



Munich Personal RePEc Archive

Bibliometric and peer review methods for research evaluation: a methodological appraisement

Cicero, Tindaro and Malgarini, Marco and Nappi, Carmela
Anna and Peracchi, Franco

ANVUR, ANVUR, ANVUR, Department of Economic and Finance,
University of Rome Tor Vergata and EIEF

6 October 2013

Online at <https://mpra.ub.uni-muenchen.de/50470/>
MPRA Paper No. 50470, posted 08 Oct 2013 19:30 UTC

Metodi bibliometrici e revisione dei pari per la valutazione della ricerca: un confronto metodologico

Tindaro Cicero (*), Marco Malgarini (*), Carmela Anna Nappi (*), Franco Peracchi (**)

Ottobre 2013

Abstract

The Italian Research Evaluation exercise for the period 2004-2010 has analyzed almost 185,000 among articles, books, patents and other scientific products submitted by Italian Universities and other public research bodies. In most cases, scientific publications have been peer reviewed; however, in hard sciences, medicines, engineering and economics, bibliometric indicators have also been used. For those areas, we have extracted a representative sample of scientific products, equal to the 10% of the reference population of submitted products, to be evaluated both with peer review and bibliometric methods. Our analysis shows that peer review and bibliometric evaluations exhibit a level of concordance higher than that observed among two different reviewers of the same article. In almost any scientific discipline, however, there is a systematic difference among peer and bibliometric evaluations: more specifically, bibliometric scores are on average significantly higher than those obtained with the peer review. Overall, our results obtained fully support the choice adopted in the Italian exercise of using both evaluation techniques in order to assess the quality of Italian research institutions.

Sintesi

L'esercizio di Valutazione della Qualità della Ricerca (VQR), svolto in Italia con riferimento al periodo 2004-2010, ha analizzato circa 185.000 articoli, monografie, brevetti e altri prodotti scientifici sottomessi per la valutazione dalle Università italiane e da altri enti pubblici di ricerca. Nella maggioranza dei casi, la valutazione si è basata sul metodo della revisione dei pari (*peer review*); nelle scienze esatte e in quelle mediche, ingegneristiche, psicologiche ed economiche si è fatto anche uso di indicatori di tipo bibliometrico, estratti dalle principali banche dati internazionali disponibili nel campo. In queste aree è stato selezionato un campione rappresentativo di articoli, pari a circa il 10% della popolazione di riferimento, che sono stati sottoposti a entrambi i metodi di valutazione: in questo modo, è stato possibile effettuare un confronto statistico approfondito tra i due diversi metodi. I risultati ottenuti evidenziano che esiste un più che adeguato grado di concordanza tra le valutazioni effettuate con i due metodi; la concordanza tra valutazione bibliometrica e *peer* è anzi leggermente superiore a quella riscontrata in media tra le valutazioni dello stesso prodotto effettuate da due diversi revisori. In pressoché tutte le aree emerge anche evidenza di differenze sistematiche, di segno positivo, tra i punteggi corrispondenti alle valutazioni *peer* e bibliometriche: ossia, la valutazione bibliometrica è in media significativamente più favorevole rispetto a quella *peer*. I risultati giustificano dunque pienamente la scelta fatta nella VQR di utilizzare entrambe le tecniche di valutazione.

Parole chiave: Indicatori bibliometrici; revisione dei pari; valutazione della ricerca

Classificazione JEL: I20; I23

* ANVUR.

** Dipartimento di Economia e Finanza, Università degli Studi di Roma Tor Vergata, e EIEF .

1. Introduzione¹

L'esercizio di Valutazione della Qualità della Ricerca (VQR) per il periodo 2004-2010 è stato curato dall'Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca (ANVUR) tra il novembre 2011 e il luglio 2013 ed ha analizzato la qualità di circa 185,000 tra articoli e monografie, brevetti e altri prodotti scientifici, sottomessi a valutazione dalle Università italiane, dagli Enti di ricerca vigilati dal Ministero dell'Università e della Ricerca (MIUR), e da altri enti e consorzi che hanno aderito volontariamente alla VQR (cfr. ANVUR, 2013).

Nella maggioranza dei casi, la valutazione si è basata sul metodo della revisione dei pari (*peer review*); nelle scienze esatte, mediche, ingegneristiche, psicologiche ed economiche si è fatto anche uso di indicatori di tipo bibliometrico, estratti dalle principali banche dati internazionali disponibili nel campo. Nelle aree in cui si sono utilizzati gli indicatori bibliometrici, è stato selezionato, con un algoritmo di estrazione casuale in grado di garantire una buona copertura statistica di tutti i settori, un numero pari a circa il 10% degli articoli valutati bibliometricamente; gli articoli estratti sono stati sottoposti a revisione dei pari, così da rendere possibile un confronto tra le due metodologie di valutazione. Tale confronto è basato su una notevole mole di informazioni (circa 10,000 articoli) ed è dunque particolarmente robusto, anche rispetto ad analoghi esercizi condotti in letteratura (cfr. ad esempio Bornmann e Leydersdorff, 2012).

Nel seguito, la sezione 2 presenta il campione statistico utilizzato, la 3 descrive nel dettaglio le metodologie utilizzate e la 4 presenta i risultati del confronto; la sezione 5 conclude il lavoro.

2. Il campione statistico

Un campione casuale di 9.199 articoli su rivista passibili di valutazione bibliometrica è stato estratto dalla popolazione di 99.005 articoli valutabili bibliometricamente e sottomessi alla valutazione nelle cosiddette "aree bibliometriche", cioè nelle aree scientifiche che hanno utilizzato indicatori bibliometrici (scienze matematiche e informatiche, scienze fisiche, scienze chimiche, scienze della terra, scienze biologiche, scienze mediche, scienze agrarie e veterinarie, ingegneria civile e architettura, ingegneria industriale e dell'informazione, e scienze economiche e statistiche). La popolazione è stata stratificata in base alla distribuzione dei prodotti all'interno dei settori individuati nelle varie aree. Ai fini della stratificazione, gli articoli sono stati attribuiti ai settori sulla base del settore scientifico-disciplinare (SSD) nel quale sono stati valutati, eliminando le duplicazioni dovute alla presentazione di uno stesso articolo da parte di autori diversi all'interno di uno stesso strato campionario. Complessivamente, il campione include il 9,3% degli articoli sottoposti a valutazione bibliometrica nelle aree bibliometriche. L'estrazione è stata effettuata nel settembre 2012, prima dell'inizio del processo di revisione *peer*, mediante una procedura casuale

¹ Una precedente versione del lavoro è stata pubblicata come Appendice del Rapporto Finale ANVUR sulla Valutazione della Qualità della Ricerca 2004-2010, disponibile all'indirizzo <http://www.anvur.org/rapporto/>. Gli autori ringraziano il Professor Sergio Benedetto, coordinatore della VQR, per le numerose utili discussioni avute nel corso del lavoro. Un sentito ringraziamento va anche ai tecnici del CINECA che hanno messo a disposizione i dati. Ogni eventuale errore rimane ovviamente di esclusiva responsabilità degli autori.

con il vincolo di selezionare una proporzione significativa di prodotti in ciascun settore. La Tabella 1 riporta l'elenco delle aree bibliometriche e, per ciascuna di esse, la dimensione della popolazione e del campione estratto in valori assoluti e in percentuale sulla popolazione.

Tabella 1: Distribuzione degli articoli su rivista nel campione e nella popolazione per ognuna delle aree bibliometriche

AREE	Popolazione	Campione	%
Scienze matematiche e informatiche	6758	631	9,3
Scienze fisiche	15029	1412	9,4
Scienze chimiche	10127	927	9,2
Scienze della terra	5083	458	9,0
Scienze biologiche	14043	1310	9,3
Scienze mediche	21191	1984	9,4
Scienze agrarie e veterinarie	6284	532	8,5
Ingegneria civile e architettura	2460	225	9,1
Ingegneria industriale e dell'informazione	12349	1130	9,2
Scienze economiche e statistiche	5681	590	10,4
Totale	99005	9199	9,3

La Tabella 2 riporta la distribuzione nelle classi di valutazione VQR, cioè Eccellente (E), Buono (B), Accettabile (A), Limitato (L), Incerto (IR), per la popolazione e per il campione di ogni area determinata dalla valutazione bibliometrica degli articoli su rivista. La distribuzione delle valutazioni bibliometriche (E/ B/ A/ L/ IR) è sufficientemente vicina nella popolazione e nel campione, sia per il totale che per le singole aree, e induce alla conclusione che il campione estratto ben rappresenti la popolazione di riferimento.

Tabella 2 Distribuzione delle valutazioni bibliometriche nel campione e nella popolazione per ogni area e per il complesso della VQR

Classe	Popolazione	%	Campione	%
Scienze matematiche e informatiche				
E	3101	45,9	277	43,9
B	822	12,2	75	11,9
A	328	4,9	34	5,4
L	593	8,8	52	8,2
IR	1914	28,3	193	30,6
Scienze fisiche				
E	8370	55,7	808	57,2
B	2410	16,0	239	16,9
A	1088	7,2	86	6,1
L	1043	6,9	79	5,6
IR	2118	14,1	200	14,2
Scienze chimiche				
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
Scienze della terra				
E	1785	35,1	178	38,9
B	1081	21,3	91	19,9
A	303	6,0	26	5,7
L	1011	19,9	82	17,9
IR	903	17,8	81	17,7
Scienze biologiche				
E	5580	39,7	527	40,2
B	2691	19,2	235	17,9
A	647	4,6	66	5,0
L	2500	17,8	230	17,6
IR	2625	18,7	252	19,2
Scienze mediche				
E	8547	40,3	774	39,0
B	3473	16,4	354	17,9
A	1029	4,9	75	3,8
L	3893	18,4	399	20,1
IR	4249	20,1	382	19,2
Scienze agrarie e veterinarie				
E	3586	57,1	295	55,5
B	702	11,2	54	10,2
A	60	1,0	6	1,1
L	892	14,2	70	13,2
IR	1044	16,6	107	20,1
Ingegneria civile e architettura				
E	1570	63,8	152	67,6
B	199	8,1	12	5,3
A	71	2,9	5	2,2
L	297	12,1	29	12,9
IR	323	13,1	27	12,0
Ingegneria industriale e dell'informazione				
E	7762	62,9	712	63,0
B	1384	11,2	125	11,1
A	485	3,9	52	4,6
L	419	3,4	30	2,7
IR	2299	18,6	211	18,7
Scienze economiche e statistiche				
E	1923	33,9	198	33,6

B	1000	17,6	102	17,3
A	856	15,1	103	17,5
L	1895	33,4	187	31,7
IR				
TOTALE				
E	47583	48,1	4 419	48,0
B	15739	15,9	1 457	15,8
A	5180	5,2	479	5,2
L	1 486	13,6	1 242	13,5
IR	17010	17,2	1 602	17,4

3. Le modalità di confronto

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

- valutazione del primo revisore (P1);
- valutazione del secondo revisore (P2);
- valutazione di sintesi dei giudizi del primo e secondo revisore (P);
- valutazione bibliometrica (F).

Le variabili P, P1 e P2 associano ciascun prodotto a 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 una *informed peer review* nel caso di risultati divergenti dei due indicatori bibliometrici (*Impact Factor* e numero di citazioni). In base alle regole VQR, le quattro classi sono definite con riferimento ai percentili della distribuzione della qualità degli articoli pubblicati nel mondo. In particolare, un articolo è classificato come E se si colloca nel 20% superiore della distribuzione mondiale della qualità degli articoli pubblicati, come B se si colloca nel successivo 20%, come A se si colloca nel successivo 10% e, infine, come L se si colloca nel 50% inferiore. Le variabili P1 e P2 sono originariamente espresse su una scala numerica compresa, a seconda dell'area, tra 3 e 27 o tra 0 e 9, con un punteggio da 1 a 9 (o da 0 a 3) assegnato ai tre criteri di rilevanza, originalità/innovazione e internazionalizzazione. Tali punteggi sono successivamente utilizzati per determinare, sulla base dei criteri fissati da ciascuna area,² la classe di valutazione *peer* di ciascun prodotto; le variabili P e F sono invece rispettivamente espresse in termini delle 4 o 5 classi di valutazione sopra elencate. 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 nei Rapporti di Area (cfr. ANVUR 2013). 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 in una valutazione finale basata su algoritmi specifici per ciascun'area. Al fine di confrontare i risultati della valutazione bibliometrica e della revisione dei pari, si procede nel seguito a confrontare gli indicatori F e P. Anche altri confronti possono essere tuttavia d'importanza

² L'etichetta "P1" e "P2" 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 delle valutazioni tra pari P1 e P2 consente di valutare il grado di corrispondenza dei giudizi tra i due revisori.

4. I risultati

4.1 Le distribuzioni delle valutazioni F e P

Le distribuzioni delle valutazioni F e P sopra descritte non sono immediatamente confrontabili, in quanto la distribuzione F delle valutazioni bibliometriche comprende una classe IR che non è invece prevista nella valutazione tra pari. È però possibile ipotizzare che una discordanza di almeno due classi tra la valutazione dei due revisori segnali un'incertezza della revisione dei pari analoga a quella che emerge dal confronto tra numero di citazioni del lavoro e *Impact Factor* della rivista nell'analisi bibliometrica; in analogia con la classificazione IR della valutazione bibliometrica, si è creata dunque una classificazione "incerta peer" (IP) per la revisione dei pari, che consente il confronto tra le distribuzioni F e P³. La Tabella 3 mostra la distribuzione in numeri assoluti e percentuali degli indicatori F e P per il totale del campione.

Tabella 3 - Confronto tra le valutazioni F e P

Valutazione bibliometrica (F)	Valutazione peer (P)					Totale
	E	B	A	L	IP	
E	1255	1828	308	212	816	4419
% rispetto al totale delle valutazioni bibliometriche di classe E	28,4	41,4	7,0	4,8	18,5	100,0
B	119	620	190	208	320	1457
% rispetto al totale delle valutazioni bibliometriche di classe B	8,2	42,6	13,0	14,3	22,0	100,0
A	13	130	89	149	98	479
% rispetto al totale delle valutazioni bibliometriche di classe A	2,7	27,1	18,6	31,1	20,5	100,0
L	19	197	153	607	266	1242
% rispetto al totale delle valutazioni bibliometriche di classe L	1,5	15,9	12,3	48,9	21,4	100,0
IR	125	546	167	345	419	1602
% rispetto al totale delle valutazioni bibliometriche IR	7,8	34,1	10,4	21,5	26,2	100,0
Totale	1531	3321	907	1521	1919	9199
% rispetto al totale delle valutazioni bibliometriche	16,6	36,1	9,9	16,5	20,9	100,0

³ I criteri adottati dall'area Scienze economiche e statistiche non prevedono una classe di assegnazione IR per l'algoritmo bibliometrico.

Gli elementi sulla diagonale principale della Tabella 3 corrispondono ai casi in cui la valutazione tra pari e quella bibliometrica coincidono. Quelli 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 principale). La tabella mostra che la maggiore discordanza tra la valutazione bibliometrica e quella dei pari è dovuta al fatto la prima tende a essere più generosa. In particolare, gli articoli classificati come eccellenti sulla base degli indicatori bibliometrici sono 4.419, quasi il triplo di quelli classificati come eccellenti dalla valutazione tra pari: solo il 28,4% degli articoli classificati come E secondo la bibliometria ottiene E anche secondo la valutazione tra pari, mentre rispettivamente nel 41,4%, 7% e 4,8% dei casi gli articoli bibliometricamente eccellenti risultano buoni, accettabili o limitati secondo la valutazione tra pari. D'altro lato, il numero di articoli classificati in B, A e L dalla valutazione tra pari (3.321, 907 e 1521 articoli rispettivamente) è nettamente maggiore di quello degli articoli che risultano in B, A e L secondo la valutazione bibliometrica (rispettivamente 1.457, 479 e 1.242 articoli). Infine, il numero di valutazioni incerte è maggiore nella revisione dei pari (1.919 articoli) rispetto a quella bibliometrica (1.602 articoli). Le valutazioni incerte secondo l'analisi bibliometrica hanno in circa il 42% dei casi una valutazione almeno pari a B secondo la *peer review*, mentre quelle incerte secondo la *peer review* hanno nel 60% circa dei casi una valutazione almeno pari a B secondo l'analisi bibliometrica.

Complessivamente, l'analisi bibliometrica e la *peer review* coincidono nel 32,5% dei casi. Se si sommano alle valutazioni coincidenti quelle che differiscono di una sola classe, si arriva a oltre il 60% del campione. Gli articoli con valutazioni che differiscono per due classi sono 726, il 7,9% del campione, quelli con massima discordanza (ossia, che differiscono per 3 classi) sono 231 (il 2,5% del campione). Il restante 29,2% del campione ha un'assegnazione incerta con uno dei due metodi, e perviene invece a una classe di assegnazione definita secondo l'altro metodo.

La Tabella 4 mostra la distribuzione degli indicatori P1 e P2. Le valutazioni dei due revisori coincidono nel 40,3% dei casi, differiscono di una classe di valutazione nel 37,8% dei casi e differiscono per due o tre classi di valutazione rispettivamente nel 17,2% e nel 4,8% dei casi. E' da notare anche che vi è accordo sul giudizio di assegnazione alla classe E in 1.127 casi, pari a circa il 47% delle valutazioni eccellenti fornite dal primo (2.431) o dal secondo (2.394) revisore.

Tabella 4: Confronto tra le valutazioni P1 e P2

P1	P2				Totale
	E	B	A	L	
E	1127	853	227	224	2431
% rispetto al totale delle valutazioni di classe E del primo revisore	46,4	35,1	9,3	9,2	100
B	834	1515	510	568	3427
% rispetto al totale delle valutazioni di classe B del primo revisore	24,3	44,2	14,9	16,6	100
A	220	501	241	399	1361
% rispetto al totale delle valutazioni di classe A del primo revisore	16,2	36,8	17,7	29,3	100
L	213	572	373	822	1980
% rispetto al totale delle valutazioni di classe L del primo revisore	10,8	28,9	18,8	41,5	100
Totale	2394	3441	1351	2013	9199
% rispetto al totale delle valutazioni del primo revisore	26,0	37,4	14,7	21,9	100

Le Tabelle 5 e 6 sono identiche alle Tabelle 3 e 4 ma si riferiscono alle singole aree. In particolare, dall'analisi dei dati della Tabella 5 emerge che il numero di valutazioni eccellenti è sempre maggiore per la valutazione bibliometrica rispetto alla revisione dei pari. Tenuto conto del numero complessivo di articoli valutati in ciascuna area, la differenza tra il numero di articoli classificati come eccellenti secondo i due metodi è particolarmente forte in Ingegneria civile e architettura (152 articoli su 225 classificati E sulla base dell'algoritmo bibliometrico, contro i soli 52 articoli eccellenti secondo la valutazione tra pari), in Ingegneria industriale e dell'informazione (712 e 239 articoli eccellenti su un totale di 1.130 articoli, rispettivamente secondo l'analisi bibliometrica e quella *peer*) e nelle Scienze fisiche (808 e 300 articoli rispettivamente, su 1.412 totali); le differenze sono invece più modeste in Scienze matematiche e informatiche (277 articoli contro 165) e nelle Scienze economiche e statistiche (198 articoli rispetto a 116). D'altra parte, il numero di valutazioni buone, accettabili o limitate è in genere maggiore per la valutazione *peer* rispetto a quella bibliometrica; fanno eccezione le Scienze chimiche e le Scienze economiche e statistiche, dove il numero di articoli classificati come limitati è maggiore secondo l'analisi bibliometrica (84 e 187 articoli rispettivamente nei due GEV) rispetto alla revisione dei pari (62 e 171 articoli rispettivamente).

Complessivamente, la tendenza della valutazione bibliometrica a essere più vantaggiosa di quella *peer* è comune a tutte le aree (si vedano le tabelle A e B nell'appendice; un test statistico che conferma tale ipotesi è presentato nella sezione 3). Le due valutazioni sono in genere concordi in circa un terzo dei casi; nelle Scienze economiche e statistiche, la concordanza è maggiore e supera il 50%. I casi di discordanza di due classi o più tra le due valutazioni sono generalmente inferiori o

al più uguali al 10%; una discordanza più elevata si registra nelle Scienze della terra e nelle Scienze mediche, dove i casi in cui le valutazioni divergono per 2 o 3 classi raggiungono il 13% circa. Dall'esame dei dati emerge anche che la convergenza nelle valutazioni dei due revisori è in genere almeno pari o comunque molto vicina nel 40% dei casi; una minore concordanza si registra invece nelle Scienze agrarie e veterinarie (32% di valutazioni concordi), una concordanza maggiore si ha nelle Scienze matematiche e informatiche e in Ingegneria civile e architettura (46% e 43% dei casi rispettivamente).

4.2 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. il grado di concordanza tra la distribuzione F e la distribuzione P, che analizza la tendenza di F e P ad assegnare lo stesso punteggio a ogni articolo;
2. il 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, sebbene il contrario non sia necessariamente vero. In generale, però, i due criteri misurano aspetti diversi 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 questo 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 (Cohen, 1960); 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.

La statistica K di Cohen è una misura del grado di concordanza tra giudizi qualitativi espressi sulla base di due diversi metodi o da due diversi revisori; rispetto al semplice calcolo della quota di valutazioni concordanti mostrato in precedenza, K tiene conto della possibile concordanza casuale esistente tra i due diversi metodi o revisori. In particolare, la statistica K è costruita in modo tale da essere pari a zero quando la concordanza tra due valutazioni è del tutto casuale, vale a dire nel caso in cui le valutazioni siano indipendenti l'una dall'altra, ed assume invece valore pari a 1 nel caso in

cui ci sia perfetta concordanza. Sulla base della stima dell'errore standard ad essa associato, ed assumendo una distribuzione Gaussiana o normale di K , è quindi possibile valutare se la statistica K è statisticamente diversa da zero ad un prescelto livello di confidenza statistica.

Quanto al grado di differenza sistematica tra le due valutazioni, si calcola in primo luogo la differenza osservata tra le due valutazioni per ciascun articolo e quindi si computa il valor medio delle differenze così calcolate. Si valuta quindi se il valor medio delle differenze tra le due distribuzioni è statisticamente pari a zero, assumendone una distribuzione t di Student; il test t si calcola quindi dividendo la media delle differenze per la sua deviazione standard: se il valore ottenuto è superiore al valore soglia della distribuzione t di Student corrispondente a una certa probabilità, si conclude che la media delle differenze è statisticamente diversa da zero, ossia che tra le due valutazioni esistono differenze sistematiche.

La Tabella 5 riporta i valori della statistica K di Cohen, calcolati per l'intero campione e separatamente per ciascuna area. I risultati sono riferiti a campioni omogenei (*paired sample*), ossia agli articoli per i quali sono disponibili sia i risultati della valutazione *peer* sia quelli relativi alla valutazione bibliometrica, eliminando cioè dal campione gli articoli per i quali la valutazione bibliometrica fornisce come risultato una classificazione IR; complessivamente, le informazioni a disposizione si riducono a 7598. E' possibile calcolare la statistica K 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 0,31; la tabella riporta anche la soglia inferiore e quella superiore della stima di K , calcolate a partire dalla stima dell'errore standard e per un livello di confidenza del 99%, assumendo una distribuzione di K Gaussiana o normale. La soglia inferiore della stima di K si avvicina in qualche caso allo zero, ma non lo raggiunge mai: è possibile concludere dunque che la concordanza registrata tra la valutazione bibliometrica e quella dei pari non è casuale ma sistematica, ossia che le due valutazioni non sono tra loro indipendenti e mostrano un grado di correlazione significativo. Come accennato sopra, il calcolo di K riportato nella prime tre colonne 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). Le colonne successive della Tabella 5 riportano i valori della statistica K calcolati utilizzando i pesi della VQR. I risultati mostrano che in questo caso la concordanza è maggiore o almeno uguale rispetto alle valutazioni basate su pesi lineari, a rafforzamento dell'ipotesi dell'esistenza di un buon grado di concordanza sia nel totale del campione che in ciascun AREE.

Tabella5: Statistica K di Cohen sul grado di concordanza

	F e P, pesi lineari			F e P, pesi VQR			P1 e P2, pesi lineari			P1 e P2, pesi VQR		
	soglia inferiore K	K	soglia superiore K	soglia inferiore K	K	soglia superiore K	soglia inferiore K	K	soglia superiore K	soglia inferiore K	K	soglia superiore K
Totale campione	0,30	0,32	0,33	0,36	0,38	0,40	0,26	0,29	0,31	0,31	0,33	0,35
Scienze matematiche e informatiche	0,24	0,32	0,40	0,23	0,32	0,40	0,27	0,36	0,45	0,26	0,35	0,44
Scienze fisiche	0,19	0,23	0,27	0,21	0,25	0,29	0,18	0,23	0,29	0,18	0,23	0,28
Scienze chimiche	0,17	0,22	0,28	0,17	0,23	0,29	0,19	0,25	0,31	0,17	0,24	0,30
Scienze della terra	0,20	0,28	0,36	0,21	0,30	0,39	0,15	0,25	0,35	0,15	0,25	0,36
Scienze biologiche	0,28	0,33	0,38	0,29	0,35	0,40	0,22	0,28	0,33	0,21	0,27	0,33
Scienze mediche	0,26	0,30	0,34	0,29	0,34	0,38	0,20	0,25	0,29	0,19	0,24	0,29
Scienze agrarie e veterinarie	0,21	0,28	0,34	0,27	0,34	0,42	0,07	0,16	0,24	0,17	0,27	0,36
Ingegneria civile e architettura	0,10	0,20	0,30	0,11	0,23	0,34	0,07	0,20	0,33	0,06	0,19	0,32
Ingegneria industriale e dell'informazione	0,12	0,16	0,20	0,13	0,17	0,21	0,13	0,19	0,25	0,12	0,18	0,24
Scienze economiche e statistiche	0,46	0,54	0,62	0,53	0,61	0,69	0,32	0,40	0,48	0,38	0,46	0,54

La Tabella 5 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. Nel complesso del campione, il grado di concordanza tra la valutazione bibliometrica (F) e la revisione peer (P) è leggermente superiore a quello esistente tra i giudizi formulati dai due revisori: in quest'ultimo caso, la statistica K calcolata sia con pesi lineari sia con i pesi VQR è pari rispettivamente a 0,29 e 0,33. Analoghi risultati si hanno a livello delle singole aree. Anche in questo caso, la statistica K è superiore allo zero considerando un intervallo di confidenza sia del 5% sia dell'1%.

La Tabella 6 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, 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 agli articoli per i quali sono disponibili sia i dati della valutazione *peer* sia quelli relativi alla valutazione bibliometrica, eliminando cioè dal campione gli articoli per i quali la valutazione bibliometrica fornisce come risultato una classificazione IR. Come ricordato sopra, gli articoli a disposizione in questo caso sono 7.598.

La terza colonna mostra che il punteggio medio finale della revisione *peer* (punteggio P) è pari a 0,626: il punteggio è superiore alla media, nell'ordine, in Scienze matematiche e informatiche, Ingegneria industriale e dell'informazione, Scienze fisiche, Ingegneria civile e architettura e Scienze chimiche ed è invece inferiore nelle altre aree. Le differenze tra aree che emergono dall'analisi dei dati della terza colonna della tabella possono essere attribuite:

- a una migliore qualità degli articoli sottomessi alla valutazione nelle aree dove il punteggio è superiore alla media complessiva;
- a una maggiore generosità dei revisori di quelle aree;
- all'intrinseca variabilità statistica nella scelta del campione.

La quarta colonna contiene il punteggio medio ottenuto nella valutazione bibliometrica: tale punteggio è pari a 0,767 per la media complessiva dei lavori valutabili bibliometricamente, risultando superiore alla media, in ordine crescente, in Scienze agrarie e veterinarie, Scienze matematiche e informatiche, Ingegneria industriale e dell'informazione, Scienze fisiche e Ingegneria civile e architettura. Le aree con punteggi superiori alla media sono dunque gli stessi sia che si guardi all'ordinamento basato sulla valutazione *peer* sia che si consideri quello basato sull'algoritmo bibliometrico.

L'ordinamento delle aree in base alla qualità degli articoli presentati è diverso: in particolare, secondo la *peer review* le Scienze chimiche ricevono le valutazioni migliori, seguite da Ingegneria civile e architettura, Scienze fisiche, Ingegneria industriale e dell'informazione, Scienze matematiche e informatiche, e Scienze agrarie e veterinarie; al di sotto della media generale si collocano in ordine decrescente di valutazione le Scienze biologiche, le Scienze economiche e statistiche, le Scienze mediche e le Scienze della terra, che ricevono le valutazioni meno favorevoli.

Secondo l'algoritmo bibliometrico, invece, le valutazioni migliori sono attribuite, nell'ordine, all'Ingegneria industriale e dell'informazione, alle Scienze fisiche, alle Scienze chimiche, all'Ingegneria civile e architettura, alle Scienze matematiche e informatiche, e alle Scienze agrarie e veterinarie; al di sotto della media generale si collocano infine, secondo l'algoritmo bibliometrico, in ordine decrescente di punteggio, le Scienze biologiche, le Scienze della terra, le Scienze mediche e le Scienze economiche e statistiche.

La quinta colonna della Tabella 6 presenta la differenza tra valutazione *peer* e bibliometrica, con le colonne 8-9 che riportano il risultato del test *t* per campioni di uguale ampiezza a 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 superiore rispetto a quella ottenuta con la valutazione *peer*. Il risultato è confermato anche dai dati riferiti a tutti le singole aree: fa eccezione la sola area di Scienze economiche e statistiche, nella quale la differenza tra la valutazione dei pari e quella bibliometrica non è statisticamente significativa agli usuali livelli di confidenza.

Tabella 6: Test t sulla differenza tra i punteggi bibliometrici e peer review

AREE	Punteggio P1	Punteggio P2	Punteggio P	Punteggio F	Diff F-P	# Osservazioni	Test t	p-value
Scienze matematiche e informatiche	0,711	0,688	0,676	0,808	0,132	438	7,828	0,000
Scienze fisiche	0,744	0,737	0,723	0,860	0,137	1212	16,407	0,000
Scienze chimiche	0,741	0,756	0,742	0,832	0,090	778	7,632	0,000
Scienze della terra	0,570	0,577	0,520	0,700	0,179	377	8,706	0,000
Scienze biologiche	0,611	0,607	0,556	0,707	0,151	1058	12,542	0,000
Scienze mediche	0,531	0,542	0,524	0,683	0,159	1603	15,938	0,000
Scienze agrarie e veterinarie	0,656	0,597	0,632	0,803	0,171	425	10,872	0,000
Ingegneria civile e architettura	0,740	0,731	0,728	0,829	0,101	198	3,895	0,000
Ingegneria industriale e dell'informazione	0,723	0,705	0,711	0,912	0,201	919	20,762	0,000
Scienze economiche e statistiche	0,498	0,528	0,542	0,561	0,019	590	1,417	0,157
Totale	0,652	0,646	0,626	0,767	0,140	7598	34,565	0,000

5. Conclusioni

Per il totale del campione di articoli sottoposti a valutazione nelle aree cosiddette bibliometriche, si riscontra una più che adeguata concordanza tra valutazioni effettuate con il metodo della revisione dei pari e con quello bibliometrico. Inoltre, il grado di concordanza tra la valutazione finale bibliometrica e quella *peer* è molto simile, anzi, leggermente superiore, al grado di concordanza tra le due valutazioni *peer*.

Emerge però in tutte le aree l'evidenza di differenze sistematiche e sempre di segno positivo (ossia, la valutazione bibliometrica è significativamente più favorevole in media rispetto a quella *peer*) tra i punteggi corrispondenti alle valutazioni *peer* e bibliometriche; fanno eccezione le Scienze economiche e statistiche, per le quali la differenza tra i risultati ottenuti con i due metodi di valutazione non è statisticamente significativa. In effetti, è possibile osservare che il numero di lavori classificati come eccellenti (E) dalla valutazione bibliometrica sia generalmente superiore a quello dei lavori classificati come eccellenti dalla valutazione tra pari.

La differenza tra *peer review* e valutazione bibliometrica in termini di valutazione media, e soprattutto la maggiore percentuale di giudizi eccellenti per la valutazione bibliometrica, non devono stupire. La *peer review* è infatti una metodologia di valutazione soggettiva, ed è prassi comune che i revisori attribuiscono in prevalenza valutazioni intermedie (buono e accettabile) piuttosto che estreme (eccellente o limitato). Ciò è dovuto anche alla percezione comune alla comunità scientifica che per eccellente si debba intendere una fascia ristretta di articoli, intorno al 10%, più bassa del 20% stabilito dalle regole della VQR.

Nella bibliometria, invece, i criteri seguiti dalle varie aree si sono per lo più basati sulla definizione del Bando VQR (20%, 20%, 10% e 50%). Inoltre, la pubblicazione dei criteri precedentemente al

conferimento dei prodotti ha consentito alle strutture di scegliere i prodotti da valutare in base a criteri precisi, che in pratica ne garantivano il risultato (ad esempio la classificazione in E).

È comunque importante sottolineare che l'analisi effettuata mostra che il grado di concordanza tra valutazione bibliometrica e valutazione *peer* è leggermente superiore al grado di concordanza tra le valutazioni *peer* di revisori diversi. Questo risultato giustifica pienamente la scelta fatta dalla VQR di utilizzare entrambe le tecniche di valutazione.

Riferimenti bibliografici

ANVUR (2013), Rapporto Finale Valutazione della Qualità della Ricerca 2004-2010, luglio, <http://www.anvur.org/rapporto/>

Bornmann, L., Leydersdorff L. (2012), The validation of (advanced) bibliometric indicators through peer assessments: A comparative study using data from InCites and F1000, *Journal of Infometrics*, vol. 7, II, Aprile

Cohen, J. (1960). "A coefficient of agreement for nominal scales". *Educational and Psychological Measurement* **20** (1): 37–46. [doi:10.1177/001316446002000104](https://doi.org/10.1177/001316446002000104).

Appendice statistica

Tabella A: Confronto tra le valutazioni F e P per area

Scienze matematiche e informatiche						
Valutazione bibliometrica (F)	Valutazione peer (P)					
	E	B	A	L	IP	Totale
E	114	92	13	8	50	277
% rispetto al totale delle valutazioni bibliometriche di classe E	41,2	33,2	4,7	2,9	18,1	100,0
B	7	28	7	22	11	75
% rispetto al totale delle valutazioni bibliometriche di classe B	9,3	37,3	9,3	29,3	14,7	100,0
A	1	7	2	14	10	34
% rispetto al totale delle valutazioni bibliometriche di classe A	2,9	20,6	5,9	41,2	29,4	100,0
L	2	13	2	22	13	52
% rispetto al totale delle valutazioni bibliometriche di classe L	3,8	25,0	3,8	42,3	25,0	100,0
IR	41	60	16	32	44	193
% rispetto al totale delle valutazioni bibliometriche IR	21,2	31,1	8,3	16,6	22,8	100,0
Totale	165	200	40	98	128	631
% rispetto al totale delle valutazioni bibliometriche	26,1	31,7	6,3	15,5	20,3	100,0
Scienze fisiche						
Valutazione bibliometrica (F)	Valutazione peer (P)					
	E	B	A	L	IP	Totale
E	252	371	45	14	126	808
% rispetto al totale delle valutazioni bibliometriche di classe E	31,2	45,9	5,6	1,7	15,6	100,0
B	28	112	32	23	44	239
% rispetto al totale delle valutazioni bibliometriche di classe B	11,7	46,9	13,4	9,6	18,4	100,0
A	2	34	10	20	20	86
% rispetto al totale delle valutazioni bibliometriche di classe A	2,3	39,5	11,6	23,3	23,3	100,0
L	3	19	11	33	13	79
% rispetto al totale delle valutazioni bibliometriche di classe L	3,8	24,1	13,9	41,8	16,5	100,0
IR	15	79	23	36	47	200
% rispetto al totale delle valutazioni bibliometriche IR	7,5	39,5	11,5	18,0	23,5	100,0
Totale	300	615	121	126	250	1 412
% rispetto al totale delle valutazioni bibliometriche	21,2	43,6	8,6	8,9	17,7	100,0
Scienze chimiche						
Valutazione bibliometrica (F)	Valutazione peer (P)					

	E	B	A	L	IP	Totale
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
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
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
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
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
Totale	197	454	66	62	148	927
% rispetto al totale delle valutazioni bibliometriche	21,3	49,0	7,1	6,7	16,0	100
Scienze della terra						
Valutazione bibliometrica (F)			Valutazione peer (P)			
	E	B	A	L	IP	Totale
E	31	73	19	16	39	178
% rispetto al totale delle valutazioni bibliometriche di classe E	17,4	41,0	10,7	9,0	21,9	100
B	4	30	12	18	27	91
% rispetto al totale delle valutazioni bibliometriche di classe B	4,4	33,0	13,2	19,8	29,7	100
A	1	6	4	9	6	26
% rispetto al totale delle valutazioni bibliometriche di classe A	3,8	23,1	15,4	34,6	23,1	100
L	0	9	13	43	17	82
% rispetto al totale delle valutazioni bibliometriche di classe L	0,0	11,0	15,9	52,4	20,7	100
IR	3	27	10	20	21	81
% rispetto al totale delle valutazioni bibliometriche IR	3,7	33,3	12,3	24,7	25,9	100
Totale	39	145	58	106	110	458
% rispetto al totale delle valutazioni bibliometriche	8,5	31,7	12,7	23,1	24,0	100
Scienze biologiche						
Valutazione bibliometrica (F)			Valutazione peer (P)			
	E	B	A	L	IP	Totale
E	143	204	34	32	114	527
% rispetto al totale delle valutazioni bibliometriche di classe E	27,1	38,7	6,5	6,1	21,6	100
B	20	86	25	38	66	235
% rispetto al totale delle valutazioni bibliometriche di classe B	8,5	36,6	10,6	16,2	28,1	100
A	0	8	9	25	24	66

% rispetto al totale delle valutazioni bibliometriche di classe A	0,0	12,1	13,6	37,9	36,4	100
L	0	38	20	106	66	230
% rispetto al totale delle valutazioni bibliometriche di classe L	0,0	16,5	8,7	46,1	28,7	100
IR	8	65	30	72	77	252
% rispetto al totale delle valutazioni bibliometriche IR	3,2	25,8	11,9	28,6	30,6	100
Totale	171	401	118	273	347	1 310
% rispetto al totale delle valutazioni bibliometriche	13,1	30,6	9,0	20,8	26,5	100
Scienze mediche						
Valutazione bibliometrica (F)			Valutazione peer (P)			
	E	B	A	L	IP	Totale
E	154	306	74	71	169	774
% rispetto al totale delle valutazioni bibliometriche di classe E	19,9	39,5	9,6	9,2	21,8	100
B	17	123	48	66	100	354
% rispetto al totale delle valutazioni bibliometriche di classe B	4,8	34,7	13,6	18,6	28,2	100
A	3	19	7	29	17	75
% rispetto al totale delle valutazioni bibliometriche di classe A	4,0	25,3	9,3	38,7	22,7	100
L	7	44	37	206	105	399
% rispetto al totale delle valutazioni bibliometriche di classe L	1,8	11,0	9,3	51,6	26,3	100
IR	9	98	40	120	115	382
% rispetto al totale delle valutazioni bibliometriche IR	2,4	25,7	10,5	31,4	30,1	100
Totale	190	590	206	492	506	1 984
% rispetto al totale delle valutazioni bibliometriche	9,6	29,7	10,4	24,8	25,5	100
Scienze agrarie e veterinarie						
Valutazione bibliometrica (F)			Valutazione peer (P)			
	E	B	A	L	IP	Totale
E	54	126	27	13	75	295
% rispetto al totale delle valutazioni bibliometriche di classe E	18,3	42,7	9,2	4,4	25,4	100
B	3	25	8	6	12	54
% rispetto al totale delle valutazioni bibliometriche di classe B	5,6	46,3	14,8	11,1	22,2	100
A	0	3	0	2	1	6
% rispetto al totale delle valutazioni bibliometriche di classe A	0,0	50,0	0,0	33,3	16,7	100
L	0	9	8	40	13	70
% rispetto al totale delle valutazioni bibliometriche di classe L	0,0	12,9	11,4	57,1	18,6	100
IR	7	38	10	25	27	107
% rispetto al totale delle valutazioni bibliometriche IR	6,5	35,5	9,3	23,4	25,2	100

Totale	64	201	53	86	128	532
% rispetto al totale delle valutazioni bibliometriche	12,0	37,8	10,0	16,2	24,1	100
Ingegneria civile e architettura						
Valutazione bibliometrica (F)			Valutazione peer (P)			
	E	B	A	L	IP	Totale
E	43	76	4	3	26	152
% rispetto al totale delle valutazioni bibliometriche di classe E	28,3	50,0	2,6	2,0	17,1	100
B	1	3	1	3	4	12
% rispetto al totale delle valutazioni bibliometriche di classe B	8,3	25,0	8,3	25,0	33,3	100
A	1	2	0	0	2	5
% rispetto al totale delle valutazioni bibliometriche di classe A	20,0	40,0	0,0	0,0	40,0	100
L	2	7	4	7	9	29
% rispetto al totale delle valutazioni bibliometriche di classe L	6,9	24,1	13,8	24,1	31,0	100
IR	3	14	0	3	7	27
% rispetto al totale delle valutazioni bibliometriche IR	11,1	51,9	0,0	11,1	25,9	100
Totale	50	102	9	16	48	225
% rispetto al totale delle valutazioni bibliometriche	22,2	45,3	4,0	7,1	21,3	100
Ingegneria industriale e dell'informazione						
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 di classe E	28,7	38,9	7,2	4,1	21,2	100
B	6	60	19	11	29	125
% rispetto al totale delle valutazioni bibliometriche di classe B	4,8	48,0	15,2	8,8	23,2	100
A	0	14	12	9	17	52
% rispetto al totale delle valutazioni bibliometriche di classe A	0,0	26,9	23,1	17,3	32,7	100
L	1	4	4	13	8	30
% rispetto al totale delle valutazioni bibliometriche di classe L	3,3	13,3	13,3	43,3	26,7	100
IR	28	84	21	29	49	211
% rispetto al totale delle valutazioni bibliometriche IR	13,3	39,8	10,0	13,7	23,2	100
Totale	239	439	107	91	254	1 130
% rispetto al totale delle valutazioni bibliometriche	21,2	38,8	9,5	8,1	22,5	100
Scienze economiche e statistiche						
Valutazione bibliometrica (F)			Valutazione peer (P)			
	E	B	A	L	IP	Totale
E	98	72	19	9		198

% rispetto al totale delle valutazioni bibliometriche di classe E	49,5	36,4	9,6	4,5		100
B	11	56	26	9		102
% rispetto al totale delle valutazioni bibliometriche di classe B	10,8	54,9	25,5	8,8		100
A	4	25	39	35		103
% rispetto al totale delle valutazioni bibliometriche di classe A	3,9	24,3	37,9	34,0		100
L	3	21	45	118		187
% rispetto al totale delle valutazioni bibliometriche di classe L	1,6	11,2	24,1	63,1		100
IR						
% rispetto al totale delle valutazioni bibliometriche IR						
Totale	116	174	129	171		590
% rispetto al totale delle valutazioni bibliometriche	19,7	29,5	21,9	29,0		100

Tabella B: Confronto tra le valutazioni P1 e P2 per area

Scienze matematiche e informatiche					
P1	P2				
	E	B	A	L	Totale
E	141	63	24	21	249
% rispetto al totale delle valutazioni di classe E del primo revisore	56,6	25,3	9,6	8,4	100,0
B	52	85	28	26	191
% rispetto al totale delle valutazioni di classe B del primo revisore	27,2	44,5	14,7	13,6	100,0
A	19	26	9	24	78
% rispetto al totale delle valutazioni di classe A del primo revisore	24,4	33,3	11,5	30,8	100,0
L	16	22	21	54	113
% rispetto al totale delle valutazioni di classe L del primo revisore	14,2	19,5	18,6	47,8	100,0
Totale	228	196	82	125	631
% rispetto al totale delle valutazioni del primo revisore	36,1	31,1	13,0	19,8	100,0
Scienze fisiche					
P1	P2				
	E	B	A	L	Totale
E	221	169	37	26	453
% rispetto al totale delle valutazioni di classe E del primo revisore	48,8	37,3	8,2	5,7	100,0
B	170	276	82	62	590
% rispetto al totale delle valutazioni di classe B del primo revisore	28,8	46,8	13,9	10,5	100,0
A	36	80	31	44	191
% rispetto al totale delle valutazioni di classe A del primo revisore	18,8	41,9	16,2	23,0	100,0
L	25	64	31	58	178
% rispetto al totale delle valutazioni di classe L del primo revisore	14,0	36,0	17,4	32,6	100,0
Totale	452	589	181	190	1 412
% rispetto al totale delle valutazioni del primo revisore	32,0	41,7	12,8	13,5	100,0
Scienze chimiche					
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
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
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
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
Totale	313	393	122	99	927
% rispetto al totale delle valutazioni del primo revisore	33,8	42,4	13,2	10,7	100
Scienze della terra					
P1		P2			
	E	B	A	L	Totale
E	31	33	11	14	89
% rispetto al totale delle valutazioni di classe E del primo revisore	34,8	37,1	12,4	15,7	100,0
B	23	76	26	27	152
% rispetto al totale delle valutazioni di classe B del primo revisore	15,1	50,0	17,1	17,8	100,0
A	12	40	14	26	92
% rispetto al totale delle valutazioni di classe A del primo revisore	13,0	43,5	15,2	28,3	100,0
L	14	32	24	55	125
% rispetto al totale delle valutazioni di classe L del primo revisore	11,2	25,6	19,2	44,0	100,0
Totale	80	181	75	122	458
% rispetto al totale delle valutazioni del primo revisore	17,5	39,5	16,4	26,6	100,0
Scienze biologiche					
P1		P2			
	E	B	A	L	Totale
E	131	113	30	32	306
% rispetto al totale delle valutazioni di classe E del primo revisore	42,8	36,9	9,8	10,5	100,0
B	101	202	59	108	470
% rispetto al totale delle valutazioni di classe B del primo revisore	21,5	43,0	12,6	23,0	100,0
A	30	49	35	68	182
% rispetto al totale delle valutazioni di classe A del primo revisore	16,5	26,9	19,2	37,4	100,0
L	41	106	50	155	352
% rispetto al totale delle valutazioni di classe L del primo revisore	11,6	30,1	14,2	44,0	100,0
Totale	303	470	174	363	1 310
% rispetto al totale delle valutazioni del primo revisore	23,1	35,9	13,3	27,7	100,0
Scienze mediche					
P1		P2			
	E	B	A	L	Totale
E	128	116	40	44	328
% rispetto al totale delle valutazioni di classe E del primo revisore	39,0	35,4	12,2	13,4	100,0

B	116	291	116	162	685
% rispetto al totale delle valutazioni di classe B del primo revisore	16,9	42,5	16,9	23,6	100,0
A	40	130	62	121	353
% rispetto al totale delle valutazioni di classe A del primo revisore	11,3	36,8	17,6	34,3	100,0
L	48	172	135	263	618
% rispetto al totale delle valutazioni di classe L del primo revisore	7,8	27,8	21,8	42,6	100,0
Totale	332	709	353	590	1 984
% rispetto al totale delle valutazioni del primo revisore	16,7	35,7	17,8	29,7	100,0
Scienze agrarie e veterinarie					
P1		P2			
	E	B	A	L	Totale
E	40	64	16	25	145
% rispetto al totale delle valutazioni di classe E del primo revisore	27,6	44,1	11,0	17,2	100,0
B	49	74	38	32	193
% rispetto al totale delle valutazioni di classe B del primo revisore	25,4	38,3	19,7	16,6	100,0
A	16	31	17	35	99
% rispetto al totale delle valutazioni di classe A del primo revisore	16,2	31,3	17,2	35,4	100,0
L	11	28	17	39	95
% rispetto al totale delle valutazioni di classe L del primo revisore	11,6	29,5	17,9	41,1	100,0
Totale	116	197	88	131	532
% rispetto al totale delle valutazioni del primo revisore	21,8	37,0	16,5	24,6	100,0
Ingegneria civile e architettura					
P1		P2			
	E	B	A	L	Totale
E	37	23	8	3	71
% rispetto al totale delle valutazioni di classe E del primo revisore	52,1	32,4	11,3	4,2	100,0
B	32	49	13	13	107
% rispetto al totale delle valutazioni di classe B del primo revisore	29,9	45,8	12,1	12,1	100,0
A	8	4	2	6	20
% rispetto al totale delle valutazioni di classe A del primo revisore	40,0	20,0	10,0	30,0	100,0
L	7	9	2	9	27
% rispetto al totale delle valutazioni di classe L del primo revisore	25,9	33,3	7,4	33,3	100,0
Totale	84	85	25	31	225
% rispetto al totale delle valutazioni del primo revisore	37,3	37,8	11,1	13,8	100,0
Ingegneria industriale e dell'informazione					
P1		P2			

	E	B	A	L	Totale
E	195	128	35	28	386
% rispetto al totale delle valutazioni di classe E del primo revisore	50,5	33,2	9,1	7,3	100,0
B	129	190	59	75	453
% rispetto al totale delle valutazioni di classe B del primo revisore	28,5	41,9	13,0	16,6	100,0
A	30	52	30	26	138
% rispetto al totale delle valutazioni di classe A del primo revisore	21,7	37,7	21,7	18,8	100,0
L	31	55	20	47	153
% rispetto al totale delle valutazioni di classe L del primo revisore	20,3	35,9	13,1	30,7	100,0
Totale	385	425	144	176	1 130
% rispetto al totale delle valutazioni del primo revisore	34,1	37,6	12,7	15,6	100,0
Scienze economiche e statistiche					
P1			P2		
	E	B	A	L	Totale
E	53	43	7	11	114
% rispetto al totale delle valutazioni di classe E del primo revisore	46	38	6	10	100
B	36	73	29	29	167
% rispetto al totale delle valutazioni di classe B del primo revisore	22	44	17	17	100
A	8	34	21	29	92
% rispetto al totale delle valutazioni di classe A del primo revisore	9	37	23	32	100
L	4	46	50	117	217
% rispetto al totale delle valutazioni di classe L del primo revisore	2	21	23	54	100
Totale	101	196	107	186	590
% rispetto al totale delle valutazioni del primo revisore	17	33	18	32	100