stimulating innovation through developing new standards, methods and tools, and sharing and transferring its know-how to the Member States and international community.

Key policy areas include: environment and climate change; energy and transport; agriculture and food security; health and consumer protection; information society and digital agenda; safety and security including nuclear; all supported through a cross-cutting and multi-disciplinary approach.

Publications Office

ISBN 978-92-79-26260-9