JRC TECHNICAL REPORTS

JRC Statistical Audit of the European Skills Index 2022

Caperna, G.

Becker, W.

2022

EUR 31019 EN

Report completed in March 2022.

Contact information
European Commission
Joint Research Centre
Directorate for Competences
Monitoring, Indicators and Impact Evaluation Unit
Competence Centre on Composite Indicators and Scoreboards

Competence Centre on Composite Indicators and Scoreboards
https://ec.europa.eu/knowledge4policy/composite-indicators

EU Science Hub
https://ec.europa.eu/jrc

JRC128829

EUR 31019 EN

PDF ISBN 978-92-76-49387-7 ISSN 1831-9424 doi:10.2760/629770

Luxembourg: Publications Office of the European Union, 2022

How to cite this report: Caperna G., Becker W., *JRC Statistical Audit of the European Skills Index 2022*, EUR 31019 EN, Publications Office of the European Union, Luxembourg, 2022, ISBN 978-92-76-49387-7, doi:10.2760/629770, JRC128829

Contents

Abstract

The European Skills Index (ESI) is a multidimensional index that ranks the skills systems of 31 countries, all EU27 countries plus Iceland, Norway, Switzerland and the United Kingdom, on the three pillars of skills development, activation and matching.

The European Commission's Competence Centre on Composite Indicators and Scoreboards (COIN) at the Joint Research Centre (JRC) in Ispra was invited by the index developers to audit the ESI for the third time. JRC-COIN aims to help ensure the transparency of the index methodology and the reliability of its results. This JRC-COIN audit focuses on data quality, the statistical soundness of the multi-level structure of the index, and the impact of key modelling assumptions on the results.

The analysis suggests that meaningful inferences from the index with special attention on countries with close scores can be drawn. The ESI is reliable, and the framework has good statistical coherence. ESI ranks are shown to be representative of a plurality of scenarios and robust to changes in the aggregation and normalisation methods and the pillar weights except for a few cases. Even though the ESI has many good statistical properties, JRC-COIN has made some suggestions to improve the definition of narratives in accordance with the methodological characteristics.

Introduction

The European Skills Index (ESI) is intended to measure the performance of EU Members States plus four countries (Iceland, Norway, Switzerland and the United Kingdom) skills formation and matching systems to enable a comparative assessment across countries. It is a multidimensional index composed of 15 indicators organised into 6 different sub-pillars. These sub-pillars are themselves aggregated three pillars. Each of these pillars corresponds to a dimension of the skills system: development, activation and matching.

The ESI framework is well constructed, and a lot of thought has clearly been put into it. However, conceptual and practical challenges are inevitable when trying to summarise with a single composite indicator the complexity of a multidimensional phenomenon. An analysis is needed to ensure and validate the statistical soundness of any composite index. The analysis performed in this audit – and discussed in this report – aims to help policymakers derive more accurate and meaningful conclusions form the Index, and to potentially guide their choices on priority setting and policy formulation.

In general, statistical soundness should be regarded as a necessary but not sufficient condition for a sound index. This is because the correlations underpinning most of the statistical analyses carried out in this report need not necessarily represent the real influence of the individual indicators on the phenomenon being measured. The development of any index must therefore be nurtured by a dynamic, iterative dialogue between the principles of statistical and conceptual soundness.

The JRC assessment of the ESI presented here focuses on two main issues: the statistical coherence of the structure, and the impact of key modelling assumptions on the ESI ranks. The statistical analysis is based on: (i) the adequacy of aggregating indicators into aggregates (sub-pillars and pillars), and aggregates into the overall index; (ii) the multidimensional structure of the ESI; and (iii) the specific impact of each element used in the computation of the Index. Finally, the JRC analysis complements the reported country rankings for the ESI with estimated intervals, in order to better appreciate the robustness of these ranks to the modelling choices.

Conceptual framework

The ESI is based on three pillars, each of which relates to one area found to be critical respect to the skills system: Skills Development; Skills Activation and Skills Matching. Each of these pillars contains 2 sub-pillars, making 6 sub-pillars in total. These 6 sub-pillars are built using 15 indicators (2 to 3 indicators for sub-pillar, see **Table 1**). The index is based on these 15 indicators and aggregated at each level using a weighted arithmetic average, except for the aggregation of the pillars that is computed as a geometric mean. The choice of the geometric mean allows for a reduced substitutability among the pillars. It means that a country with unbalance among the values of the pillars will be penalised by the aggregation formula, respect to one whit balanced profile. Since the 2018 ESI edition, the weights are meant to balance the role of the elements of an aggregate. The aim is to obtain balanced groups where all elements have, approximately, similar correlation with the aggregate. The weight used in the Index are listed in following table.

Table 1: Conceptual framework of the ESI

Pillar (weight)	Sub-Pillar (weight)	Indicator	Code	Ind. Weight	Direction
Skills Development (0.3)	Basic education (0.5)	Reading, maths & science scores (aged 15)	RMS	0.4	-1
		Upper secondary attainment (and above)	SecAttain	0.3	1
		Pre-primary pupil-to-teacher ratio	PupTechRat	0.3	1
	Training and other education (0.5)	Recent training	Training	0.3	1
		High digital skills (substitutes High Computer skills)	Digital	0.35	1
		VET students	VETs	0.35	1
Skills Activation (0.27)	Transition to work (0.5)	Early leavers from training	EarLeavTrain	0.6	-1
		Recent graduates in employment	Grads	0.4	1
	Labour market participation (0.5)	Activity rate (aged 25-54)	ActRate20-24	0.5	1
		Activity rate (aged 20-24)	ActRate25-54	0.5	1
Skills Matching (0.43)	Skills utilisation (0.4)	Long-term unemployment	LTUnemp	0.4	-1
		Underemployed part-time workers	UndEmpPart	0.6	-1
	Skills mismatch (0.6)	Over-qualification rate (tertiary graduates)	OverQual	0.4	-1
		Low-wage workers (ISCED 5-8)	LoWage	0.1	-1
		Qualification mismatch	QualMisMat	0.5	-1

Source: Developers of the index and the European Commission's Joint Research Centre, 2022.

Data quality and availability

Management of missing data

The 2022 European Skills Index draws on annual data, up to 2020. Whenever data were missing, the developer followed the rule of the last available year, replacing missing values with previous year values. The main year used for each indicator is presented in **Table 2**. The indicator showing older data is Qualification Mismatch, where all values are relative to 2016. This indicators is relatively the most important (approximately 13% of the weight of ESI). The concept described by this indicator require probably to keep into account some lag respect to the qualification of single individuals. As a consequence, JRC-COIN suggests to have a special focus on its coverage and timeliness.

The data used in this audit is the result of the imputation described above. After this step, data coverage is extremely high with only three values missing in the entire dataset. For remaining missing values, the developers opted for an implicit imputation at the aggregate level. In practice, the choice was to not impute the values. Because of this, the score of the aggregate containing the missing value is based on the other elements which are observed. This approach is usually selected to improve transparency and avoid any methodological black box. The developers set other coverage criteria for the inclusion of countries in the final dataset. Fortunately, this edition of the Index did not require the exclusion of any country because of lack of data.

Treatment of outliers

The audit also investigated the presence of outliers that could potentially bias the effect of the indicators on the aggregates. The JRC recommends an approach for outlier identification based on skewness and kurtosis values[1], i.e. when the variables simultaneously have absolute skewness higher than 2.0 and kurtosis higher than 3.5.

The results of the analysis on the presence of outliers are presented in **Table 2**. Despite two indicators show high values of skewness and kurtosis, the issue of outliers is implicitly solved in the normalisation step, which is based on goalposts (see below). The normalisation method based on goalposts can be effective in reducing outliers.

Normalisation

The indicators are rescaled to a 0-100 scale, using a special case of min-max normalisation, where the minimum and maximum are substituted by lower and upper bounds. The limits are representative of absolute "worst" and "best" cases. If a country happens to achieve a values outside the bounds, the values 0 and 100 are attributed as normalised value (respectively for values below the minimum and for cases above the maximum).Moreover, all the intermediate values are computed with the following two formulas:

Indicator with positive direction:

$$score = \frac{value - lower\ bound}{upper\ bound - lower\ bound} * 100$$

Indicator with negative direction:

$$score = \frac{upper\ bound - value}{upper\ bound - lower\ bound} * 100$$

An indicator is intended to be positive when higher values indicate better performance (it is negative if higher values indicate worse performance). The direction of all the indicators

[1] Groeneveld, R. A. and Meeden, G., 'Measuring Skewness and Kurtosis', *Journal of the Royal Statistical Society*, Series D, vol. 33, pp. 391–399, 1984.

is represented in **Table 1**. The bounds considered by the developers, together with their rationale, are presented in the report of the ESI2022.[2]

Table 2: Summary statistics of the indicators included in the ESI2022 (raw values).

Indicator	N. missing	Missing (%)	Main data year	Mean	Min	Max	Skew	Kurtosis
RMS	0	0	2018	485.78	426.65	525.51	-1.05	0.82
SecAttain	0	0	2020	77.76	55.50	89.20	-1.19	1.06
PupTechRat	1	3.23	2019	13.60	5.00	39.70	2.71	10.48
Training	0	0	2020	11.33	1.00	28.60	0.83	-0.08
Digital	0	0	2019	34.65	10.00	62.00	0.08	-0.15
VETs	0	0	2019	47.87	16.90	70.80	-0.03	-1.15
EarLeavTrain	0	0	2020	4.60	1.80	9.70	1.06	0.73
Grads	0	0	2020	80.62	54.90	92.20	-1.50	2.95
ActRate20-24	0	0	2020	60.50	39.70	78.90	-0.17	-1.38
ActRate25-54	0	0	2020	87.01	76.50	92.40	-1.18	3.35
LTUnemp	0	0	2020	2.03	0.50	10.90	3.46	14.41
UndEmpPart	0	0	2020	2.84	0.25	7.30	0.36	-0.19
OverQual	0	0	2019	23.64	9.71	48.01	1.13	1.51
LoWage	0	0	2019	8.39	2.36	15.59	0.30	-0.55
QualMisMat	2	6.45	2016	33.58	17.10	44.00	-0.68	-0.07

Source: European Commission's Joint Research Centre, 2022.

Note: The values of skewness and kurtosis exceeding the threshold are written in red.

[2] https://www.cedefop.europa.eu/en/tools/european-skills-index

Statistical coherence

The assessment of statistical coherence consists of a multi-level analysis of the correlations of variables, and a comparison of ESI rankings with their constituent pillars.

Correlation analysis

The statistical coherence of an index should be considered a necessary but not-sufficient condition for a sound index. Given that the statistical analysis is mostly based on correlations, the correspondence of every index to a real-world phenomenon needs to be critically addressed by developers and experts, because 'correlations do not necessarily represent the real influence of the individual indicators on the phenomenon being measured' (OECD & JRC, 2008)[3]. This influence relies on the interplay between both conceptual and statistical soundness. The degree of coherence between the conceptual framework and the statistical structure of the data is an important factor for the reliability of an index.

The correlation analysis is used to assess the extent to which the observed data support the conceptual framework. Ideally, there should be positive significant correlations within every level of the index, JRC-COIN suggest between 0.3 and 0.92. This effectively ensures that the overall index scores adequately reflect the underlying indicator values.

Redundancy, which could be identified by very high correlations (>0.92), should be avoided in the framework. This is because if two indicators are collinear, this may amount to double counting (and therefore over-weighting) of the same phenomenon.

Correlation analysis between indicators and aggregates

Table 3 shows the correlation coefficients between indicators, highlighting in boxes those belonging to the same sub-pillar. While **Table 4** shows the correlation between indicators and the aggregates. Many correlations are significant and positive (>0.30). However, a few problematic cases are identified in the paragraphs below.

- PupTeachRat and VETs, respectively from sub-pillar 1.1. and 1.2, show shallow, when not negative, correlations with the other indicators in their sub-pillars. This may suggest that these indicators do not entirely cooperate with the others, and this may cause a conflict in results and reduce the impact of the aggregate to which they belong in the following aggregations.

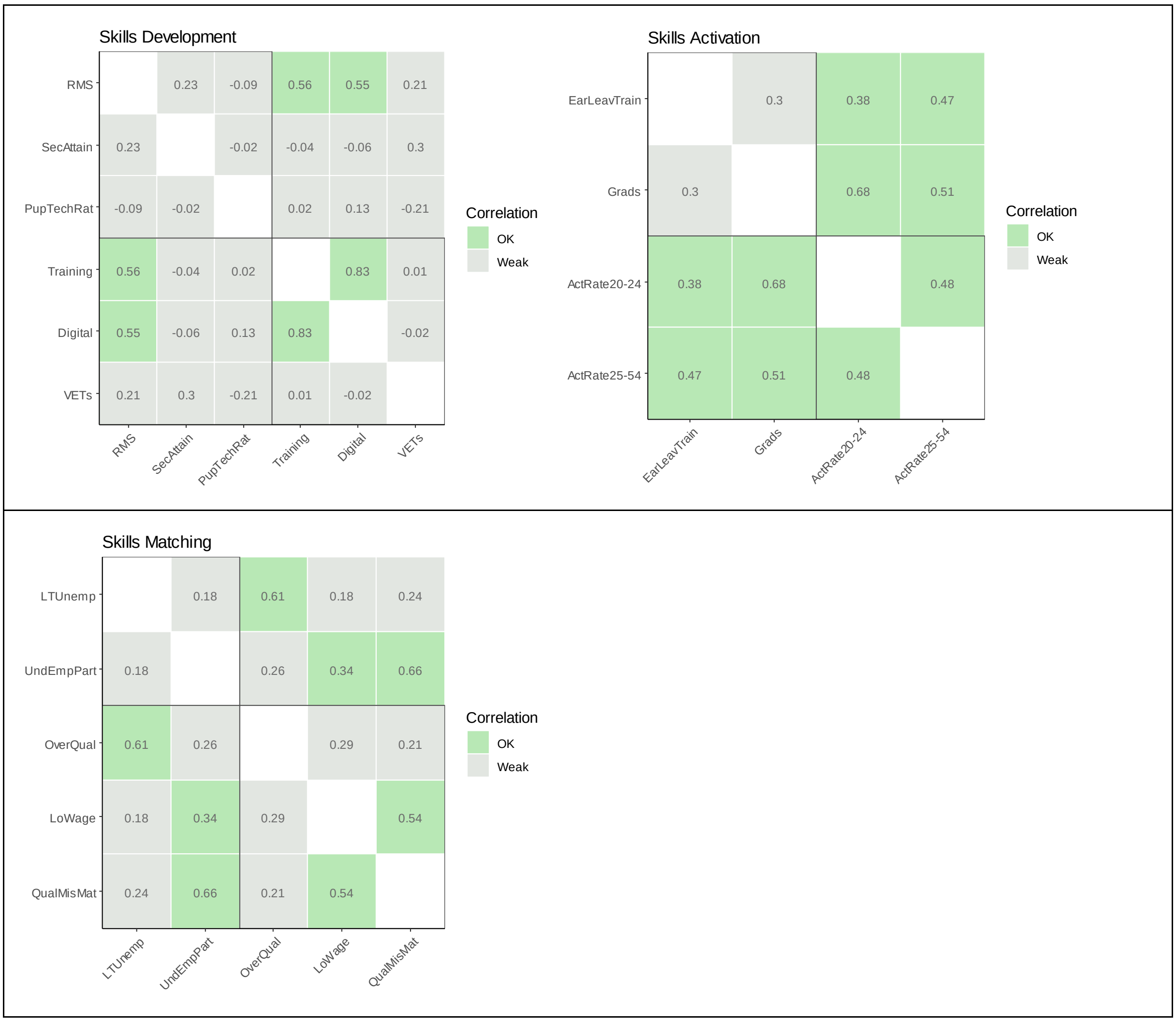
- Other correlations are below the conceptual threshold of 0.3. Among those, is good to give attention to the LTUnemp and UndEmpPart (sp3.1) and OverQual and QualiMisMat (sp3.2).

- Nevertheless, only PupTeachRat shows a low correlations (0.18) with the Index. Moreover, all indicators show relatively balanced correlations with the aggregates, meaning that they are represented in the higher levels. See Table 4 for details.

A suggestion would be to keep monitoring these specific indicators and their position in the framework for future editions of the index in order to check their behaviour and modify them if appropriate.

[3] OECD/EC JRC (Organisation for Economic Co-operation and Development/European Commission, Joint Research Centre). 2008. Handbook on Constructing Composite Indicators: Methodology and User Guide. Paris: OECD.

Table 3: Correlations between indicators in the same pillar



Skills Development

	RMS	SecAttain	PupTechRat	Training	Digital	VETs
RMS		0.23	-0.09	0.56	0.55	0.21
SecAttain	0.23		-0.02	-0.04	-0.06	0.3
PupTechRat	-0.09	-0.02		0.02	0.13	-0.21
Training	0.56	-0.04	0.02		0.83	0.01
Digital	0.55	-0.06	0.13	0.83		-0.02
VETs	0.21	0.3	-0.21	0.01	-0.02	

Correlation: OK / Weak

Skills Activation

	EarLeavTrain	Grads	ActRate20-24	ActRate25-54
EarLeavTrain		0.3	0.38	0.47
Grads	0.3		0.68	0.51
ActRate20-24	0.38	0.68		0.48
ActRate25-54	0.47	0.51	0.48	

Correlation: OK / Weak

Skills Matching

	LTUnemp	UndEmpPart	OverQual	LoWage	QualMisMat
LTUnemp		0.18	0.61	0.18	0.24
UndEmpPart	0.18		0.26	0.34	0.66
OverQual	0.61	0.26		0.29	0.21
LoWage	0.18	0.34	0.29		0.54
QualMisMat	0.24	0.66	0.21	0.54	

Correlation: OK / Weak

Note: Numbers represent the Pearson correlation coefficients. Good correlations (i.e. Pearson correlation coefficients greater than 0.30 and lower than 0.92) are highlighted in light green. Correlations with low values (here <0.30) are written in grey. Correlations at risk of redundancy (here >0.91) are written in darker green. Correlations with meaningful negative value (here -0.30), when present, are highlighted in purple.

Source: European Commission's Joint Research Centre, 2022.

Table 4: Correlations between indicators and their aggregates

	SubPillar	Pillar	Index
RMS	0.6	0.74	0.48
SecAttain	0.61	0.42	0.48
PupTechRat	0.59	0.29	0.18
Training	0.81	0.7	0.4
Digital	0.8	0.72	0.47
VETs	0.54	0.42	0.39
EarLeavTrain	0.89	0.74	0.55
Grads	0.7	0.76	0.77
ActRate20-24	0.92	0.84	0.51
ActRate25-54	0.77	0.75	0.58
LTUnemp	0.67	0.61	0.81
UndEmpPart	0.85	0.76	0.34
OverQual	0.73	0.69	0.72
LoWage	0.64	0.56	0.31
QualMisMat	0.82	0.8	0.43

Correlation: High / OK / Weak

Note: Numbers represent the Pearson correlation coefficients. Good correlations (i.e. Pearson correlation coefficients greater than 0.30 and lower than 0.92) are highlighted in light green. Correlations with low values (here <0.30) are written in grey. Correlations at risk of redundancy (here >0.91) are written in darker green. Correlations with meaningful negative value (here -0.30) are highlighted in purple.

Source: European Commission's Joint Research Centre, 2022.

Correlation analysis between sub-pillars

Table 5 shows the correlation coefficients between sub-pillars, highlighting in boxes those of the same pillar. The statistical results confirm the conceptual framework:

- Within each pillar, correlations are positive (>0.30). Only pillar 1 shows smaller correlations among its elements, but always above the threshold;
- The elements of the first two pillars are all correlated positively, confirming their closeness (as also highlighted in developers report). The sub-pillars of the third pillar are correlated but more independent respect to the others belonging to P1 and P2. This is, probably, the reason of the behaviour of P3 respect to the others (see below for details).

In general, these results reassure on the coherence between the conceptual framework and the statistical structure of the pillars.

Table 5: Correlations between sub-pillars in the same pillar

ESI

	sp1.1	sp1.2	sp2.1	sp2.2	sp3.1	sp3.2
sp1.1		0.34	0.44	0.33	0.34	0.18
sp1.2	0.34		0.53	0.54	-0.02	0.3
sp2.1	0.44	0.53		0.69	0.29	0.28
sp2.2	0.33	0.54	0.69		0.14	0.03
sp3.1	0.34	-0.02	0.29	0.14		0.73
sp3.2	0.18	0.3	0.28	0.03	0.73	

Correlation: OK / Weak

Note: Numbers represent the Pearson correlation coefficients. Good correlations (i.e. Pearson correlation coefficients greater than 0.30 and lower than 0.92) are highlighted in light green. Correlations with low values (here <0.30) are written in grey. Correlations at risk of redundancy (here >0.91) are written in darker green.

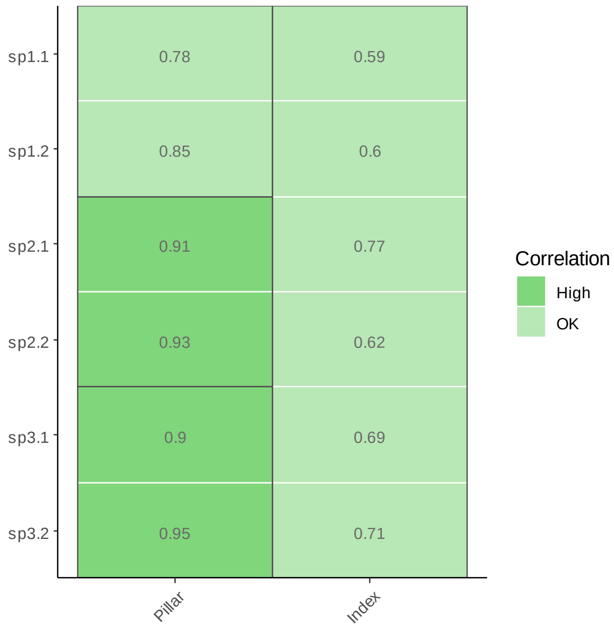
Source: European Commission's Joint Research Centre, 2022.

Correlation analysis between aggregates and index

The values in **Table 6** represent the correlation between the sub-pillars and their respective aggregates. This level of aggregation is the most important as it represents the consistency of the general concepts.

All the sub-pillars appear consistent and well allocated within their pillar, with very high correlation levels. The only note could be raised respect to the excessive level of correlation. In particular, the Sub-pillars of the second and third pillar show very high relation with their pillar. The excessive correlation is absent when observing the relation between sub-pillars and the index and between the sub-pillars of the same aggregate, resulting in a fairly balanced representation of all six sub-pillars.

Table 6: Correlations of sub-pillars with pillars and the index



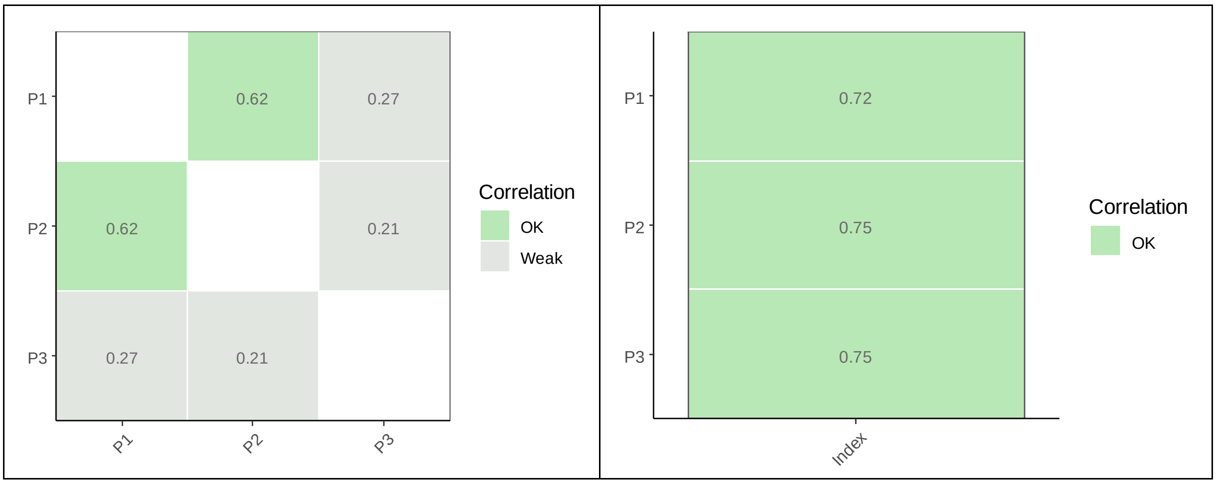
Note: Numbers represent the Pearson correlation coefficients. Good correlations (i.e. Pearson correlation coefficients greater than 0.30 and lower than 0.92) are highlighted in light green. Correlations with low values (here <0.30) are written in grey. Correlations at risk of redundancy (here >0.91) are written in darker green.

Source: European Commission's Joint Research Centre, 2022.

The direct correlation between pillars and with the index (**Table 7**) describes the balance at the top level of aggregation. Pillars are positively correlated, except for pillar 3 (Skills matching), which is the only pillar that has a correlation below 0.3 with all the others.

The weak correlation would have a direct effect on the relation of the pillar with the Index, but the conceptual framework is explicitly meant to balance the elements. Pillar 3 has higher weight and hence all the pillars have balanced correlation with the European Skills Index. Without using extreme weights, the developers achieve their aim of giving each pillar the same impact and proportion of representation in the ESI.

Table 7: Correlations between the pillars and with the index



Note: Numbers represent the Pearson correlation coefficients. Good correlations (i.e. Pearson correlation coefficients greater than 0.30 and lower than 0.92) are highlighted in light green. Correlations with low values (here <0.30) are written in grey.
Source: European Commission's Joint Research Centre, 2022.

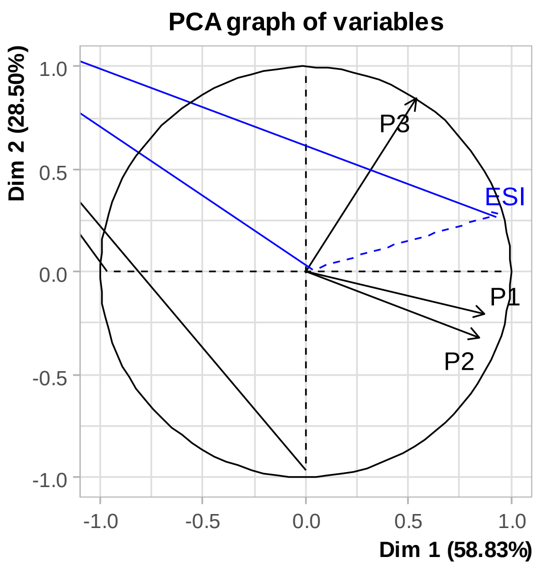
Principal components analysis of the ESI

As a further step in the analysis of statistical coherence, principal components analysis (PCA), was used to confirm the presence of one single statistical dimension among the three ESI pillars. Technically, the expectation here is that there is only one principal component with an eigenvalue greater than 1, or explaining more than 70% of the variance. In practice, the achievement of these thresholds suggests the presence of a common, unidimensional phenomenon underlying the pillars.

In the case of the ESI, the first principal component (PC1 or Dim 1) is the only one with an eigenvalue significantly higher than 1 (PC1 = 1.75, PC2 = 0.86) and explains about 59% of the total variation, while the second principal component explains an additional 29%. Figure **1** illustrates the projections of the ESI pillars onto the plane spanned by the first two principal components in a 'factor map'.

It is clear how pillar 3 shows a partially different concept respect to the other two (Figure 1). The correlation analysis show how this pillar is related only partially with the others, the decision of the developers of increasing its weight is the central for the role of the Matching pillar. The JRC-COIN can only suggest to give a specific attention to Skills Matching pillar, and embrace its difference respect to the other. It is important to notice how the developers are already doing it from previous editions, also thanks to previous collaboration with JRC-COIN.

Figure 1: Factor map of the three pillars and comparison with the overall ESI



Source: European Commission's Joint Research Centre, 2022.

Added value of the ESI

Sometimes a high statistical association among the main components of an index can be due to the redundancy of information. This is not the case with the ESI. For 32% or more of the countries included in the index, the ESI ranking, and any of the three pillar rankings, differ by 5 positions or more, which is 1/6 of the entire range (see This result suggests that the ESI ranking highlights aspects of countries' efforts that do not emerge by looking at the three pillars separately. In particular, pillar one is confirmed to be the most aligned component. Almost 70% of the countries show a difference in rank of 5 or fewer positions compared to the ESI.

Table 8). This result suggests that the ESI ranking highlights aspects of countries' efforts that do not emerge by looking at the three pillars separately. In particular, pillar one is confirmed to be the most aligned component. Almost 70% of the countries show a difference in rank of 5 or fewer positions compared to the ESI.

Table 8: Distribution of rank differences between pillar and ESI rankings

Shift respect to ESI	Pillar 1	Pillar 2	Pillar 3
More than 10 positions	9.7%	16.1%	16.1%
6-10 positions	22.6%	45.2%	35.5%
Above 5	32.3%	61.3%	51.6%
4-5 positions	12.9%	6.5%	9.7%
Up to 3 positions	48.4%	25.8%	29.0%
0 positions	6.5%	6.5%	9.7%

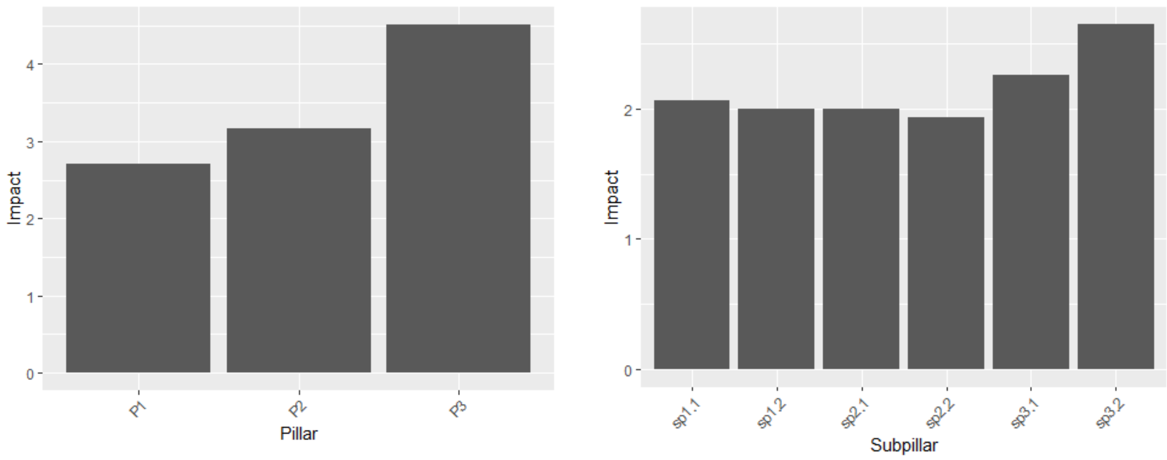
Source: European Commission's Joint Research Centre, 2022.

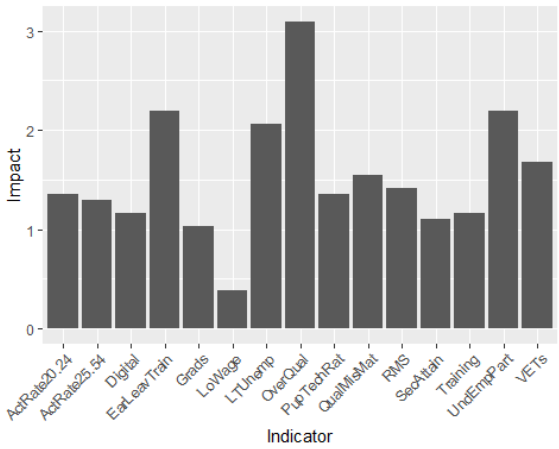
Impact of the components of the ESI

The study of the impact of the components (underlying indicator or aggregates) on the index is conducted by observing alternative simulated rankings based on the omission of one component at a time. One would typically expect to find some variability in rankings in such simulations. Otherwise, the omitted component would be proven to be irrelevant, adding no significant valuable information to the index.

Figure **2** outlines the average shifts in the ESI country rankings when one element is omitted at a time.

Figure 2: Average shifts in ESI country rankings when one element is omitted at a time





Source: European Commission’s Joint Research Centre, 2022.

- Among the elementary indicators, OverQual, EarLeavTrain, LTUnemp and UndEmpPart have the most significant impacts on the rankings, with an average shift of the absolute rank of 3.1, 2.2, 2.0 and 2.2 positions respectively. The omission of one of these indicators would cause a relevant change in the rankings of countries. The least impactful indicator is LoWage, this is probably due to its low weight (3% of weight in the computation of the ESI).

- The analysis of aggregates shows the importance of the third pillar showing how the pillar itself and its elements are the most impactful in the ESI. For example, exclusion of P3 would cause an average rank shift of 4.5. This result is partially due to the weight of the pillar and partially to its specificity.

Impact of modelling assumptions on the ESI results

A fundamental step in the statistical analysis of a composite indicator is to assess the effect of different modelling assumptions on the country rankings. Despite the efforts in the development process, there is an unavoidable subjectivity (or uncertainty) in the resulting choices. This subjectivity can be explored by comparing the results obtained under different – alternative – assumptions.

The literature on this topic[4] suggests assessing the robustness of the index by means of a Monte Carlo simulation and by applying a multi-modelling approach. This also assumes 'error-free' data as possible errors have already been corrected in the preliminary stage of the index construction before the audit.

The ESI, like most composite indicators, is the outcome of several choices. Among other things, these choices include: (i) the underlying theoretical framework; (ii) the indicators selected; (iii) the imputation of missing values; (iv) the weighting scheme; (v) the normalisation method; and (vi) the aggregation method. Some of these choices may be based on expert opinion or other consideration driven by statistical analysis or the need to ease communication or draw attention to specific issues.

This section aims to test the impact of varying some of these assumptions within a range of plausible alternatives in an uncertainty analysis. The objective is therefore to try to quantify the uncertainty in the ranks of the ESI, which can demonstrate the extent to which countries can be differentiated by their scores.

The modelling issues considered in the robustness assessment of the ESI are the aggregation formula, the data treatment (outliers and normalisation), and the pillars' weights. The following paragraphs deal with each of these in turn.

Aggregation formula. The ESI team opted for the geometric averaging of the three pillars, which implies a limited compensability, penalising all countries showing unbalanced performances. This approach can reward a country with generalised average results respect to countries with outstanding achievements in one pillar accompanied by under-performing values in the others. To assess the impact of this choice, the JRC included in the analysis a comparison with the arithmetic mean, which allows, on the contrary, perfect compensability between outstanding performance and weak results. The comparison of the two aggregation approaches should be able to highlight countries with unbalanced profiles, since the geometric mean tends to penalise low values, especially in the presence of other values that are not so low (unbalanced profiles).

Data treatment. In the ESI no normalised indicators shows the presence of outliers, and it is due to the normalisation based on conceptual bounds. To assess the impact of this choice, the JRC included in the analysis a comparison with a more common approach. That is, the treatment of outliers with windsorization, followed by minmax normalisation.

Weights. Monte Carlo simulation comprised 1 000 runs of different sets of weights for the three pillars constituting the ESI. The weights are the result of a random extraction based on uniform continuous distributions centred in the reference values plus or minus 20% of these values.

[4] Saisana, M., B. D'Hombres, and A. Saltelli. 2011. 'Rickety Numbers: Volatility of University Rankings and Policy Implications'. *Research Policy*, 40: pp. 165–177.
Saisana, M., A. Saltelli, and S. Tarantola. 2005. 'Uncertainty and Sensitivity Analysis Techniques as Tools for the Analysis and Validation of Composite Indicators', *Journal of the Royal Statistical Society* A 168 (2): pp. 307–323.

Four models were tested combining the different aggregation formulas and data treatment methods, which resulted in a total of 4 000 runs of simulations (1 000 simulated sets of weights for each combination of aggregation and treatment).

Table 9: Alternative assumptions considered in the robustness analysis

	Reference	Alternative
I. Aggregation formula	Arithmetic average	Geometric average
II. Data treatment	Goalposts	Windsorisation and Min-Max
III. Weighting system of pillars	Fixed weights	Varying up to 20%
Skills Development	0.30	U[0.240;0.360]
Skills Activation	0.27	U[0.216;0.324]
Skills Matching	0.43	U[0.344;0.516]

Source: European Commission's Joint Research Centre, 2022.

Note: while JRC-COIN suggests avoiding the use of weights above 0.5, here values up to 0.516 are allowed for sake of completeness of the simulation.

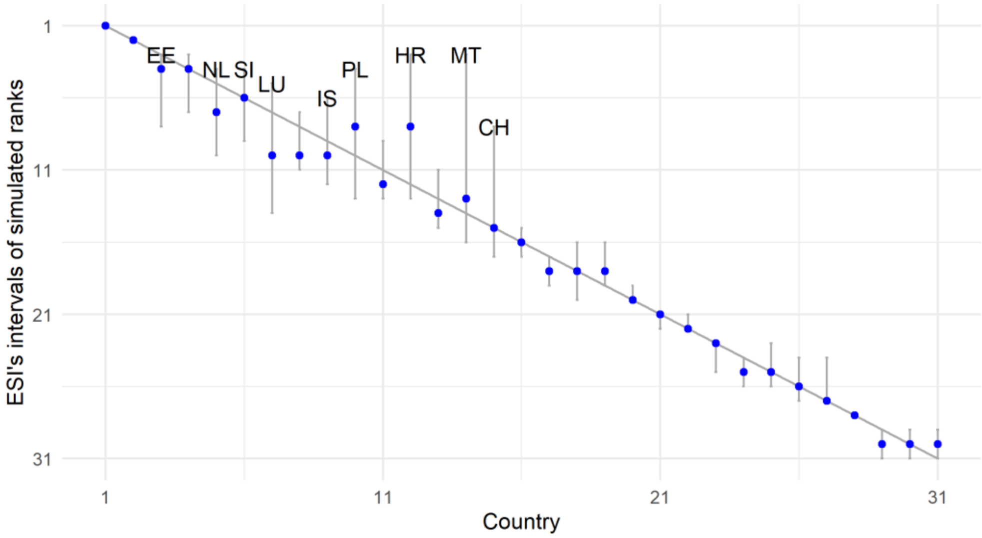
The main results obtained from the robustness analysis are shown in **Figure 3**, with median ranks and 90% intervals computed across the 4 000 Monte Carlo simulations. Countries are ordered from first rank to last according to their original ESI rank, where each blue dot represents the median rank among the iterations for each country, and error bars represent the 90% interval across all simulations, i.e. from the 5th to the 95th percentile.

ESI ranks are shown to be representative of a plurality of scenarios and robust to changes in the aggregation method, data treatment and pillar weights for most of the countries. Suppose one considers the median rank across the simulated scenarios as being representative of these scenarios. In this case, the fact that the ESI rank is close to the median rank (less than three positions away) for 94% of the countries suggests that the ESI represents a suitable summary measure of the four scenarios tested. Only Luxemburg and Croatia's median rank is more than three positions away from the nominal. Furthermore, the reasonable narrow intervals for most of the countries' ranks (less than 5 positions for about 73% of countries) imply that the ranks are also, for most countries, robust to changes in the pillars' weights and other modelling assumptions.

Nine countries show a simulated interval larger than – or equal to – 5 positions: Estonia, Netherlands, Slovenia, Luxemburg, Iceland, Poland, Croatia, Malta and Switzerland. The possible source of their instability is investigated in the following paragraphs. It is crucial to note how, in the ESI, nine countries show an ESI value between 59.8 and 61.4 (9 countries in 1.6% of the range), these makes those countries particularly susceptible to rank instability. Among the nine countries with similar score, ranking between the 7^{th} and 16^{th} position of the ESI, six are also among the nine with large interval, while the other three have rank interval of four.

Excluding the abovementioned aspect, country ranks in ESI are robust to changes in the pillar weights, data treatment, and aggregation formula for most of the countries considered. These ranks are robust enough to allow for meaningful inferences to be drawn, except for the nine countries showing very similar score, which may be taken into account when discussing the rankings of the Index. For full transparency and information, Table 9 reports the country ranks together with the simulated intervals (central 90 percentiles observed among the 4 000 scenarios) to appreciate better the robustness and behaviour of specific countries with respect to perturbations.

Figure 3: Robustness analysis on ranks (ESI rank vs median rank and 90% intervals)



Source: European Commission's Joint Research Centre, 2022.
Note: Labelled countries show an interval of at least five positions.

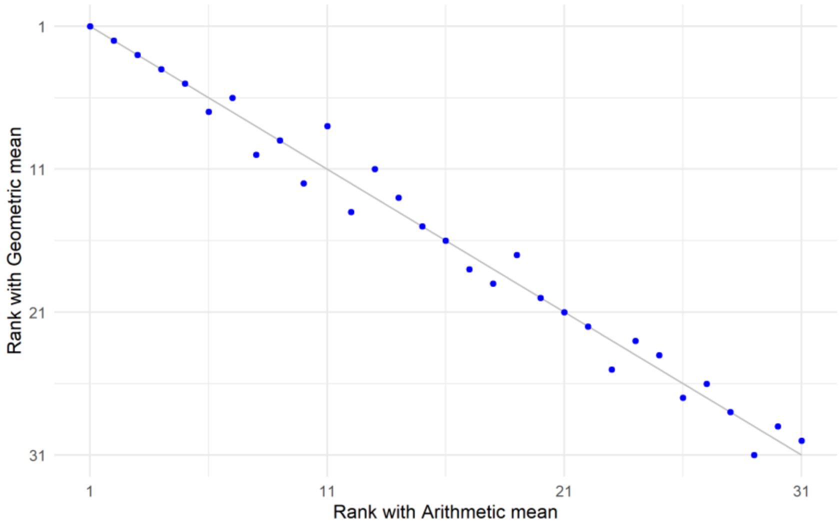
Table 10: ESI ranks and 90% intervals

Country	ESI Rank	Interval	Country	ESI Rank	Interval
Czechia	1	[1-1]	Latvia	17	[17-19]
Finland	2	[2-2]	Hungary	18	[16-20]
Estonia	3	[3-8]	Slovakia	19	[16-19]
Denmark	4	[3-7]	Lithuania	20	[19-20]
Netherlands	5	[4-10]	Belgium	21	[21-22]
Slovenia	6	[4-9]	United Kingdom	22	[21-22]
Luxembourg	7	[5-14]	France	23	[23-25]
Norway	8	[7-11]	Portugal	24	[24-26]
Iceland	9	[6-12]	Romania	25	[23-26]
Poland	10	[4-13]	Ireland	26	[24-27]
Sweden	11	[9-13]	Bulgaria	27	[24-27]
Croatia	12	[3-13]	Cyprus	28	[28-28]
Germany	13	[11-15]	Greece	29	[29-31]
Malta	14	[3-16]	Spain	30	[29-31]
Switzerland	15	[8-17]	Italy	31	[29-31]
Austria	16	[15-17]			

Source: European Commission's Joint Research Centre, 2022.

The uncertainty analysis is also complemented by a sensitivity exercise, in which the ESI rankings are compared with the rankings resulting from specific changes in the modelling assumptions. In **Figure 4**, it is possible to compare the ranks derived from ESI with the ranks that would have been obtained by changing the aggregation procedure from geometric to arithmetic mean. This comparison makes it possible to inquire whether the variability in the rank intervals originates from the modelling assumptions underlying the aggregation procedure. In the figure, the countries placed under the diagonal decrease in rank positions with the geometric mean. They are probably penalised by the geometric mean for their unbalanced profiles. No country show at least five positions of difference when comparing the two alternative formulas.

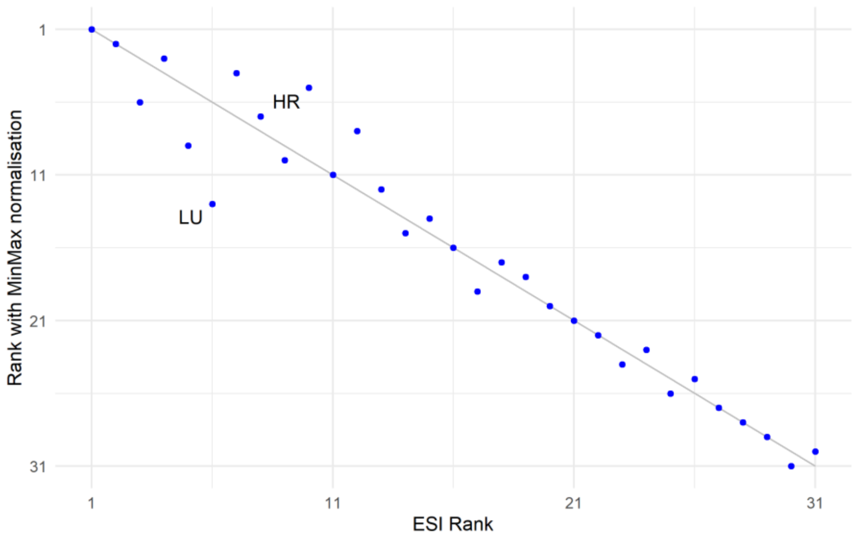
Figure 4: Sensitivity Analysis: Comparison of ranks according to arithmetic and geometric means



Source: European Commission's Joint Research Centre, 2022.
Note: Labelled countries show a shift of at least five positions between the two aggregation formulas.

Similarly, in **Figure 5** it is possible to compare the original ESI ranks with the ranks that would have been obtained by changing the normalisation method and outliers management. This comparison makes it possible to further investigate the source of the variability in the rank intervals. The ESI is clearly influenced by the data treatment, in particular in the case of Luxemburg and Croatia which are respectively penalised (-7 positions) and rewarded (+5 positions) by the MinMax approach. Nevertheless, this results is completely in line with the analysis suggested above. The robustness of the ranking is directly influenced by the closeness of the scores of countries in the upper half of the index. This is not true for the two best achievers, who keep their position in all alternative scenarios.

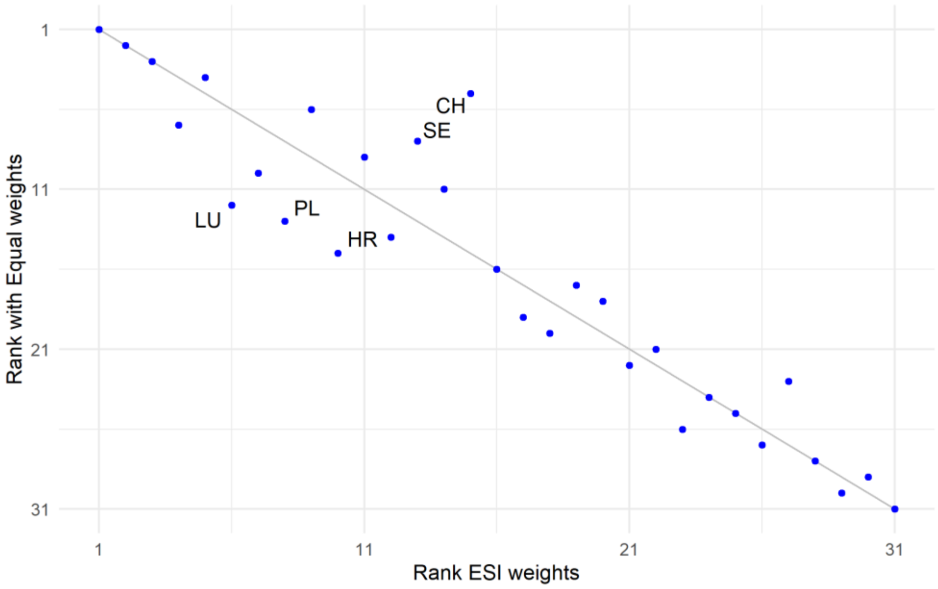
Figure 5: Sensitivity Analysis: Comparison of ranks according to Normalisation formula



Source: European Commission's Joint Research Centre, 2022.
Note: Labelled countries show a shift of at least five positions between the two data treatment methods.

Another test has been performed as check. In **Figure 6** it is possible to compare the original ESI ranks with the ranks that would have been obtained by using an Equal weighting scheme on the pillars. This comparison makes it possible to further investigate the source of the variability in the rank intervals. Changing the weights of the pillars, all the countries of the group highlighted above change their rank of at least 2-3 positions. With Luxemburg, Poland, Croatia, Sweden and Switzerland showing a change of five or more. It is obvious that such a change in weights has to have an effect. Nevertheless, it would be unfair to identify this instability as a structural problem of the ESI. There is no change to suggest in this sense, but a very special care in the interpretation of those ranks.

Figure 6: Sensitivity Analysis: Comparison of ranks considering ESI weights vs equal weights



Source: European Commission's Joint Research Centre, 2022.
Note: Labelled countries show a shift of at least five positions between the two weighting schemes.

Conclusions

The JRC statistical audit delves into the extensive work carried out by the developers of the ESI to suggest improvements in terms of data characteristics, structure and methods used. In addition, the analysis aims to ensure the transparency of the index methodology and the reliability of the results.

The data coverage of the framework is excellent. Most indicators contain no missing values for this edition because the developers rely on data from previous years. Only one indicator uses data from 2016.

A few indicators present outliers that are implicitly treated with goalpost normalisation by the developers. The analysis suggests that the ESI is statistically balanced within its pillars. There are mostly positive correlations between indicators and their corresponding sub-pillar, thus suggesting that most of the indicators provide meaningful information on the variation of the scores. This result is due to the decision to use weights to balance each element's role in the composite indicator, especially the pillars.

Indicators Pre-primary pupil-to-teacher ratio and VET students, respectively from sub-pillar 1.1. and 1.2, show shallow, when not negative, correlations with the other indicators in their sub-pillars. This may suggest that these indicators do not entirely cooperate with the others, and this may cause a conflict in results and reduce the impact of the aggregate to which they belong in the following aggregations.

The JRC analysed a series of different choices made during the construction of the index. The uncertainty analysis results reveal that the ESI is a robust summary measure for many countries. The simulated intervals are narrow enough for meaningful inferences to be drawn from the index on most countries; there is a shift of fewer than five positions for about 73% of the countries included in the index. This means nine countries have 90% confidence interval widths of at least five positions. Thus, their ranks vary significantly with changes in weights, data treatment and aggregation method, as also observed in the sensitivity analysis. Nevertheless, it would be unfair to identify this instability in the nine countries as a structural problem of the ESI since most have very similar scores (a difference of 1.6 or less on a 0-100 scale). JRC-COIN has no change to suggest in this sense, but very special care in interpreting those ranks.

Taking into account the points above, this audit confirms that the ESI is reliable and that the framework has good statistical coherence. The audit also acknowledges the significant efforts by the developers' team to obtain a balanced and transparent result.

References

Groeneveld, R. A. and G. Meeden. 1984. 'Measuring Skewness and Kurtosis'. *The Statistician* 33: 391–399.

OECD/EC JRC (Organisation for Economic Co-operation and Development/European Commission, Joint Research Centre). 2008. *Handbook on Constructing Composite Indicators: Methodology and User Guide.* Paris: OECD.

Saisana, M., B. D'Hombres, and A. Saltelli. 2011. 'Rickety Numbers: Volatility of University Rankings and Policy Implications'. *Research Policy* 40: pp. 165–177.

Saisana, M., A. Saltelli, and S. Tarantola. 2005. 'Uncertainty and Sensitivity Analysis Techniques as Tools for the Analysis and Validation of Composite Indicators'. *Journal of the Royal Statistical Society* A 168 (2): pp307–323.

List of Figures

KJ-NA-31019-EN-N

The European Commission's science and knowledge service

Joint Research Centre

JRC Mission

As the science and knowledge service of the European Commission, the Joint Research Centre's mission is to support EU policies with independent evidence throughout the whole policy cycle.

EU Science Hub
ec.europa.eu/jrc

@EU_ScienceHub

EU Science Hub - Joint Research Centre

Joint Research Centre

EU Science Hub

Publications Office
of the European Union

doi:10.2760/629770

ISBN 978-92-76-49387-7