

National Agency for the Evaluation of
Universities and Research Institutes



Agenzia Nazionale di Valutazione del
sistema Universitario e della Ricerca

Evaluation of Research Quality



Valutazione Qualità della Ricerca

Valutazione della Qualità della Ricerca 2004-2010 (VQR 2004-2010)

Rapporto finale di area

Gruppo di Esperti della Valutazione dell'Area 09 (GEV09)



Appendice A. Il confronto tra valutazione *peer* e valutazione bibliometrica

A.1. Il campione casuale

Un campione casuale di 1.130 articoli su rivista passibili di valutazione bibliometrica è stato estratto dalla popolazione di 12.349 articoli sottomessi alla valutazione al GEV09. La popolazione è stratificata in base alla distribuzione dei prodotti all'interno dei Settori Scientifico-Disciplinari (SSD nel seguito) dell'area e poi raggruppati, per esigenze di significatività in 9 macro-settori. La classificazione degli articoli all'interno dei settori si basa sul settore di effettiva valutazione dell'articolo ed è calcolata escludendo i casi di articoli duplicati presentati da diversi autori all'interno di uno stesso strato campionario. I macro-settori considerati sono: Ingegneria meccanica, aeronautica e spaziale; Ingegneria manifatturiera, impiantistica e gestionale; Ingegneria energetica, termo-meccanica e nucleare; Ingegneria chimica e dei materiali; Ingegneria elettrica, elettronica e misure; Ingegneria delle telecomunicazioni e campi elettromagnetici; Ingegneria dei sistemi e bio-ingegneria; Ingegneria informatica; Ingegneria delle infrastrutture e del territorio (in comune con l'Area CUN 08). Il campione include il 9,0% dei prodotti di Ingegneria meccanica, aeronautica e spaziale, il 9,7% di prodotti di Ingegneria manifatturiera, impiantistica e gestionale, il 9,3% di Ingegneria energetica, termo-meccanica e nucleare, il 9,2% in Ingegneria chimica e dei materiali, l'8,9% in Ingegneria elettrica, elettronica e misure, il 9,2% in Ingegneria delle telecomunicazioni e campi elettromagnetici, il 9,5% in Ingegneria dei sistemi e bio-ingegneria; l'8,9% in Ingegneria informatica (Tabella A.1). Inoltre, data la scarsa numerosità dei dati del macro-settore Ingegneria delle infrastrutture e del territorio, i 6 prodotti campionati verranno considerati nel campione totale ma non costituiranno oggetto di analisi dettagliata come per gli altri macro-settori. Il campione è stato estratto ai primi di settembre 2012, prima dell'inizio del processo di revisione *peer*, mediante una procedura casuale con il vincolo di selezionare una proporzione significativa di prodotti in ciascun SSD.

Macro-settori	Popolazione	Campione	%
Ingegneria meccanica,	1390	125	9,0

aeronautica e spaziale			
Ingegneria manifatturiera, impiantistica e gestionale	837	81	9,7
Ingegneria energetica, termo-meccanica e nucleare	1259	117	9,3
Ingegneria chimica e dei materiali	2186	201	9,2
Ingegneria elettrica, elettronica e misure	2359	210	8,9
Ingegneria delle telecomunicazioni e campi elettromagnetici	1469	135	9,2
Ingegneria dei sistemi e bio-ingegneria	1158	110	9,5
Ingegneria informatica	1632	145	8,9
Ingegneria delle Infrastrutture e del territorio*	59	6*	10,2
Totale	12349	1130	9,1

Tabella A.1: Distribuzione degli articoli su rivista nel campione e nella popolazione.

Classe	Popolazione	%	Campione	%
Ingegneria meccanica, aeronautica e spaziale				
E	786	56,55%	68	54,40
B	227	16,33%	20	16,00
A	69	4,96%	9	7,20
L	54	3,88%	7	5,60
IR	254	18,27%	21	16,80
Ingegneria manifatturiera, impiantistica e gestionale				
E	470	56,15%	44	54,32
B	136	16,25%	12	14,81
A	54	6,45%	6	7,41
L	43	5,14%	4	4,94
IR	134	16,01%	15	18,52
Ingegneria energetica, termo-meccanica e nucleare				
E	730	57,98%	69	58,97



B	139	11,04%	15	12,82
A	75	5,96%	6	5,13
L	53	4,21%	5	4,27
IR	262	20,81%	22	18,80
Ingegneria chimica e dei materiali				
E	1592	64,82%	150	74,63
B	148	6,03%	10	4,98
A	49	2,00%	3	1,49
L	34	1,38%	3	1,49
IR	633	25,77%	35	17,41
Ingegneria elettrica, elettronica e misure				
E	1441	61,09%	121	57,62
B	329	13,95%	31	14,76
A	109	4,62%	12	5,71
L	53	2,25%	2	0,95
IR	427	18,10%	44	20,95
Ingegneria delle telecomunicazioni e campi elettromagnetici				
E	1016	69,16%	94	69,63
B	99	6,74%	9	6,67
A	28	1,91%	6	4,44
L	39	2,65%	1	0,74
IR	287	19,54%	25	18,52
Ingegneria dei sistemi e bio-ingegneria				
E	798	68,91%	76	69,09
B	92	7,94%	10	9,09
A	29	2,50%	1	0,91
L	33	2,85%	1	0,91
IR	206	17,79%	22	20,00
Ingegneria informatica				
E	907	55,58%	87	60,00
B	202	12,38%	18	12,41
A	67	4,11%	9	6,21
L	104	6,37%	6	4,14

IR	352	21,57%	25	17,24
Totale				
E	7762*	62,86%*	712	63,01
B	1384*	11,21%*	125	11,06
A	485*	3,93%*	52	4,60
L	419*	3,39%*	30	2,65
IR	2299*	18,62%*	211	18,67

Tabella A.2: Distribuzione delle valutazioni bibliometriche nel campione e nella popolazione.

La Tabella A.2 riporta la distribuzione nelle classi di valutazione VQR (Eccellente, Buono, Accettabile, Limitato, Incerto – IR) ottenuta utilizzando la valutazione bibliometrica degli articoli su rivista nelle quattro aree dei subGEV, per il campione e per la popolazione. Come si può osservare la distribuzione delle valutazioni bibliometriche (E, B, A, L, IR) è sufficientemente vicina nella popolazione e nel campione, sia per il totale sia per i macro-settori così da concludere che il campione estratto è rappresentativo della popolazione di riferimento.

Per ciascun articolo su rivista incluso nel campione casuale sono disponibili le seguenti informazioni:

- Rapporto del primo revisore (P1)
- Rapporto del secondo revisore (P2)
- Rapporto di un eventuale terzo e quarto revisore (P3 e P4)
- Valutazione di sintesi dei giudizi del primo e secondo revisore (P)
- Valutazione bibliometrica (F)

Le variabili P e P1-P4 assumono come valore una delle 4 classi di valutazione E, B, A, L; la valutazione bibliometrica F ha inoltre come possibile risultato la classe di valutazione “IR”, ossia il suggerimento di procedere con la “*informed peer review*” nel caso di risultati molto diversi tra i due indicatori bibliometrici (Impact Factor e numero citazioni, cfr. i criteri del GEV descritti nell’Appendice B). Le quattro classi, secondo il Bando VQR, corrispondono rispettivamente al 20% superiore della distribuzione della qualità degli articoli pubblicati nel mondo, al successivo 20%, al successivo 10% e al 50% inferiore della distribuzione della qualità degli articoli pubblicati. Più precisamente, le variabili P1-P4 sono originariamente misurate su una scala



numerica compresa tra 0 e 9, con un punteggio da 0 a 3 assegnato a 3 diversi criteri; tali punteggi sono successivamente utilizzati per determinare per ciascun prodotto sottomesso a valutazione la classe di valutazione *peer* del prodotto, sulla base dei criteri fissati dal GEV¹; le variabili P ed F sono invece rispettivamente espresse in termini delle 4 o 5 classi di valutazione sopra elencate. Sulla base del Bando VQR, alle quattro classi E, B, A, L corrispondono rispettivamente i punteggi 1; 0,8; 0,5; 0. La classificazione adottata nell'analisi bibliometrica si basa sui criteri descritti nella Sezione 2 di questo rapporto. Nella revisione dei pari, ai revisori esterni è stato richiesto di valutare ciascun prodotto sulla base della loro percezione soggettiva della qualità del prodotto rispetto alla distribuzione mondiale dei prodotti della ricerca nel settore scientifico a cui il prodotto faceva riferimento. La valutazione dei revisori è stata quindi sintetizzata sulla base di un algoritmo specifico al GEV09, secondo il quale, rispettivamente: i prodotti di classe E erano quelli con un punteggio complessivo pari a 8-9; i prodotti di classe B avevano un punteggio complessivo pari a 6-7; i prodotti di classe A un punteggio complessivo pari a 5 e i prodotti di classe L un punteggio inferiore o uguale a 4. Al fine di confrontare i risultati della valutazione bibliometrica e della revisione tra pari, si procede nel seguito a confrontare gli indicatori F e P. Anche altri confronti possono essere tuttavia di importanza significativa: in particolare il confronto tra le valutazioni tra pari P1 e P2 consente di valutare il grado di corrispondenza dei giudizi tra i due revisori².

A.2. Le distribuzioni F e P

Le distribuzioni F e P sopra descritte non sono immediatamente confrontabili, dato che la distribuzione F delle valutazioni bibliometriche comprende una classe IR che non è invece prevista nella valutazione dei pari. È però possibile ipotizzare che una discordanza di almeno due classi tra la valutazione del primo e secondo revisore segnali un'incertezza della revisione dei pari del tutto analoga a quella che emerge dal confronto tra numero di citazioni e fattore di impatto della sede di pubblicazione nell'analisi bibliometrica: in analogia con la classificazione IR della valutazione bibliometrica, si è creata dunque una corrispondente classificazione "Incerta *Peer*" (IP) per la valutazione dei pari, al fine di consentire il confronto tra le distribuzioni F e P.

¹ L'etichetta "P1", "P2", "P3" e "P4" assegnata ai due revisori è puramente convenzionale e riflette esclusivamente l'ordine di accettazione della proposta di revisione avanzata al potenziale revisore.

² Per 33 prodotti è stata utilizzata anche una terza revisione, non utilizzata per il confronto.

La Tabella A.3 mostra la distribuzione in numeri assoluti e percentuali degli indicatori F e P sopra descritti per il totale del campione del GEV09.

Valutazione Bibliometrica (F)	Valutazione Peer (P)					
	E	B	A	L	IP	Totale
E	204	277	51	29	151	712
% rispetto al totale delle valutazioni bibliometriche in classe E	28,65	38,90	7,16	4,07	21,21	100,00
B	6	60	19	11	29	125
% rispetto al totale delle valutazioni bibliometriche in classe B	4,80	48,00	15,20	8,80	23,20	100,00
A	0	14	12	9	17	52
% rispetto al totale delle valutazioni bibliometriche in classe A	0,00	26,92	23,08	17,31	32,69	100,00
L	1	4	4	13	8	30
% rispetto al totale delle valutazioni bibliometriche in classe L	3,33	13,33	13,33	43,33	26,67	100,00
IR	28	84	21	29	49	211
% rispetto al totale delle valutazioni bibliometriche IR	13,27	39,81	9,95	13,74	23,22	100,00
Totale	239	439	107	91	254	1130
% rispetto al totale delle valutazioni bibliometriche	21,15	38,85	9,47	8,05	22,48	100,00

Tabella A.3: Confronto tra F e P – totale del campione.

Gli elementi sulla diagonale principale della Tabella A.3 corrispondono ai casi in cui la valutazione dei pari e quella bibliometrica coincidono. Gli elementi al di fuori della diagonale principale corrispondono invece ai casi di non coincidenza tra F e P, o perché la valutazione F è migliore della P (elementi al di sopra della diagonale principale) o viceversa (elementi al di sotto della diagonale). La Tabella A.3 mostra che nella maggior parte dei casi la discordanza tra la valutazione bibliometrica e quella dei pari è dovuta al fatto che la valutazione bibliometrica è più generosa. In particolare, gli articoli classificati come eccellenti sulla base degli indicatori bibliometrici sono 712, contro i soli 239 prodotti eccellenti della valutazione tra pari: solo circa il 29% degli articoli classificati come E secondo la bibliometria ottiene E anche secondo la revisione tra pari, mentre rispettivamente nel 39%, 7% e 4% dei casi i prodotti bibliometricamente eccellenti risultano buoni, accettabili o limitati nella valutazione tra pari. D'altro lato, il numero di articoli che sono classificati in B, A e L dalla valutazione tra pari (439, 107 e 91 articoli rispettivamente) è nettamente più elevato rispetto agli articoli che risultano in B,



A ed L secondo la valutazione bibliometrica (125, 55 e 30 articoli rispettivamente). Infine, la numerosità di valutazioni incerte è maggiore nella revisione tra pari (254 articoli) rispetto a quella bibliometrica (211 articoli). Le valutazioni bibliometriche incerte hanno in circa il 53% dei casi una valutazione almeno pari a B nell'analisi *peer*, mentre in circa il 71% dei casi le valutazioni incerte *peer* ricevono un punteggio bibliometrico almeno pari a B.

Complessivamente, l'analisi bibliometrica e la revisione tra pari coincidono nel 30% dei casi. Se si sommano alle valutazioni coincidenti quelle che differiscono di una sola classe, si arriva al 59% del campione. Gli articoli con valutazioni che differiscono per due classi sono 66, il 6% del campione, quelli con massima discordanza (ossia, che differiscono per 3 classi) sono 30 (il 3% del campione). Il restante 32% del campione ha una assegnazione IR o IP con uno dei due metodi, e perviene ad una classe di assegnazione definita secondo l'altro metodo.

La Tabella A.4 mostra la distribuzione degli indicatori P1 e P2. Le valutazioni dei due revisori coincidono nel 41% dei casi, sono diverse per una classe di valutazione nella stessa percentuale dei casi e divergono invece rispettivamente per 2 o 3 classi di valutazione nel 17% e nel 5% dei casi.

P1	P2				
	E	B	A	L	Totale
E	195	128	35	28	386
% rispetto alle valutazioni del primo revisore in classe E	50,52	33,16	9,07	7,25	100
B	129	190	59	75	453
% rispetto alle valutazioni del primo revisore in classe B	28,48	41,94	13,02	16,56	100
A	30	52	30	26	138
% rispetto alle valutazioni del primo revisore in classe A	21,74	37,68	21,74	18,84	100
L	31	55	20	47	153
% rispetto alle valutazioni del primo revisore in classe L	20,26	35,95	13,07	30,72	100
Totale	385	425	144	176	1130
% rispetto alle valutazioni del primo revisore	34,07	37,61	12,74	15,58	100

Tabella A.4: Confronto tra le valutazioni P1 e P2 – totale del campione.



È da notare anche che le valutazioni su un giudizio di assegnazione alla classe E sono convergenti in 195 casi, pari a circa il 51% del totale delle valutazioni eccellenti fornite dal primo revisore e dal secondo revisore.

In sintesi, è possibile definire il grado di accordo raccogliendo i casi in cui si ha un completo accordo, un disaccordo di una, due o tre classi, ed un'incertezza della bibliometria (IR), della *peer* (IP) oppure di entrambi i metodi di valutazione (IR&IP). I risultati riportati nella tabella A.5 testimoniano che il completo accordo tra la biblio e la *peer* si verifica soltanto nel 25,6% dei casi ma, se si sommano le prime due righe della tabella si ottiene un valore pari a 54,7%. I casi dove si ottiene un disaccordo maggiore di una classe sono l'8,5%. In molti casi abbiamo un'incertezza sia della bibliometria (14,3% dei casi) sia della *peer review* (18,1% dei casi).

GRADO ACCORDO	%
Completo accordo	25.6%
Discordanza 1 classe	29.1%
Discordanza 2 classi	5.8%
Discordanza 3 classi	2.7%
Biblio incerta (IR)	14.3%
Peer incerta (IP)	18.1%
Biblio e <i>peer</i> incerta (IR&IP)	4.3%

Tabella A.5: grado di accordo peer-biblio. Biblio Incerta (IR) significa che il GEV ha ritenuto la valutazione bibliometrica incerta ed ha quindi fatto ricorso alla peer review. Peer incerta (IP) si verifica quando i pareri dei revisori differiscono di due o tre classi.

Le Tabelle A.6 e A.7 estendono i risultati delle Tabelle A.3 e A.4 ai singolimacro-settori. In particolare, dall'analisi dei dati della Tabella A.6 emerge che in tutti i macro-settori il numero di valutazioni eccellenti è nettamente maggiore secondo la valutazione bibliometrica rispetto alla *peer*. D'altra parte, il numero di valutazioni buone e accettabili è in genere maggiore secondo l'analisi *peer* rispetto a quella bibliometrica.



Ingegneria meccanica, aeronautica e spaziale						
Valutazione Bibliometrica (F)	Valutazione Peer (P)					
	E	B	A	L	IP	Totale
E	19	33	2	0	14	68
% rispetto al totale delle valutazioni bibliometriche in classe E	27,9	48,5	2,9	0,0	20,6	100
B	4	7	2	0	7	20
% rispetto al totale delle valutazioni bibliometriche in classe B	20,0	35,0	10,0	0,0	35,0	100
A	0	4	0	0	5	9
% rispetto al totale delle valutazioni bibliometriche in classe A	0,0	44,4	0,0	0,0	55,6	100
L	0	3	0	2	2	7
% rispetto al totale delle valutazioni bibliometriche in classe L	0,0	42,9	0,0	28,6	28,6	100
IR	2	10	3	2	4	21
% rispetto al totale delle valutazioni bibliometriche IR	9,5	47,6	14,3	9,5	19,0	100
Totale	25	57	7	4	32	125
% rispetto al totale delle valutazioni bibliometriche	20,0	45,6	5,6	3,2	25,6	100
Ingegneria manifatturiera, impiantistica e gestionale						
Valutazione Bibliometrica (F)	Valutazione Peer (P)					
	E	B	A	L	IP	Totale
E	9	14	6	3	12	44
% rispetto al totale delle valutazioni bibliometriche in classe E	20,5	31,8	13,6	6,8	27,3	100
B	0	3	4	1	4	12
% rispetto al totale delle valutazioni bibliometriche in classe B	0,0	25,0	33,3	8,3	33,3	100
A	0	3	0	1	2	6
% rispetto al totale delle valutazioni bibliometriche in classe A	0,0	50,0	0,0	16,7	33,3	100
L	0	1	0	2	1	4
% rispetto al totale delle valutazioni bibliometriche in classe L	0,0	25,0	0,0	50,0	25,0	100



bibliometriche in classe L						
IR	0	6	1	4	4	15
% rispetto al totale delle valutazioni bibliometriche IR	0,0	40,0	6,7	26,7	26,7	100
Totale	9	27	11	11	23	81
% rispetto al totale delle valutazioni bibliometriche	11,1	33,3	13,6	13,6	28,4	100
Ingegneria energetica, termo-meccanica e nucleare						
Valutazione Bibliometrica (F)	Valutazione Peer (P)					
	E	B	A	L	IP	Totale
E	15	32	6	1	15	69
% rispetto al totale delle valutazioni bibliometriche in classe E	21,7	46,4	8,7	1,4	21,7	100
B	0	9	2	0	4	15
% rispetto al totale delle valutazioni bibliometriche in classe B	0,0	60,0	13,3	0,0	26,7	100
A	0	2	2	1	1	6
% rispetto al totale delle valutazioni bibliometriche in classe A	0,0	33,3	33,3	16,7	16,7	100
L	0	0	0	1	4	5
% rispetto al totale delle valutazioni bibliometriche in classe L	0,0	0,0	0,0	20,0	80,0	100
IR	3	8	4	2	5	22
% rispetto al totale delle valutazioni bibliometriche IR	13,6	36,4	18,2	9,1	22,7	100
Totale	18	51	14	5	29	117
% rispetto al totale delle valutazioni bibliometriche	15,4	43,6	12,0	4,3	24,8	100
Ingegneria chimica e dei materiali						
Valutazione Bibliometrica (F)	Valutazione Peer (P)					
	E	B	A	L	IP	Totale
E	37	58	7	7	41	150
% rispetto al totale delle valutazioni bibliometriche in classe E	24,7	38,7	4,7	4,7	27,3	100
B	0	4	2	0	4	10
% rispetto al totale delle valutazioni bibliometriche in classe B	0,0	40,0	20,0	0,0	40,0	100
A	0	1	1	1	0	3
% rispetto al totale delle valutazioni bibliometriche in classe A	0,0	33,3	33,3	33,3	0,0	100



bibliometriche in classe A						
L	0	0	1	2	0	3
% rispetto al totale delle valutazioni bibliometriche in classe L	0,0	0,0	33,3	66,7	0,0	100
IR	2	15	3	5	10	35
% rispetto al totale delle valutazioni bibliometriche IR	5,7	42,9	8,6	14,3	28,6	100
Totale	39	78	14	15	55	201
% rispetto al totale delle valutazioni bibliometriche	19,4	38,8	7,0	7,5	27,4	100
Ingegneria elettrica, elettronica e misure						
Valutazione Bibliometrica (F)	Valutazione Peer (P)					
	E	B	A	L	IP	Totale
E	35	49	11	5	21	121
% rispetto al totale delle valutazioni bibliometriche in classe E	28,9	40,5	9,1	4,1	17,4	100
B	1	16	3	5	6	31
% rispetto al totale delle valutazioni bibliometriche in classe B	3,2	51,6	9,7	16,1	19,4	100
A	0	4	4	0	4	12
% rispetto al totale delle valutazioni bibliometriche in classe A	0,0	33,3	33,3	0,0	33,3	100
L	1	0	1	0	0	2
% rispetto al totale delle valutazioni bibliometriche in classe L	50,0	0,0	50,0	0,0	0,0	100
IR	5	18	4	5	12	44
% rispetto al totale delle valutazioni bibliometriche IR	11,4	40,9	9,1	11,4	27,3	100
Totale	42	87	23	15	43	210
% rispetto al totale delle valutazioni bibliometriche	20,0	41,4	11,0	7,1	20,5	100
Ingegneria delle telecomunicazioni e campi elettromagnetici						
Valutazione Bibliometrica (F)	Valutazione Peer (P)					
	E	B	A	L	IP	Totale
E	20	37	7	4	26	94
% rispetto al totale delle valutazioni bibliometriche in classe E	21,3	39,4	7,4	4,3	27,7	100
B	0	3	2	3	1	9

% rispetto al totale delle valutazioni bibliometriche in classe B	0,0	33,3	22,2	33,3	11,1	100
A	0	0	1	2	3	6
% rispetto al totale delle valutazioni bibliometriche in classe A	0,0	0,0	16,7	33,3	50,0	100
L	0	0	0	1	0	1
% rispetto al totale delle valutazioni bibliometriche in classe L	0,0	0,0	0,0	100,0	0,0	100
IR	4	7	2	4	8	25
% rispetto al totale delle valutazioni bibliometriche IR	16,0	28,0	8,0	16,0	32,0	100
Totale	24	47	12	14	38	135
% rispetto al totale delle valutazioni bibliometriche	17,8	34,8	8,9	10,4	28,1	100
Ingegneria dei sistemi e bio-ingegneria						
Valutazione Bibliometrica (F)	Valutazione Peer (P)					
	E	B	A	L	IP	Totale
E	27	27	7	4	11	76
% rispetto al totale delle valutazioni bibliometriche in classe E	35,5	35,5	9,2	5,3	14,5	100
B	1	5	1	1	2	10
% rispetto al totale delle valutazioni bibliometriche in classe B	10,0	50,0	10,0	10,0	20,0	100
A	0	0	0	1	0	1
% rispetto al totale delle valutazioni bibliometriche in classe A	0,0	0,0	0,0	100,0	0,0	100
L	0	0	0	0	1	1
% rispetto al totale delle valutazioni bibliometriche in classe L	0,0	0,0	0,0	0,0	100,0	100
IR	4	12	2	2	2	22
% rispetto al totale delle valutazioni bibliometriche IR	18,2	54,5	9,1	9,1	9,1	100
Totale	32	44	10	8	16	110
% rispetto al totale delle valutazioni bibliometriche	29,1	40,0	9,1	7,3	14,5	100
Ingegneria informatica						
Valutazione Bibliometrica (F)	Valutazione Peer (P)					
	E	B	A	L	IP	Totale
E	41	26	5	5	10	87

% rispetto al totale delle valutazioni bibliometriche in classe E	47,1	29,9	5,7	5,7	11,5	100
B	0	13	3	1	1	18
% rispetto al totale delle valutazioni bibliometriche in classe B	0,0	72,2	16,7	5,6	5,6	100
A	0	0	4	3	2	9
% rispetto al totale delle valutazioni bibliometriche in classe A	0,0	0,0	44,4	33,3	22,2	100
L	0	0	1	5	0	6
% rispetto al totale delle valutazioni bibliometriche in classe L	0,0	0,0	16,7	83,3	0,0	100
IR	8	7	2	5	3	25
% rispetto al totale delle valutazioni bibliometriche IR	32,0	28,0	8,0	20,0	12,0	100
Totale	49	46	15	19	16	145
% rispetto al totale delle valutazioni bibliometriche	33,8	31,7	10,3	13,1	11,0	100

Tabella A.6: Confronto tra F e P per macro-settori.

Complessivamente, la tendenza della valutazione bibliometrica a essere più vantaggiosa rispetto a quella *peer* è comune a tutti i macro-settori (un test formale di tale ipotesi è presentato nella sezione A.3) I casi di discordanza di due classi o più tra le due valutazioni sono più frequenti in Ingegneria manifatturiera, impiantistica e gestionale (14%) e Ingegneria elettrica, elettronica e misure e Ingegneria dei sistemi e bio-ingegneria (11%).

La Tabella A.7 estende i risultati della Tabella A.4 ai singoli macro-settori. In particolare, dall'analisi emerge che la convergenza nelle valutazioni dei due revisori è nettamente più elevata in Ingegneria informatica (63% dei casi), mentre è superiore al 40% in Ingegneria elettrica, elettronica e misure e Ingegneria dei sistemi e bio-ingegneria. Le divergenze per almeno 2 classi di valutazione sono particolarmente frequenti in Ingegneria manifatturiera, impiantistica e gestionale (29%) e in Ingegneria delle telecomunicazioni e campi elettromagnetici (28%). È da notare anche che le valutazioni su un giudizio di assegnazione alla classe E sono particolarmente convergenti nel macro-settore Ingegneria informatica con l'85% delle valutazioni rispetto al primo revisore e il 78% rispetto al secondo revisore.



Ingegneria meccanica, aeronautica e spaziale					
P1	P2				
	E	B	A	L	Totale
E	12	21	3	4	40
% rispetto al totale delle valutazioni di classe E del primo revisore	30,0	52,5	7,5	10,0	100,0
B	25	21	3	14	63
% rispetto al totale delle valutazioni di classe B del primo revisore	39,7	33,3	4,8	22,2	100,0
A	3	5	2	0	10
% rispetto al totale delle valutazioni di classe A del primo revisore	30,0	50,0	20,0	0,0	100,0
L	2	6	1	3	12
% rispetto al totale delle valutazioni di classe L del primo revisore	16,7	50,0	8,3	25,0	100,0
Totale	42	53	9	21	125
% rispetto al totale delle valutazioni del primo revisore	33,6	42,4	7,2	16,8	100,0
Ingegneria manifatturiera, impiantistica e gestionale					
P1	P2				
	E	B	A	L	Totale
E	8	7	2	0	17
% rispetto al totale delle valutazioni di classe E del primo revisore	47,1	41,2	11,8	0,0	100,0
B	9	9	6	6	30
% rispetto al totale delle valutazioni di classe B del primo revisore	30,0	30,0	20,0	20,0	100,0
A	2	4	4	1	11
% rispetto al totale delle valutazioni di classe A del primo revisore	18,2	36,4	36,4	9,1	100,0
L	7	6	3	7	23
% rispetto al totale delle valutazioni di classe L del primo revisore	30,4	26,1	13,0	30,4	100,0
Totale	26	26	15	14	81
% rispetto al totale delle valutazioni del primo revisore	32,1	32,1	18,5	17,3	100,0
Ingegneria energetica, termo-meccanica e nucleare					
P1	P2				



	E	B	A	L	Totale
E	14	13	2	6	35
% rispetto al totale delle valutazioni di classe E del primo revisore	40,0	37,1	5,7	17,1	100,0
B	15	22	11	9	57
% rispetto al totale delle valutazioni di classe B del primo revisore	26,3	38,6	19,3	15,8	100,0
A	2	4	3	1	10
% rispetto al totale delle valutazioni di classe A del primo revisore	20,0	40,0	30,0	10,0	100,0
L	2	8	2	3	15
% rispetto al totale delle valutazioni di classe L del primo revisore	13,3	53,3	13,3	20,0	100,0
Totale	33	47	18	19	117
% rispetto al totale delle valutazioni del primo revisore	28,2	40,2	15,4	16,2	100,0
Ingegneria chimica e dei materiali					
P1			P2		
	E	B	A	L	Totale
E	33	28	15	9	85
% rispetto al totale delle valutazioni di classe E del primo revisore	38,8	32,9	17,6	10,6	100,0
B	19	31	10	10	70
% rispetto al totale delle valutazioni di classe B del primo revisore	27,1	44,3	14,3	14,3	100,0
A	7	8	2	4	21
% rispetto al totale delle valutazioni di classe A del primo revisore	33,3	38,1	9,5	19,0	100,0
L	5	9	5	6	25
% rispetto al totale delle valutazioni di classe L del primo revisore	20,0	36,0	20,0	24,0	100,0
Totale	64	76	32	29	201
% rispetto al totale delle valutazioni del primo revisore	31,8	37,8	15,9	14,4	100,0
Ingegneria elettrica, elettronica e misure					
P1			P2		
	E	B	A	L	Totale
E	35	23	4	3	65



% rispetto al totale delle valutazioni di classe E del primo revisore	53,8	35,4	6,2	4,6	100,0
B	26	38	10	15	89
% rispetto al totale delle valutazioni di classe B del primo revisore	29,2	42,7	11,2	16,9	100,0
A	6	16	4	2	28
% rispetto al totale delle valutazioni di classe A del primo revisore	21,4	57,1	14,3	7,1	100,0
L	6	9	2	11	28
% rispetto al totale delle valutazioni di classe L del primo revisore	21,4	32,1	7,1	39,3	100,0
Totale	73	86	20	31	210
% rispetto al totale delle valutazioni del primo revisore	34,8	41,0	9,5	14,8	100,0
Ingegneria delle telecomunicazioni e campi elettromagnetici					
P1	P2				
	E	B	A	L	Totale
E	20	12	5	4	41
% rispetto al totale delle valutazioni di classe E del primo revisore	48,8	29,3	12,2	9,8	100,0
B	12	21	9	9	51
% rispetto al totale delle valutazioni di classe B del primo revisore	23,5	41,2	17,6	17,6	100,0
A	4	3	5	6	18
% rispetto al totale delle valutazioni di classe A del primo revisore	22,2	16,7	27,8	33,3	100,0
L	6	10	4	5	25
% rispetto al totale delle valutazioni di classe L del primo revisore	24,0	40,0	16,0	20,0	100,0
Totale	42	46	23	24	135
% rispetto al totale delle valutazioni del primo revisore	31,1	34,1	17,0	17,8	100,0
Ingegneria dei sistemi e bio-ingegneria					
P1	P2				
	E	B	A	L	Totale
E	26	17	3	1	47
% rispetto al totale delle valutazioni di classe E del primo revisore	55,3	36,2	6,4	2,1	100,0
B	14	17	5	5	41

% rispetto al totale delle valutazioni di classe B del primo revisore	34,1	41,5	12,2	12,2	100,0
A	3	4	3	5	15
% rispetto al totale delle valutazioni di classe A del primo revisore	20,0	26,7	20,0	33,3	100,0
L	1	3	1	2	7
% rispetto al totale delle valutazioni di classe L del primo revisore	14,3	42,9	14,3	28,6	100,0
Totale	44	41	12	13	110
% rispetto al totale delle valutazioni del primo revisore	40,0	37,3	10,9	11,8	100,0
Ingegneria informatica					
P1	P2				
	E	B	A	L	Totale
E	47	7	1	0	55
% rispetto al totale delle valutazioni di classe E del primo revisore	85,5	12,7	1,8	0,0	100,0
B	8	29	5	7	49
% rispetto al totale delle valutazioni di classe B del primo revisore	16,3	59,2	10,2	14,3	100,0
A	3	8	6	7	24
% rispetto al totale delle valutazioni di classe A del primo revisore	12,5	33,3	25,0	29,2	100,0
L	2	3	2	10	17
% rispetto al totale delle valutazioni di classe L del primo revisore	11,8	17,6	11,8	58,8	100,0
Totale	60	47	14	24	145
% rispetto al totale delle valutazioni del primo revisore	41,4	32,4	9,7	16,6	100,0

Tabella A.7: Confronto tra le valutazioni P1 e P2 – suddivise per subGEV.

A.3. Il confronto tra le distribuzioni di F (bibliometrica) e P (peer)

Il confronto tra la valutazione dei pari e quella bibliometrica si può basare su due criteri fondamentali:



1. Grado di concordanza tra la distribuzione F e la distribuzione P, ossia se F e P tendono ad assegnare lo stesso punteggio ad ogni articolo.
2. Grado di differenza sistematica esistente tra F e P misurata mediante la differenza media del punteggio assegnato da F e P sulla base dei pesi attribuiti alle classi della VQR.

Ovviamente, una perfetta concordanza implica anche la non esistenza di differenze sistematiche tra F e P, ma il contrario non è necessariamente vero, e in generale i due criteri misurano due diversi aspetti della differenza esistente tra le due distribuzioni. Si consideri ad esempio una distribuzione con un basso grado di concordanza tra F e P (molti articoli ricevono differenti valutazioni F e P). Anche in tale caso può accadere che, in media, F e P forniscano un punteggio complessivo simile. Questa distribuzione sarebbe caratterizzata da un basso livello di concordanza e da un basso grado di differenza sistematica: adottare uno dei due metodi di valutazione (per esempio quella bibliometrica, F) comporterebbe una frequente differenza di valutazione degli articoli sulla base della bibliometria e della valutazione *peer* (ossia, si avrebbero molti articoli con una buona valutazione in base a F, ma una peggiore valutazione in base a P, o viceversa).

Alternativamente, si consideri un caso di elevata (ma non perfetta) concordanza tra F e P. In questo caso, potrebbe ancora succedere che, per esempio, il numero di articoli con classificazione elevata sia sistematicamente maggiore in F che in P. In questo caso si avrebbe un elevato grado di concordanza, ma anche un alto grado di differenza sistematica tra le due distribuzioni, dato che il punteggio medio attribuito da F differirebbe dal punteggio medio di P. Adottare uno dei due metodi di valutazione può risultare in una sopravvalutazione (o sottovalutazione) in relazione all'altro criterio: ossia, gli articoli riceverebbero un punteggio notevolmente diverso se valutati con F o con P.

Da un punto di vista statistico, il grado di concordanza tra F e P può essere misurato utilizzando la statistica K di Cohen; differenze sistematiche tra F e P possono invece essere misurate guardando alle differenze tra le medie delle distribuzioni e valutandone la significatività con un test *t* di Student.

A.3.1 Il grado di concordanza tra le distribuzioni F e P

La Tabella A.8 riporta i valori della statistica K di Cohen, calcolati per l'intero campione e separatamente per ciascun macro-settore. La statistica K è costruita in modo tale da essere pari a zero quando la concordanza tra due (o più) valutazioni è del tutto casuale, vale a dire nel caso in



cui le valutazioni siano indipendenti l'una dall'altra; la statistica assume invece valore pari ad 1 nel caso in cui ci sia perfetta concordanza. È possibile calcolare il test utilizzando una matrice standard di pesi lineari (1; 0,67; 0,33; 0) attribuiti ai casi di concordanza, discordanza di una classe e così via, rispettivamente. Le analisi riportate sono riferite a campioni omogenei (*paired sample*), ossia ai prodotti del campione per i quali sono disponibili sia i dati della valutazione *peer* sia quelli relativi alla valutazione bibliometrica, eliminando cioè dal campione i prodotti per i quali la valutazione bibliometrica fornisce come risultato una classificazione IR. In questo caso, nel totale del campione, K è uguale a 0,16, un valore statisticamente diverso da zero agli usuali livelli di significatività. Il grado di concordanza negli 8 macro-settori presenta valori vicini a quelli registrati per l'intero campione, con l'unica eccezione di Ingegneria Informatica, dove la concordanza è al disopra di ogni settore (0,41). I valori sono comunque sempre statisticamente diversi da zero ad un livello di confidenza dell'1%.

Come accennato sopra, il calcolo di K riportato nella prima riga della tabella usa pesi lineari. È possibile argomentare che nel nostro caso i pesi appropriati da utilizzare debbano però essere quelli suggeriti dalle regole della VQR. Sulla base di questi pesi, è possibile calcolare la distanza tra le valutazioni utilizzando i punteggi numerici della VQR (1, 0,8, 0,5, 0), associati con le valutazioni qualitative (E, B, A, L). La seconda riga della Tabella A.7 riporta i valori della statistica K calcolati utilizzando i pesi della VQR. I risultati sono piuttosto simili ma superiori a quelli ottenuti usando pesi lineari, e indicano un buon grado di concordanza sia nel totale del campione che in ciascunmacro-settore.

Nel complesso del campione, il grado di concordanza tra la valutazione bibliometrica (F) e la revisione *peer* (P) è leggermente inferiore a quello esistente tra i giudizi formulati tra i due revisori esterni. Tuttavia emergono informazioni abbastanza eterogenee all'interno dei macro-settori. In Ingegneria meccanica, aeronautica e spaziale e in Ingegneria energetica, termomeccanica e nucleare il grado di concordanza è maggiore tra la valutazione bibliometrica (F) e la valutazione *peer* (P). In Ingegneria chimica e dei materiali la concordanza è significativa in egual misura in entrambi i casi ma moderatamente più bassa rispetto agli altri macro-settori. Nei rimanenti 5 macro-settori invece il grado di concordanza è maggiore tra i due giudizi dei revisori. A questo proposito, la Tabella A.8 riporta anche la statistica K per il grado di concordanza tra i due revisori (P1 e P2), sia per il totale del campione che per le singole aree di ricerca. Nel totale del campione, la statistica K calcolata con pesi lineari è pari a 0,19 (0,18 con i pesi della VQR), un valore superiore di quello calcolato per la concordanza tra valutazione *peer* e valutazione bibliometrica. Nei singoli macro-settori il grado di concordanza tra le due valutazioni *peer* si conferma significativo: la statistica K calcolata con pesi lineari assume valori compresi tra 0,09

in Ingegneria chimica e dei materiali e 0,56 in Ingegneria informatica. Invece nei macro-settori di Ingegneria meccanica, aeronautica e spaziale e Ingegneria energetica, termo-meccanica e nucleare tale concordanza non è significativa. Negli altri casi il test z ad essa associato conduce sempre a rifiutare l'ipotesi nulla di non concordanza ad un livello di significatività dell'1%, tranne nei macro-settori di Ingegneria manifatturiera, impiantistica e gestionale e Ingegneria chimica e dei materiali dove i valori sono significativi al 5%. Risultati analoghi si trovano utilizzando i pesi della VQR al posto dei pesi lineari. L'unica differenza è la non significatività del valore K nel macro-settore Ingegneria chimica e dei materiali.

Area/Test	F e P, pesi lineari	F e P, pesi VQR	P1 e P2, pesi lineari	P1 e P2, pesi VQR
Totale	0.1615 (10.56)**	0.1710 (10.91)**	0.1935 (8.30)**	0.1818 (7.77)**
Ingegneria meccanica, aeronautica e spaziale	0.1355 (2.44)**	0.1401 (2.44)**	-0.0469 (-0.69)	-0.0428*(-0.63)
Ingegneria manifatturiera, impiantistica e gestionale	0.1325 (2.33)**	0.1514 (2.50)**	0.1731 (2.03)*	0.1835 (2.10)*
Ingegneria energetica, termo-meccanica e nucleare	0.1606 (3.47)**	0.1688 (3.47)**	0.0466 (0.67)	0.0217 (0.31)
Ingegneria chimica e dei materiali	0.0996 (4.06)**	0.1186 (4.50)**	0.0947 (1.75)*	0.0883 (1.64)
Ingegneria elettrica, elettronica e misure	0.1105 (2.99)**	0.0904 (2.54)**	0.2021 (3.71)**	0.2071 (3.81)**
Ingegneria delle telecomunicazioni e campi elettromagnetici	0.1117 (3.52)**	0.1203 (3.80)**	0.1845 (2.73)**	0.1674 (2.42)**
Ingegneria dei sistemi e bio-ingegneria	0.1214 (2.97)**	0.1332 (3.44)**	0.2708 (3.56)**	0.2545 (3.45)**
Ingegneria informatica	0.4052 (7.50)**	0.4204 (7.83)**	0.5566 (8.57)**	0.5120 (7.99)**

Nota: La tabella riporta la statistica K e in parentesi il valore z ad essa associato³. La presenza di un asterisco indica la significatività del test al livello dell'5%, la presenza di due asterischi la significatività dell'1%.

Tabella A.8: Statistica K di Cohen sul grado di concordanza.

³Il test z verifica se K è statisticamente pari a zero assumendone una distribuzione Gaussiana, o normale. Si calcola dividendo il valore di K per il suo errore standard. Se il valore di z è superiore al valore-soglia della distribuzione normale corrispondente a una certa probabilità, si conclude che la statistica K è statisticamente diversa da zero, ossia che le due valutazioni non sono indipendenti l'una dall'altra e mostrano quindi un grado statisticamente significativo di concordanza.



A.3.2 Il grado di differenza sistematica tra le distribuzioni F e P

La Tabella A.9 riporta il punteggio medio risultante dalle valutazioni F e P. I valori numerici sono ottenuti sommando i pesi assegnati dalla VQR alle quattro classi di merito (rispettivamente, 1; 0,8; 0,5; 0) e dividendo per il numero degli articoli valutati. Si noti ancora una volta come, date le regole della VQR, gli scarti tra F e P non abbiano lo stesso peso: ad esempio, la differenza tra L e A ha un peso di 0,5, mentre la differenza tra E e B ha un peso pari solo a 0,2. I punteggi riportati sono riferiti a campioni omogenei (*paired sample*), ossia ai prodotti del campione per i quali sono disponibili sia i dati della valutazione *peer* sia quelli relativi alla valutazione bibliometrica, eliminando cioè dal campione i prodotti per i quali la valutazione bibliometrica fornisce come risultato una classificazione IR.

La terza colonna mostra che il punteggio medio finale della revisione *peer* (punteggio P) è pari a 0,711: soltanto nei macro-settori di Ingegneria manifatturiera, impiantistica e gestionale e di Ingegneria delle telecomunicazioni e campi elettromagnetici il punteggio è inferiore. Le differenze tra i subGEV che emergono dall'analisi dei dati della terza colonna della tabella possono essere attribuite:

- A una peggiore qualità degli articoli sottomessi alla valutazione in Ingegneria manifatturiera, impiantistica e gestionale e in Ingegneria delle telecomunicazioni e campi elettromagnetici rispetto ai rimanenti
- A una maggiore generosità dei revisori nei rimanenti 6 macro-settori rispetto a quelli degli ultimi due
- All'intrinseca variabilità statistica nella scelta del campione

La quarta colonna contiene il punteggio medio ottenuto nella valutazione bibliometrica: tale punteggio è pari a 0,912 per la media complessiva dei lavori sottoposti al GEV09, risultando un'equi-ripartizione dei macro-settori al di sopra e al di sotto della media. Il valore più alto si osserva in Ingegneria chimica e dei materiali (0,961) mentre quello più basso in Ingegneria meccanica, aeronautica e spaziale (0,851). L'ordinamento dei macro-settori in base alla qualità degli articoli presentati varia a seconda che si consideri l'analisi bibliometrica oppure quella *peer*. I macro-settori che presentano la maggior variabilità nell'ordinamento, variando di cinque posizioni sono Ingegneria delle telecomunicazioni e campi elettromagnetici e Ingegneria



meccanica, aeronautica e spaziale. Il macro-settore Ingegneria energetica, termo-meccanica e nucleare invece non presenta alcuna variazione nell'ordinamento. I restanti macro-settori presentano variazioni al massimo di due posizioni. Il risultato più interessante dell'analisi è mostrato nella quinta colonna, che presenta la differenza tra valutazione *peer* e bibliometrica, con le colonne 7-8 che riportano il risultato del test *t* per campioni di uguale ampiezza ad esso associato. Nel totale del campione, emerge una differenza sistematica tra la valutazione bibliometrica e la valutazione *peer*: più precisamente, la valutazione media ottenuta con l'analisi bibliometrica è migliore di quella ottenuta con la valutazione *peer*. Il risultato è confermato anche per i dati riferiti ai tutti gli 8 macro-settori.

Macro-settori	Punteggio P1	Punteggio P2	Punteggio P	Punteggio F	Diff. F-P	# osservazioni	test t	p-value
Ingegneria meccanica, aeronautica e spaziale	0,778	0,728	0,724	0,851	0,127	104	3,850	0,000
Ingegneria manifatturiera, impiantistica e gestionale	0,597	0,702	0,600	0,858	0,258	66	5,906	0,000
Ingegneria energetica, termo-meccanica e nucleare	0,746	0,694	0,716	0,884	0,169	95	5,929	0,000
Ingegneria chimica e dei materiali	0,772	0,724	0,732	0,961	0,229	166	11,422	0,000
Ingegneria elettrica, elettronica e misure	0,736	0,749	0,727	0,914	0,188	166	8,072	0,000
Ingegneria delle telecomunicazioni e campi elettromagnetici	0,687	0,681	0,670	0,947	0,277	110	10,192	0,000
Ingegneria dei sistemi e bio-ingegneria	0,786	0,759	0,747	0,960	0,214	88	7,604	0,000
Ingegneria informatica	0,744	0,735	0,717	0,883	0,166	120	6,707	0,000
Totale	0,739	0,723	0,711	0,912	0,201	919	20,762	0,000

Tabella A.9: Test *t* sulla differenza tra i punteggi bibliometrici e *peer review*.



A.4. Conclusioni

Nel totale del campione dei prodotti del GEV09 conferiti per la valutazione, si riscontra un'ampia e adeguata concordanza tra valutazioni effettuate con il metodo della revisione tra pari e con quello bibliometrico. Inoltre, il grado di concordanza tra valutazione finale bibliometrica e *peer* è molto simile al grado di concordanza tra le due valutazioni *peer*. D'altro lato, però, emerge evidenza di differenze sistematiche tra i punteggi corrispondenti alle valutazioni *peer* e bibliometriche. In effetti, è possibile osservare che il numero di prodotti della ricerca classificati come eccellenti (E) con l'algoritmo di valutazione bibliometrica sia superiore a quello dei prodotti "eccellenti" secondo la valutazione tra pari.

Il grado di concordanza tra valutazioni *peer* e valutazioni bibliometriche è moderato in quasi tutti i macro-settori, risulta invece piuttosto elevato in Ingegneria informatica. Le differenze sistematiche tra i punteggi medi sono statisticamente significative e sempre di segno positivo (ossia, la valutazione bibliometrica è significativamente più favorevole in media rispetto a quella *peer*).



Appendice B. Criteri per la valutazione dei prodotti di ricerca

Salvo restando che la responsabilità finale della valutazione è stata affidata al GEV, la valutazione ha impiegato la tecnica della *informedpeer review*, vale a dire un mix di criteri bibliometrici e di revisione *peer*.

I prodotti di cui alla tipologia **a)** della sezione 2.3 del Bando ANVUR VQR 2004-2010 che sono indicizzati nelle banche dati Web of Science di Thomson Reuters (WoS) e/o Scopus di Elsevier (Scopus) sono stati valutati utilizzando i criteri descritti nella Sezione 2.3.

I prodotti diversi dalla tipologia **a)** elencati nella sezione 2.3 del Bando ANVUR VQR 2004-2010 sono stati valutati utilizzando la valutazione *peer*. Essa prevede l'invio a revisori esterni oppure, in alcuni casi limitati per i quali vi siano le specifiche competenze richieste all'interno e non sussistano conflitti di interesse, la valutazione diretta *peer* all'interno del GEV.

I prodotti valutati mediante *peer review* appartengono quindi a 5 categorie:

- Articoli di cui alla tipologia **a)** della sezione 2.3 del Bando ANVUR VQR 2004-2010 che non sono indicizzati nelle banche dati WoS e/o Scopus;
- Articoli che sono indicizzati nelle banche dati WoS e/o Scopus che richiedono la *peer review* in base all'algoritmo bibliometrico descritto più avanti;
- Articoli che sono indicizzati nelle banche dati WoS e/o Scopus per i quali la *peer review* è richiesta dall'autore o dalla struttura, oppure è ritenuta opportuna dal GEV;
- Articoli che sono indicizzati nelle banche dati WoS e/o Scopus e che sono stati valutati utilizzando sia l'algoritmo bibliometrico sia la *peer review* al fine di studiare la correlazione tra i due metodi di valutazione: tali articoli sono stati individuati tramite un algoritmo di campionamento casuale stratificato studiato da un Gruppo di lavoro dell'ANVUR.
- Prodotti di ricerca di cui alle tipologie **b), c), d), e)** elencate nella sezione 2.3 del Bando ANVUR VQR 2004-2010;



B.1. Peer review

La *peer review* prevede l'acquisizione del parere di due revisori esterni, oppure, in alcuni casi limitati per i quali vi siano le specifiche competenze all'interno del GEV e non sussistano conflitti di interesse, la valutazione diretta da parte di un membro del GEV e di un revisore esterno, o di due diversi membri del GEV.

La selezione dei revisori esterni, preferibilmente stranieri, attese le sue rilevanti finalità di pubblico interesse, si informa al principio di leale cooperazione istituzionale ed è retta da criteri di correttezza, obiettività e imparzialità. Ciascun componente del GEV è tenuto a garantire l'anonimato degli esperti cui è affidata la valutazione dei prodotti. Le informazioni acquisite ai fini della selezione dei revisori possono essere utilizzate esclusivamente ai fini della valutazione della qualità della ricerca.

La scelta dei revisori esterni è stata effettuata evitando conflitti di interesse tra i revisori stessi e gli autori e/o la struttura di affiliazione. Inoltre, è stata garantita l'indipendenza dei revisori ponendo attenzione alla sede di affiliazione, alla collaborazione scientifica, e, ove possibile, alla nazionalità. Per minimizzare i conflitti di interesse, si sono privilegiati i revisori operanti al di fuori dei confini nazionali. L'individuazione dei revisori è stata fatta, ove possibile, indipendentemente da due diversi membri GEV.

B.1.1 L'individuazione dei revisori peer

Il GEV ha scelto di coinvolgere revisori esterni con un profilo di ricerca internazionale, un curriculum di alta levatura, testimoniato, in particolare negli ultimi anni, da un elevato numero di pubblicazioni nelle sedi di riferimento della comunità scientifica internazionale del settore, un significativo numero di citazioni e la necessaria competenza nella specifica area di valutazione. Il GEV ha preparato un elenco di revisori esterni, stabilendo standard minimi di qualità scientifica, di impatto sulla comunità scientifica internazionale e di esperienza nella valutazione.

In una prima fase il GEV ha operato per predisporre l'elenco dei revisori. A tale scopo, il Presidente del GEV ha consultato la lista di revisori della propria area, resa disponibile dal CINECA, e ha chiesto ai componenti GEV, tramite i coordinatori dei subGEV, di suggerire un numero significativo di esperti che soddisfano i criteri indicati nel paragrafo precedente.



Il Presidente GEV ha raccolto le indicazioni corredate di informazioni fornite sulla base di una scheda condivisa, e, anche con l'ausilio dei coordinatori di subGEV, ha provveduto a modificare la lista CINECA con integrazioni e/o cancellazioni.

Il processo di integrazione della lista è continuato per tutta la durata della valutazione, sulla base delle necessità emerse a valle della trasmissione dei prodotti da parte delle strutture.

B.2. Assegnazione dei livelli di merito VQR sulla base delle valutazioni peer

La valutazione da parte dei revisori *peer* si è basata su una apposita scheda revisore predisposta dal GEV, costituita da una serie di domande a risposta multipla e da un campo libero con numero limitato di parole. Il GEV ha trasformato le indicazioni contenute nella scheda revisore in una delle 4 classi finali di merito. Nel caso di valutazioni non convergenti dei revisori "*peer*" o, nel caso di disponibilità di entrambe, tra *peer review* e analisi bibliometrica, il subGEV ha creato al suo interno un Gruppo di Consenso con il compito di proporre al GEV il punteggio finale del prodotto oggetto del giudizio difforme mediante la metodologia del *consensus report*. Il Gruppo di Consenso ha potuto avvalersi anche del giudizio di un terzo esperto. In ogni caso la responsabilità della valutazione conclusiva è dell'intero GEV.

B.2.1 Valutazione dei Brevetti

I prodotti della tipologia *d* del Bando VQR sono stati sottoposti a *peer review* con le seguenti limitazioni di classificazione. I livelli di merito VQR A (eccellente) e B (buono) possono essere assegnati esclusivamente a brevetti internazionali o che siano già stati ceduti o dati in licenza a un'azienda.

B.3. Analisi bibliometrica

B.3.1 Basi di dati

Il GEV ha utilizzato le basi di dati Web of Science di Thomson Reuters (WoS) e Scopus di Elsevier (Scopus).



B.3.2 Finestra temporale delle citazioni

Nel calcolo dell'indicatore bibliometrico il GEV ha utilizzato le citazioni fino al 31 dicembre 2011.

B.3.3 Auto-citazioni

L'opportunità di includere o escludere le autocitazioni nella valutazione bibliometrica è tuttora oggetto di dibattito nella comunità scientifica. Nella VQR 2004-2010, per motivi di ordine tecnico legati soprattutto all'uso di WoS, che non lo consente direttamente, e ai problemi legati alla disambiguazione dei nomi degli autori nel caso in cui si tentasse di realizzare tale operazione a partire dai dati grezzi, non verranno escluse le autocitazioni.

B.3.4 Gli indicatori bibliometrici

La valutazione ha utilizzato, per tutti gli articoli pubblicati su riviste indicizzate nelle basi di dati WoS e Scopus, un algoritmo che tiene conto, in misura diversa a seconda della data di pubblicazione dell'articolo, sia del numero di citazioni che dell'indicatore bibliometrico della rivista ospitante. Per quanto riguarda quest'ultimo, per le riviste indicizzate in Scopus è stato impiegato SJR; per le riviste indicizzate in WoS, per gli anni da 2004 a 2006 compresi è stato impiegato il solo IF, mentre per gli anni da 2007 a 2010 compresi verrà effettuata una combinazione mediante PCA (Principal Component Analysis) di IF, 5YIF, AI e EF, come descritto in dettaglio nella Sezione B.2.4.

L'algoritmo utilizzato per la classificazione degli articoli nelle 4 classi di merito della VQR è il seguente:

1. Dato l'articolo e la rivista che lo ha pubblicato, si identifica la corrispondente *subjectcategory*(SC) in WoS e la science journal classification (ASJC) in Scopus; nel seguito l'algoritmo verrà descritto con riferimento alle sole SC e all'IF, essendo implicito che identica procedura viene seguita per le ASJC e per altri indicatori bibliometrici;
2. Se la rivista appartiene a più di una SC, si utilizza, ai fini dell'individuazione univoca della SC, l'indicazione della struttura che ha proposto l'articolo, o, se necessario, l'eventuale modifica da parte del GEV;



3. Sia in WoS che in Scopus esiste la categoria: “multidisciplinary science”, che include riviste caratterizzate da una pluralità di argomenti scientifici, quali Nature, Science, ecc.:
 - a. Gli articoli pubblicati su una rivista che compare solo in tale categoria saranno riassegnati ad un'altra SC sulla base delle citazioni contenute nell'articolo. In particolare, per ognuna delle riviste citate si individuerà una (o più) SC di appartenenza, e si sceglierà la SC finale con una regola di decisione maggioritaria. Nell'assegnazione alla nuova SC, l'articolo porterà con sé l'IF della rivista e il numero di citazioni ricevute;
4. Nel caso dei subGEV 09a, 09b e nel caso del subGEV 09c, con l'esclusione del SSD ING/INF-05, si considera la funzione di distribuzione cumulativa empirica dell'indicatore bibliometrico per le riviste appartenenti alla SC individuata, per l'anno di pubblicazione dell'articolo da valutare; si divide poi la funzione di distribuzione cumulativa in 4 quartili;
5. Nel caso del SSD ING/INF-05 si assumono le classi di qualità definite dal GEV01 per le riviste appartenenti alla SC individuata per l'articolo da valutare (vedi Sezione B.4.2);
6. Si calcola la funzione di distribuzione cumulativa empirica del numero di citazioni di tutti gli articoli (dalla data di pubblicazione al 31 dicembre 2011) pubblicati dalle riviste appartenenti alla SC individuata, per l'anno di pubblicazione dell'articolo da valutare;
7. Si divide la funzione di distribuzione cumulativa del numero di citazioni in 4 classi, corrispondenti al 20% di articoli più citati (classe 1), al successivo 20% (classe 2), al successivo 10% (classe 3) e al 50% degli articoli meno citati (classe 4);
8. Dato il valore dell'indicatore bibliometrico, si calcola il quartile di appartenenza della rivista e dato il numero di citazioni dell'articolo da valutare, si determina la classe di appartenenza dell'articolo. Il prodotto da valutare viene attribuito ad una delle 16 coppie di classi rappresentate in una matrice 4x4, che contiene sulle ascisse i quartili dell'indicatore bibliometrico (o le classi di qualità del GEV01 per il SSD ING/INF-05) e sulle ordinate le classi della distribuzione del numero di citazioni;
9. L'attribuzione della classe finale di merito avviene secondo l'algoritmo seguente, nel quale la lettera “E” si riferisce alla classe finale “eccellente”, la “B” a “buono”, la “A” a accettabile, e la “L” a “limitato”:
 - a. Quando le coordinate dell'articolo lo posizionano in una delle 4 caselle della diagonale principale, e quindi le 2 indicazioni basate su indicatore bibliometrico e su citazioni coincidono, la classe finale è la stessa (vedi Figura B1);
 - b. Quando le coordinate danno indicazioni diverse (elemento non appartenente alla diagonale principale), occorre distinguere due casi a seconda della data di pubblicazione dell'articolo.
 - c. Se l'articolo è stato pubblicato nel quadriennio 2004-2007, potendosi considerare sufficientemente consolidata la distribuzione di citazioni ad esso associata, si utilizza la matrice riportata in Figura B2 per l'attribuzione della classe finale di

merito, nella quale la classificazione finale è ottenuta privilegiando l'indicazione fornita dalla classificazione citazionale.

		Indicatore bibliometrico			
		1	2	3	4
n. di citazioni	1	E			
	2		B		
	3			A	
	4				L

Figura B.1: Matrice di corrispondenza tra classi iniziali della rivista (colonne) e citazioni (righe) e classe finale VQR.

		Indicatore bibliometrico			
		1	2	3	4
n. di citazioni	1	E	E	E	IR
	2	E	B	B	IR
	3	IR	A	A	L
	4	IR	IR	L	L

Figura B.2: Matrice di corrispondenza tra classi iniziali della rivista (colonne) e citazioni (righe) e classe finale VQR per articoli pubblicati nel quadriennio 2004-2007.

- b. Se l'articolo è stato pubblicato nel triennio 2008-2010, e quindi la sua "storia" citazionale è, al contrario, ritenuta non sufficientemente consolidata, si utilizza la matrice seguente per l'attribuzione della classe finale di merito (vedi Figura 3), nella quale si privilegia l'indicazione fornita dalla classificazione basata sull'indice di qualità della rivista.
- c. Le caselle delle matrici di Figura B2 e Figura B3 etichettate "IR" si riferiscono ai casi nei quali la valutazione non avverrà in maniera automatica, ma mediante la "informed peer review". In tali casi, il GEV affiderà l'articolo alla *peer review* oppure, in casi limitati, lo valuterà direttamente.

		Indicatore bibliometrico			
		1	2	3	4
n. di citazioni	1	E	E	IR	IR
	2	E	B	A	IR
	3	E	B	A	L
	4	IR	B	A	L

Figura B.3: Matrice di corrispondenza tra classi iniziali della rivista (colonne) e citazioni (righe) e classe finale VQR per articoli pubblicati nel triennio 2008-2010.

- d. L'algoritmo distingue, calcolando distribuzioni cumulative empiriche separate per le citazioni, gli articoli "scientifici" da quelli di rassegna, che ricevono notoriamente un maggior numero di citazioni. Quando il numero di articoli di rassegna in una SC non è tale da fornire distribuzioni significative, si ricorre alla *peer review*.

B.3.5 Gestione dei conflitti

Nel caso di articoli sottoposti a valutazione bibliometrica che ottengano una valutazione diversa utilizzando i dati contenuti in WoS e Scopus, si adotta la valutazione più favorevole se la classe



finale è individuata in modo univoco con ciascuno dei due data base, mentre si ricorre alla valutazione *peer* quando almeno una delle due valutazioni bibliometriche lo suggerisca.

Nel caso di disponibilità di una o più valutazioni *peer* e della valutazione bibliometrica per lo stesso prodotto, eventuali conflitti di attribuzione vengono risolti dal GEV su proposta del subGEV tramite un gruppo di consenso.

B.4 Risoluzione dei conflitti

I membri dei GEV si astengono dal valutare o dall'assegnare ad altri membri dei GEV o a esperti esterni:

- a. prodotti di cui siano autori o co-autori
- b. prodotti di cui siano autori o co-autori parenti o affini fino al 4° grado
- c. prodotti presentati da università presso cui i membri stessi abbiano o abbiano avuto un rapporto di lavoro o con le quali abbiano svolto incarichi o collaborazioni ufficiali, a partire dal 1/1/2007
- d. prodotti presentati da enti di ricerca vigilati dal MIUR e da altri soggetti pubblici e privati sottoposti volontariamente alla VQR presso cui i membri stessi abbiano o abbiano avuto un rapporto di lavoro o con le quali abbiano svolto incarichi o collaborazioni ufficiali, inclusa l'affiliazione a enti di ricerca, a partire dal 1/1/2007.

Nei casi di cui al punto d) precedente, esiste conflitto di interesse:

- i. nel caso in cui la struttura abbia una permanente strutturazione interna di tipo territoriale o disciplinare (es. sezione locale di ente di ricerca, istituto, dipartimento), limitatamente ai prodotti presentati dalla stessa articolazione;
- ii. nel caso in cui la struttura non abbia una permanente strutturazione interna di tipo territoriale o disciplinare (es. sezione locale di ente di ricerca, istituto, dipartimento), in riferimento a tutti i prodotti presentati nei limiti in cui ciò sia possibile senza precludere la possibilità di valutare il prodotto;
- iii. nel caso in cui la strutturazione interna abbia luogo a più livelli gerarchici (es. più istituti riuniti sotto un dipartimento) il conflitto di interesse sorge al livello più basso (es. membri GEV affiliati ad istituti diversi di uno stesso dipartimento, sono in conflitto di interesse soltanto rispetto a prodotti presentati da autori appartenenti allo stesso istituto).

Nei casi di conflitto di interesse, il Presidente del GEV incarica delle procedure di valutazione un altro membro del GEV per il quale non vi siano conflitti di interesse.

Nel caso di conflitti di interesse che coinvolgano il Presidente del GEV, l'assegnazione dei prodotti relativi è fatta dal coordinatore del VQR o da persona da lui incaricata.



B.5 Classificazione delle riviste

B.5.1 Classificazione riviste del GEV09 escluso il SSD ING-INF/05

Per ciascuna SubjectCategory (SC) WoS di interesse per il GEV09, e per gli anni 2007, 2008, 2009 e 2010 si considerano i quattro indicatori bibliometrici disponibili atti a misurare la “qualità” di una rivista, ovvero Impact Factor (IF) e Impact Factor a cinque anni (5YIF), EigenFactor (EF) e ArticleInfluence (AI).

L’algoritmo di combinazione e classificazione si articola nei seguenti passi

- i. Ciascuno dei quattro parametri bibliometrici disponibili viene normalizzato al valore massimo corrispondente nella SC, ottenendo dei parametri normalizzati all’intervallo [0,1].
- ii. Se una particolare rivista non possiede 5YIF e AI (come accade per riviste recenti), ad esse viene attribuito un valore pari alla media campionaria dei valori del corrispondente parametro nella SC;
- iii. I valori dei parametri normalizzati così determinati sono raccolti in quattro vettori IF, 5YIF, EF e AI
- iv. Si effettua una Principal Component Analysis (PCA) sfruttando la matrice di covarianza dei vettori IF, 5YIF, EF e AI e si calcola l’autovettore $PC_4=[PC_{41},PC_{42},PC_{43},PC_{44}]$ corrispondente all’autovalore di modulo massimo, ed avente norma L_1 unitaria;
- v. Si effettua la PCA della matrice sfruttando la matrice di covarianza dei vettori IF e EF e si calcola l’autovettore $PC_2=[PC_{21},PC_{22}]$ corrispondente all’autovalore di modulo massimo ed avente norma L_1 unitaria;
- vi. Per ciascuna rivista R_i che possiede tutti e 4 indici bibliometrici si calcola un coefficiente di classificazione (compreso tra 0 e 1) pari a:

$$CC_i = PC_{41} \times IF_i + PC_{42} \times 5YIF_i + PC_{43} \times EF_i + PC_{44} \times AI_i;$$

- vii. Per ciascuna rivista R_k che possiede solo IF e EF si calcola un coefficiente di classificazione (compreso tra 0 e 1) pari a:

$$CC_k = PC_{21} \times IF_k + PC_{22} \times EF_k;$$

- viii. Si ordinano i coefficienti di classificazione di tutte le riviste che possiedono tutti e quattro gli indicatori bibliometrici e se ne divide la distribuzione in quartili rispetto all’indicatore definito al punto vi. Si fa la stessa operazione per tutte le riviste che possiedono IF ed EF (cioè tutte le riviste della SC), utilizzando questa classificazione per valutare solo le riviste che possiedono l’indicatore definito al punto vii.



B.5.2 Classificazione riviste valida per l'ISSD ING-INF/05

Nel caso del SSD ING-INF/05 si sono utilizzati l'indice bibliometrico IF a 5 anni fornito dalla banca dati ISI WoS, l'indice bibliometrico 2010 SJR fornito dalla banca dati Scopus, la classificazione effettuata dall'*Australian Research Council* nel 2010 all'interno del processo di valutazione *Excellence in Research for Australia (ERA)* e che prevedeva le classi di merito A*/A/B/C, e la classificazione effettuata dal GRIN (GRuppo di INformatica) nel 2004 che prevedeva le classi di merito A/B/C/D, completata con ACM e IEEE Transactions/Journals del settore stabilite dopo il 2004 e presenti in ISI WoS e/o Scopus. Più precisamente, si è proceduto come segue:

- all'interno di ciascuna sotto-categoria ISI WoS e Scopus di *Computer Science* è stato calcolato per ogni rivista il *percentrank*, cioè il rango del suo IF a 5 anni (per ISI WoS) o di SJR 2010 (per Scopus) come percentuale sugli altri valori presenti nella stessa sotto-categoria, non considerando le riviste con IF a 5 anni o SJR pari a 0;
- a ciascuna rivista sono stati assegnati tre coefficienti:
 - o $R_i = 3x + 1$, dove x è il massimo *percentrank* fra le sotto-categorie ISI WoS o Scopus di *Computer Science* che la contenevano, oppure $R_i = 0$ se la rivista aveva sia IF a 5 anni che SJR pari a 0;
 - o R_a pari rispettivamente a 4, 3, 2, 1, 0 se la classificazione ERA è rispettivamente A*, A, B, C o assente;
 - o R_g pari rispettivamente a 4, 3, 2, 1, 0 se la classificazione GRIN è rispettivamente A, B, C, D o assente;
- è stato calcolato un indice R_f come segue:
 - o $R_f = (R_i + R_a + R_g) / 3$ se tutti e tre i coefficienti sono diversi da zero;
 - o $R_f = (R_i + 2R_a + 2R_g) / 3$ se R_i è diverso da zero e esattamente uno fra R_a e R_g è pari a zero;
 - o $R_f = 2R_i / 3$ se R_i è diverso da zero e $R_a = R_g = 0$;
 - o $R_f = (R_a + R_g) / 2$ se $R_i = 0$ e sia R_a che R_g sono diversi da zero;
 - o $R_f = R_a + R_g$ se $R_i = 0$ e almeno uno fra R_a e R_g è zero;
- sono state escluse le riviste con $R_f = 0$;
- le riviste rimaste sono state ordinate in base a R_f , e la classe di merito "1" è stata attribuita alle riviste nel primo 20% di questo ordinamento, la classe "2" alle riviste nel secondo 20%, e la classe "3" alle riviste nel successivo 10%.