

National Agency for the Evaluation of
Universities and Research Institutes



Agenzia Nazionale di Valutazione del
sistema Universitario e della Ricerca

Evaluation of Research Quality



Valutazione Qualità della Ricerca

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

Rapporto finale di area

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



1	INTRODUZIONE	3
1.1	LA COMPOSIZIONE DEL GRUPPO DI ESPERTI DELLA VALUTAZIONE (GEV)	6
1.2	I TEMPI.....	7
1.3	DESCRIZIONE DELL'AREA	9
2	LA VALUTAZIONE DEI LAVORI	12
2.1	LA CLASSIFICAZIONE DELLE RIVISTE	12
2.2	LA VALUTAZIONE BIBLIOMETRICA	13
2.3	I REVISORI ESTERNI.....	16
2.4	LA VALUTAZIONE <i>PEER</i>	18
2.5	I LAVORI NON VALUTABILI.....	20
3	I LAVORI VALUTATI	20
4	SINTESI DEI RISULTATI DI AREA.....	22
4.1	LA GRADUATORIA DELLE STRUTTURE	25
4.2	LA VALUTAZIONE DEI DIPARTIMENTI	28
5	ANALISI DEI RISULTATI.....	33
6	CONCLUSIONI	36
	APPENDICE A. IL CONFRONTO TRA VALUTAZIONE <i>PEER</i> E VALUTAZIONE BIBLIOMETRICA	38
	APPENDICE B. CRITERI PER LA VALUTAZIONE DEI PRODOTTI DI RICERCA	61



1 Introduzione

L'Area 09 (Ingegneria Industriale e dell'Informazione) dal punto di vista dei temi di ricerca, delle metodologie di studio e della collocazione editoriale dei lavori scientifici, si colloca tra le aree scientifiche (supponendo che sia accettabile una dicotomia tra aree scientifiche e aree umanistiche), con una consolidata tradizione di adozione di parametri bibliometrici, in particolare l'*Impact Factor* (IF). Ciononostante, i molti Settori Scientifico-Disciplinari (SSD) compresi all'interno dell'Area 09 (ben 42) presentano una notevole eterogeneità delle sedi privilegiate di pubblicazione, con una nutrita presenza sia di articoli sulle riviste internazionali censite dalle maggiori banche dati, sia di contributi a congressi internazionali e nazionali, ma anche di articoli su riviste non presenti in banche dati, di monografie e capitoli di libri. Un valore significativo hanno inoltre nel settore dell'ingegneria i brevetti italiani o internazionali. L'importanza relativa delle pubblicazioni su rivista o a congresso è variabile in funzione del SSD, ma in tutti i casi la collocazione internazionale delle pubblicazioni costituisce un elemento di pregio rilevante.

Il compito affidato al Gruppo di Esperti della Valutazione (GEV) è consistito nella valutazione delle pubblicazioni dei ricercatori dell'area utilizzando le griglie di classificazione definite dal bando ministeriale. Le griglie prevedono che ciascun lavoro inviato alla VQR sia collocato in una classe di merito (Eccellente – E, Buono – B, Accettabile – A, Limitato – L). Per individuare le classi di merito, il bando prevede che il GEV possa utilizzare sia l'analisi bibliometrica sia la valutazione di pari, o *peer review*.

La riflessione preliminare del GEV ha affrontato i vantaggi e gli svantaggi che ciascuno dei due metodi di valutazione presenta.

La valutazione bibliometrica è stata considerata possibile e affidabile prevalentemente per articoli pubblicati su riviste indicizzate nelle banche dati di riferimento della VQR, categoria nella quale ricade circa il 77% dei lavori complessivamente conferiti al GEV. La valutazione bibliometrica non è esente da limiti, poiché assegna una fascia di merito elevata a un articolo per



il solo fatto di essere pubblicato su una rivista con parametri bibliometrici elevati, e/o avere ricevuto numerose citazioni, indipendentemente dal contenuto scientifico dell'articolo. assumendo che ciò costituisca una metrica indiretta della qualità del contenuto. Questa ipotesi può essere considerata vera in molti casi, ma sicuramente esistono eccezioni. A questo proposito occorre, infatti, notare che qualunque indicatore bibliometrico di una rivista è una grandezza media che si ottiene elaborando (in modo più o meno complesso) una distribuzione di citazioni che è fortemente asimmetrica con pochi articoli con moltissime citazioni e con moltissimi articoli che sono citati poco o per nulla. In altri termini, solo un numero molto limitato di articoli in una rivista è responsabile del valore dell'indicatore bibliometrico. Di conseguenza, nonostante esso fornisca senza dubbio indicazioni importanti sull'impatto di una rivista scientifica, non è corretto usarlo come unico indicatore di qualità di ogni singolo articolo pubblicato nella rivista: è fondamentale tener conto anche del numero di citazioni e del posizionamento dell'articolo nella distribuzione citazionale delle pubblicazioni che appartengono alla stessa omogenea area scientifica. Bisogna infine osservare che l'impossibilità di eliminare le autocitazioni e di filtrare alcuni noti fenomeni di manipolazione dei parametri bibliometrici di alcune riviste inducono a considerare i risultati attendibili in media, ma non rispetto ad un singolo articolo o autore. Considerando che la VQR ha come obiettivo la valutazione di strutture che espongono numeri elevati di prodotti, il GEV ha scelto di procedere con la valutazione bibliometrica per gli articoli che possedevano parametri bibliometrici adeguati. Con l'obiettivo di non incoraggiare fenomeni di manipolazione degli indici, anche e soprattutto con riferimento a valutazioni future, che si auspica possano adottare metriche migliorate rispetto a quelle attualmente disponibili, il GEV ha inoltre deciso di non adottare un solo indice bibliometrico per la graduatoria delle riviste (il candidato naturale sarebbe stato l'IF, che è però anche il più soggetto a manipolazioni), ma di costruire un mix degli indicatori più comunemente disponibili. Nel caso particolare del SSD ING-INF/05 (Sistemi di Elaborazione delle Informazioni), la graduatoria è stata definita congiuntamente dai membri dei GEV delle aree 01 e 09, utilizzando criteri differenti da quelli degli altri SSD dell'Area 09, ma comuni con il SSD INF/01 (Informatica), con il ragionevole



obiettivo di adottare lo stesso metodo di valutazione per tutte le pubblicazioni scientifiche del settore dell'informatica.

La valutazione di pari (*peer review*) in linea di principio consente di valutare nel migliore dei modi il contenuto di un lavoro scientifico mediante il giudizio di uno o più esperti del settore. D'altro canto, la dimensione stessa della VQR (più di 16.000 lavori inviati al GEV) ne riduce l'affidabilità. Esiste, infatti, un chiaro conflitto tra qualità e quantità della *peer review*; tale conflitto è causato dalla necessità di individuare un gran numero revisori qualificati e dall'attenzione che ciascun revisore può dedicare alla valutazione di un lavoro, specie nel caso in cui venga sovraccaricato di valutazioni. Inoltre, la valutazione di pari, pur adottando le cautele del caso, non esclude del tutto la possibilità di conflitti d'interesse, né assicura una valutazione uniforme tra i diversi lavori, settori disciplinari, temi di ricerca, ecc. Infine, non è sempre facile ricondurre a una sintesi i giudizi difformi di due revisori, soprattutto quando i pareri sono molto discordi.

Come descritto di seguito, la soluzione adottata dal GEV09 è stata di utilizzare la *peer review* per tutti i lavori scientifici non pubblicati su riviste indicizzate nelle banche dati di riferimento per la VQR (pubblicazioni in atti di congressi, monografie, capitoli di libro, brevetti); invece, gli articoli su rivista che dispongono di parametri bibliometrici adeguati sono stati valutati utilizzando la bibliometria. Un campione statistico significativo di questo secondo gruppo di prodotti è stato inoltre valutato con entrambi i metodi di valutazione, con l'obiettivo di verificare la congruenza tra essi, al fine di poter integrare i risultati dei due metodi nella valutazione complessiva. Inoltre, tutti quei prodotti che dalla valutazione bibliometrica hanno ricevuto una classificazione IR (che stava a indicare un'incertezza nell'indicazione fornita dalla combinazione dei parametri bibliometrici della rivista e dell'analisi citazionale e una necessità di procedere a una *peer review*) hanno ricevuto un'ulteriore valutazione mediante *peer review*. Complessivamente, circa il 45% dei prodotti conferiti al GEV ha avuto una valutazione con *peer review* e circa il 22% ha avuto entrambe le valutazioni.



Il confronto statistico tra i due metodi di valutazione (*peer review* da un lato, basata sul giudizio di due revisori cui si aggiunge, nel caso di giudizi discordanti dei revisori, quello di un gruppo di consenso formato da due o più membri del GEV; dall'altro lato valutazione bibliometrica usando gli indici definiti dal GEV) per il campione considerato rappresenta un elemento di grande importanza, che sarà trattato nell'Appendice A di questo documento.

Vale la pena sottolineare che ogni valutazione ha margini di errore. La valutazione tramite *peer review* perché intrinsecamente soggettiva, ma anche, ad esempio, perché un giudizio potrebbe essere dato frettolosamente o senza adeguata competenza da parte del revisore; l'analisi bibliometrica perché le banche dati contengono inevitabilmente errori e perché le citazioni sono indicatori imperfetti della qualità di un lavoro scientifico.

Il GEV conferma la fiducia nei metodi adottati per gli scopi della VQR, cioè la valutazione di raggruppamenti consistenti di docenti che operano all'interno di strutture di ricerca con una dimensione significativa; ritiene però anche doveroso evidenziare gli oggettivi limiti di validità dei metodi adottati, che non si possono applicare a singoli individui o singole pubblicazioni.

1.1 La composizione del Gruppo di Esperti della Valutazione (GEV)

Il GEV ha valutato i lavori di tutti i 42 settori scientifico-disciplinari (SSD) dell'Area 09 (cfr. Tabella 1.1, Settori scientifico-disciplinari (SSD) dell'Area 09 – Ingegneria Industriale e dell'Informazione). Il GEV è composto di 40 membri (Tabella. 1.2, Composizione del Gruppo di Esperti della Valutazione – Area 09), organizzati in tre gruppi più ristretti (i subGEV, elencati in Tabella 1.3, Organizzazione degli esperti in subGEV, corrispondenti SSD e distribuzione dei prodotti della ricerca).

Complessivamente, 5 membri del GEV lavorano in università straniere (Erlangen-Nuremberg, ETH Zurigo, Louvain, UCLA); i rimanenti 34 provengono da 22 atenei italiani, tra cui 6 del Mezzogiorno (Calabria, Napoli "Federico II", Seconda Università di Napoli, Palermo, Salento, Sannio di Benevento); 8 dell'Italia centrale (Cassino, Firenze, Politecnica delle Marche,



L'Aquila, Pisa, S. Anna di Pisa, Roma "La Sapienza" e "Tor Vergata"); 8 dell'Italia settentrionale (Bologna, Ferrara, Genova, Padova, Parma, Trento, e i Politecnici di Milano e Torino). Si tratta di un gruppo molto diversificato sul piano delle competenze e delle sedi di affiliazione, che ha affrontato la VQR con grande impegno, estremo rigore, totale trasparenza e fortissimo spirito di collaborazione. Per organizzare i propri lavori, il GEV ha creato gruppi di lavoro più ristretti (i subGEV). La numerosità dei componenti dei subGEV riflette approssimativamente il numero di lavori scientifici attesi per ciascun gruppo di lavoro.

Il GEV ha tenuto complessivamente sei riunioni plenarie (cfr. Tabella 1.4, Elenco delle riunioni del GEV), ma la gran parte del lavoro è stata svolta per via telematica, tramite gli strumenti messi a disposizione dal CINECA e la posta elettronica (sono state scambiate diverse migliaia di messaggi tra i componenti del GEV). Tutte le decisioni rilevanti sono state adottate tramite approvazione dei membri GEV. I membri del GEV sono stati costantemente informati, con cadenza settimanale, degli sviluppi della valutazione e dei problemi incontrati e sono stati consultati per ogni decisione di rilievo. Tutte le decisioni più importanti sono state prese all'unanimità e in piena armonia all'interno del gruppo. Il Gruppo è stato coadiuvato da un Assistente, il Dottore di Ricerca Filippo Costa, che ha lavorato con notevolissima competenza, enorme dedizione, grande passione e infinita pazienza; a lui va il sentitissimo ringraziamento di tutti i membri del gruppo.

1.2 I tempi

Il GEV ha lavorato complessivamente 18 mesi, da dicembre 2011 a giugno 2013. Dopo la riunione d'insediamento del 12 dicembre 2011, nei mesi di gennaio e febbraio 2012 è stato discusso e approvato il documento "Criteri per la valutazione dei prodotti di ricerca, Gruppo di Esperti della Valutazione dell'Area 09", pubblicato sul sito ANVUR il 29 febbraio 2012 e riportato nell'Appendice B di questo documento.



Tra marzo e aprile 2012 il GEV ha predisposto l'elenco delle classificazioni delle riviste, utilizzando la combinazione di indici bibliometrici definita nel documento dei criteri e la procedura concordata per l'area dell'informatica. Negli stessi mesi è stato predisposto un elenco di revisori esterni, con l'obiettivo di ottenere la quasi totalità di revisioni da parte di esperti che operano fuori dai confini nazionali, al fine di minimizzare la possibilità di conflitti di interesse. Sempre nello stesso periodo sono state messe a punto le domande da sottoporre ai revisori, in collaborazione con gli altri GEV.

Tra maggio e settembre 2012 i lavori da sottoporre alla valutazione *peer* sono stati assegnati a uno o due membri del GEV in base all'argomento principale della pubblicazione e al SSD di competenza dei membri del GEV. Come previsto dai criteri di valutazione del GEV, i prodotti sono stati esaminati dai membri del GEV per individuare i lavori non valutabili e quelli non ammissibili.

La valutazione tramite *peer review*, dall'invio dei lavori ai revisori alla formulazione del giudizio di consenso da parte dei due membri GEV responsabili del lavoro, si è svolta da ottobre 2012 a maggio 2013. I tempi sono stati più o meno equivalenti per tutti i SSD, nonostante le notevoli differenze nel numero di lavori da valutare.

A febbraio 2013 sono stati classificati in via preliminare i lavori "non valutabili". A marzo 2013 l'ANVUR, nello spirito di trasparenza e collaborazione con gli atenei e gli enti di ricerca, ha inviato a ciascuna struttura l'elenco dei lavori giudicati preliminarmente "non valutabili" insieme con la motivazione e la richiesta di trasmettere eventuali controdeduzioni. I dati delle valutazioni sono stati elaborati nei mesi di aprile e maggio 2013; è stato quindi redatto il Rapporto finale di Area, discusso nella riunione plenaria del 28 maggio e successivamente approvato per via telematica.



1.3 Descrizione dell'area

L'Area 09 (Ingegneria Industriale e dell'Informazione) è composta di circa 5.500 ricercatori che afferiscono complessivamente a 91 strutture, tra cui 73 università, 10 enti di ricerca e 8 consorzi (si vedano le Tabelle 1.5a, 1.5b, 1.5c). L'area comprende otto macro-settori: 09/A –Ingegneria Meccanica, Aerospaziale e Navale (circa il 14% dei ricercatori che operano in ambito universitario), 09/B – Ingegneria Manifatturiera, Impiantistica e Gestionale (circa il 9%), 09/C – Ingegneria Energetica, Termomeccanica e Nucleare (12%), 09/D – Ingegneria Chimica e dei Materiali (9%), 09/E –Ingegneria Elettrica, Elettronica e Misure (circa il 16% dei ricercatori), 09/F – Ingegneria delle Telecomunicazioni e Campi Elettromagnetici (circa il 9%), 09/G – Ingegneria dei Sistemi e Bioingegneria (8%), 09/H – Ingegneria Informatica (12%), distribuiti in 42 settori scientifico-disciplinari, che includono anche ING-IND/28, 29 e 30 (0.9%) che appartengono all'Area 08, ma che per la VQR sono stati inclusi nell'Area 09.

In termini di dimensioni, come evidenziato dalle Tabelle 1.5 (a, b e c), in 51 strutture la presenza dell'Area 09 è numericamente limitata (meno di 100 lavori conferiti – strutture di piccola dimensione). Delle rimanenti 40 strutture, 27 hanno conferito tra 100 e 350 lavori (strutture di media dimensione), 13 strutture oltre 350 lavori (strutture di grande dimensione). I Politecnici di Milano e Torino e il CNR hanno conferito oltre mille lavori ciascuno. Le sei strutture più grandi includono, oltre ai Politecnici di Milano e Torino e al CNR, le Università di Bologna, Roma “La Sapienza” e Napoli “Federico II”.

L'Area 09 è caratterizzata da una notevole eterogeneità e, allo stesso tempo, da una sovrapposizione di temi di ricerca tra alcuni SSD, con una presenza non trascurabile di lavori multidisciplinari. L'eterogeneità riguarda non solo i temi di ricerca, ma anche le tipologie e le collocazioni delle pubblicazioni di maggior pregio; ad esempio, studiosi di ingegneria nucleare pubblicano molto spesso i propri lavori migliori su riviste internazionali in lingua inglese, mentre in altri settori, prevalentemente in alcuni dell'area dell'informazione, soprattutto dell'ingegneria informatica, sono considerate di grande rilievo le pubblicazioni nei congressi internazionali più selettivi del settore, che in alcuni casi hanno percentuali di accettazione inferiori al 10%.



In secondo luogo, in molti casi vi sono sovrapposizioni di interessi scientifici tra docenti che appartengono a settori disciplinari diversi. Per la natura oggi pervasiva delle tecnologie dell'informatica e più in generale delle ICT è naturale che molti ricercatori di quell'area abbiano interessi che spaziano su moltissimi settori scientifici, sia all'interno dell'Area 09, sia in numerose altre aree, tecnologiche come pure umanistiche.

Infine, molti lavori dell'Area 09 hanno una forte connotazione multidisciplinare e sono svolti in collaborazione non solo tra ricercatori che appartengono a diversi settori della stessa area, ma anche tra docenti di aree diverse. Il numero di prodotti in comune con altre aree è riportato nella Tabella 1.6. È frequente, infatti, la collaborazione con ricercatori delle aree della matematica, della fisica e della chimica, come pure con ricercatori dell'area economica e della statistica. Nella tabella sono riportati, per ogni area, il numero di prodotti in comune sottoposti alla VQR.

Per quanto riguarda la tipologia delle pubblicazioni, è interessante notare che vi sono differenze non solo tra SSD diversi, ma anche all'interno di ciascun SSD. In ciascuno dei SSD sono presenti pubblicazioni su riviste, monografie, capitoli di libro e atti di congresso, anche se le proporzioni cambiano significativamente secondo i settori.

Per quanto riguarda la disponibilità di dati bibliometrici, l'Area 09 si trova in una posizione analoga a quella delle altre aree scientifiche, e da tempo dispone di banche dati che censiscono gli articoli sulle principali riviste internazionali.

Le Tabelle 1.7 (Distribuzione dei prodotti di ricerca conferiti per tipologia e SSD), 1.8 (Distribuzione dei prodotti di ricerca per lingua di pubblicazione e SSD), 1.9 (Distribuzione dei lavori conferiti per tipologia, lingua di pubblicazione e subGEV) e 1.10 (Distribuzione dei lavori conferiti per anno di pubblicazione, tipologie e SSD) riportano la distribuzione dei lavori sottomessi alla VQR per tipologia di pubblicazione, SSD, lingua di pubblicazione e anno di pubblicazione. Nel complesso dell'area, quasi l'82% dei lavori è rappresentato da articoli su rivista, seguito da articoli a congressi (14%), da contributi in volume (2,5%) e monografie



(0,75%). La percentuale di articoli su riviste raggiunge un minimo nel settore ING-IND/02 (37,5%) e un massimo nei settori ING-IND/18 e ING-IND/23 (100%). Al contrario, la percentuale di articoli a congressi è massima nel settore ING-IND/01 (56%). La distribuzione dei prodotti nelle diverse categorie, differenziati secondo la lingua, per i 3 subGEV è riportata nella Tabella 1.8. Di nuovo si può osservare la percentuale molto elevata di prodotti in lingua inglese, specie con riferimento a articoli su rivista e atti di convegni; per i primi la percentuale è sempre vicina al 90%, mentre per i secondi la percentuale è prossima all'80% per i subGEV 09a e 09b, mentre sale al 94% per il subGEV 09c.

La Tabella 1.9 riporta, per ogni SSD e per ogni anno, le percentuali di prodotti conferiti nelle diverse categorie. Si può osservare come la distribuzione del numero totale di prodotti pubblicati nei sette anni della VQR presenti un andamento quasi costante, cui si sovrappone una lieve crescita con la riduzione dell'età del prodotto (1874 prodotti del 2004 – 11,5%, 1918 prodotti del 2005 – 11,7%, 2121 del 2006 – 13%, 2309 del 2007 – 14%, 2627 del 2008 – 16%, 2726 del 2009 – 16,7%, 2772 del 2010 – 17%). Questo testimonia un buon bilanciamento delle diverse metriche adottate per prodotti con età diverse.

Non per tutti i lavori la scheda riporta la lingua di pubblicazione. È interessante notare che, come prevedibile, la lingua prevalente dei lavori inviati al GEV è l'inglese (88,6%), seguita dall'italiano (2,7%). Il numero di lavori pubblicati in altre lingue è trascurabile; per l'8,4% dei lavori non è stata fornita l'informazione sulla lingua di pubblicazione.

All'interno dei SSD, la percentuale di lavori pubblicati in lingua inglese è superiore al 90% nei settori ING-IND/01, ING-IND/05, ING-IND/18, ING-IND/21, ING-IND/29, ING-IND/32, ING-IND/33, ING-IND/34, ING-INF/01, ING-INF/02, ING-INF/03, ING-INF/05, ING-INF/06, ING-INF/07; raggiunge valori poco inferiori al 75% solo nel settore ING-IND/28. La percentuale di lavori pubblicati in italiano supera il 10% solo nei settori ING-IND/11, ING-IND/20, ING-IND/28. In generale, si nota una notevolissima propensione a pubblicare articoli su rivista in lingua inglese in tutti i settori dell'ingegneria.



2 La Valutazione dei lavori

Per la valutazione dei prodotti di ricerca il GEV ha adottato sia indicatori bibliometrici sia valutazione tramite *peer review*. La procedura ha seguito strettamente quanto previsto dai “Criteri di valutazione”, pubblicati il 29 febbraio 2012 (cfr. Appendice B).

Tutti i prodotti di ricerca diversi da articoli su riviste classificate (atti di congresso, monografie, capitoli di libro, ecc.) e tutti gli articoli su riviste non classificate sono stati valutati con *peer review*. Gli articoli pubblicati su riviste classificate sono stati valutati in base agli indicatori bibliometrici scelti e al numero di citazioni ricevute nel 2004-2010 nelle banche dati ISI WoS e/o Scopus. Un campione casuale significativo di articoli pubblicati su riviste classificate è stato valutato anche con *peer review*.

2.1 La classificazione delle riviste

La procedura di classificazione delle riviste ha riguardato tutte le riviste su cui sono stati pubblicati articoli dai ricercatori dell’Area 09 e ha seguito la metodologia descritta in dettaglio nell’Appendice B e qui richiamata in sintesi.

L’elenco delle riviste considerate dal GEV 09 include le riviste contenute in *Web of Science* (WoS) e nel *Journal of Citation Report* (JCR), gestiti da Thomson, e in Scopus, gestito da Elsevier, su cui sono stati pubblicati articoli dai ricercatori dell’area nel periodo d’interesse della VQR: 2004-2010.

Le riviste sono state classificate secondo le *subject category* (SC) JCR-WoS e le ASJC Scopus utilizzando una combinazione basata su *Principal Component Analysis* (PCA) di 4 indicatori bibliometrici di uso consolidato: *Impact Factor* (IF), *Impact Factor* a cinque anni (5YIF), *EigenFactor* (EF) e *ArticleInfluence* (AI).

Il GEV ha classificato le riviste per ciascuna area di classificazione in quattro classi di merito (1, 2, 3, 4) sulla base dei quartili della distribuzione del risultato della PCA.



Nel caso particolare dell'area informatica, come già detto, si è sviluppata una classifica comune con il GEV 01, utilizzando il 5YIF fornito dalla banca dati Thomson, l'indice bibliometrico 2010 SJR fornito dalla banca dati Scopus, la classificazione effettuata dall'*Australian Research Council* nel 2010 all'interno del processo di valutazione *Excellence in Research for Australia (ERA)* e che prevedeva le classi di merito A*/A/B/C, e la classificazione effettuata dal GRIN (GRuppo di INformatica) nel 2004 che prevedeva le classi di merito A/B/C/D, completata con ACM e *IEEE Transactions/Journals* del settore stabilite dopo il 2004 e presenti in Thomson e/o Scopus.

Gli elenchi delle classifiche delle riviste sono stati pubblicati nei mesi di marzo e aprile 2012.

2.2 La valutazione bibliometrica

Come previsto dai criteri di valutazione, la classe di merito finale di ciascun articolo dipende dalla classe di merito iniziale della rivista (calcolata sulla base di indicatori bibliometrici) e dal numero di citazioni ricevute da ciascun articolo in WoS o Scopus. Vi è una forte correlazione tra la graduatoria basata sugli indicatori bibliometrici delle riviste, e il numero di citazioni. Facendo riferimento al database WoS, circa il 77% degli articoli con almeno 5 citazioni complessive nel 2004-2010, riceve valutazione finale eccellente (E). Dei circa 1.300 articoli con almeno 5 citazioni medie annue, circa il 75% è pubblicato in riviste classificate in classe 1 sulla base degli indicatori bibliometrici delle riviste e circa il 16% in classe 2.

Facendo unicamente riferimento alle citazioni ricevute dagli articoli, è possibile individuare la percentuale degli articoli presentati alla VQR e oggetto di valutazione bibliometrica che si collocano nel top 1% o nel top 5% della distribuzione mondiale calcolata per l'anno di pubblicazione nella *Subject Category* (SC per WoS o ASJC per Scopus) scelta per ogni articolo. La distribuzione è stata calcolata sia per WoS sia per Scopus ed è stato considerato il miglior risultato. Il calcolo considera tutti gli articoli pubblicati in un anno in una determinata *subject category*, identifica gli insiemi che comprendono l'1% e il 5% dei più citati e calcola il numero



di articoli sottomessi alla VQR presenti in tali insiemi. Tali articoli rappresentano i punti di eccellenza di ogni SSD. Il numero di lavori compresi in questi insiemi e sottomessi alla VQR e la percentuale rispetto al numero di articoli valutati bibliometricamente all'interno di ogni SSD sono riportati nella Tabella 1.11 (Tabella 1.11: Numero e percentuale di lavori che si collocano nel top 1% o nel top 5% della distribuzione mondiale delle citazioni). È interessante osservare che complessivamente l'Area 09 colloca nel top 1% o nel top 5% della distribuzione mondiale una quota rispettivamente pari a circa il 2% e 10% dei prodotti sottomessi alla VQR, con una presenza quindi nelle due fasce di eccellenza circa doppia rispetto a un campione casuale. Se poi si analizzano le percentuali relative ai singoli SSD, si può vedere che nel top 1% della produzione internazionale alcuni SSD dell'ingegneria industriale arrivano a collocare circa il 7% dei prodotti (ING/IND-20 e ING/IND-27). Tra gli SSD dell'ingegneria dell'informazione, solo 2 superano il 3% (ING-INF/03 e ING-INF/04).

Come già ricordato, per verificare la congruenza tra la valutazione bibliometrica e la valutazione tramite *peer review* è stato estratto un campione casuale di articoli pubblicati su riviste classificate e sottomessi alla valutazione bibliometrica; gli articoli estratti sono stati valutati con il metodo della *peer review*. Il confronto tra valutazione bibliometrica e valutazione tramite *peer review* è descritto in dettaglio nell'Appendice A di questo Rapporto e qui richiamato in sintesi (Appendice A: Confronto tra valutazione *peer* e valutazione bibliometrica). Il campione comprende il 10% degli articoli su rivista, per un complesso di 1.130 osservazioni. Il campione è stato estratto all'inizio di settembre 2012, prima dell'inizio della *peer review*. Gli articoli selezionati sono stati valutati in *peer review* con le stesse modalità di tutti gli altri lavori assoggettati alla valutazione *peer*. In particolare, gli articoli sono stati collocati in una classe di merito da due revisori esterni al GEV scelti da uno o due membri GEV, oppure da un revisore esterno e un membro GEV o da due membri GEV. In caso di discordanza di valutazione dei due revisori, un Gruppo di Consenso formato da due o più membri GEV ha provveduto alla valutazione di sintesi.



Il confronto tra la distribuzione bibliometrica nelle quattro classi di merito e le valutazioni *peer* di sintesi mostra che la valutazione bibliometrica tende a classificare in classe E un maggior numero di articoli rispetto alla *peer review*. In generale la *peer review* tende ad abbassare la valutazione bibliometrica nel 35% dei casi. Si verifica spesso che la *peer review* classifichi con maggiore frequenza gli articoli in classe B rispetto all'analisi bibliometrica. Viceversa, la *peer review* tende a elevare la valutazione della bibliometria solo in pochi casi (2,6%). Questo evidentemente significa che l'algoritmo bibliometrico individua con una buona affidabilità un *limite superiore* della valutazione. La frequenza di attribuzione alle classi A e L è simile con entrambi i metodi di valutazione. Complessivamente è possibile definire il grado di accordo raccogliendo i casi in cui si ha un completo accordo, un disaccordo di una, due o tre classi, un'incertezza della bibliometria (IR), della *peer* (IP), oppure di entrambi i metodi di valutazione (IR&IP). Si verifica un completo accordo tra la bibliometria e la *peer review* soltanto nel 25,6% dei casi ma, se si sommano i casi di completo accordo e di differenza di una classe si ottiene un valore pari a 54,7%. I casi dove si ottiene un disaccordo maggiore di una classe sono soltanto l'8,5%. In molti casi si verifica un'incertezza sia della bibliometria (valutazioni IR – corrispondenti al 14,3% dei casi), della *peer review* (18,1% dei casi), oppure di entrambe (4,3% dei casi).

Il confronto tra valutazione bibliometrica e *peer review* (si veda l'Appendice A: Confronto tra valutazione *peer* e valutazione bibliometrica) è stato inoltre condotto considerando due criteri: (1) il grado di concordanza tra valutazione bibliometrica e valutazione *peer*, misurato con l'indice Kappa di Cohen; (2) la verifica di differenze sistematiche tra le medie campionarie della valutazione bibliometrica e *peer review*.

Nel totale del campione dei prodotti del GEV09 conferiti per la valutazione, si riscontra una più che adeguata concordanza tra valutazioni effettuate con il metodo della *peer review* e con quello bibliometrico. Inoltre, il grado di concordanza tra valutazione finale bibliometrica e valutazione tramite *peer review* è molto simile al grado di concordanza tra le due valutazioni *peer*. Dall'altro lato, però, emerge evidenza di differenze sistematiche tra i punteggi



corrispondenti alle valutazioni *peer* e bibliometriche. In effetti, è possibile osservare che il numero di prodotti della ricerca classificati come eccellenti (E) con l’algoritmo di valutazione bibliometrica sia superiore a quello dei prodotti “eccellenti” secondo la *peer review*.

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

2.3 I revisori esterni

L’elenco dei revisori è stato costruito considerando standard di qualità scientifica, impatto sulla comunità scientifica internazionale, esperienza nella valutazione, competenza nelle rispettive aree di valutazione, e copertura delle aree di ricerca dell’Area 09. Per limitare conflitti d’interesse, il GEV ha cercato di coinvolgere nel massimo grado possibile revisori di istituzioni straniere.

Dal punto di vista operativo, il Presidente del GEV e i coordinatori dei subGEV hanno inizialmente consultato la lista dei revisori di Area 09 del CINECA e l’elenco dei revisori che si sono autonomamente proposti attraverso il sito del CINECA; hanno inoltre chiesto ai membri GEV e alle società scientifiche di proporre, indipendentemente gli uni dagli altri, nominativi per i rispettivi settori di competenza.

Grande attenzione è stata posta all’anonimato dei revisori, sia nella fase di predisposizione dell’elenco dei revisori, sia nella fase operativa della valutazione. I singoli membri GEV hanno potuto consultare, per scegliere i revisori, solo l’elenco dei revisori relativi ai propri SSD o *Subject Categories* di competenza.

La lista dei revisori è stata integrata durante il processo di valutazione sulla base delle necessità sorte a valle della trasmissione dei lavori da parte delle strutture, quando sono emerse



aree di competenza non adeguatamente coperte dai revisori inizialmente contattati, la necessità di integrare il numero di revisori per argomenti in cui sono stati presentati molti lavori o sostituire revisori non più disponibili o che si sono resi disponibili a valutare un numero limitato di lavori. Le integrazioni dei revisori sono avvenute su proposta dei membri GEV, previa approvazione, caso per caso, da parte del Presidente e dei coordinatori subGEV, con la stessa procedura utilizzata per la composizione della lista di partenza.

Complessivamente, il GEV ha utilizzato 1.236 revisori esterni. Per ciascuno dei 3 subGEV circa l'80% dei revisori è affiliato a un'istituzione straniera. La Tabella 2.1 (Tabella 2.1: Numero di revisori e di revisioni effettuate per SSD del prodotto) riporta il numero e la percentuale di revisori esterni con affiliazione italiana o estera. Le percentuali variano molto, ma sono inferiori al 50% solo per 2 SSD. La tabella riporta inoltre il numero di revisioni effettuate da revisori con affiliazione italiana e non, per ogni SSD. Anche in questo caso le percentuali sono molto variabili, ma sono inferiori al 50% solo per 5 SSD. La Tabella 2.2 (Tabella 2.2: Revisioni assegnate, effettuate e rifiutate per nazionalità del revisore) contiene il numero di revisioni assegnate, effettuate o rifiutate da revisori esterni con affiliazione italiana e non, mostrando le informazioni aggregate per l'intero GEV. Sono state complessivamente assegnate 18.185 revisioni di cui ne sono state effettuate 12.425 (con una percentuale di accettazione pari al 68,3%). Il valore medio del tempo intercorso tra l'assegnazione della revisione e la compilazione del modulo con il giudizio per le revisioni effettuate è di circa 3 settimane. La Tabella 2.3 (Tabella 2.3: Quantità di revisioni effettuate da ciascun revisore) mostra il dettaglio relativo al numero di revisioni effettuate da ciascun revisore esterno.

Il punteggio medio (normalizzato tra 0 e 1) attribuito dai revisori è di 0,597 con valori compresi tra 0,58 (subGEV 09c) e 0,625 (subGEV 09b) nei 3 subGEV. Anche la deviazione standard del punteggio attribuito dai revisori è simile nelle tre aree. Questi dati indicano un comportamento sostanzialmente uniforme dei revisori all'interno del GEV.



2.4 La valutazione *peer*

Il GEV09 ha valutato in *peer review* oltre 7.000 lavori (il 45% circa dei lavori complessivamente valutati). Essi comprendono:

- tutti i lavori scientifici diversi da articoli su riviste (monografie, capitoli di libro, atti di congresso, ecc.);
- tutti gli articoli su riviste non classificate;
- tutti gli articoli pubblicati su riviste classificate che hanno ottenuto una valutazione bibliometrica IR;
- un campione casuale degli articoli su riviste classificate (1.152 articoli su rivista).

Ciascun lavoro da valutare in *peer review* è stato assegnato a uno o due membri del GEV in base all'argomento principale della ricerca. L'assegnazione dei prodotti ai membri GEV ha evitato tutti i conflitti d'interesse con gli autori e la sede di ricerca degli autori. Sono stati inoltre esclusi casi di potenziali conflitti d'interesse per progetti di ricerca comuni e affiliazioni precedenti dei membri GEV con la sede degli autori.

I prodotti di ricerca dei membri GEV e del Presidente GEV sono stati valutati seguendo le regole previste dalla VQR (Linee Generali per i Gruppi di Esperti della VQR, Sezione 5). In particolare, i coordinatori dei subGEV hanno assegnato a revisori esterni i prodotti di ricerca dei componenti del GEV e il coordinatore della VQR ha gestito i prodotti di ricerca del Presidente del GEV.

In media, ciascun membro del GEV è stato responsabile di circa 440 lavori; il numero è maggiore nell'area dell'ingegneria chimica e dei materiali (662 lavori). L'area di Ingegneria dell'informazione ha registrato il numero maggiore di lavori valutati in *peer review*. I membri del GEV hanno effettuato una prima valutazione interna dei prodotti per individuare i "prodotti non valutabili" (lavori appartenenti a tipologie escluse dalla VQR, con allegati mancanti o non adeguati per la valutazione, pubblicati in anni precedenti o successivi al periodo di riferimento della VQR, lavori non pubblicati, lavori senza alcun contenuto di ricerca, come libri con contenuti didattici di base, ecc.).



I lavori sono stati valutati da due revisori differenti, esterni o interni al GEV. Per una parte dei prodotti (23,6%), due membri GEV che avevano in carico il lavoro hanno scelto ciascuno un revisore esterno in maniera indipendente. La scelta dei revisori è stata fatta evitando conflitti d'interesse tra revisori e autori e tra revisori e sede di ricerca degli autori, anche in questo caso seguendo le regole della VQR. L'indipendenza dei revisori è stata garantita prestando attenzione alla sede di ricerca dei revisori, alle collaborazioni di ricerca e, ove possibile, alla nazionalità. Nei casi in cui il revisore abbia evidenziato la presenza di conflitti d'interessi non noti al GEV, la non competenza sullo specifico lavoro, o la mancata disponibilità per motivi di tempo, il membro GEV che aveva proceduto all'assegnazione ha provveduto a modificarla e assegnare nuovamente il lavoro a un altro revisore.

I revisori hanno fornito una valutazione quantitativa e qualitativa dei prodotti di ricerca utilizzando una scheda di valutazione predisposta dal GEV e costituita da tre domande tendenti a valutare le caratteristiche di rilevanza, originalità e internazionalizzazione previste dal bando e da un campo libero. Il punteggio di sintesi è stato ottenuto sommando i punteggi delle tre risposte e attribuendo una classe di merito complessiva. Dato l'elevato numero di revisori stranieri, la scheda di valutazione è stata predisposta in inglese per tutti i prodotti valutati.

Una volta ricevute le schede di valutazione compilate dai valutatori, il GEV ha proceduto alle valutazioni di sintesi, costituendo Gruppi di Consenso nei casi in cui la classe di merito assegnata dai due valutatori differiva di più di una classe. Sono stati costituiti in totale 1.610 gruppi di consenso, formati, per ciascun lavoro, da due o più componenti del GEV. I Gruppi di Consenso hanno definito la valutazione di sintesi all'unanimità oppure ricorrendo (in casi limitati) a un terzo revisore e sono stati integrati, solo in caso di conflitto di valutazioni tra i componenti del Gruppo di Consenso, dal coordinatore del subGEV relativo e/o dal Presidente del GEV. Particolare attenzione da parte dei Gruppi di Consenso e dei coordinatori dei subGEV è stata posta all'esame di lavori con pareri discordi dei revisori, specie nel caso di forte disaccordo. I casi di conflitto di valutazione di sintesi sono stati complessivamente 1.435 su 7.500 lavori



valutati. Il ricorso al terzo revisore esterno è stato limitato grazie all'ampio consenso sulle valutazioni di sintesi nei Gruppi.

2.5 I lavori non valutabili

La definizione dei lavori “non valutabili” è stata affrontata con grande attenzione da parte del GEV, con l'obiettivo di limitare al minimo indispensabile l'uso di questa classificazione e di cercare di sanare tutti i casi possibili. Questo perché il bando VQR prevede linee guida di carattere generale circa la tipologia e l'ammissibilità dei prodotti ed è stato interpretato in modo diverso dai singoli atenei e ricercatori. In secondo luogo, non conoscendo ex-ante la tipologia di lavori presentati alla VQR, i criteri stessi di valutazione del GEV 09 non avevano delineato in maniera esauriente e sufficientemente chiara i casi in cui un lavoro sarebbe stato considerato “non valutabile”. Questo potrebbe avere generato incertezze negli atenei e tra i ricercatori circa le tipologie “ammissibili”.

I lavori individuati come “non valutabili” dal GEV sono stati segnalati alle strutture per consentire le opportune verifiche e controdeduzioni e, se possibile, sanare i problemi rilevati. Dopo aver esaminato le controdeduzioni, il GEV ha considerato non valutabili 77 lavori. Tali casi si riferiscono ad articoli pubblicati in date diverse dal 2004-2010, mancanza del file pdf anche in seguito alla richiesta da parte di ANVUR, tipologie non assimilabili a prodotti di ricerca (ad esempio, libri didattici di base). Oltre a questi hanno subito penalizzazioni i lavori inviati più di una volta da una stessa struttura, o i lavori inviati a strutture diverse da parte dello stesso autore. Il GEV 09, come già detto, ha cercato di ridurre al minimo i casi di non ammissibilità.

3 I lavori valutati

Il numero totale di prodotti conferiti al GEV 09 è pari a 16.347. Questo numero corrisponde a circa il 97% dei 16.858 lavori attesi (Tabella 3.1: Numero di prodotti conferiti e attesi, per SSD). Non sono quindi stati presentati alla valutazione solo 511 lavori attesi. La distribuzione del



numero di lavori attesi per i ricercatori che hanno partecipato alla VQR è riportata nella Tabella 3.2 (Tabella 3.2: Numero di soggetti valutati e relativo numero di prodotti attesi); come prevedibile, per la grande maggioranza dei ricercatori universitari (4.141 ricercatori su 4.843) erano attesi 3 prodotti, mentre per la maggioranza dei ricercatori degli enti di ricerca (490 su 611) erano attesi 6 prodotti.

Osservando la Tabella 3.1 si può vedere che la percentuale dei prodotti conferiti è di poco inferiore al 90% solo in un caso, ed è superiore al 95% per ben 33 SSD (in 8 di questi casi la percentuale è pari al 100%). Si noti che nel caso dei ricercatori non afferenti a università e quindi non inquadrati in un SSD, si è preferito non forzare una distribuzione sui SSD; per questi ricercatori la percentuale di prodotti presentati è circa il 95%.

La gran parte dei prodotti conferiti al GEV 09 è stata valutata all'interno del GEV (15.550 prodotti su 16.347, il 95%). I prodotti rimanenti sono stati trasferiti ai GEV con competenze più adatte alla valutazione; la Tabella 3.3 (Tabella 3.3: Numero e percentuale di prodotti di ricerca conferiti da soggetti afferenti all'area, per area di valutazione) riporta il numero di prodotti presentati al GEV 09 che sono stati valutati dai diversi GEV. Una percentuale rilevante di lavori è stata valutata dalle aree di Matematica, di Fisica e di Chimica (GEV1, GEV2 e GEV3).

La Tabella 3.4 (Tabella 3.4: Percentuale di prodotti di ricerca valutati attraverso la *peer review*, la bibliometria e con entrambi i metodi di valutazione) riporta per i diversi SSD il numero di prodotti valutati mediante bibliometria, mediante *peer review* o mediante entrambi i metodi. Circa il 77% dei prodotti è stato valutato mediante gli indicatori bibliometrici e il 45% circa mediante *peer review*. Il 22% circa dei prodotti ha quindi ricevuto una doppia valutazione, bibliometrica e *peer*. Dalla Tabella 3.4 si può vedere che la percentuale di valutazione bibliometrica è compresa tra il 30% e il 100% e che la percentuale di valutazione con *peer review* varia tra un minimo dell'11% e un massimo pari a 87.5%.



4 Sintesi dei risultati di area

Le tabelle commentate nel seguito presentano i risultati di sintesi della valutazione, con la distribuzione delle 4 classi finali di merito nel complesso dell'Area 09, per tipo di valutazione (*peer* o bibliometrica), per tipologia di lavoro (articolo su rivista, congresso, monografia, ecc.), per ciascuno dei 3 subGEV e per ciascuno dei 42 SSD.

Sulla base del bando VQR, i singoli lavori ricevono un punteggio pari a 1, 0,8, 0,5 e 0 a seconda che il lavoro sia valutato Eccellente, Buono, Accettabile o Limitato; ai lavori mancanti è assegnato un punteggio pari a -0,5, ai lavori non valutabili è assegnato un punteggio -1, e in casi accertati di plagio o frode -2.

Indicando rispettivamente con n_{iE} , n_{iB} , n_{iA} , n_{iL} , n_{iMIS} , n_{iNV} e n_{iPL} il numero di lavori Eccellenti (E), Buoni (B), Accettabili (A), Limitati (L), Mancanti (MIS), Non Valutabili (NV) e Plagio (PL) della struttura i -esima, si ottiene la valutazione complessiva v_i della struttura nell'Area 09 come:

$$v_i = n_{iE} + 0.8n_{iB} + 0.5n_{iA} - 0.5n_{iMIS} - n_{iNV} - 2n_{iPL}$$

Un primo indicatore di qualità della produzione scientifica è il punteggio medio della struttura i , $I_i = \frac{v_i}{n_i}$, dove $n_i = n_{iE} + n_{iB} + n_{iCA} + n_{iD} + n_{iMIS} + n_{iNV} + n_{iPL}$. L'indice è pari a 1 nel caso in cui la struttura abbia presentato tutti i lavori attesi, e tutti abbiano ottenuto una valutazione eccellente.

Ciascuna delle tabelle che seguono riporta il punteggio medio e la distribuzione delle valutazioni.

La Tabella 4.1 (Tabella 4.1: Punteggi ottenuti e distribuzione dei prodotti nelle classi di merito) riporta i valori relativi al punteggio totale di tutti i prodotti conferiti al GEV 09 e il punteggio medio ricavato dividendo il punteggio totale per il numero totale di prodotti attesi. Inoltre, la stessa tabella mostra anche la percentuale di prodotti che hanno ottenuto le valutazioni



E, B, A, L, o che sono stati penalizzati. Si può osservare che il voto medio è quasi 0,7 e che più della metà dei prodotti sono stati considerati eccellenti. Più del 70% dei prodotti ha ricevuto una valutazione almeno B. Più dell'80% dei prodotti è stato considerato almeno accettabile. Circa il 15% dei prodotti è stato considerato limitato (L). Meno del 4% dei prodotti è stato penalizzato (prodotti mancanti, non valutabili o plagi).

La tabella mostra anche gli indicatori riferiti ai 3 subGEV e ai prodotti dei ricercatori degli enti di ricerca, che non sono stati inquadrati in alcun SSD. Si possono osservare differenze significative tra gli indici dei diversi gruppi: il punteggio medio varia da meno di 0,6 per gli enti di ricerca a più di 0,85 per il subGEV 09b e la percentuale di prodotti eccellenti varia dal 39% circa (enti) al 73% circa (09b). La percentuale di prodotti almeno accettabili (E+B+A) varia dal 72% circa (enti) al 92% circa (09b). La percentuale di prodotti penalizzati è compresa tra un minimo pari all'1,7% (09b) e un massimo del 7% circa (enti).

La Tabella 4.2 (Tabella 4.2: Punteggi ottenuti e distribuzione dei prodotti nelle classi di merito) riporta ancora gli stessi indicatori, ma questa volta riferiti ai 42 SSD del GEV 09. Le differenze diventano ancora più significative: il punteggio medio varia da meno di 0,35 a quasi 0,89 e la percentuale di prodotti eccellenti varia dal 9% circa all'80%. La percentuale di prodotti almeno buoni (E+B) varia dal 29% circa al 91% circa. La percentuale di prodotti penalizzati è compresa tra un minimo pari a zero e un massimo pari al 12,5%.

La Tabella 4.3 (Tabella 4.3: Punteggi ottenuti e distribuzione dei prodotti conferiti non penalizzati nelle classi di merito) riporta per i 3 subGEV e gli enti di ricerca, per ciascun anno della VQR, il punteggio medio dei prodotti conferiti non penalizzati e la percentuale di prodotti nelle classi E, B, A, L. Si può osservare che gli indicatori rimangono abbastanza stabili nel corso degli anni, con il punteggio medio che non scende mai sotto al 90% del valore massimo tranne che per gli enti di ricerca nel 2010, e la percentuale di prodotti eccellenti che non scende mai sotto all'85% del valore massimo tranne che per gli enti di ricerca nel 2009 e nel 2010.

La Tabella 4.4 (Tabella 4.4: Punteggi ottenuti e distribuzione dei prodotti conferiti non penalizzati nelle classi di merito per tipologia di prodotto (indicizzato o non indicizzato))



presenta un confronto tra la valutazione bibliometrica dei prodotti indicizzati e la valutazione con *peer review* dei prodotti non indicizzati. I dati mostrano che la valutazione ottenuta dai prodotti indicizzati è assai migliore da quella dei prodotti non indicizzati, come prevedibile, poiché i prodotti migliori sono quelli che vengono indicizzati dalle banche dati utilizzate dalla VQR. A ciò si aggiunge il fatto che la valutazione bibliometrica effettuata con i criteri scelti dal GEV09 è più generosa della valutazione *peer*, come dimostrano i risultati dell'Appendice A. Il voto medio dei prodotti non penalizzati risultante dalla valutazione bibliometrica dei prodotti indicizzati è più del doppio del voto medio risultante dalla valutazione *peer* dei prodotti non indicizzati (0,86 contro 0,34) e il numero di prodotti indicizzati buoni o eccellenti secondo la valutazione bibliometrica è tre volte superiore rispetto al risultato dei prodotti non indicizzati secondo la *peer review*. Coerentemente, il numero di prodotti limitati risultante dalla valutazione bibliometrica dei prodotti indicizzati (6%) è di gran lunga inferiore al risultato dei prodotti non indicizzati secondo la *peer review* (quasi il 50%). Un confronto equo tra valutazione bibliometrica e *peer review*, riguardante un campione statisticamente significativo di prodotti che hanno ottenuto entrambi le valutazioni, è presentato nell'Appendice A e mostra una maggior generosità della valutazione bibliometrica.

La Tabella 4.5 (Tabella 4.5: Punteggi ottenuti e distribuzione dei prodotti conferiti non penalizzati nelle classi di merito per tipologia di pubblicazione) permette un confronto tra le valutazioni riportate dalle diverse tipologie di prodotto. Si può vedere che quasi l'82% dei prodotti sono articoli su rivista e quasi il 96% sono articoli su rivista o congressi, le tipologie di pubblicazione più usuali e con più alta reputazione nel settore dell'ingegneria. I capitoli di libro sono meno del 2,5% e i brevetti meno dello 0,5%. Come prevedibile, per gli articoli su rivista il voto medio è molto più alto che per le altre tipologie di prodotto e la percentuale di eccellenti è approssimativamente un ordine di grandezza maggiore. Questo risultato può essere in parte attribuibile alla maggior generosità della valutazione bibliometrica, ma è sicuramente coerente con la reputazione che all'interno del settore si attribuisce alle diverse tipologie di prodotto scientifico.



La Tù

Tabella 4.6 (Tabella 4.6: Punteggi ottenuti e distribuzione dei prodotti conferiti non penalizzati nelle classi di merito per tipologia di pubblicazione e per SSD) mostra un confronto tra le valutazioni riportate dalle diverse tipologie di prodotto nei diversi SSD. Due SSD, ING-IND/18 e ING-IND/23, hanno presentato esclusivamente pubblicazioni su rivista, con punteggi medi superiori a 0,8. Per gli altri SSD, i punteggi medi delle pubblicazioni su rivista sono compresi tra 0,58 e 0,95 circa. La percentuale di eccellenti per le pubblicazioni su rivista varia tra il 23% circa e l'84% circa. Sempre con riferimento alle pubblicazioni su rivista, la percentuale di prodotti E o B varia tra un minimo pari al 45% circa e un massimo pari al 96,5% circa. Si noti che il voto medio dei contributi a congresso è sempre inferiore a quello degli articoli su rivista, e analogamente per la percentuale di prodotti eccellenti. Per gli atti di convegno, i valori della percentuale di prodotti E o B superano il 45% solo per i SSD ING-IND/01 e ING-IND/33 e sono compresi tra il 30% e il 45% per i SSD ING/IND/04, ING/IND/05, ING/IND/07, ING/IND/19, ING/IND/26, ING/INF/02, ING/INF/04, ING/INF/05.

4.1 La graduatoria delle strutture

Per la graduatoria delle strutture si propone di seguito un indicatore di qualità della ricerca indipendente dalla numerosità dei soggetti valutati della struttura afferenti all'area. Si presenta inoltre l'indicatore IRAS1 definito nel decreto ministeriale e nel Bando VQR, che tiene conto sia della qualità della ricerca sia della numerosità dei soggetti valutati della struttura afferenti all'area. Gli indicatori si riferiscono a una struttura e un'area particolare, e non sono quindi influenzati da eventuali difformità nelle valutazioni inter-area. Fatti salvi i 797 lavori di soggetti che afferiscono all'Area 09 ma valutati da GEV responsabili di altre aree (il 4,87% del totale) le graduatorie sono ottenute da una valutazione omogenea all'area.



Indicando con $N = \sum_i n_i$ e $V = \sum_i v_i$, l'indicatore $R_i = \frac{\frac{v_i}{V}}{\frac{n_i}{N}}$ indica la posizione della

struttura rispetto alla media di area. Se il valore di R_i è maggiore di 1, la struttura ha una qualità superiore alla media di area, se è minore di 1, inferiore alla media.

L'indicatore $IRAS1 = \frac{v_i}{V} = R_i \times \frac{n_i}{N}$, riflette sia la valutazione qualitativa sia le dimensioni della struttura, e può essere utilizzato per una distribuzione delle risorse che sia vista come un'alterazione, basata sulla qualità, della distribuzione basata solo sulla numerosità dei lavori. Infatti, se in tutte le strutture i lavori avessero la stessa valutazione media (lo stesso R_i) l'indicatore IRAS1 rifletterebbe soltanto il numero relativo di lavori attesi, e quindi il peso numerico della struttura nella specifica area valutata.

Poiché la VQR non può diffondere i risultati delle singole valutazioni, le tabelle prendono in considerazione solo strutture con almeno 10 lavori attesi (e quindi conferiti da almeno 3 soggetti valutati) per le università e i consorzi, e con almeno 19 lavori attesi per gli enti di ricerca. Inoltre, per motivi di affidabilità statistica, le tabelle distinguono tra strutture con numero di lavori attesi compreso tra 10 e 100 (piccole), strutture con più di 100 e meno di 350 prodotti attesi (medie) e strutture con oltre 350 lavori attesi (grandi). Le tabelle evidenziano anche gli enti di ricerca e consorzi che superano la soglia dei 19 e 10 lavori, rispettivamente.

La Tabella 4.7a (Tabella 4.7a: Graduatorie degli atenei ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito) presenta per ciascuna struttura, separando le strutture universitarie di piccola dimensione (con un numero di lavori attesi tra 10 e 100), le strutture universitarie di media dimensione (con un numero di lavori attesi tra 101 e 350), e quelle di grande dimensione (con un numero di lavori attesi superiore a 350), la percentuale di lavori nelle varie classi di merito e gli indicatori I , R , $IRAS1$. La graduatoria è ordinata per voto medio ($I=v/n$) decrescente. Per “somma punteggi (v)” si intende la valutazione complessiva dei prodotti appartenenti alla struttura, ottenuta sommando i punteggi dei singoli prodotti. Per #



prodotti attesi si intende il numero di prodotti attesi calcolato sulla base del SSD di afferenza dei soggetti valutati e del numero di prodotti che da bando questi erano tenuti a inviare alla VQR. La categoria “prodotti penalizzati” contiene i prodotti non valutabili e casi accertati di plagio o frode così come previsto dal bando VQR del 7 Novembre 2011, i prodotti mancanti (cioè attesi e non inviati), i prodotti identici presentati più volte dalla stessa struttura e i prodotti identici presentati più volte dallo stesso soggetto valutato per due strutture di tipologia differente (es. Università ed Ente di Ricerca). R indica il voto medio normalizzato. $(n/N) \times 100$ rappresenta la percentuale dei prodotti attesi della struttura rispetto ai prodotti attesi dell’area (N : numero prodotti attesi dell’area; n : numero prodotti attesi della struttura). IRAS1 è il primo indicatore di qualità della ricerca definito nel Bando VQR come rapporto tra il punteggio complessivo raggiunto da una struttura in una data area e la valutazione complessiva dell’area stessa.

La Tabella 4.7b (Tabella 4.7b: Graduatorie degli enti di ricerca ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito) presenta ancora gli stessi dati per gli enti di ricerca, e la Tabella 4.7c (Tabella 4.7c: Graduatorie dei consorzi ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito) per i consorzi.

Nel caso delle strutture universitarie si rileva anzitutto che le 4 prime posizioni dell’ordinamento globale in funzione dell’indicatore R_i sono occupate da strutture di piccola dimensione e le prime 11 posizioni sono occupate da strutture di dimensione media o piccola. Ciò induce a ipotizzare una maggior facilità a raggiungere valori elevati per strutture piccole o per qualità intrinseca, o per una miglior scelta dei prodotti, o per una differente distribuzione delle età dei ricercatori. Tra le strutture universitarie di piccola dimensione, i 5 valori maggiori dell’indicatore R_i sono ottenuti nell’ordine da: Torino, Pisa S. Anna, Roma Biomedico, Bolzano, Basilicata. Le prime 5 strutture di media dimensione risultano essere Sannio, Trento, Ferrara, Siena e Cassino. Le prime 5 di grande dimensione sono Padova, Milano Politecnico, Bologna, Torino Politecnico e Roma Tor Vergata. Tra le strutture universitarie di piccola dimensione, ben 11 presentano valori dell’indicatore R_i superiori a 1, come 13 strutture di media dimensione e



solo 4 di grande dimensione. Lo stesso vale per 5 enti di ricerca su 8 (ASI, IIT, INRIM, Elettra e CNR) e per 2 consorzi su 5 (INSTM e INCA).

La Tabella 4.8 (Tabella 4.8: Graduatorie di strutture per subGEV ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito) presenta, per ciascuna struttura universitaria di piccola, media o grande dimensione (con un numero di lavori attesi tra 10 e 100, oppure tra 101 e 350, oppure superiore a 350), la percentuale di lavori nelle varie classi di merito per ciascun subGEV. Per molte strutture si osservano variazioni significative di voto medio e di posizione in graduatoria nei diversi subGEV. È interessante il confronto dei valori delle singole strutture con i valori complessivi per i diversi subGEV, riportati nella stessa Tabella 4.8. Per esempio la posizione in graduatoria di Padova, che risulta essere la prima in graduatoria tra le strutture di grande dimensione e che quindi fornisce l'elemento di paragone più affidabile, è 4 per il subGEV 09a, 22 per il subGEV 09b, e 13 per il subGEV 09c.

Infine, la Tabella 4.9 (Tabella 4.9: Graduatorie di strutture per SSD ordinate per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito) presenta, per ciascuna struttura universitaria di piccola, media o grande dimensione, la percentuale di lavori nelle varie classi di merito per ciascun SSD. Anche in questo caso, per molte strutture si osservano variazioni significative di voto medio e di posizione in graduatoria nei diversi SSD. Per esempio la posizione in graduatoria di Padova è 1 per i SSD ING-IND/16, ING-IND/17 e ING-INF/02, ING-INF/06, mentre risulta essere 22 per il SSD ING-INF/05.

4.2 La valutazione dei dipartimenti

La VQR ha tra i suoi compiti anche l'elaborazione di una graduatoria dei dipartimenti che possa essere utilizzata come informazione dagli organi decisionali delle strutture. I dipartimenti istituiti a valle della Legge 240 presentano tipologie diverse. Le più frequenti sono dipartimenti con docenti e ricercatori che appartengono tutti (o in gran parte) a una stessa area, e dipartimenti più compositi, con docenti e ricercatori di due (o più) aree diverse. In entrambi i casi, occorre



costruire indicatori di dipartimento a partire dalle valutazioni dei lavori associati ai soggetti valutati di ciascun dipartimento. Anche in questo caso, come per le strutture, è importante che il risultato finale non sia influenzato da difformità delle valutazioni inter-area.

Di seguito si presentano gli indicatori per i dipartimenti, calcolati con le medesime modalità delle strutture (l'indicatore IRD1 per i dipartimenti sostituisce IRAS1). Si noti che gli indicatori si riferiscono esclusivamente ai ricercatori dell'Area 09. Pertanto, un dipartimento presente in un'unica area, sarà presente solo nella graduatoria di quell'area; invece, un dipartimento con soggetti valutati di più aree diverse sarà presente nella graduatoria di due o più aree diverse.

Poiché la VQR non può diffondere i risultati delle singole valutazioni, le tabelle prendono in considerazione solo dipartimenti con almeno 10 lavori attesi (e quindi conferiti da almeno 3 soggetti valutati). Per motivi di affidabilità statistica, le tabelle distinguono tra dipartimenti *piccoli*, con numero di lavori attesi compreso tra 10 e 50 e dipartimenti *grandi*, con oltre 50 lavori attesi. Le tabelle evidenziano anche gli enti di ricerca e consorzi che superano la soglia dei 19 e 10 lavori.

La Tabella 4.10a (Tabella 4.10a: Graduatorie dei dipartimenti post L. 240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito) presenta, per ciascun dipartimento universitario post L.240 con oltre 10 lavori attesi, la percentuale di lavori nelle varie classi di merito e gli indicatori I , R , $IRD1$. La tabella è ordinata per valori decrescenti del voto medio. Bisogna osservare che un numero non trascurabile di prodotti (628 su 13.121 – quasi il 5%) non è stato associato dalle università ad alcun dipartimento e quindi compare con la dicitura *n.d.*. Per alcuni atenei questo indica l'esistenza di un solo dipartimento; per gli altri atenei il numero di prodotti non è tale da alterare le classifiche se uniformemente distribuito sui dipartimenti, ma potrebbe causare perturbazioni nel caso in cui i prodotti non associati fossero concentrati su pochi dipartimenti.

Anche in questo caso si vede come la valutazione tenda a favorire strutture di piccola dimensione. Infatti, le prime 5 posizioni della classifica complessiva sono occupate da



dipartimenti piccoli. Per i dipartimenti piccoli si osserva, però, una maggior dispersione di valori, infatti, i valori dell'indicatore R_i sono compresi tra un minimo pari a 0,08 circa e un massimo pari a 1,38 circa. Invece, per i dipartimenti grandi i valori dell'indicatore R_i sono compresi tra un minimo pari a 0,6 circa e un massimo pari a 1,3 circa. Esistono inoltre significative differenze tra dipartimenti all'interno dello stesso ateneo. Per esempio, nel caso dell'Università di Padova (che presenta solo 21 prodotti non attribuiti ad alcun dipartimento su 597), i 3 dipartimenti considerati compaiono in classifica in posizioni 12, 25 e 49, con valori dell'indicatore R_i pari a circa 1,24, 1,16 e 1,07.

La Tabella 4.11a (Tabella 4.11a: Graduatorie dei dipartimenti post L.240 ordinati per voto medio ($I=v/n$) all'interno delle strutture (elencate in ordine alfabetico) e distribuzione dei prodotti nelle classi di merito) presenta gli stessi dati, ma raggruppando i dipartimenti degli stessi atenei in righe contigue, in modo da facilitare confronti all'interno degli atenei.

Le Tabelle 4.10b e 4.11b presentano gli stessi dati, per le sottostrutture degli enti di ricerca. Nel caso del CNR si osservano notevoli differenze tra i dipartimenti, con valori dell'indicatore R_i compresi tra un minimo pari a 0,16 e un massimo superiore a 0,8.

Le Tabelle 4.12 (Tabella 4.12: Graduatorie dei dipartimenti ante L.240 ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito) e 4.13 (Tabella 4.13: Graduatorie dei dipartimenti ante L.240 ordinati per voto medio ($I=v/n$) all'interno delle strutture (elencate in ordine alfabetico) e distribuzione dei prodotti nelle classi di merito) presentano, per ciascun dipartimento universitario pre L.240 con oltre 10 lavori attesi la percentuale di lavori nelle varie classi di merito e gli indicatori I , R , $IRDI$. Le tabelle sono ordinate per valori decrescenti del voto medio.

Le Tabelle 4.14 (Tabella 4.14: Graduatorie dei dipartimenti post L.240 per subGEV ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito) e 4.15 (Tabella 4.15: Graduatorie dei dipartimenti post L.240 per SSD ordinati per voto medio ($I=v/n$) e distribuzione dei prodotti nelle classi di merito) presentano i dati disaggregati a livello di subGEV e di SSD.



La Tabella 4.16 (Tabella 4.16: Graduatoria delle strutture per voto medio (VM) dei soggetti valutati (SV) e distribuzione in classi di voto medio) presenta la graduatoria delle strutture per voto medio (VM) dei soggetti valutati (SV) e la distribuzione in classi di voto medio. Per VM si intende il voto medio dei ricercatori della struttura, calcolato come rapporto tra la somma del punteggio medio dei prodotti presentati da ciascun ricercatore e il numero di soggetti valutati della struttura. Le 8 classi di voto medio contengono rispettivamente il numero dei soggetti valutati con $VM < 0$, $VM = 0$, $0 < VM \leq 0,2$, $0,2 < VM \leq 0,4$, $0,4 < VM \leq 0,6$, $0,6 < VM \leq 0,8$, $0,8 < VM \leq 0,99$, $VM = 1$. Inoltre la tabella riporta il numero di soggetti valutati i cui prodotti hanno tutti classe finale E (eccellente) assunti prima e dopo il 2006. La tabella non include le strutture con meno di 4 soggetti valutati. Si può osservare come in un caso la percentuale di ricercatori con $VM < 0$ raggiunga ben il 33%. La percentuale di ricercatori con $VM = 1$ varia su tutto lo spettro da 0% a 100%. Il numero totale di ricercatori valutati è pari a 5.599. Il numero totale di ricercatori assunti dopo il 2006 con tutti i prodotti E è pari a 369, mentre il numero totale di ricercatori assunti prima del 2006 con tutti i prodotti E è pari a 1.426. In totale, la percentuale di ricercatori con tutti i prodotti E è pari al 32% circa. La percentuale di ricercatori con tutti prodotti E nelle diverse strutture varia su tutto il campo tra 0 e 100%.

La Tabella 4.17 (Tabella 4.17: Graduatoria dei dipartimenti ex L.240 per voto medio (VM) dei soggetti valutati (SV) e distribuzione in classi di voto medio) presenta la graduatoria dei dipartimenti ex L.240 per voto medio (VM) dei soggetti valutati (SV) e distribuzione in classi di voto medio. Inoltre la tabella riporta il numero di soggetti valutati i cui prodotti hanno tutti classe finale E (eccellente) assunti prima e dopo il 2006. Sulla base del Bando, e per i soli ricercatori delle università e ricercatori e tecnologi degli enti di ricerca, il numero di prodotti attesi subiva delle riduzioni in funzione dell'anno della presa di servizio nei rispettivi ruoli. Per tutte le altre categorie, non erano previste riduzioni se non legate a determinate categorie di congedo. Nel costruire la tabella, si è identificato l'anno della presa di servizio con il numero di prodotti attesi sulla base del Bando, senza tenere conto di eventuali congedi. La tabella non include i dipartimenti con meno di 4 soggetti valutati.



La Tabella 4.18 (Tabella 4.18: Numero soggetti valutati attivi, non attivi, parzialmente attivi e parzialmente inattivi per struttura) mostra il numero di soggetti valutati attivi, non attivi, parzialmente attivi e parzialmente inattivi per struttura. La colonna “# SV” indica il numero di soggetti valutati (SV) della struttura. Per soggetti valutati attivi (SV Attivi) si intendono i soggetti valutati che hanno conferito un numero di prodotti uguale al numero di prodotti attesi. Per soggetti valutati non attivi (SV NA) si intendono i soggetti valutati che non hanno presentato alcun prodotto. Per soggetti valutati parzialmente attivi (SV PA) si intendono i soggetti valutati che non hanno presentato tutti i prodotti attesi ma un numero superiore alla metà. Per soggetti valutati parzialmente inattivi (SV PI) si intendono i soggetti valutati che hanno presentato al più la metà dei prodotti attesi. La colonna “% (NA+PA+PI)” rappresenta la quota di soggetti valutati non pienamente attivi ottenuta considerando i non attivi (NA), i parzialmente attivi (PA) e i parzialmente inattivi (PI). La tabella non include i dipartimenti con meno di 4 soggetti valutati. La percentuale di soggetti valutati non pienamente attivi raggiunge un massimo pari al 22% circa. Nel caso dell’Università di Padova questo valore è inferiore al 2,8%.

La Tabella 4.19 (Tabella 4.19: Punteggio medio dei soggetti valutati per subGEV e anno di assunzione in ruolo) presenta il punteggio medio dei soggetti valutati per subGEV e anno di assunzione in ruolo. Il punteggio medio è calcolato come rapporto del punteggio complessivo e il numero dei prodotti attesi. Il numero di prodotti attesi è stato calcolato sulla base del SSD di afferenza dei soggetti valutati e del numero di prodotti che da bando questi erano tenuti a sottomettere. In generale si osserva una crescita del punteggio medio al diminuire dell’età dei ricercatori, con le eccezioni dei ricercatori più giovani del subGEV 09c e dei ricercatori degli enti non universitari.

Bisogna osservare come le valutazioni relative ai ricercatori, specie nei casi di prodotti con autori multipli, siano almeno in parte arbitrarie, dal momento che in molte strutture la scelta dell’associazione dei prodotti ai ricercatori è stata operata dall’ente, che ha mirato all’ottimizzazione del proprio punteggio totale, in coerenza con gli obiettivi della VQR, senza porre attenzione all’attribuzione dei lavori al singolo ricercatore.



5 Analisi dei risultati

I 40 membri del GEV 09 (diventati 39 a partire da Maggio 2012), 5 dei quali operano in atenei esteri, hanno valutato 16.347 prodotti, pari a circa il 97% dei 16.858 lavori attesi per i 42 SSD dell'Area 09 e presentati da oltre 5.500 ricercatori di 91 strutture: 73 università, 10 enti di ricerca e 8 consorzi. Il fatto che la percentuale di prodotti conferiti alla VQR dalle strutture sia così elevata testimonia la vivacità della ricerca italiana nel settore dell'ingegneria. Per ben 8 SSD la percentuale di prodotti valutati sale al 100% dei prodotti attesi, dando evidenza di un importante impegno scientifico di tutti i ricercatori di quei SSD.

Oltre l'80% dei prodotti presentati sono articoli su rivista; considerando sia articoli su rivista sia contributi a congresso, le due tipologie di pubblicazione più rilevanti per il settore dell'ingegneria, si arriva oltre il 95%. Dei prodotti per cui è indicata la lingua, più del 95% sono in inglese. In generale, si osserva una notevolissima propensione a pubblicare articoli su rivista in lingua inglese (la tipologia di prodotto scientifico con la maggiore reputazione) in tutti i settori dell'ingegneria.

Circa il 77% dei prodotti presentati è stato valutato mediante parametri bibliometrici; circa il 45% è stato valutato mediante *peer review* e quindi circa il 22% ha ricevuto entrambe le valutazioni. Circa la metà di questo 22% è stata scelta in modo da avere un campione statisticamente significativo per confrontare i risultati dei due metodi di valutazione. Il confronto ha mostrato una maggior generosità dell'algorithm bibliometrico scelto rispetto alla *peer review*.

La valutazione dei prodotti effettuata dal GEV utilizzando i parametri definiti nel decreto ministeriale e nel bando VQR identifica come eccellenti i prodotti riferibili al primo 20% della produzione mondiale complessiva. Questa fascia è assai più ampia di quella utilizzata in altri contesti. Per questo motivo, nell'analisi dell'area si sono esaminati i prodotti che, facendo unicamente riferimento alle citazioni ricevute dagli articoli, si collocano nel primo 1% e nel



primo 5% della produzione mondiale complessiva. I valori sono calcolati rispetto all'anno e alla *Subject Category* in cui la pubblicazione si colloca. Complessivamente, l'Area 09 colloca nel top 1% o nel top 5% della distribuzione mondiale una quota rispettivamente pari a circa il 2% e 10% dei prodotti sottomessi alla VQR, con una presenza quindi nelle due fasce di eccellenza circa doppia rispetto a un campione casuale. Questo risultato indica chiaramente che la ricerca italiana nel settore dell'ingegneria è di ottima qualità. Se poi si analizzano le percentuali relative ai singoli SSD, si può vedere che nel top 1% della produzione internazionale alcuni SSD dell'ingegneria industriale arrivano a collocare circa il 7% dei prodotti sottomessi alla VQR (ING/IND-20 e ING/IND-27). Tra gli SSD dell'ingegneria dell'informazione, due superano il 3% (ING-INF/03 e ING-INF/04). Questo indica una notevole presenza di gruppi di ricerca di assoluta eccellenza a livello internazionale.

Sulla base del bando VQR, i singoli lavori ricevono un punteggio pari a 1, 0,8, 0,5 e 0 a seconda che il lavoro sia valutato Eccellente (E), Buono (B), Accettabile (A) o Limitato (L); ai lavori mancanti è assegnato un punteggio pari a -0,5, ai lavori non valutabili è assegnato un punteggio -1, e in casi accertati di plagio o frode -2. Complessivamente, per l'Area 09 le percentuali dei prodotti nelle classi sono: E circa 50%, B quasi 20%, A oltre 10% e L oltre 15%. Poco più del 3,5% dei prodotti è stato penalizzato. La gran parte dei "prodotti" penalizzati è stata valutata negativamente in quanto mancante (ovvero non presentata alla VQR dalla struttura valutata). Solo 77 prodotti presentati sono stati considerati non ammissibili.

Vi è una forte correlazione tra la graduatoria basata sugli indicatori bibliometrici delle riviste e il numero di citazioni. Facendo riferimento al database WoS, circa il 77% degli articoli con almeno 5 citazioni complessive nel 2004-2010, riceve valutazione finale eccellente (E). Dei circa 1.300 articoli con almeno 5 citazioni medie annue, circa il 75% è pubblicato in riviste classificate nel gruppo migliore.

Complessivamente, per la valutazione tramite *peer review* il GEV ha utilizzato oltre 2.000 revisori (dei quali l'82% sono ricercatori operanti all'estero) che hanno generato oltre 12.000 review (82% da ricercatori operanti all'estero). Il punteggio medio attribuito dai revisori è di



0,597 con valori compresi tra 0,58 e 0,625 nei 3 subGEV. Anche la deviazione standard del punteggio attribuito dai revisori è simile nelle tre aree. Questi dati indicano un comportamento sostanzialmente uniforme dei revisori all'interno del GEV.

Esaminando i risultati della valutazione dei prodotti complessivamente conferiti al GEV si può osservare che il voto medio è quasi 0,7 e che più della metà dei prodotti sono stati considerati eccellenti (E). Più del 70% dei prodotti ha ricevuto una valutazione almeno buona (B). Più dell'80% dei prodotti è stata considerata almeno accettabile. Questi dati testimoniano un'ottima qualità dei prodotti presentati alla VQR nel settore dell'ingegneria. Si possono osservare differenze significative tra gli indici dei diversi gruppi: il punteggio medio varia da meno di 0,6 per gli enti di ricerca a più di 0,85 per il subGEV 09b e la percentuale di prodotti eccellenti (E) varia dal 39% circa (enti) al 73% circa (09b). La percentuale di prodotti almeno accettabili (E+B+A) varia dal 72% circa (enti) al 92% circa (09b). La percentuale di prodotti penalizzati è compresa tra un minimo pari all'1,7% (09b) e un massimo del 7% circa (enti). Le differenze diventano ancora più significative se si considerano aggregazioni a livello di SSD: il punteggio medio varia da meno di 0,35 a quasi 0,89 e la percentuale di prodotti eccellenti (E) varia dal 9% circa all'80%. La percentuale di prodotti almeno buoni (E+B) varia dal 29% circa al 91% circa. La percentuale di prodotti penalizzati è compresa tra un minimo pari a zero e un massimo pari al 12,5%. Questi dati mostrano una realtà piuttosto diversificata, ma di ottimo livello nella grande maggioranza dei casi.

Nel caso delle strutture universitarie si rileva una maggior facilità a raggiungere valori elevati degli indicatori di qualità per strutture piccole. Questo può essere dovuto o a una maggiore qualità intrinseca, o a una miglior scelta dei prodotti (facilitata dai numeri contenuti), o a una differente distribuzione delle età dei ricercatori (si osserva, infatti, un incremento del voto medio per ricercatore al diminuire dell'età).

Analizzando la percentuale del numero di ricercatori che hanno presentato tutti prodotti eccellenti (E) nelle diverse strutture, si osserva che questa varia su tutto lo spettro da 0% a 100%. Il numero totale di ricercatori assunti dopo il 2004 con tutti i prodotti E è pari a 369, mentre il



numero totale di ricercatori assunti prima del 2004 con tutti i prodotti E è pari a 1.426. In totale, la percentuale di ricercatori con tutti i prodotti E è pari al 32% circa. La percentuale di soggetti valutati non pienamente attivi è superiore al 10% solo in 5 strutture. Ancora una volta, questi indici riflettono una qualità di elevatissimo livello.

6 Conclusioni

I risultati della VQR sono complessivamente molto lusinghieri per l'Area 09 (Ingegneria Industriale e dell'Informazione). La quantità di prodotti presentati è molto vicina alla quantità di prodotti attesi, più della metà dei prodotti presentati è stata considerata eccellente (E) e per circa un terzo dei ricercatori per i quali le strutture hanno presentato prodotti, questi prodotti hanno tutti ricevuto una valutazione eccellente (E).

I meccanismi di valutazione dei prodotti scientifici si sono dimostrati adeguati agli scopi della VQR, ma nel caso di valutazioni future richiederanno qualche revisione. I suggerimenti che il GEV 09 ritiene di poter trasmettere ad ANVUR sono i seguenti.

- È importante definire un algoritmo che consenta di considerare in modo differente le citazioni in funzione della provenienza, con l'obiettivo di valorizzare le citazioni provenienti da ricercatori o gruppi di ricerca con eccellente reputazione internazionale e di escludere le citazioni provenienti dallo stesso autore o dallo stesso gruppo di ricerca, o dalla stessa struttura dell'autore. Per un numero limitato di prodotti di ricerca è stata osservata, infatti, una quasi totalità di autocitazioni, così che una loro eliminazione avrebbe portato a un cambiamento radicale della valutazione finale del prodotto.
- È auspicabile incrementare l'intervallo tra la fine del periodo coperto dalla valutazione e la valutazione stessa, in modo da permettere agli indici bibliometrici



di stabilizzarsi e quindi differenziare in modo significativo le soglie sui numeri di citazioni. Infatti, dato che la distribuzione delle citazioni degli articoli pubblicati in aree omogenee è tipicamente molto asimmetrica ed evidenzia un numero relativamente basso di lavori molto citati e un numero molto alto di lavori poco o per nulla citati, potrebbe essere elevata la probabilità di avere il passaggio da una classe di merito a quella successiva nella distribuzione delle citazioni a causa una o due citazioni, o addirittura la fusione di due classi di merito.

- È auspicabile l'introduzione di meccanismi più sofisticati per la composizione della valutazione bibliometrica della sede di pubblicazione (rivista) e dell'informazione ottenuta dalla distribuzione citazionale, al fine di ottenere la valutazione del singolo prodotto. Inoltre, nel caso in cui, nei prossimi anni, sia dimostrata la validità bibliometrica degli indicatori a livello di singolo articolo (le cosiddette “*altmetrics*” o “*article level metrics*”) si raccomanda di considerarne l'adozione, almeno come ulteriore parametro di valutazione.
- È necessario migliorare la facilità d'interazione con gli strumenti informatici per la valutazione delle pubblicazioni (in particolare quelli per la *peer review*) e per il monitoraggio del processo di valutazione.
- È importante stabilire un contatto preliminare con i revisori, specie se stranieri, per informarli degli obiettivi, delle metodologie e dei tempi della valutazione.
- È indispensabile definire tempi ragionevoli e certi per l'intero processo di valutazione.