

Agenzia Nazionale di Valutazione del
sistema Universitario e della Ricerca



National Agency for the Evaluation of
Universities and Research Institutes

How robust is journal rating in HSS?

ANTONIO FERRARA

**National Agency for the Evaluation of Universities and Research Institutes (ANVUR)-
Italy**

ANDREA BONACCORSI

**University of Pisa
Former member of Board, ANVUR**

RESSH 2015 Conference

Maison des Sciences de l'Homme en Bretagne, Rennes

June 4-6, 2015

Journal classification

- Journal evaluation was needed to identify scientific journals (as opposed to non-scientific) and classify A-rated journals, within the framework of the Habilitation procedure (see below)
- Expert panels were created in July 2012 (overall n= 24 members)
- Coverage aimed at all journals in which “Italian scholars have published”
- Triangulation approach
 - VQR (Evaluation of Research Quality exercise; see below)
 - Scientific societies
 - Expert panels

Framework

- all universities and Public Research Organizations (PROs)
- all researchers at universities submit 3 products (6 for PROs)
- over 184.000 products evaluated
- evaluation mix
 - bibliometrics
 - peer review

Evaluation panels

- 14 disciplinary panels, of which 9 in STEM and 5 in HSS (plus Architecture, minus Psychology)
- 450 experts appointed, selected by ANVUR on the basis of appr. 3,000 applications
- coverage of disciplines + international expertise
- over 14.500 referees involved (1/3 from abroad)

Peer review procedure

- Peer review is adopted for all products in Arts and Humanities (area 10), History and Philosophy (area 11), Law (area 12) and Political and Social Sciences (area 14)
- In Economics and Management peer review is used only for books and book chapters, while journal articles are evaluated through bibliometrics (WoS)
- Each “research outcome” was submitted to two referees
- Efforts have been made to submit the three “research outcomes” of a single researcher to six *different* referees (whenever possible)
- Consensus rule: if the scores of both referees overlap, the final score is assigned; otherwise, the final decision is made by the panel
- If there is a disagreement between referees of more than one merit class, a third opinion is called for

Product score

- Excellent (score 1)
 - Good (score 0.8)
 - Acceptable (score 0.5)
 - Limited (score 0)
-
- Penalty scores:
 - submission of less than 3 products (- 0.5)
 - product not admissible to evaluation (-1)
 - plagiarism (-2)

Peer review process

- Criteria for peer review
 - Originality (ability to advance knowledge)
 - Relevance (importance for the scientific community)
 - Internationalization (potential for impact on the international scientific community)
- Informed peer review:
 - evaluation of journal articles was based on informed peer review – referees were informed about the score of journal (when available)
- Journal evaluation approach
 - small number of top journals- class A
 - good quality journals- class B
 - all others- class C
- Limited coverage of the journal list (aimed only at identifying the relative quality of journals from which researchers would submit their best articles)

«Habilitation» and academic promotion

- New legislation on academic recruitment and promotion: in 2012 the legislation was drastically modified
- Transition from a decentralized promotion system at university level to a dual layer system
 - scientific habilitation at national level
 - local recruitment at department level
- National Scientific Habilitation (ASN, Abilitazione Scientifica Nazionale) procedures were opened in 2012 and 2013 to examine applications of candidates for
 - Full professor
 - Associate professor
- Five-member national committees (including an “external” one from OECD countries) oversaw procedures in each of 184 separate broad scientific disciplines (*Settori Concorsuali*)
- Committee members were extracted randomly from lists of full professors who applied for the position, after a pre-screening carried out by ANVUR

Indicators of scientific productions (2002-2012) used in the Habilitation procedure

- “Bibliometric sectors” (STEM)
 - Number of articles in indexed journals
 - Number of citations received by articles
 - Contemporary H-index
- “Non bibliometric sectors” (HSS)
 - Number of books
 - Number of book chapters and journal articles
 - Number of articles in A-rated journals
- Full counting of articles
- Self-citations included (to be modified in the future)
- Indexed journals: Wos or Scopus

The principle of median

- distribution of indicators for all Italian professors (full and associate)
- publication of the median value of the distribution for all disciplines
- candidates to Habilitation committees should have indicators beyond the median value of Full professors (non-normalized)
- candidates to Habilitation should have normalized indicators beyond the median value of Full professors or Associate professors (i.e. the upper echelon)
- rule made flexible by a provision that each committee may decide to adopt different sets of indicators before accessing the list of candidates
- Indicators' values higher than median ones as a necessary but not sufficient condition

Outcomes of the Habilitation procedures

- **overall share of Habilitation granted 40%**
- curbing the relational bias documented in past procedures at local level and in other countries (e.g. Spain) (Zynovieva, Sylos Labini et al. 2015)
- reducing the gender gap in academic promotions (De Maria and Scoppa, 2014)
- strong correlation between habilitation and scores received in the Evaluation of Quality of Research (Bonaccorsi, Costantini and Setti, 2015a; 2015b)

What is an A-class journal?

- **Legislative criteria for A-rated journals**
 - *International recognition* (within the relevant scientific community)
 - Rigorous manuscript *selection*
 - *Diffusion* (within the relevant scientific community)
 - *Esteem* (within the relevant scientific community)
- **Operationalization of A-rating criteria by ANVUR**
 - Regular publication
 - Presence in international databases (WoS-Scopus, but also disciplinary databases + international journal lists, such as ERIH, Latindex, CIRC/MESH)
 - Diffusion in academic libraries and in a significant number of foreign ones
 - Double or single blind peer review
 - Rate of acceptance of manuscripts
 - International editorial board
 - Source of contributions (at least nation-wide)
 - Expert opinion based on a standardized questionnaire (1-2 min per journal)

Journal classification process

Stage 1. Initial evaluation (July-September 2012)

Lists were published in September 2012 and used for the calculation of median values

Methodology: Triangulation between

- previous evaluation during VQR (2011)
- opinions of scientific societies
- own expert assessment

Evaluation process

Stage 2. Candidates submission

- all journals in candidates' publication lists not already covered in the initial list
- First wave of candidate submissions - October 2012
- Second wave of candidate submissions - October 2013

Evaluation process

Stage 3. Annual revision

- procedure opened to editors of journals
- only upgrading: from non-scientific to scientific journal, from scientific to class A
- $n > 500$ for the first annual revision; reviewed lists published in February 2014
- second annual revision ($n > 500$) just completed; reviewed lists to be published in June 2015
- **A general revision of the lists is scheduled for the second half of 2015. Downgrades and not just upgrades will be contemplated**

Data on the journal classification process (updated to 2014)

AREA CUN	# Riviste esaminate	# Riviste scientifiche	% su esaminate	# Riviste in classe A (con ripetizioni per Settore concorsuale)	# Riviste in classe A (senza ripetizioni per Settore concorsuale)	di cui italiane	% A italiane	% A su esaminate	% A su scientifiche
8 architettura	5674	2068	36,4	775	155	69	44,5%	2,7	7,5
10	14757	5803	39,3	2195	1227	422	34,4%	8,3	21,1
11	15571	5829	37,4	1694	591	180	30,5%	3,8	10,1
12	6898	2216	32,1	4576	286	109	38,1%	4,1	12,9
13	15649	6691	42,7	9045	767	4	0,5%	4,9	11,5
14	8489	3392	40	1160	653	88	13,5%	7,7	19,2
Totale	67038	25999	38,8	19445	3679	872	23,7%	5,5	14,1

The legal battle

- Italy as a Roman-law country: each act of the Public Administration can be subject to appeal to a special Tribunal (TAR, Tribunale Amministrativo Regionale)
- Three types of appeals against journal classification (2012-2015)
 - (a) lack of formal legal legitimation of ANVUR in journal classification*

Association of Professors in Constitutional Law
argument that the competence to classify journals is only with scientific societies- no legitimation of ANVUR
Tribunal decision: rejected
 - (b) lack of motivation of individual decisions*

journals not accepted in class A may apply to the Tribunal and invoke “lack of motivation”
Tribunal decision (in most cases): accepted, obligation for ANVUR to replicate the procedure and fully motivate the decision
 - (c) legal basis of peer review*

two contradictory decisions: TAR fully confirmed the overall procedure (March 2014), while the Consiglio di Stato (April 2015) called for a formal definition of criteria for the selection of experts and of anonymous referees in order to avoid conflicts of interest or “conflicts between schools”

Research questions/1

Q1 How robust is the journal classification?

- large literature on journal rating (Type I and Type II errors)
 - *Does journal rating predict correctly the quality of articles?*
 - answer based on analysis of citations of articles- not suitable for HSS
- quasi-natural experiment
- evaluation under VQR and under ASN carried out by two *separate* and *independent* panels
- different coverage (self-selection of 3 products under VQR vs submission of all research products in the career under ASN)
- expert panels under ASN knew about the journal score of VQR and generally followed the initial classification
- individual articles evaluated by > 14,000 referees
- criteria for articles refer to originality, relevance and internationalization; criteria for journals refer to rigorous selection, diffusion and reputation
- clear instructions about the need to fully read the article and give an assessment which reflects its quality- knowledge of the VQR rating of the journal only support information

Research questions/2

Q2 What is the role of scientific societies?

- submission of list of journals to ASN panel in July-August 2012
- designation not necessarily followed by the expert panel
- large differences in the attitude of scientific societies
 - **opposition** (all societies in Law refused to deliver their lists)
 - **non-selectivity** (few societies listed up to 300-400 titles to be promoted in class A)
 - **niche protection** (societies in small fields tended to promote many journals)
 - **selectivity** (majority of societies delivered well balanced lists)

Preliminary analysis of association between the evaluation of research outcomes and that of journals (for all areas – from Bonaccorsi-Cicero-Ferrara-Malgarini 2015)

		Evaluation of journal			Total
		A	Not A	Not academic	
Evaluation of research product	A	1,344	573	20	1,937
	B	3,184	1,743	92	5,019
	C	1,322	1,096	80	2,498
	D	837	1,176	150	2,163
	Non-academic and other	14	21	8	43
	Total	6,701	4,609	350	11,660

Description of variables

Variable	Description of variables	Measure
Lan_ita	Prevailing language of the journal	Dummy (1= Italian language prevailing; 0=otherwise)
Lan_en		Dummy (1= English language prevailing; 0=otherwise)
Lan_fra		Dummy (1= French language prevailing; 0=otherwise)
Lan_ger		Dummy (1= German language prevailing; 0=otherwise)
Lan_other		Dummy (1= Other languages prevailing; 0=otherwise)
Sub_gev_his	Disciplinary nature of the Expert sub-panel (sub-GEV)	Dummy (1= History; 0=otherwise)
Sub_gev_phil		Dummy (1= Philosophy; 0=otherwise)
Sub_gev_geo		Dummy (1= Geography; 0=otherwise)
Sub_gev_anth		Dummy (1= Anthropology; 0=otherwise)
Sub_gev_edu		Dummy (1= Education; 0=otherwise)
Coauthor	Presence of a co-author affiliated to foreign institutions in the article	Dummy (1= at least one foreign coauthor; 0=otherwise)
Foreign	Presence of a foreign referee within the referees evaluating the article	Dummy (1= at least one foreign referee; 0=otherwise)
Size_field	Number of full professors in the scientific discipline in which the article has been evaluated	Ratio scale
Score_product	Score received by the individual article	Ordinal scale Score product= 1 (Excellent), 0.8 (Good), 0.5 (Acceptable), 0 (Limited), -1 (Product non admitted to evaluation), -2 (Plagiarism)
Average VQR score	Average score received by all articles published in the journal	Ratio scale
Rat_VQR	Ordinal transformation of the rating of journals carried out within the VQR exercise Journals have been first classified as National (NAT) or International (INT). Within these classes, they were further classified as Class B and Class A (the latter as the top). Professional journals have been classified separately.	Ordinal scale Rat_VQR= 0 if rating not available Rat_VQR= 1 if journal is professional Rat_VQR= 2 if journal is national and class B Rat_VQR= 3 if journal is international and class B Rat_VQR= 4 if journal is national and class A Rat_VQR= 5 if journal is international and class A
Rat_ASN	Rating of journals carried out within the ASN procedure	Dummy (1= A-rated journal; 0= non-A journal)
Scientific society	Rating of journals suggested by the relevant scientific society	Dummy (1= the relevant scientific society has suggested the journal for the A-rating; 0=otherwise)

Journal rating in the VQR and ASN exercises

Field	Rating in the VQR procedure						Of which Rating in the ASN procedure		
	A/INT	A/NAT	B/INT	B/NAT	Not classified	Total	Class A	Scientific	Not scientific
Anthropology	9	7	8	1	59	84	23	53	8
Philosophy	48	34	33	30	202	347	99	221	27
Geography	1	4	6	5	79	95	6	66	23
Education	4	20	2	12	120	158	40	77	41
History	18	26	21	22	202	289	48	205	36
Library sciences	6	8	70	6	46	66	19	43	4
Total	86	99	70	76	708	1039	235	665	139

Results of regression models – 1

(dependent variable: Score_product)

	Model 1 Rat_VQR	Model 2 Rat_ASN	Model 3 Scientific society	Model 4 All
Lan_ita	.1598143 .1328623	-.1019263 .0723108	-.0871846 .0723453	.1314206 .1327271
Lan_eng	.2252314 .132157	.1183209 .0728002	.1160646 .0729394	.2151653 .1317579
Lan_fra	.1148588 .1380771	.1213632 .0784685	.1057815 .0786472	.1162312 .1377582
Lan_ger	.1481307 .1536858	.1409903 .0926245	.1421113 .092779	.1664392 .1532748
Lan_other	.1869151 .1345508	.0015461 .0750889	.0068984 .0751919	.1627253 .1343035
Sub_gev_his	.0947625 * .0458309	-.0141634 .0366416	-.0220987 .0366942	.0949965 * .045732
Sub_gev_phil	.0595791 .0409581	-.0354297 .0327806	-.0133455 .0328525	.0725486 .0410388
Sub_gev_geo	-.0425687 .045241	-.1097789 ** .0365366	-.1320458 *** .0366921	-.0561481 .0453225
Sub_gev_anth	-.0339417 .0541141	-.1218611 ** .0416891	-.145461 ** .0418003	-.0439871 .054129
Sub_gev_edu	-.0128572 .0441087	-.089673 * .0355206	-.0793601 * .0355524	-.0177304 .0439861
Coauthor	.0124408 .0662018	.02678 .0445109	.0229446 .0445848	.0058304 .0660125
Foreign	-.0317383 .0172833	-.0129417 .0145888	-.018646 .0146358	-.0335519 .017247
Size_field	-.0006474 * .0002657	-.0005683 * .000226	-.0006693 *** .0002259	-.0006432 * .0002662
Rat_VQR	.1049925 *** .0116672			.088297 *** .0127056
Rat_ASN		.1493224 *** .0138161		.0205014 .0224948
Scientific society			.143757 *** .013752	.0569681 ** .021749
Constant	.1864044 .1488206	.6749509 .078169	.6786939 .0782906	.2026508 .1484694
R-square	0.1493	0.1316	0.1287	0.1563
Adj R-square	0.1405	0.1259	0.1230	0.1462
Number of observations	1363	2152	2152	1363

Results of regression models: Model 1 Logistic regression model (Dependent variable: Rat_ASN); Model 2 Ordered logistic regression model (Dependent variable: Rat_VQR)

	Model 1 Rat_ASN	Model 2 Rat_VQR
Lan_ita	.581518 ** .2249017	-5.228166 *** .9157619
Lan_eng	-.5383972 * .2472696	-2.379211 ** .9119048
Lan_fra	-1.481857 ** .3718787	-1.282264 .9659073
Lan_ger	-1.004917 .5547407	-.8486412 1.030403
Lan_other	-1.144953 .7601758	-4.377333 *** .9337629
Sub_gev_his	.0944993 .3479711	-1.189352 ** .3769124
Sub_gev_phil	1.435008 *** .315991	-.6008526 .3317426
Sub_gev_geo	-3.049853 *** .5615696	-1.595078 *** .4284762
Sub_gev_anth	-6.159329 *** .9221272	-.6688773 .5430735
Sub_gev_edu	1.431866 *** .341981	-1.164294 ** .3556083
Size_field	-.0044959 * .0020837	.0009011 .0021486
Average_VQR_score	2.437634 *** .3163482	6.904628 *** .4726353
Scientific_society_1	3.411207 *** .1557751	1.756716 *** .1615415
Constant	-3.724984 *** .4203032	n.a.
Pseudo R-square	0.4725	0.3694
Number of observations	2152	1363

Results of logistic regression model – Dependent variable: Scientific society

	Model 1 - Scientific society
Lan_ita	-.0876093 .2783897
Lan_eng	-.56949 .3422812
Lan_fra	-.7131457 .5483816
Lan_ger	-2.422244 * .8705261
Lan_other	-1.334214 1.263638
Sub_gev_his	.4637697 .4503733
Sub_gev_phil	-1.199583 * .394836
Sub_gev_geo	2.09677 *** .5153204
Sub_gev_anth	-.655673 .4567345
Sub_gev_edu	.3808408 .4285492
Size_field	.0061408 * .0025399
Average_VQR_score	1.542401 ** .5286377
Rat_ASN	2.294372 *** .2031527
Rat_VQR	.9335866 *** .1379128
Constant	-4.614796 *** .6377709
Pseudo R-square	0.3137
Number of observations	1302

Findings

- strong and significant impact of journal rating on the average score of articles published in the journal- *overall robustness of the rating*
- small value of R2- variability in quality scores of articles *is not fully determined* by the merit class of the journal
- larger R2 in the journal rating model- *experts correctly classify journals* based on the (assessed) average quality of the articles published in them
- in a few cases, significant impact of disciplinary differences- importance to have a *balanced representation of disciplines in expert panels* in order to avoid negative and positive biases
- overall good ability of *scientific societies* to identify top class journals
- *small size effect*: small scientific communities protect themselves by giving higher scores to products in the peer review process and supporting a larger number of journals in the top class

Overall assessment of the experience

Initial classification

- very tight deadline (60 days) to publish the initial list; thus, a great deal of work had to be performed in a limited time
- mistaken inclusion of a few dozen non-scientific journals, resulting in mostly negative media coverage

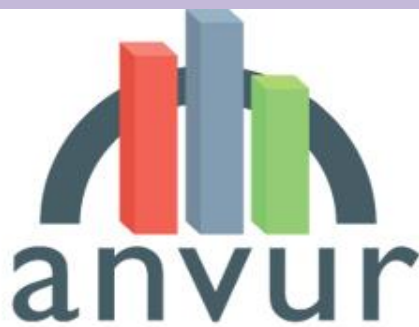
Annual revision

- welcome by the HSS community
- based on peer review by anonymous referees
- overall confirmation of reputational criteria
- average acceptance class A appr. 12%, showing the overall robustness of previous classification

Impact on the system

- pressure towards the systematic adoption of peer review by all academic journals
- shift of junior researchers's preferences to A-rated journal
- restructuring of the academic publishing sector

Agenzia Nazionale di Valutazione del
sistema Universitario e della Ricerca



National Agency for the Evaluation of
Universities and Research Institutes

Thank you for your attention