



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

A.1 Il campione casuale

Un campione casuale di 927 articoli su rivista passibili di valutazione bibliometrica è stato estratto dalla popolazione di 10127 articoli, valutabili bibliometricamente, sottomessi alla valutazione nel GEV03. Il campione è stato stratificato sulla base della distribuzione dei prodotti all'interno dei 3 sub-GEV identificati nell'Area, ossia: Chimica analitica e chimico-fisica; Chimica inorganica e industriale; Chimica organica e farmaceutica. Il campione include il 9,2 % dei prodotti di Chimica analitica e chimico-fisica e di Chimica inorganica e industriale e il 9,1% di quelli di Chimica organica e farmaceutica (Tabella A.1). 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 sub-GEV.

| Sub-GEV | Popolazione | Campione | % |
|------------------------------------|-------------|----------|-----|
| Chimica analitica e chimico-fisica | 3013 | 276 | 9,2 |
| Chimica inorganica e industriale | 3076 | 283 | 9,2 |
| Chimica organica e farmaceutica | 4038 | 368 | 9,1 |
| Totale | 10127 | 927 | 9,2 |

Tabella A.1: Distribuzione degli articoli su rivista 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 nei tre sub-GEV, per il campione e per la popolazione. Come si vede, la distribuzione delle valutazioni bibliometriche (E/ B/ A/ L/ IR) è sufficientemente vicina nella popolazione e

nel campione in tutti i sub-GEV del GEV03, così da concludere che il campione estratto è rappresentativo della popolazione di riferimento.

| Classe | Popolazione | % | Campione | % |
|---|-------------|------|----------|------|
| Chimica analitica e Chimica-fisica | | | | |
| E | 1624 | 53,9 | 147 | 53,3 |
| B | 561 | 18,6 | 46 | 16,7 |
| A | 75 | 2,5 | 5 | 1,8 |
| L | 252 | 8,4 | 20 | 7,2 |
| IR | 501 | 16,6 | 58 | 21,0 |
| Chimica inorganica e industriale | | | | |
| E | 1828 | 59,4 | 171 | 60,4 |
| B | 579 | 18,8 | 49 | 17,3 |
| A | 77 | 2,5 | 8 | 2,8 |
| L | 200 | 6,5 | 20 | 7,1 |
| IR | 392 | 12,7 | 35 | 12,4 |
| Chimica organica e farmaceutica | | | | |
| E | 1907 | 47,2 | 180 | 48,9 |
| B | 837 | 20,7 | 75 | 20,4 |
| A | 161 | 4,0 | 13 | 3,5 |
| L | 491 | 12,2 | 44 | 12,0 |
| IR | 642 | 15,9 | 56 | 15,2 |
| Totale | | | | |
| E | 5359 | 52,9 | 498 | 53,7 |
| B | 1977 | 19,5 | 170 | 18,3 |
| A | 313 | 3,1 | 26 | 2,8 |
| L | 943 | 9,3 | 84 | 9,1 |
| IR | 1535 | 15,2 | 149 | 16,1 |

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

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 come possibile risultato anche 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, sono definite con riferimento ai percentili della distribuzione della qualità degli articoli pubblicati nel mondo. In particolare, la qualifica di eccellente corrisponde a un articolo che si colloca nel 20% superiore della distribuzione della qualità degli articoli pubblicati nel mondo, quella di buono nel successivo 20%, di accettabile nel successivo 10% e, infine, quella di limitato nel 50% inferiore. Le variabili P1-P4 sono originariamente misurate su una scala numerica compresa tra 3 e 27, con un punteggio da 1 a 9 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 nell’Appendice B 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 GEV03, secondo il quale, rispettivamente: i prodotti di classe E erano quelli con un punteggio complessivo uguale o superiore a 23; i prodotti di classe B avevano un punteggio complessivo compreso tra 18 e 22; i prodotti di classe A un punteggio complessivo compreso tra 15 e 17 e i prodotti di classe L un punteggio inferiore o uguale 15. 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

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

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. E' 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. 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 GEV03.

| Valutazione bibliometrica (F) | Valutazione peer (P) | | | | | Totale |
|---|----------------------|------|------|------|------|--------|
| | E | B | A | L | IP | |
| E | 162 | 231 | 22 | 17 | 66 | 498 |
| % rispetto al totale delle valutazioni bibliometriche di classe E | 32,5 | 46,4 | 4,4 | 3,4 | 13,3 | 100,0 |
| B | 22 | 97 | 12 | 12 | 27 | 170 |
| % rispetto al totale delle valutazioni bibliometriche di classe B | 12,9 | 57,1 | 7,1 | 7,1 | 15,9 | 100,0 |
| A | 1 | 12 | 6 | 6 | 1 | 26 |
| % rispetto al totale delle valutazioni bibliometriche di classe A | 3,8 | 46,2 | 23,1 | 23,1 | 3,8 | 100,0 |
| L | 1 | 33 | 9 | 19 | 22 | 84 |
| % rispetto al totale delle valutazioni bibliometriche di classe L | 1,2 | 39,3 | 10,7 | 22,6 | 26,2 | 100,0 |
| IR | 11 | 81 | 17 | 8 | 32 | 149 |
| % rispetto al totale delle valutazioni bibliometriche IR | 7,4 | 54,4 | 11,4 | 5,4 | 21,5 | 100,0 |
| Totale | 197 | 454 | 66 | 62 | 148 | 927 |

² Nel GEV03, per 7 prodotti è stato necessario procedere anche ad una terza valutazione dei pari. La terza valutazione non è stata qui considerata.



| | | | | | | |
|---|------|------|-----|-----|------|-------|
| % rispetto al totale delle valutazioni bibliometriche | 21,3 | 49,0 | 7,1 | 6,7 | 16,0 | 100,0 |
|---|------|------|-----|-----|------|-------|

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 la discordanza tra la valutazione bibliometrica e quella dei pari è dovuta ad una maggiore polarizzazione dei giudizi della bibliometria rispetto alla revisione tra pari: infatti, sia gli articoli classificati eccellenti che quelli valutati limitati sono in numero maggiore in base alla valutazione bibliometrica (498 e 84 articoli rispettivamente) rispetto a quella dei pari (197 e 62 articoli). Dei 498 articoli classificati come E dall'analisi bibliometrica, solo il 32,5% ottiene E anche secondo la revisione tra pari. Il numero di articoli che sono classificati in B ed A dalla valutazione tra pari è particolarmente elevato (454 e 66 articoli rispettivamente) a confronto con gli articoli che risultano in B ed A secondo la valutazione bibliometrica (170 e 26 articoli rispettivamente). Infine, la numerosità di valutazioni incerte è analoga nei due casi: 149 e 148 casi rispettivamente secondo la bibliometria e la revisione tra pari. Le valutazioni bibliometriche incerte hanno in circa il 62% dei casi una valutazione almeno pari a B nell'analisi *peer*, e viceversa.

Complessivamente, l'analisi bibliometrica e la revisione tra pari coincidono nel 34% dei casi. Se si sommano alle valutazioni coincidenti quelle che differiscono di una sola classe, si arriva al 66% del campione. Gli articoli con valutazioni che differiscono per due classi sono 68, il 7,3% del campione, quelli con massima discordanza (ossia, che differiscono per 3 classi) sono 18 (l'1,9% del campione). Il restante 25% del campione ha una assegnazione IR o IP con uno dei due metodi, e arriva invece 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 42,5% dei casi, sono diverse per una classe di valutazione nel 41,5% dei casi e divergono invece rispettivamente per 2 o 3 classi di valutazione nel 12,1% e nel 3,9% dei casi. E' da notare anche che le valutazioni su un giudizio di assegnazione alla classe E sono convergenti in 150 casi, pari a circa il 52% del totale delle valutazioni eccellenti (290) fornite dal primo revisore e a circa il 48% delle valutazioni eccellenti (313) del secondo revisore.

Le Tabelle A.5 e A.6 estendono i risultati delle Tabelle A.3 e A.4 ai singoli sub-GEV. In particolare, dall'analisi dei dati della Tabella A.5 emerge che in tutti i sub-GEV il numero di valutazioni eccellenti è maggiore secondo la valutazione bibliometrica rispetto alla *peer*; anche le valutazioni limitate sono più frequenti secondo l'analisi bibliometrica nella Chimica analitica e

chimico-fisica e nella Chimica organica-industriale, mentre la frequenza dei giudizi di limitato si equivale tra i due metodi di valutazione nella Chimica inorganica e industriale. D'altra parte, il numero di valutazioni buone e accettabili è sempre maggiore secondo l'analisi *peer* rispetto a quella bibliometrica. Complessivamente, la tendenza della valutazione bibliometrica ad essere più vantaggiosa rispetto a quella *peer* è comune a tutti i sub-GEV (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 Chimica organica e farmaceutica (10,3%) rispetto a Chimica analitica e Chimica-fisica (8,3%) e a Chimica inorganica e industriale (8,9%).

| P1 | P2 | | | | |
|---|------|------|------|------|--------|
| | E | B | A | L | Totale |
| E | 150 | 101 | 19 | 20 | 290 |
| % rispetto al totale delle valutazioni di classe E del primo revisore | 51,7 | 34,8 | 6,6 | 6,9 | 100,0 |
| B | 126 | 199 | 60 | 34 | 419 |
| % rispetto al totale delle valutazioni di classe B del primo revisore | 30,1 | 47,5 | 14,3 | 8,1 | 100,0 |
| A | 21 | 55 | 20 | 20 | 116 |
| % rispetto al totale delle valutazioni di classe A del primo revisore | 18,1 | 47,4 | 17,2 | 17,2 | 100,0 |
| L | 16 | 38 | 23 | 25 | 102 |
| % rispetto al totale delle valutazioni di classe L del primo revisore | 15,7 | 37,3 | 22,5 | 24,5 | 100,0 |
| Totale | 313 | 393 | 122 | 99 | 927 |
| % rispetto al totale delle valutazioni del primo revisore | 33,8 | 42,4 | 13,2 | 10,7 | 100,0 |

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

| Chimica analitica e Chimica-fisica | | | | | | |
|---|-----------------------------|------|-----|-----|------|--------|
| Valutazione bibliometrica (F) | Valutazione <i>peer</i> (P) | | | | | Totale |
| | E | B | A | L | IP | |
| E | 56 | 59 | 7 | 4 | 21 | 147 |
| % rispetto al totale delle valutazioni bibliometriche di classe E | 38,1 | 40,1 | 4,8 | 2,7 | 14,3 | 100,0 |



| | | | | | | |
|---|-----------------------------|------|------|------|------|--------|
| B | 7 | 28 | 3 | 1 | 7 | 46 |
| % rispetto al totale delle valutazioni bibliometriche di classe B | 15,2 | 60,9 | 6,5 | 2,2 | 15,2 | 100,0 |
| A | 0 | 1 | 1 | 3 | 0 | 5 |
| % rispetto al totale delle valutazioni bibliometriche di classe A | 0,0 | 20,0 | 20,0 | 60,0 | 0,0 | 100,0 |
| L | 0 | 11 | 0 | 3 | 6 | 20 |
| % rispetto al totale delle valutazioni bibliometriche di classe L | 4 | 28 | 10 | 2 | 13 | 57,0 |
| IR | 4 | 29 | 10 | 2 | 13 | 58 |
| % rispetto al totale delle valutazioni bibliometriche IR | 6,9 | 50,0 | 17,2 | 3,4 | 22,4 | 100,0 |
| Totale | 67 | 128 | 21 | 13 | 47 | 276 |
| % rispetto al totale delle valutazioni bibliometriche | 24,3 | 46,4 | 7,6 | 4,7 | 17,0 | 100,0 |
| Chimica inorganica e industriale | | | | | | |
| Valutazione bibliometrica (F) | Valutazione peer (P) | | | | | |
| | E | B | A | L | IP | Totale |
| E | 55 | 81 | 8 | 8 | 19 | 171 |
| % rispetto al totale delle valutazioni bibliometriche di classe E | 32,2 | 47,4 | 4,7 | 4,7 | 11,1 | 100,0 |
| B | 8 | 26 | 4 | 4 | 7 | 49 |
| % rispetto al totale delle valutazioni bibliometriche di classe B | 16,3 | 53,1 | 8,2 | 8,2 | 14,3 | 100,0 |
| A | 0 | 5 | 2 | 1 | 0 | 8 |
| % rispetto al totale delle valutazioni bibliometriche di classe A | 0,0 | 62,5 | 25,0 | 12,5 | 0,0 | 100,0 |
| L | 1 | 4 | 6 | 6 | 3 | 20 |
| % rispetto al totale delle valutazioni bibliometriche di classe L | 5 | 20 | 30 | 30 | 15 | 100,0 |
| IR | 5 | 21 | 1 | 1 | 7 | 35 |
| % rispetto al totale delle valutazioni bibliometriche IR | 14,3 | 60,0 | 2,9 | 2,9 | 20,0 | 100,0 |
| Totale | 69 | 137 | 21 | 20 | 36 | 283 |
| % rispetto al totale delle valutazioni bibliometriche | 24,4 | 48,4 | 7,4 | 7,1 | 12,7 | 100,0 |
| Chimica organica e farmaceutica | | | | | | |
| Valutazione bibliometrica (F) | Valutazione peer (P) | | | | | |

| | E | B | A | L | IP | Totale |
|---|------|------|------|------|------|--------|
| E | 51 | 91 | 7 | 5 | 26 | 180 |
| % rispetto al totale delle valutazioni bibliometriche di classe E | 28,3 | 50,6 | 3,9 | 2,8 | 14,4 | 100,0 |
| B | 7 | 43 | 5 | 7 | 13 | 75 |
| % rispetto al totale delle valutazioni bibliometriche di classe B | 9,3 | 57,3 | 6,7 | 9,3 | 17,3 | 100,0 |
| A | 1 | 6 | 3 | 2 | 1 | 13 |
| % rispetto al totale delle valutazioni bibliometriche di classe A | 7,7 | 46,2 | 23,1 | 15,4 | 7,7 | 100,0 |
| L | 0 | 18 | 3 | 10 | 13 | 44 |
| % rispetto al totale delle valutazioni bibliometriche di classe L | 0 | 41 | 7 | 23 | 30 | 100,0 |
| IR | 2 | 31 | 6 | 5 | 12 | 56 |
| % rispetto al totale delle valutazioni bibliometriche IR | 3,6 | 55,4 | 10,7 | 8,9 | 21,4 | 100,0 |
| Totale | 61 | 189 | 24 | 29 | 65 | 368 |
| % rispetto al totale delle valutazioni bibliometriche | 16,6 | 51,4 | 6,5 | 7,9 | 17,7 | 100,0 |

Tabella A.5: Confronto tra F e P per sub-GEV

La Tabella A.6 estende i risultati della Tabella A.4 ai singoli sub-GEV. In particolare, dall'analisi emerge che la convergenza nelle valutazioni dei due revisori è più elevata nella Chimica organica e farmaceutica (44,6% dei casi) rispetto alla Chimica analitica e chimico-fisica (39,8%) e alla Chimica inorganica e industriale (42,4%). Le divergenze per almeno 2 classi di valutazione sono più elevate nella Chimica organica e farmaceutica e alla Chimica analitica e chimico-fisica (17,6% e 17% dei casi rispettivamente) rispetto alla Chimica Inorganica e Industriale (12,7%). E' da notare anche che le valutazioni su un giudizio di assegnazione alla classe E sono convergenti in più della metà dei casi in tutti e tre i sub-GEV considerati.

| Chimica analitica e Chimica-fisica | | | | | |
|------------------------------------|----|----|---|---|--------|
| P1 | P2 | | | | |
| | E | B | A | L | Totale |
| E | 53 | 30 | 9 | 8 | 100 |



| | | | | | |
|---|-----------|------|------|------|--------|
| % rispetto al totale delle valutazioni di classe E del primo revisore | 53,0 | 30,0 | 9,0 | 8,0 | 100 |
| B | 37 | 46 | 17 | 6 | 106 |
| % rispetto al totale delle valutazioni di classe B del primo revisore | 34,9 | 43,4 | 16,0 | 5,7 | 100 |
| A | 4 | 24 | 6 | 7 | 41 |
| % rispetto al totale delle valutazioni di classe A del primo revisore | 9,8 | 58,5 | 14,6 | 17,1 | 100 |
| L | 8 | 12 | 4 | 5 | 29 |
| % rispetto al totale delle valutazioni di classe L del primo revisore | 27,6 | 41,4 | 13,8 | 17,2 | 100 |
| Totale | 102 | 112 | 36 | 26 | 276 |
| % rispetto al totale delle valutazioni del primo revisore | 37,0 | 40,6 | 13,0 | 9,4 | 100 |
| Chimica inorganica e industriale | | | | | |
| P1 | P2 | | | | |
| | E | B | A | L | Totale |
| E | 51 | 40 | 5 | 3 | 99 |
| % rispetto al totale delle valutazioni di classe E del primo revisore | 51,5 | 40,4 | 5,1 | 3,0 | 100 |
| B | 40 | 53 | 22 | 8 | 123 |
| % rispetto al totale delle valutazioni di classe B del primo revisore | 32,5 | 43,1 | 17,9 | 6,5 | 100 |
| A | 8 | 16 | 4 | 3 | 31 |
| % rispetto al totale delle valutazioni di classe A del primo revisore | 25,8 | 51,6 | 12,9 | 9,7 | 100 |
| L | 4 | 8 | 6 | 12 | 30 |
| % rispetto al totale delle valutazioni di classe L del primo revisore | 13,3 | 26,7 | 20,0 | 40,0 | 100 |
| Totale | 103 | 117 | 37 | 26 | 283 |
| % rispetto al totale delle valutazioni del primo revisore | 36,4 | 41,3 | 13,1 | 9,2 | 100 |
| Chimica organica e farmaceutica | | | | | |
| P1 | P2 | | | | |
| | E | B | A | L | Totale |

| | | | | | |
|---|------|------|------|------|-----|
| E | 46 | 31 | 5 | 9 | 91 |
| % rispetto al totale delle valutazioni di classe E del primo revisore | 50,5 | 34,1 | 5,5 | 9,9 | 100 |
| B | 49 | 100 | 21 | 20 | 190 |
| % rispetto al totale delle valutazioni di classe B del primo revisore | 25,8 | 52,6 | 11,1 | 10,5 | 100 |
| A | 9 | 15 | 10 | 10 | 44 |
| % rispetto al totale delle valutazioni di classe A del primo revisore | 20,5 | 34,1 | 22,7 | 22,7 | 100 |
| L | 4 | 18 | 13 | 8 | 43 |
| % rispetto al totale delle valutazioni di classe L del primo revisore | 9,3 | 41,9 | 30,2 | 18,6 | 100 |
| Totale | 108 | 164 | 49 | 47 | 368 |
| % rispetto al totale delle valutazioni del primo revisore | 29,3 | 44,6 | 13,3 | 12,8 | 100 |

Tabella A.6: Confronto tra le valutazioni P1 e P2 per sub-GEV

A.3 Il confronto tra le distribuzioni di F e P

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.7 riporta i valori della statistica K di Cohen, calcolati per l'intero campione e separatamente per ciascun sub-GEV. I risultati sono riferiti a campioni omogenei (*paired sample*), ossia ai prodotti del campione per i quali sono disponibili sia i risultati 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; nel caso del GEV di Chimica, le osservazioni a disposizione si riducono a 778. 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. E' 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. In questo caso, nel totale del campione, K è uguale a 0,22, un valore statisticamente diverso da zero agli usuali livelli di significatività. Il grado di concordanza nei 3 sub-GEV presenta valori vicini a quelli registrati per l'intero campione, e 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. E' possibile argomentare che nel nostro caso i pesi appropriati da utilizzare debbano però essere quelli suggeriti dalle regole della VQR. In particolare, 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 mostrano che in questo caso la concordanza è maggiore o uguale rispetto alle valutazioni basate su pesi lineari, a conferma del buon grado di concordanza sia nel totale del campione che in ciascun sub-GEV.

| Test | Totale campione | Chimica analitica e chimico-fisica | Chimica inorganica e industriale | Chimica organica e farmaceutica |
|-----------------------|-----------------|------------------------------------|----------------------------------|---------------------------------|
| F e P, pesi lineari | 0,2246 (10,67)* | 0,2261 (5,65)* | 0,2024 (5,62)* | 0,2304 (6,87)* |
| F e P, pesi VQR | 0,2296 (10,42)* | 0,2192 (5,47)* | 0,2158 (5,73)* | 0,2368 (6,59)* |
| P1 e P2, pesi lineari | 0,2501 (10,02)* | 0,2143 (4,47)* | 0,2824 (6,35)* | 0,2394 (6,18)* |
| P1 e P2, pesi VQR | 0,2381 (9,60)* | 0,1974 (4,22)* | 0,2996 (6,84)* | 0,2107 (5,34)* |

Nota: La tabella riporta la statistica K e in parentesi il valore z^3 ad essa associato. La presenza di un asterisco indica la significatività del test al livello dell'1%

Tabella A.7: Statistica K di Cohen sul grado di concordanza

La Tabella A.7 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 i singoli sub-GEV. Nel complesso del campione, il grado di concordanza tra la valutazione bibliometrica (F) e la revisione *peer* (P) è molto simile a quello esistente tra i giudizi formulati tra i due revisori esterni: in quest'ultimo caso, la statistica K calcolata con pesi lineari è pari a 0,25 (0,24 con i pesi della VQR). Analoghi risultati si hanno a livello dei singoli sub-GEV. Il test z associato conduce sempre a rifiutare l'ipotesi nulla di non concordanza ad un livello di significatività dell'1%.

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

La Tabella A.8 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 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,

³ Il 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.



mentre la differenza tra E e B ha un peso pari solo a 0,2. Come nel caso delle analisi contenute nella sezione precedente, i risultati 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. Come ricordato sopra, gli articoli a disposizione in questo caso sono 778.

La terza colonna mostra che il punteggio medio finale della revisione *peer* (punteggio P) è pari a 0,742: il punteggio medio in Chimica inorganica e industriale (0,741) è sostanzialmente in linea con quello medio complessivo, mentre in Chimica organica e farmaceutica (0,719) e in Chimica analitica e chimico-fisica (0,776) i punteggi sono rispettivamente inferiori o superiori alla media generale. Le differenze tra i sub-GEV che emergono dall'analisi dei dati della terza colonna della tabella possono essere attribuite:

- A una migliore qualità degli articoli sottoposti alla valutazione in Chimica analitica e chimico-fisica rispetto a quella degli articoli degli altri sub-GEV
- A una maggiore generosità dei revisori di Chimica analitica e chimico-fisica rispetto agli altri
- All'intrinseca variabilità statistica nella scelta del campione.

La quarta colonna contiene il punteggio medio ottenuto nella valutazione bibliometrica: tale punteggio è pari a 0,832 per la media complessiva dei lavori sottoposti al GEV03, risultando inferiore alla media nella Chimica organica e farmaceutica (0,790) e superiore alla media nella Chimica analitica e chimico-fisica (0,855) e nella Chimica inorganica e industriale (0,864). L'ordinamento dei sub-GEV in base alla qualità degli articoli presentati è diverso a seconda che si consideri l'analisi bibliometrica e quella *peer*: in ambedue i casi, la Chimica organica e farmaceutica riceve la valutazione peggiore, ma la Chimica inorganica e industriale prevale sulla Chimica analitica e chimico-fisica secondo la valutazione bibliometrica, e viceversa secondo la valutazione dei pari. 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 statisticamente maggiore a quella ottenuta con la valutazione *peer*. Il risultato è confermato anche per i dati riferiti ai tre sub-GEV considerati.

| Sub-GEV | Punteggio P1 | Punteggio P2 | Punteggio P | Punteggio F | Diff F-P | # Osservazioni | Test t | p-value |
|------------------------------------|--------------|--------------|-------------|-------------|----------|----------------|--------|---------|
| Chimica analitica e chimico-fisica | 0,767 | 0,775 | 0,776 | 0,855 | 0,078 | 218 | 3,764 | 0,000 |
| Chimica inorganica e industriale | 0,753 | 0,763 | 0,741 | 0,864 | 0,123 | 248 | 6,195 | 0,000 |
| Chimica organica e farmaceutica | 0,714 | 0,737 | 0,719 | 0,790 | 0,071 | 312 | 3,580 | 0,000 |
| Totale | 0,741 | 0,756 | 0,742 | 0,832 | 0,090 | 778 | 7,632 | 0,000 |

Tabella A.8: Test t sulla differenza tra i punteggi bibliometrici e peer review

A.4 Conclusioni

Nel totale del campione dei prodotti del GEV03 conferiti per la valutazione, si riscontra una più che 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 una chiara evidenza di differenze sistematiche tra i punteggi corrispondenti alle valutazioni *peer* e bibliometriche. In effetti, è possibile osservare come 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 elevato in tutti i sub-GEV. 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: Documento sui criteri dell'area 03

1. Introduzione

In linea con gli obiettivi generali dell'esercizio di Valutazione della Qualità della Ricerca (VQR) 2004-2010, i criteri adottati del GEV03 mirano a fornire un panorama qualitativo della ricerca nel campo delle scienze chimiche in Italia. Nel complesso, i prodotti che il GEV03 esaminerà per questo esercizio sono principalmente articoli su riviste scientifiche. Questi articoli saranno classificati prevalentemente utilizzando indicatori bibliometrici integrati allo scopo di valutare sia la qualità della rivista in cui sono pubblicati sia il loro impatto specifico, quantificato in base al numero citazioni ricevute dall'articolo fino al dicembre 2011. Chiaramente, articoli pubblicati all'inizio del periodo di valutazione, nel 2004 e 2005, hanno avuto più tempo per accumulare citazioni e raggiungere un valore statisticamente significativo dell'impatto, rispetto ai lavori pubblicati nel 2009 o 2010. Questi ultimi potranno richiedere valutazioni aggiuntive secondo la metodologia peer review. Pertanto, si raccomanda che i soggetti valutati e le strutture di ricerca forniscano nella scheda descrittiva che accompagna i prodotti, in particolare per quelli pubblicati nel 2009 e nel 2010, tutte le informazioni utili alla valorizzazione del prodotto, quali, ad esempio, l'indicazione che il prodotto ha aperto un nuovo campo di ricerca o possiede un carattere profondamente interdisciplinare, al fine di facilitare l'identificazione dei prodotti da sottoporre a peer review.

2. Delimitazione dell'area GEV

Il GEV03 (Scienze Chimiche) include i SSD compresi da CHIM01 a CHIM12.



3. Organizzazione del GEV

3.1. Sub-GEV

Il GEV dell'area 03 è organizzato nei seguenti 3 sottogruppi (sub-GEV):

Sub-GEV03-01: AREA ANALITICO – CHIMICO-FISICA

SSD: CHIM01, CHIM02, CHIM12

Coordinatore: Luisa Torsi (Bari)

Altri membri: Angela Agostiano (Bari), Vincenzo Barone (Pisa), Perla Colombini (Pisa), Elena Ibáñez (Madrid), Stefana Milioto (Palermo), Patrick Sandra (Ghent), Antonio Tilocca (London), Francesco Zerbetto (Bologna)

Sub-GEV03-02: AREA INORGANICO – INDUSTRIALE

SSD: CHIM03, CHIM04, CHIM05, CHIM07

Coordinatore: Gianfranco Pacchioni (Milano)

Altri membri: Lucia Banci (Firenze), Gabriele Centi (Messina), Gaetano Guerra (Salerno), Roberto Purrello (Catania), Christian Serre (Versailles)

Sub-GEV03-03: AREA ORGANICO – FARMACEUTICA

SSD: CHIM06, CHIM08, CHIM09, CHIM10, CHIM11

Coordinatore: Maurizio Prato (Trieste)

Altri membri: Maurizio Botta (Siena), Alberto Brandi (Firenze), Horst Kessler (München), Henriette Molinari (Verona), Ferdinando Pizzo (Perugia), Patrizia Santi (Parma), Leonardo Scapozza (Genève)

3.2. Allocazione dei prodotti all'interno del GEV

L'allocazione dei prodotti ai sub-GEV avverrà sulla base del SSD indicato dal soggetto valutato così come trasmesso dalle strutture. Il significato del SSD associato al prodotto, che può anche essere diverso dal SSD di appartenenza del soggetto valutato, si riferisce al GEV che con maggior competenza, secondo il soggetto valutato, può valutare il prodotto in questione.

Nel caso in cui un prodotto sia assegnato a più GEV per il suo carattere interdisciplinare saranno adottati identici criteri di valutazione concordati tra i vari GEV. A tale scopo, i Presidenti dei GEV interessati istituiscono specifici Gruppi di Consenso Inter-Area.



3.3. Regole di funzionamento del GEV

- La convocazione del GEV avviene almeno 15 giorni prima della riunione. La riunione è convocata dal Presidente, che fissa anche l'ordine del giorno.
- Le decisioni all'interno del GEV vengono prese a maggioranza semplice. Per partecipare alla votazione non è necessario essere fisicamente presenti alle riunioni, purché si sia presenti in modalità telematica.
- Alle riunioni del GEV partecipa, con funzioni di segretario senza diritto di voto, anche l'assistente al GEV attribuito da ANVUR al GEV. Al termine di ciascuna riunione viene redatto un resoconto della seduta in italiano, e un verbale succinto che riporta le conclusioni principali in lingua italiana e inglese. I verbali vengono fatti circolare tra i membri del GEV e approvati tramite email o utilizzando l'ambiente SW predisposto dal CINECA.

4. Mix valutativo

Salvo restando che la responsabilità finale della valutazione è affidata al GEV, il GEV03 utilizzerà per la valutazione la tecnica della “informed peer review”, vale a dire un mix di criteri bibliometrici e di revisione “peer”.

I prodotti di cui alle tipologie a della sezione 2.3 del Bando ANVUR VQR 2004-2010 identificati nelle banche dati *Web of Science di Thomson Reuters (WoS)* e/o Scopus verranno valutati utilizzando i criteri bibliometrici descritti nella sezione 3.2.

I prodotti delle tipologie b,c,d,e elencate nella sezione 2.3 del Bando ANVUR VQR 2004-2010 verranno valutati utilizzando la valutazione “peer”.

I prodotti valutati mediante “peer review” (che prevede l'invio a revisori esterni oppure, in alcuni casi limitati per i quali vi siano le competenze richieste all'interno, la valutazione diretta “peer” all'interno del GEV) appartengono a 3 categorie:

1. Prodotti di ricerca di cui alle tipologie b,c,d,e elencate nella sezione 2.3 del Bando ANVUR VQR 2004-2010;
2. Articoli che sono indicizzati nelle banche dati ISI WoS e/o Scopus che richiedono la “peer review” in base all'algoritmo bibliometrico descritto nella sezione 3.2;
3. Articoli che sono indicizzati nelle banche dati ISI WoS e/o Scopus e che saranno valutati utilizzando l'algoritmo bibliometrico e la “peer review” al fine di studiare la correlazione tra i due metodi di valutazione: tali articoli saranno individuati



tramite un algoritmo di campionamento casuale stratificato studiato da un Gruppo di lavoro dell'ANVUR.

5. Peer Review

Nel caso di utilizzo della “peer review” i prodotti saranno inviati a due revisori esterni, oppure valutati, sussistendo le condizioni di competenza e di assenza di conflitti, all'interno del GEV utilizzando le stesse procedure e la stessa scheda di revisione.

La selezione dei revisori esterni, italiani e 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 verrà effettuata evitando conflitti di interesse tra i revisori stessi e gli autori e/o la struttura di affiliazione. Inoltre, verrà 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 privilegeranno, ove possibile, i revisori operanti al di fuori dei confini nazionali.

5.1. L'individuazione dei revisori “peer”

Il GEV intende coinvolgere revisori esterni con un profilo di ricerca internazionale, un curriculum di altro profilo, 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 citazione e la necessaria competenza nella specifica area di valutazione. Il GEV preparerà un elenco di revisori esterni, stabilendo standard minimi di qualità scientifica, di impatto sulla comunità scientifica internazionale e di esperienza nella valutazione.

Grande attenzione verrà posta al mantenimento dell'anonimato dei revisori, sia nella fase di predisposizione dell'elenco dei revisori, che nella fase operativa di valutazione.



Per quanto attiene alla prima, il Presidente GEV consulterà la lista di revisori della propria area resa disponibile dal CINECA, e chiederà ai componenti GEV, tramite i coordinatori dei sub-GEV, di suggerire un numero significativo di esperti che soddisfano ai criteri indicati nel paragrafo precedente.

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

Il processo di integrazione della lista continuerà per tutta la durata della valutazione, sulla base delle necessità che dovessero emergere a valle della trasmissione dei prodotti da parte delle strutture.

5.2. Assegnazione dei livelli di merito VQR sulla base delle valutazioni “peer”

La valutazione dei revisori “peer” si baserà 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 trasformerà 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 sub-GEV creerà al suo interno un Gruppo di Consenso con il compito di proporre al GEV il punteggio finale del prodotto oggetto del giudizio difforme dei revisori esterni mediante la metodologia del “consensus report”. Il Gruppo di Consenso potrà avvalersi anche del giudizio di un terzo esperto. In ogni caso la responsabilità della valutazione conclusiva è dell'intero GEV.

6. Analisi bibliometrica

6.1. Basi di dati

Il GEV03 utilizzerà come basi di dati di riferimento WoS e Scopus.



6.2. Finestra temporale delle citazioni

Nel calcolo degli indicatori bibliometrici il GEV utilizzerà le citazioni fino al 31 dicembre 2011.

6.3. 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 se lo si volesse realizzare a partire dai dati grezzi, non verranno escluse le autocitazioni.

6.4. Gli indicatori bibliometrici

La valutazione utilizzerà, 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 (IF).

Tale scelta è dettata dalle seguenti considerazioni:

Il puro indicatore citazionale assume valori anche molto piccoli a seconda della disciplina e dell'età dell'articolo, rendendo difficile una discriminazione effettiva tra classi di merito; queste considerazioni sono il risultato di una significativa sperimentazione effettuata utilizzando le basi di dati acquisite per l'esercizio VQR;

L'uso del solo indicatore citazionale costituisce un obiettivo facile per possibili future manipolazioni, inducendo comportamenti fuorvianti quali l'estensiva autocitazione e la citazione mutua all'interno di un gruppo ristretto, non giustificate da considerazione di natura tecnica;

Si ritiene importante comunicare ai giovani ricercatori il messaggio che esiste un livello qualitativo anche profondamento diverso tra le varie riviste e che ci si debba cimentare con severe procedure di "peer review" e avere, quale obiettivo, quello di pubblicare nelle migliori riviste del proprio settore.

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



Dato l'articolo e la rivista che lo ha pubblicato, si identifica la corrispondente “subject category (SC)” in ISI WoS e la “science journal classification (ASJC) in Scopus; nel seguito l'algoritmo verrà descritto con riferimento alle sole SCs e all'IF, essendo implicito che identica procedura viene seguita per le ASJCs e per altri indicatori bibliometrici;

Se la rivista appartiene a più di una SC, si utilizza, ai fini dell'individuazione univoca della SC, l'indicazione del soggetto valutato che ha proposto l'articolo, o, se necessario, l'eventuale modifica da parte del GEV;

Sia in ISI che in Scopus esiste la categoria: “multidisciplinary science”, che include riviste caratterizzate da una pluralità di argomenti scientifici, quali Nature, Science, ecc.. Gli articoli pubblicati su una rivista di tale categoria saranno riassegnati ad un'altra SC sulla base delle citazioni contenute nell'articolo. In particolare, per ognuno 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;

Si calcola la funzione di distribuzione cumulativa empirica dell'IF delle riviste appartenenti alla SC individuate per l'anno di pubblicazione dell'articolo da valutare;

Si divide la funzione di distribuzione cumulativa in 4 classi, caratterizzate dai valori di probabilità 0.2, 0.2, 0.1, 0.5;

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;

Si divide la funzione di distribuzione cumulativa del numero di citazioni in 4 classi, caratterizzate dai valori di probabilità 0.2, 0.2, 0.1, 0.5;

Dati l'IF e il numero di citazioni dell'articolo da valutare, esso viene attribuito ad una delle 16 coppie di classi rappresentate in una matrice 4x4, che contiene sulle ascisse le classi della distribuzione dell'IF e sulle ordinate le classi della distribuzione del numero di citazioni;

L'attribuzione della classe finale di merito avviene secondo l'algoritmo seguente, nel quale la lettera "A" si riferisce alla classe finale "eccellente", la "B" a "buono", la "C" a accettabile, e la "D" a "limitato":

Quando le coordinate dell'articolo lo posizionano in una delle 4 caselle della diagonale principale, e quindi le 2 indicazioni basate su IF e su citazioni coincidono, la classe finale è la stessa (vedi Figura 1).

Bibliometry (IF, SJR,...)

| | A | B | C | D |
|---|---|---|---|---|
| A | A | | | |
| B | | B | | |
| C | | | C | |
| D | | | | D |

Figura 1. Matrice di corrispondenza tra classi iniziali di IF e citazioni e classe finale VQR

Quando le coordinate danno indicazioni diverse (casella non appartenente alla diagonale principale), occorre distinguere due casi a seconda della data di pubblicazione dell'articolo.

Se l'articolo è stato pubblicato nel quinquennio 2004-2008, a potendo quindi considerare sufficientemente consolidata la distribuzione di citazioni ad esso associata, si utilizza la matrice riportata in Figura 2 per l'attribuzione della classe finale di merito, nella quale si privilegia l'indicazione fornita dalla classificazione citazionale.

2004-2008
Bibliometry (IF)

| | | A | B | C | D |
|-----------|---|----|---|---|----|
| Citations | A | A | A | A | PR |
| | B | B | B | B | PR |
| | C | PR | C | C | C |
| | D | PR | D | D | D |

Figura 2. Matrice di corrispondenza tra classi iniziali di IF e citazioni e classe finale VQR per articoli pubblicati nel quinquennio 2004-2008

Se l'articolo è stato pubblicato nel biennio 2009-2010, 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'IF.

2009-2010

Bibliometry (IF)

| | | A | B | C | D |
|----------|---|----|----|----|----|
| Citation | A | A | PR | PR | PR |
| | B | A | B | C | D |
| | C | A | B | C | D |
| | D | PR | PR | PR | D |

Figura 3. Matrice di corrispondenza tra classi iniziali di IF e citazioni e classe finale VQR per articoli pubblicati nel biennio 2009-2010

Le caselle delle matrici di Figura 2 e Figura 3 etichettate “IR” si riferiscono ai casi nei quali la valutazione non verrà fatta in maniera automatica, ma mediante la “informed peer review”. In tali casi, il GEV valuterà direttamente l’articolo, o, in mancanza al suo interno delle competenze necessarie, lo affiderà alla “peer review”.

L’algoritmo distinguerà, calcolando distribuzioni cumulative empiriche separate per le citazioni, gli articoli “scientifici” da quelli di rassegna, che ricevono notoriamente un maggior numero di citazioni.



6.5. Gestione dei conflitti

Nel caso di articoli sottoposti a valutazione bibliometrica utilizzando sia ISI WoS che Scopus che ottengono una valutazione diversa si adatterà la valutazione più favorevole.

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

6.6. Assenza di indicatori di citazione

Tutti i prodotti non contenuti nelle basi di dati citazionali ISI WoS e Scopus saranno sottoposti a valutazione diretta da parte del GEV, o, in assenza delle competenze necessarie al suo interno, a “peer review” da parte di revisori esterni selezionati dal GEV.

7. Altri prodotti

Come già detto in precedenza, tutti i prodotti privi di indicatori di citazione (tra cui le tipologie c,d,e del Bando) verranno inviati alla peer review.

Gli abstract relativi ad atti di congresso pubblicati su riviste con codice ISSN saranno sottoposti a peer review, ma non potranno ricevere una classificazione VQR migliore di C.

I lavori di rassegna privi di indici bibliometrici saranno sottoposte a peer review, ma non potranno ricevere una classificazione VQR migliore di C.

I brevetti saranno sottoposti a peer review. I livelli di merito VQR A (eccellente) e B (buono) potranno essere assegnati esclusivamente a brevetti internazionali approvati in Europa, USA o Giappone e che siano già stati ceduti o dati in licenza a un'azienda.

8. Risoluzione dei conflitti di interesse

Sulla base delle linee guida GEV dell'ANVUR, i membri dei GEV si asterranno 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 negli anni 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, negli anni 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 incaricherà delle procedure di valutazione un altro membro del GEV per i quali non vi siano conflitti di interesse.

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



Appendice C. Considerazioni e analisi specifiche dell'area 03.

Ai membri del GEV03 è sembrato opportuno dedicare quest'appendice a due aspetti particolari. Il primo è il problema delle aree molto interdisciplinari che sarà qui declinato con esplicito riferimento al settore dei beni artistici e culturali (SSD Chim12 ma non solo). Il secondo aspetto è una breve valutazione della procedura di valutazione da parte dei membri stranieri del GEV in modo da avere un punto di vista più 'internazionale'. Partiamo dunque dal primo aspetto

L'analisi chimico-fisica dei materiali d'interesse per i beni artistici e culturali è un perfetto modello di collaborazione con curatori di musei, storici, antropologi e/o con altri specialisti che partecipano assieme ai chimici nel fornire risposte a questioni interessanti e importanti che riguardano lavori archeologici o opere d'arte o altri meritevoli di studio: quali sono i materiali? Com'è stata eseguita l'opera? Com'è cambiata o si è deteriorata l'opera nel tempo? Poiché nessun singolo esperto è in grado di fornire una risposta a tutte queste domande, esperti di molte aree diverse che utilizzano vari tipi di tecniche strumentali, analitiche, spettroscopiche e informatiche sono accomunati nella pubblicazione dei risultati in modo da mettere in comune le loro conoscenze ed *expertise*.

Quest'area multidisciplinare si fonda su applicazioni innovative di tecniche più o meno tradizionali di analisi chimico/fisica e sta acquisendo un'importanza sempre crescente. I continui miglioramenti della tecnologia stanno facendo sì che la chimica analitica, la spettroscopia, l'*imaging*, la modellistica computazionale e la grafica/visualizzazione computerizzata rappresentino strumenti sempre più efficienti e integrati per lo studio della provenienza di oggetti culturalmente importanti, e per la soluzione dei problemi connessi alla degradazione di tali oggetti a causa delle più recenti variazioni ambientali che includono, tra l'altro, i cambiamenti climatici e l'inquinamento urbano.

Sebbene il ruolo di quest'area di ricerca sia molto importante ed in continua crescita, la possibilità di pubblicare lavori in giornali del settore con alto IF è tuttora piuttosto ridotta. E' interessante notare che solo pochi lavori multidisciplinari sono pubblicati su riviste generaliste ad alto IF come i Proceedings of the National Academy of Sciences (IF 9,681), Nature (IF 36.280), Analytical Chemistry (IF 5.856), Langmuir (IF 4.186): questi lavori riguardano nuovi sviluppi di strumentazioni, tecnologie o materiali d'interesse nel campo dei beni culturali oppure argomenti

particolarmente attraenti, spesso collegati a scoperte archeologiche. In realtà, anche se il numero di riviste dedicate ai beni culturali è sempre più rilevante, la tabella seguente, che raccoglie le riviste più accreditate del settore, mostra che solo poche di esse ha un IF.

| <i>Rivista</i> | <i>IF, 2011</i> |
|--|-----------------|
| Journal of Archaeological Science | 1,914 |
| Archaeometry | 1,183 |
| Journal of Cultural Heritage | 1,079 |
| Studies in Conservation | 0,4 |
| Archaeological and Anthropological Science | |
| e-preservation science (open access) | |
| Conservation Science in Cultural Heritage | |
| Heritage Science (open access) | |
| Canadian Journal of Cultural Studies | |
| Kermes | |
| Science and Technology for Cultural heritage | |
| National Gallery Technical Bulletin | |
| | |

Ad esempio, il National Gallery Technical Bulletin, pubblicato dal 1977, non ha un IF, anche se la rivista ha raggiunto una posizione leader nello studio dei materiali e delle tecniche della pittura e nell'analisi scientifica dei quadri: essa rappresenta una lettura indispensabile per conservatori, scienziati della conservazione, storici dell'arte, collezionisti e curatori. Inoltre, numerosi Congressi Internazionali pubblicano anteprime o atti dei lavori (pre- o post-prints) che sono molto utili per la disseminazione della conoscenza scientifica verso le comunità dei conservatori e dei restauratori, ma il valore di tali pubblicazioni per la nostra comunità scientifica è pressoché nullo. Un esempio specifico è rappresentato dalla pubblicazione della Conferenza triennale dell'ICOM-CC (International Council of Museums- Committee for Conservation) con lo scopo

di promuovere l'investigazione, analisi e conservazione di opere particolarmente rilevanti da un punto di vista culturale o storico. Tali pubblicazioni sono sottomesse a una revisione particolarmente severa da parte di esperti del campo ed hanno un'importante diffusione a livello mondiale: purtroppo, tuttavia, non è disponibile un IF. Allo stesso modo, tutti i lavori riguardanti le applicazioni di analisi strumentale e/o di procedure analitiche specifiche alla soluzione di problemi concernenti la caratterizzazione dei materiali utilizzati in oggetti d'interesse storico o artistico appaiono per la maggior parte su riviste a basso IF, con una conseguente sottostima del loro valore nella nostra VQR, a dispetto del loro alto valore intrinseco.

Gli aspetti appena analizzati spiegano perché la maggior parte dei lavori valutati dal GEV03 e dedicati ai beni culturali ha ricevuto una classificazione C o D. Per questo e altri motivi la definizione attuale degli SSD appare in qualche senso limitativa e non del tutto soddisfacente. Solo per fornire un altro esempio, le riviste che pubblicano lavori nel campo delle nanoscienze o dei materiali avanzati hanno IF molto più alti di quelle dedicate alla termodinamica chimica. E pur tuttavia tutte queste aree appartengono in maggioranza allo SSD Chim02. Ancora una volta, ci sembra di poter concludere che una valutazione comparativa più approfondita richiede analisi complesse e articolate condotte da specialisti dei vari settori e un confronto tra settori omologhi nel panorama internazionale, in sostanza una qualche forma di normalizzazione rispetto al livello di citazioni, IF, ecc. del corrispondente (ristretto) settore a livello internazionale.

Per quanto riguarda il secondo punto, riportiamo qui sotto alcune considerazioni elaborate dai colleghi stranieri partecipanti al GEV3.

In termini generali la procedura è stata abbastanza complessa fin dall'inizio e ciò non è sorprendente in vista delle dimensioni del processo di valutazione, essendo questa la prima volta che una valutazione di queste dimensioni è stata condotta a livello globale. Nel seguito sono riportati alcuni commenti sulle varie fasi del processo di valutazione.

- Procedura di assegnazione ai membri del GEV: l'assegnazione dei prodotti a 2 membri diversi del GEV è sicuramente necessaria ed è stata condotta in maniera corretta. Il numero di prodotti assegnati a ciascun membro del GEV è stato probabilmente troppo elevato, ma il tempo supplementare concesso alla fine della procedura valutativa ha reso possibile di condurre a buon fine il lavoro.
- Assegnazione dei prodotti ai valutatori: il processo di selezione dei valutatori esterni è stato abbastanza complesso e in questo caso l'uso esclusivo di parole chiave (alcune in Italiano, alcune in Inglese, acronimi, ecc.) ha complicato il lavoro. Ad esempio, è stato in



pratica impossibile ritrovare nel sistema un valutatore ben noto al membro del GEV con un nome incompleto o senza indicazione precisa dell'area. A questo proposito il nostro suggerimento sarebbe di non assegnare i revisori ad aree specifiche, ma di lasciarli tutti in un database unico cui sia possibile accedere ogni volta che sia necessario eseguire un'assegnazione. Inoltre la selezione delle parole chiave richiede, a nostro avviso, una profonda revisione.

- Dopo la selezione di un revisore è stato necessario accedere al database un numero eccessivo di volte per accertare l'effettiva accettazione. Inoltre il tempo concesso ai revisori per accettare di esaminare ogni prodotto è stato eccessivo. Un possibile miglioramento potrebbe consistere nell'invio automatico da parte del sistema di una segnalazione o un messaggio al membro del GEV che ha gestito l'assegnazione per aggiornare l'informazione sul prodotto assegnato. Tale approccio potrebbe essere utilizzato anche in altri passaggi della procedura.
- Il sito su cui è stata fatta la valutazione dei prodotti è stato in generale robusto e facile da utilizzare: l'unico problema rilevante è stato legato alla necessità di chiudere definitivamente la valutazione per conservare i commenti in essa contenuti. Sarebbe particolarmente utile poter salvare una bozza dei giudizi prima di chiudere definitivamente la valutazione.
- Talvolta la classificazione bibliometrica fornita dal sistema è stata difforme da quella condotta autonomamente sui diversi database soprattutto in termini di numero di citazioni.
- Il resto della procedura (accordo tra i membri del GEV, differenze segnalate da diversi colori, ecc.) è stato ben congegnato e facile da seguire.
- I criteri utilizzati per definire e risolvere i disaccordi sono stati molto validi.
- In termini generali l'intera procedura è risultata ben congegnata e condotta con competenza.