



Pitfalls in data gathering to assess judiciaries

Marco Fabri¹

Abstract

This paper is divided into two parts plus some concluding remarks. The first one deals with some problems in comparing the number of judges, court personnel, and caseload in European judiciaries. Data come from the Commission for the Efficiency of Justice (CEPEJ) of the Council of Europe. The second part deals with some pitfalls in the data gathering carried out by the European Network of Councils for the Judiciary (ENCJ) in the attempt to measure judicial independence and accountability. Each case study brings some hints, summed up in the concluding remarks, that may be useful to improve both exercises.

Keywords: Comparative judicial systems, Judicial independence and accountability, Commission for the Efficiency of Justice (CEPEJ), European Network of Councils for the Judiciary (ENCJ).

1. Introduction

“When you can measure what you are speaking about, and express it in numbers, you will know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfactory kind; it may be the beginning of knowledge, but you have scarcely in your thoughts advanced to the state of Science, whatever the matter may be” (Kelvin 1883, p. 73).

Maybe this is a little extreme. This famous quote comes from a physicist who also stated that: *“I have not the smallest molecule of faith in aerial navigation other than ballooning”*², however the issue to measuring phenomena not only in the so-called hard sciences, but also in the social sciences is of paramount importance.

Court's values and performances are not exceptions. They are difficult to measure, but in recent years some more attention has been paid to *try to say something about them*. This is due to several reasons, such as, just to mention a few: a more managerial approach following the “new” public management wave with more pressure for better courts' performance and the assessment of judges' productivity, the enlargement of the European Union and the hope to share common values and similar performance, the economic crisis, the introduction of performance-based-budgets, the excessive length of the judicial proceedings and the need to understand the reasons behind it.

Measuring cannot disregard reliable data, which also are the basis for an informed fact based policy making at the supranational, national, or court level.

This paper analyses two remarkable efforts to collect data on courts' functioning and values. The first part of this work addresses the comparability of some data coming from the most comprehensive data collection on the functioning of European judiciaries carried out by the Commission for the Efficiency of Justice (CEPEJ) of the Council of Europe.³ These data are often used to make comparisons across countries, and they are also used in the European Union Justice Scoreboard. The second part takes into consideration the work carried out by the European Network of Councils for the Judiciary (ENCJ) on judicial independence and accountability, which has been finalized in the 2017 Report: *Independence, Accountability and Quality of the Judiciary*. This report is also at the basis of the paper by Frans Van Dijk and Geoffrey Vos (2018), which is the backbone of this special issue.⁴

1 Marco Fabri is Senior researcher at the Research Institute on Judicial Systems, National Research Council of Italy (IRSIG-CNR), Via Zamboni, 26, 40126 Bologna e-mail: marco.fabri@irsig.cnr.it. IRSIG's mission is to study and carry out research in the domain of justice, both as a scientific endeavour and as a service for policies to improve the administration of justice. IRSIG's scientific effort is also directed at facilitating the convergence of basic features of European Union justice systems. IRSIG's research deals with the “law in action”, governance, court management and ICT, working practices, design and assessment of judicial institutions and reforms in comparative perspective. www.irsig.cnr.it.

2 <https://zapatopi.net/kelvin/papers/letters.html#baden-powell> last visited November 2018.

3 The first part of this paper has also been presented at the seminar “Too few judges? Regulating the number of judges in society”, held in Onati the 30th of June 2016, which brought to the publication: *Methodological issues in the comparative analysis of the number of judges in European countries*, Onati Socio-Legal Series, v.7, n. 4.

4 Frans van Dijk and Geoffrey Vos, A Method for Assessment of the Independence and Accountability of the Judiciary, *International Journal for Court Administration*, Volume 9, nr. 3, 2018, p. 1-21. The ENCJ's report also deals with a section on “quality”, which will not be dealt with in this paper.

The CEPEJ and ENCJ exercises are different in many ways, but they have some similarities in the methodological approach that makes it interesting to analyze them both, in particular as far as the measurement problems are concerned.⁵

In the concluding remarks I will emphasize the importance of reliable datasets for making meaningful comparisons, and I will give some suggestions for improving both exercises. Sophisticated and fancy statistical or economic analysis can be fascinating, but they can be dramatically wrong due to the poor reliability of the data set that they may use. Comparative quantitative analyses in the judicial context should be approached carefully and humbly. They usually need to be supported by qualitative analyses that can improve the interpretation of data.

2. CEPEJ data on judges and court staff

The Committee of Ministers of the Council of Europe established the Commission for the Efficiency of Justice (CEPEJ, *Commission européenne pour l'efficacité de la justice*)⁶ in 2002 “for improving the quality and efficiency of the European judicial systems and strengthening the court users’ confidence in such systems”.⁷ CEPEJ’s mission is to propose pragmatic solutions as regards judicial organization, to enable a better implementation of the Council of Europe’s standards in the justice field, to contribute toward relieving the caseload of the European Court of Human Rights by providing States with effective solutions to prevent violations of the right to a fair trial within a reasonable time.⁸

CEPEJ’s work is organized in plenary sessions and ‘working groups’ that deal with some specific issues.⁹ In particular, one group deals with the “Evaluation of European Judicial Systems”, which collects quantitative and qualitative data on several topics of the judicial systems of the Member States. The collection is carried out through a questionnaire with more than 200 questions, which is filled in by the Member States national correspondents every two years. This is a unique collection of data and information about the functioning of European judicial systems since 2004, which has no equal in any other study carried out by international organizations or researchers.

The data collection organized by CEPEJ also is the basis for the ‘European Union Justice Scoreboard’. The Scoreboard is published every year since 2013 and compiles data only from the European Union countries. The Scoreboard is “an information tool aiming at assisting the EU and the Member States to achieve more effective justice by providing objective, reliable and comparable data on quality, independence and efficiency justice systems in all Member States”. (European Commission 2016, p. 1).¹⁰ Data on judges, court personnel, resources, and performance are usually provided for by CEPEJ.

Over the years, the CEPEJ Evaluation working group, the secretariat, and the countries’ national correspondents have constantly worked together to improve the reliability and consistency of the data collected. A quite detailed ‘Explanatory note’ has been drafted, and periodically amended, to “assist the national correspondents and other persons entrusted with replying to the questions”¹¹ and, in so doing, to ensure that concepts are addressed according to a common understanding.

Country replies to the questionnaire often come with further comments that are very informative and valuable to give a better interpretation of the numbers. This allows grasping several important features of the different systems. However, as I will discuss later, quite often, they are still not good enough to make ‘safe comparisons’ across countries.

Some other collections of data also have been carried out in the criminal field¹² but, as of today, the most current data on the functioning of European judicial systems is the CEPEJ collection.

The CEPEJ reports over the years are full of warnings to avoid superficial comparisons or, even worse, meaningless ranking of the Council of Europe Member States. In CEPEJ’s words: “Comparing quantitative figures from different States or entities, with different geographical, economic, and judicial background is a difficult task which must be addressed cautiously [...] Data cannot be read as they are but must be interpreted in the light of the methodological notes and comments. Comparing is not ranking” (CEPEJ 2016, p. 6).

⁵ This was the topic that it was asked me to address during a workshop to debate and assess the ENCJ 2017 Report. The workshop was organized by the ENCJ and the University of Utrecht; it was held in Utrecht, The Netherlands, 12-13 April 2018.

⁶ CEPEJ has 47 members appointed by the 47 Member States, and it is supported by a Secretariat within the Directorate General of Human Rights and Legal Affairs of the Council of Europe

⁷ http://www.coe.int/t/dghl/cooperation/Cepej/presentation/CEPEJ_depliant_en.pdf

⁸ http://www.coe.int/t/dghl/cooperation/Cepej/presentation/CEPEJ_depliant_en.pdf

⁹ Another working group is ‘The Saturn Centre for Judicial Time Management’, which is charged with collecting data and information about the length of judicial proceedings in the Member States, sharing practices, and developing tools and innovative ideas to improve the pace of litigation, and to prevent violations of the right for a fair trial within a reasonable time. A third working group is dealing with the broad subject of ‘quality’ of judicial systems, promoting practices and customers’ satisfaction surveys for the improvement of court functioning. CEPEJ also has an intense international cooperation activity in several European and not European States to support reform processes and implement Cepej recommendations. This activity has been constantly increasing in the last few years.

¹⁰ See also http://ec.europa.eu/justice/newsroom/effective-justice/news/160411_en.htm

¹¹ Cepej (2013), *Explanatory note to the scheme for evaluating judicial systems*, Cepej (2012)13 Rev, Strasbourg, France. Cepej (2015), *Explanatory note to the scheme for evaluating judicial systems 2014-2016 cycle*, Cepej (2015)2, Strasbourg, France.

¹² For example, European Institute for Crime, Prevention and Control, *European Sourcebook of Crime and Criminal Justice Statistics*, Helsinki, Finland, 2014. United Nations office on Drugs and Crime, *Developing Standards in Justice and Home Affairs Statistics. International and EU Acquis*, Vienna, Austria, UNODC, 2010. www.unodc.org/documents/southeasterneurope/CARDS_Developing_Standards_in_JHA_Statistics.pdf. Download 16 September 2016.

In this large data collection from different countries, consistency is a constant problem, and many efforts have to be done to define a clear and common 'unit of analysis' of what is required to be counted. Over the years, in the constant attempt to improve the consistency of the data collected, several definitions have been refined during the meetings among the national correspondents. Usually, these definitions are the result of a difficult process of including a large variety of differences in common categories concerning the legal system in the Member States.

Only an informed reading of the figures collected can avoid flaws; data cannot be passively taken but must be interpreted in the light of qualitative information that can better explain the meaning of that number. Large comments sometimes included in the replies are useful and informative for a better understanding of 'cold figures'.

Among the data collected there are the number of judges and court personnel employee in each judiciary. CEPEJ defines, a 'judge' as "an entrusted person with giving, or taking part in, a judicial decision opposing parties who can be either natural or physical persons, during a trial" (CEPEJ 2016, p. 81).

Figures about judges are collected by dividing them into three categories: *Professional judges* "those who have been trained and who are paid as such", and whose main function is to work as a judge and not as a prosecutor. The fact of working full-time or part-time has no consequence on their status"; *Professional judges practicing on an occasional basis* "paid as such", *Non-professional judges* "volunteers who are compensated for their expenses and who give binding decisions in courts" (CEPEJ 2016, p. 81).

Judges, and other court personnel are counted using the "Full-Time Equivalent" (FTE) method,¹³ to try to ensure consistent data and provide a starting point for any comparative analysis.

Unfortunately, the FTE method does not seem to have been applied by all the countries. As the CEPEJ report states: "only some states have indicated details (judges seconded to the ministries, judges on maternity leave, for instance)" (CEPEJ 2014, p. 156). Doubts about the appropriate use of the FTE method are evident from the comments about the data supplied by several countries. Inconsistency in the application of the FTE counting of judges can generate serious problems in any comparative analysis.

Another problem is related to the structure of the different European judiciaries and their jurisdictions. Courts' jurisdiction can be quite different in the various countries, and this can affect the counting of both judges and other court personnel significantly.

Usually, there are four major jurisdictions: civil, criminal, administrative, financial/tax. Generally speaking, in many countries civil and criminal matters are called 'ordinary jurisdictions', dealt with by 'ordinary courts' which can be internally divided into two or more specialized branches or sectors. Administrative and tax matters can be quite often two autonomous jurisdictions dealt with by different specialized courts, with dedicated judges and court personnel.

The jurisdiction considered in the counting of judges and court personnel are not always specified neither in the data collected nor in the comments. In some cases, countries have interpreted the counting in different ways. For example, France included in the total number of judges the administrative judges, while they have not been counted in Italy, although both countries have similar jurisdictions for administrative matters.

The three sub-categories of judges ("professional judge", "professional judges sitting occasionally", "non-professional judges") also raise some concern, because they have not been interpreted consistently. For example, in the 2014 report, Germany also included as 'professional judges' the number of professional judges sitting in courts part-time occasionally, and it is not clear whether they have been calculated using the Full-Time Equivalent.

The further problem is that in some countries, 'specialized' civil or criminal cases (e.g. labor, small claims, commercial, misdemeanor) can be dealt with by different kinds of judges, and it is not always clear if and how these judges have been counted. For large countries, this can alter the numbers reported significantly and jeopardize any cross-country analysis, unless an in-depth qualitative analysis is carried out.

For example, in France, the numerous commercial and labor cases are not dealt with by the civil courts but by specialized courts with 'judges/adjudicators' appointed by the business community. In France, the number of 'non-professional judges' does not include 'jurors or lay judges'; however in other countries such as, for example, Germany and Slovenia jurors or lay judges were included in the numbers.¹⁴ Norway reported both in 2012 and 2014, the same number of 43,000 "non-professional judges", but it is not explained why they are so many. Several others examples can be given from other countries on the same note.

Also, non-professional judges are indicated in gross numbers and not in Full-Time Equivalent. It could happen that a non-professional judge works only a few hours per year, whereas others can serve almost full time.

13 "The full-time equivalent indicates the number of persons working the standard number of hours; the number of persons working part time is converted to full-time equivalent. For instance, when two people work half the standard number of hours, they count for one "full-time equivalent", one half-time worker should count for 0.5 of a full-time equivalent" Cepej (2013), Explanatory Note to the Scheme for Evaluating Judicial Systems. 2014-2016 Cycle, Strasbourg, France, p. 2.

14 The CEPEJ *Explanatory note to the scheme for evaluating judicial systems 2014-2016 cycle*, CEPEJ (2015)2, p. 12, it is quite clear: "*non-professional judges* are those who sit in courts (as defined in question 46) and whose decisions are binding but who do not belong to the categories mentioned in questions 46 and 48 above. This category includes namely lay judges and the (French) 'juges consulaire'. Neither the arbitrators, nor the persons who have been sitting in a jury (see question 50) are subject to this question", but not all the National correspondents used this definition.

As shown, if it is difficult to count the number of judges in a consistent way across the different countries, the counting of “non-judges” is even more difficult.

The CEPEJ collects data about the ‘Rechtspfleger’¹⁵ in the countries that have such a position.¹⁶ Data are also collected on the number of ‘judicial advisors or registrars’, and ‘administrative staff’, however, these categories are not always very easy to differentiate from the other two categories listed in the CEPEJ questionnaire: ‘technical staff’, and ‘other kind’ of staff.¹⁷ It is also not clear in which category the ‘law assessors’ or ‘law clerks’ are reported to fit. They are judges’ assistants, assigned with for example legal research or drafting court decisions. Such functionaries are employed in several countries such as the Nordic countries and Switzerland.

To sum up, on the one hand, CEPEJ data collection on the number of judges and court personnel is an outstanding effort and it gives a first interesting overview of the different judicial settings and basic figures. On the other hand, some major problems in the comparability of data have to be mentioned. In particular: a) FTE counting does not seem to be used consistently by all the reporting countries; b) the figures on judges can mix up different jurisdictions which are not consistent across countries; c) the counting of judges in the three categories proposed (professional, professional practicing on occasional basis, non-professional judges) is not consistent across countries, d) all these problems are even more severe in the counting of “non-judges”.

3. CEPEJ data on caseload

Courts’ caseload data are collected for incoming, pending, and disposed cases. These data are the basis for three basic indicators of court functioning: Clearance rate (disposed cases/incoming cases x 100); Case turnover ratio (disposed cases/pending cases x 100), Forecast disposition time (365/case turnover ratio) (CEPEJ 2014, p. 191 and CEPEJ 2016, p. 185).

Also for this data collection, the main concern is to have a clear definition about incoming, disposed, and pending cases, and it is not that simple as it might appear. There is not a common consolidated definition of what a civil case or a criminal case is, as well as there is not a commonly shared interpretation of disposed and pending cases across countries.

Data on civil cases are also collected for seven case categories: civil and commercial litigious; civil and commercial non-litigious; non-litigious enforcement; non-litigious land registry; non-litigious business; administrative law cases; other cases.

Once again, the different settings of European judiciaries generate some significant problems. For example, in Austria, court statistics do not allow a distinction between litigious and non-litigious cases, and then figures supplied are just an estimation. In the Netherlands, it is not possible to make a distinction between litigious and non-litigious incoming cases, but this is only possible after the case has been disposed of. The Czech Republic reported the number of electronic payment orders in the category ‘other cases’, while other countries in the ‘non-litigious civil and commercial’. In Norway, courts also have functions of public notaries and marriages that have been estimated at 25,000 per year. These cases have been included in the ‘other cases’ category. In Lithuania and Denmark, administrative cases have been included in the ‘other cases’ category. The same category has been used for insolvency registry and labor cases in Hungary, but not in several other countries.

In general, it is not always clear if and in which category administrative cases have been counted. In some countries, civil cases can include administrative cases, while in others they are counted in a different category.

15 As reported in the CEPEJ report 2014 (p. 175), the *Rechtspfleger* is “an independent judicial body, anchored in the constitution and performing the tasks assigned to it by law. [...] The *Rechtspfleger* does not assist the judge, but works alongside the latter and may carry out various legal tasks, for example in the areas of family and guardianship law, the law of succession, the law of land registry and commercial registers. He/she also has the competence to making judicial decisions independently on granting nationality, payment orders, execution of court decisions, auctions of immovable goods, criminal cases, the enforcement of judgments in criminal cases (including issuing arrest warrants), orders enforcing non-custodial sentences or community service orders, prosecution in district courts, decisions concerning legal aid, etc.; the *Rechtspfleger*, to a certain extent, falls between judges and non-judge staff, but does not have the status of judge;

16 16 countries reported to have some kind of staff close to the definition given to *Rechtspfleger*, these countries are mainly in central Europe and the Balkans, in somehow was inspired by the German tradition.

17 *Explanatory note to the scheme for evaluating judicial systems 2014-2016 cycle*, CEPEJ (2015)2, p. 13 “*Non-judge (judicial) staff* directly assist a judge with judicial support (assistance during hearings, (judicial) preparation of a case, court recording, judicial assistance in the drafting of the decision of the judge, legal counselling - for example court registrars). If data has been given under the previous category (*Rechtspfleger*), please do not add this figure again under the present category. *Administrative staff* are not directly involved in the judicial assistance of a judge, but are responsible for administrative tasks (such as the registration of cases in a computer system, the supervision of the payment of court fees, administrative preparation of case files, archiving) and/or the management of the court (for example a head of the court secretary, head of the computer department of the court, financial director of a court, human resources manager, etc.). *Technical staff* are staff in charge of execution tasks or any technical and other maintenance related duties such as cleaning staff, security staff, staff working at the courts’ computer departments or electricians. *Other non-judge staff* include all non-judge staff that aren’t included under the categories 1-4”.

Comments reported by the countries in the CEPEJ reports show how hard it would be to make any comparison about the figures reported in each category based on the different definitions and then the counting, given by each country.¹⁸

These few examples show that it is misleading to make comparisons across countries without considering the specificity of the court's performance figures. This can be even more misleading if courts' performance is put in relation to personnel resources, due to the difficulties in comparing the number of judges and court staff mentioned in the previous section.

In addition, even though the reliability of data has improved, there are still some concerns about the quality of the data collected. Some checks in each country should be done to verify the correspondence to what is supposed to be reported and what is done in practice, also to assess the possible error rates.

Therefore, notwithstanding the remarkable efforts carried out by CEPEJ, there are still some significant problems in the collection of data about the number of judges, court personnel, and caseload, which jeopardize a reliable comparative analysis across countries.

4. The ENCJ's exercise to measure judicial independence and accountability

The second part of this paper deals with the ENCJ's recent exercise to measure judicial independence and accountability in the EU Member States through a scorecard. In particular, I will focus, as requested by the editors of this special issue, on some pitfalls related to the measurement method used to assess the variables 'independence' and 'accountability'.

The ENCJ "unites the national institutions in the member States of the European Union which are independent of the executive and legislature, and which are responsible for the support of the judiciaries in the independent delivery of justice". The strategic objectives for 2018-2021 are to "provide support for the independence, accountability, and quality of judiciaries in Europe [...] to promote access to justice in a digital age [...] to strengthen mutual trust among the judiciaries in Europe".¹⁹

Among the 2016-2017 ENCJ activities, a specific group dealt with the challenge to measure judicial independence and accountability in each participating country with a scorecard. The scorecard is calculated by adding the points that each judiciary scored on several indicators and sub-indicators. The scorecards, according to the ENJC, should stimulate judiciaries and governments to policies' design and implementation on judicial independence and accountability.

The 2017 report by the ENCJ, along with the paper by Van Dijk and Vos, are the backbone of this special issue. Van Dijk and Vos' paper is based on the 2017 ENCJ report; but it further explains the approach in a broader theoretical framework.

The work carried out is now in the planned phase to be externally reviewed by the scientific community, and by ENCJ's partners (ENCJ 2017, p. 7). This paper is part of this external review.

The ENCJ work is mainly based on two questionnaires designed, submitted, and analyzed by the ENCJ. The first one was submitted to Judicial Councils, or other administration of justice bodies, and it was aimed to collecting data on the formal legal setting of judicial independence and accountability. The second questionnaire aimed to gather data on the judges' self-perceptions of their independence. It was submitted to EU judges, through the national Judicial Councils. 11,712 judges from 26 countries participated in the survey (ENCJ 2017, p. 4 and 34).²⁰

Data from other European surveys were also used to calculate an indicator about the citizens' perception of independence. Quite interestingly: "The correlation between this indicator and the perceived independence by judges is high, showing that the perceptions of judges of their actual independence are fairly in agreement with those of citizens" (ENCJ 2017, p. 21).

The ENCJ effort is different from the CEPEJ's exercise for objectives, magnitude, and resources, but they have quite a few similarities. For example, it is similar the approach to data collection. CEPEJ employs national correspondents, who are mainly members of Ministries of Justice or dedicated Courts' agencies (Northern European model). ENJC has collected data through Judicial Councils, or other institutional bodies including sometimes Ministries of Justice in countries without a Council.

Both institutions share the same concern about the use of the data collected. ENCJ has produced "country profiles" with the data collected, and it warns that they "must be used with circumspection, due to the unavoidable arbitrariness of some categorizations and scoring [...] determining what is good and what is less good practice is based on shared values and ideas within the ENCJ, and such is not absolute science. Still, the profiles need to be taken seriously to set priorities for change" (ENCJ 2017, p. 84).

18 Cepej (2014), p. 219-220 "Austria: misdemeanours and /or minor criminal cases include all offences fined or punish with a prison sentence up to one year and must not be decided by a jury [...] Belgium: severe criminal offences include cases dealt with by first instance ordinary criminal courts. Misdemeanours/ minor criminal cases include cases dealt with by the Police Court [...] Denmark: severe criminal cases are defined as those cases where a lay assessor participates or cases dealt with by a jury; non-contested plea cases (guilty plea) are included as severe criminal cases. Misdemeanours and/or minor criminal cases are typically cases where the maximum sentence is a fee [...] Poland: misdemeanours cases includes the offences punishable by a maximum penalty up to 1 month of detention or a fine (or both). All other criminal cases are severe cases. Statistics contain also the so called 'organisation cases' which do not deal directly with crimes". See also comments reported in the 2016 dynamic data-base (2014 data).

19 ENCJ members are 23 national institutions. There are also 14 observers Institutions, mainly Ministries of justice and Court Administration organization of countries that do not have a judicial council. <https://www.encj.eu/> last visited August 2018.

20 This exercise is planned to take place every two years. (ENCJ 2017, p. 7).

As it is for CEPEJ: “The indicators have not been developed to create rankings of judicial systems, but can be used to discuss the strengths and weaknesses of judicial systems. Readers of the report are advised to treat the comparison of data from different countries with various geographical, economic and legal backgrounds with great caution” (ENCJ 2017, p. 19).

The answers collected have been codified by the ENCJ working group with a score for each ‘option’ (sub-indicator), following some self-elaborated guidelines. The scorecard for each judiciary is the sum of the points of each indicator.²¹

More in detail, scores in the questionnaire answered by Judicial Councils have different minimum and maximum (e.g., 0-1, 0-3, 0-5, 0-10, 0-15), depending on the indicator. Therefore, “for all the indicators a high score is good and a low score is bad” (ENCJ 2017, p. 19). For example, question 8e “Can the management of the court exert pressure in individual cases on the way judges handle their cases with respect to the timeliness/efficiency of judicial decisions?” scores 3 points if the answer is ‘no’, it scores 0 points if it is ‘yes’. Question 7a “Can a judge be transferred (temporarily or permanently) to another judicial office (to other judicial duties, court or location) without his/her consent?” Scores 15 points if the answer is ‘no’, it scores 0 points if the answer is ‘yes’.

These ranges (min-max points) decided by the ENCJ’ working group, which are a kind of weight for each indicator, may be further debated because they are the key factors to determine the final score of each judiciary. For example, some particular issues for their relevance to judicial independence could have a negative score, while in the current point system the minimum score is zero.

It is also a point of attention that some questions (for example number 6b)²² can have multiple answers,²³ and it is not clear how the score has been calculated.

As already mentioned, also this data collection relies on self-compilation, which raises some concerns about the genuineness, and then reliability, of some answers.

Van Dijk and Vos are aware of this risk (p. 23): “In some situations, this may lead to self-serving bias. This is difficult to avoid, but a group of experts from *within the ENCJ* responded to queries about the interpretation of the questions, checked the logic and plausibility of the answers, and resolved *ambiguities* [roman added], ensuring as far as possible that the indicators were measured uniformly and correctly”.

It would be interesting to know more about the *ambiguities* and the problems encountered, to understand better which items were more difficult to interpret. In the future, it may be beneficial to have in the group of experts someone who is not necessarily *within the ENCJ*, who may help to improve the quality of the questions and to check the consistency of the answers.

A couple of specific examples from the Italian case, can be useful to show the pitfalls that may be due to the self-compilation of the questionnaire by Judicial Councils without any further check.

Some questions are just related to legal matters about the judiciary and, in this respect, the margins of interpretation are limited, although still possible. For example, question 7a “Can a judge be transferred (temporarily or permanently) to another judicial office (to other duties, court or location) without his/her consent?” was answered “yes” by the Italian Judicial Council. However, as far as I know, the cases in which this could happen are exceptional and very rarely occur (Contini et al. 2017, Di Federico 2012 and 2005). The answer given by the Council was just legally oriented, missing the ultimate issue raised by the question.

Other questions asked to make an assessment about different issues. In these cases, answers are more discretionary, and maybe questionable, as it is the related score of the indicator. For example, question number 3a asked: “Is the funding of the judiciary sufficient as to allow the courts to handle their caseload” (and other items). The Italian Council answered that the funding of the judiciary is sufficient to handle their caseload but, unfortunately, it is well known that many Italian courts still suffer a dramatic excessive length of case disposition and a huge number of pending cases, so it is debatable that funding is sufficient. A similar comment can also be made about the Council’s answer that there is sufficient funding to “facilitate judges and other personnel in matters of IT systems, building, etc.”.

Another example comes from question 5e item 3: “Is the promotion of judges is solely based on merit?” The Council institutional and formal answer was “yes”. However, it is well known that the powerful fractions of the Italian Magistrates Association (*Associazione Nazionale Magistrati*) play a very significant role in judges’ and public prosecutors’ promotions, in particular for the selection of Court Presidents and Chief Prosecutors (Fabri 2016, Di Federico 2012).

This (inter)subjective elements as pointed out by Van Dijk and Vos in their conclusion (p.31) “could make outcomes dependent on the incentives of those who conduct the evaluation”, only partially mitigated by the fact that “the indicators are based on the formal arrangements in a country and otherwise observable phenomena, and can be readily checked by any knowledgeable observer”. I think this can contribute to an improper scorecard assessment of each judiciary.

These examples raise two concerns about the questionnaire submitted. The first one is more specifically related to the formulation of questions, which may lead to different interpretations, and then to improper and a wrong final scores.

21 See for an overview of the questionnaire and the indicators the appendix of the article of Van Dijk and Vos in this Issue.

22 Question 6b is: “Which is the competent body to make the following decisions in the context of disciplinary procedures against judges? It has 6 different items, and for each item the competent body can be “the judiciary”, “the executive”, “the legislature”.

23 For example, this is the case of Italy where in two answers (item d and f) two competent bodies were indicated.

The second concern is broader. Generally speaking, questions posed through the questionnaire focussed on the formal legal setting of the judiciary. Therefore, answers do not take into consideration the real practices, which should be the core of the measurement of judicial independence and accountability.

Data collected through the survey submitted to EU judges also raise some concern about the methodology used and therefore the reliability of its outcome. For example, it is not clear who are the “judges” who answered the questionnaire. As the previous section showed, the definition of a judge is quite different in every country and different kind of “adjudicators” may have a different opinion about judicial independence and accountability. As the data show, the number of replies for each country is quite different, in particular considering the various sizes of each judiciary. General “averages” are necessarily affected by these differences, as well as the reliability of some country scorecards.

Some concerns also deal with the relationship between concepts and indicators proposed by the report and by Van Dijk and Vos’s paper. On the one hand, the distinction between “formal requirements” of independence (usually arranged by law), and “perception of independence” (how judges and the citizens perceive their judiciary independent) is convincing. Convincing also is the association between “formal requirements”, and the so-called “*de jure*” independence, which the authors of the report call “objective independence”. On the other hand, the association between the “perceptions of independence” is much less convincing as a proxy of the “*de facto*” independence, which the authors call “subjective independence” (Hayo and Voigt 2007). How the formal requirements are met in practice (*de facto* independence) is of paramount importance as mentioned in Van Dijk and Vos’s introduction (p. 1): “it is important to have a clear understanding not only of what is required for judicial independence and accountability but also to what extent these requirements are met in practice”. At this stage, the ENCJ study does not allow to say much on how judicial independence and accountability are dealt with in practice, also due to the methodology used to collect data.

Alternative ways to “measure” *de facto* independence are briefly presented and then excluded, with good reasons, by Van Dijk and Vos’s paper, p. 9), but still to base the measurement of *de facto* independence only on perceptions does not match the variable that ENCJ wants to measure. Day by day practices have to be observed to check if and how the legal framework is applied in practice. As Van Dijk and Vos point out (p. 25) “In Bulgaria, many of the formal arrangements are state of the art [...] There is a strong legal basis for an independent judiciary [however] The perceptions of independence lack behind, which is likely to reflect reality”. On the contrary, (p. 26) countries such as Denmark, Sweden, and Finland with “weak formal arrangements of independence and accountability go together with positive perceptions in society and among judges about independence”. Judges’ and citizens’ perceptions are not good enough proxies to explain how judicial independence and accountability work in practice (*de facto*). Perceptions and practices are simply two different things that need to be investigated in different ways.

I share what is written in the ENCJ conclusions (p. 31): “As to the way forward, making a *systematic assessment of the level of independence and accountability achieved in practice* [roman added] by the national legal system is a crucial starting point for improving justice systems across the EU”. This systematic assessment of the level *achieved in practice* still need some more efforts by the ENCJ.

I also have some concerns about the definition and the measurement of accountability. The ENCJ report uses the concept of “accountability” and “transparency” as interchangeable. It is understandable that for the sake of simplicity and measurability, the two concepts have been associated, but it is intuitive that transparency is just one part of the broader concept of accountability (Contini and Mohr 2007).²⁴ Therefore, the current measurement of accountability may consider some more indicators in the future.²⁵

5. Concluding remarks

I believe that both parts of this paper show the importance to found any analysis on robust and reliable data, whatever they may be quantitative or qualitative. If data are not reliable, any analysis is undermined in its foundation, and it will be necessary result in misleading outputs and outcomes. This is particularly risky for analyses that should support the policymakers, as argued by ENCJ, which may be erroneously induced to design and implement policies based on flawed data. This is even truer for comparative analyses, which are appealing but complex, due to the variety of institutional settings, legal rules, and practices.

As this paper has shown, there are still some severe problems in the comparability of data about the number of judges, court personnel, and court-performance across European countries. CEPEJ is still struggling to improve the consistent interpretation of its questionnaire through a constant revision of the so-called ‘Explanatory note’, and the organization of meetings among the

²⁴ Contini and Mohr (2007, p. 30) quote Herbert Simon definition of accountability as: “the combination of methods, procedures and forces determining which values are to be reflected in administrative decisions”. As they argue in their paper, transparency is a pre-requisite of accountability, which has other components such as the correct application of law and procedures, the correct use of resources, a fair treatment of parties.

²⁵ Van Dijk and Vos in section IV of their paper already wider the concept of accountability: “Systematic accountability requires transparency of procedure (e.g., case allocation and complaint procedures) and transparency of performance (e.g., timeliness and efficiency). Individual judges are accountable if their decisions are publicly available, properly reported and explained”.

national correspondents. A similar approach, on a limited scale, is also used by the ENCJ.²⁶ Data collected should be supplemented by more qualitative information to better understand their comparability.

A possible development towards a prudent and progressive comparability of the data across countries could be the creation of clusters of judiciaries, to have groups that are less different in, at least, some of their constitutive features.

Due to the complexity of the different justice systems and their context conditions, the focus of the comparative analysis should be on a limited number of 'comparable cases', meaning cases that are less different among themselves (Lijphart 1971). This does not necessarily imply to compare judiciaries that are in a same geographical area because the emphasis is on the features of each judiciary and not on their geographical position. The grouping process should adopt a 'fuzzy logic',²⁷ to be flexible in the establishment of comparable clusters.²⁸ As a starting point, the cluster could include just a few countries and then progressively be enlarged, applying a "fuzzy membership" (Ragin C. and P. Pennings 2005).

The key variables considered to select the 'unit of analysis' to be compared across judiciaries have to go through a qualitative and flexible 'calibration process'²⁹ that allow identifying the judiciaries that can be grouped in the same cluster. This process is necessarily qualitative, it uses substantive knowledge, with approximate reasoning to make explicit.

In both exercises, data have been collected through national institutions/correspondents. It could be useful to involve researchers and academics in the validation process of the data collected. In particular, the ENCJ should avoid the risk of insularity in their research activities. There is always a resource problem, but I believe the ENCJ' activities will benefit from the involvement of scholars or other experts, as well as researchers can learn from the practitioners' perspective.

The ENCJ effort is remarkable; however, it also suffers from quite a few specific methodological problems. Some of them can probably be solved with some tune-up work (e.g., ambiguity in the formulation of the questions or some pre-set answers, scores coding, involvement of "external" experts), some other will request more resources and much more effort (e.g., sample representativeness, weak connection between the indicators and the variable that has to be measured).

In this regard, the questionnaire filled out by Judicial Councils, or other institutions should be made public for two main reasons. The first one is that they are full of interesting pieces of information that can be very useful for researchers, practitioners, and policymakers. The second one is that this will add transparency to the study, giving the possibility to everyone to assess the reliability of data and analyses. Comments on single countries by researchers or other experts should be welcomed by the ENCJ, since they may improve the quality of the study.

Some more emphasis should be put on the different institutional settings of judiciaries and Judicial Councils, because most probably they affect the scoring and the level of independence and accountability. For example, there are countries such as Italy and Romania where the Judicial Council serves for both judges and public prosecutors. If and how different settings should be considered in the questionnaire, in the scoring, and in the data analysis, are matters of further investigation.

The ENCJ report (p.20), as well as Van Dijk's and Vos's paper (p.22), state that "judicial self-government, balanced by accountability, is desirable". I think this statement is questionable, also based on the data collected with this exercises by the ENCJ. There is no evidence that judicial self-government is per se a guarantee of higher judicial independence and, even less, of higher judicial accountability (Bobek and Kosar 2013, Di Federico 2012). Maybe it is *desirable* for the sake of judges, but this is a different argument.

This brings me to a more general comment on the usage in the ENCJ report of the terms "justice system", "judicial system" and "judiciary". I cannot deal here with this subject in detail, but these terms should be framed better in the ENCJ's report, because they may lead to some misunderstandings. Some explanation is needed, similar to the one that CEPEJ gives about the budget of the judiciary, in which 'the judiciary' refers only to courts.

26 A similar approach has been followed by the "Court Statistic Project" in the United States (<http://www.bjs.gov/index.cfm?ty=dcdetail&iid=283>). "The Court Statistics Project (CSP) provides a systematic means to develop a valid, uniform, and complete statistical database that details the operation of state court systems. It provides high-quality, baseline information on state court structure, jurisdiction, reporting practices, and caseload volume and trends [...] The CSP fulfills the vital role of translating diverse state court caseload statistics into a common framework that all states use when establishing their respective goals and policies. Information for the CSP's national caseload databases comes from published and unpublished sources supplied by state court administrators and appellate court clerks. The CSP has evolved since 1975 by providing more consistent definitions of key terms and parameters for counting. The State Court Model Statistical Dictionary (updated version published in 1989) provided the first set of common terminology, definitions, and usage for reporting appellate and trial court caseloads".

27 I use here the theory of 'fuzzy logic' in a broad sense. It means that clusters of countries should be classified without sharply defined boundaries but based on some constitutive features and the qualitative information collected to establish quite consistent groups of countries for each issue to compare. These cluster groups can vary if the issue to compare varies because the constitutive features of the countries can vary as well. For example, on the number of judges could be possible to create a cluster of 3 or more countries that have several similarities on that issue, but the same 3 countries could not be compared on the number of court personnel.

28 "Comparatists have certainly learned that legal principles are not absolute [...] and the conflict of values has to be reconciled not by the rigor of artificial logic, but by a flexible and pragmatic recognition that [...] a compromise solution has to be formed" (Cappelletti 1983, p. 13).

29 Ragin C., *Fuzzy Sets: Calibration Versus Measurement*, University of Arizona Paper, www.u.arizona.edu/cragin/fsqca/download/calibration. Download 12 October 2016, p. 1 "Calibration is a necessary and routine research practice in such fields as chemistry, astronomy, and physics. In these and other natural sciences, researchers calibrate their measuring devices and the readings these instruments produce by adjusting them so that they match or conform to dependably known standards".

Some more attention should also be addressed to collecting more data on “internal independence” which seems quite neglected in comparison to indicators and sub-indicators used to assess “external independence”.

Moreover, some more information about the sample of respondents to the questionnaire is needed to assess the findings and the score reliability of the survey.

Finally, in some countries more than in others, it would be interesting to carry out a similar but reverse exercise, *to analyze the independence of the executive and the legislative from their judiciary and public prosecutor's office*, but this is something that, I suppose, it is not going to be carried out by the ENCJ.

References

M. Bobek and D. Kosar, *Global Solutions, Local Damages: A critical Study in Judicial Councils in Central and Eastern Europe*, Research Paper in Law n. 7, College of Europe, Bruges, 2013.

M. Cappelletti M., *Who Watches the Watchmen? A comparative study on Judicial Responsibility*, “American Journal of Comparative Law”, 183, 1983, vol. 31, pp. 1-62.

Cepej, *European Judicial Systems. Efficiency and Quality of Justice*, Cepej Study n. 20, Strasbourg, France, Council of Europe, 2014.

Cepej, *Explanatory note to the scheme for evaluating judicial systems 2014-2016 cycle*, Cepej (2015)2, Strasbourg, France, 2015.

Cepej, *European Judicial Systems. Efficiency and Quality of Justice*, Cepej Study n. 23, Strasbourg, France, Council of Europe, 2016.

F. Contini and R. Mohr, *Reconciling independence and accountability in judicial systems*, *Utrecht Law Review*, 3, 2, 2007. pp. 26-43.

G. Di Federico, *Recruitment, Professional Evaluation and Career of Judges and Prosecutors in Italy*. in G. Di Federico (ed.) *Recruitment, Professional Evaluation and Career of Judges and Prosecutors in Europe: Austria, France, Germany, Italy, The Netherlands and Spain*. Bologna, Lo Scarabeo, 2005.

G. Di Federico, *Judicial Independence in Italy*, in A. Seibert-Fohr (ed.), *Judicial Independence in Transition*, Springer, Heidelberg, 2012, pp. 357-401.

European Commission, Directorate General for Justice, *The 2016 EU Justice Scoreboard*, Bruxelles, 2016. p. 1 See also http://ec.europa.eu/justice/newsroom/effective-justice/news/160411_en.htm

European Network of Council for the Judiciary, *Independence, Accountability and Quality of the Judiciary. Performance indicators 2017*, www.encj.eu/node/250, Download 4 April 2017.

European Institute for Crime, Prevention and Control, *European Sourcebook of Crime and Criminal Justice Statistics*, Helsinki, Finland, 2014.

M. Fabri, *Clash of visions: regulating judges and prosecutors in Italy*, in R. Devil and A. Dodek (eds.), *Regulating Judges. Beyond Independence and Accountability*, Elgar, Cheltham, 2016, pp. 245-261.

B. Hayo and S. Voigt, *Explaining de facto judicial independence*, “International Review of Law and Economics”, 27, 2007, pp. 269–290.

W. T. Kelvin, *Electrical Units of Measurement*, A Lecture delivered at the Institution of Civil engineers, May 3, 1883, in *Popular lectures and Addresses*, London, MacMillan, 1883, pp. 73-136, available at: <https://archive.org/details/popularlecturesa01kelvuoft/page/n7>

A. Lijphart A., *Comparative Politics and Comparative Method*, “The American Political Science Review”, vol. 65, n. 3, 1971, pp. 682-693.

R. Mohr and F. Contini. *Conflicts and Commonalities in Judicial Evaluation*, “Oñati Socio-Legal Series”, vol. 4, n. 5, 2014, available at SSRN: <https://ssrn.com/abstract=2537860>

C. Ragin C., *Fuzzy Sets: Calibration Versus Measurement*, University of Arizona Paper, www.u.arizona.edu/cragin/fsqca/download/calibration. Download 12 October 2016.

C. Ragin C. and P. Pennings, *Fuzzy Sets and Social Research*, “Sociological Methods & Research”, vol. 33, n. 4, 2005, pp. 423-430.

United Nations Office on Drugs and Crime (2010), *Developing Standards in Justice and Home Affairs Statistics. International and EU Acquis*, Vienna, Austria, UNODC, 2010.

www.unodc.org/documents/southeasterneurope/CARDS_Developing_Standards_in_JHA_Statistics.pdf. Download 16 September 2016.