



3 Il processo di valutazione

La valutazione dei prodotti è stata effettuata dai GEV utilizzando la bibliometria e la *peer review*. Ogni prodotto è stato assegnato a due membri GEV responsabili del relativo processo di valutazione. Nel seguito, il processo sarà descritto separatamente per ciascuna metodologia.

3.1 La valutazione mediante *peer review*

La procedura connessa alla valutazione mediante *peer review* ha avuto inizio a Febbraio 2015 con la costituzione di un albo di revisori ANVUR-VQR2 suddiviso per GEV. Pur avendo a disposizione l'albo di revisori REPRISE del MIUR, utilizzato dal MIUR per la valutazione *ex ante*, tra gli altri, dei progetti PRIN, si è ritenuto opportuno costituire un nuovo albo, tenuto conto del fatto che i revisori presenti in REPRISE non erano mai stati sottoposti a valutazione preventiva sulla base delle loro credenziali scientifiche, e che il numero degli esperti stranieri era piuttosto limitato.

I GEV hanno quindi operato prima una selezione dei revisori dell'albo REPRISE in base a criteri di merito scientifico (indice h di Hirsch, numero di citazioni, produzione scientifica recente) e, successivamente, lo hanno integrato con un numero elevato di esperti scelti sulla base degli stessi criteri e interpellati individualmente al fine di valutarne la disponibilità a partecipare alla VQR2. Ovviamente, la scelta dei criteri di merito è stata modulata dai vari GEV a seconda della disponibilità o meno di informazioni di natura bibliometrica.

Nel solo caso del GEV12 è stato pubblicato un modulo di autocandidatura, compilabile da coloro che, non essendo già presenti nell'Albo REPRISE, intendessero contribuire al processo di valutazione in qualità di revisori.

Integrando le liste dei revisori dell'albo REPRISE con quelle costruite dai GEV, si è pervenuto all'albo iniziale ANVUR-VQR2, costituito da quasi 14.000 nomi. Il processo di selezione dei revisori è proseguito anche durante la fase di valutazione per coinvolgere competenze non coperte dalle liste definite fino a quel momento e resesi necessarie per la valutazione di prodotti specifici.

I due membri GEV responsabili di ogni prodotto hanno scelto separatamente i due revisori, ponendo una particolare attenzione a evitare i conflitti d'interesse sulla base delle indicazioni contenute nei documenti sui criteri di valutazione.



Nelle Tabella 3.1, Tabella 3.2 Tabella 3.3 e nella Figura 3.1 sono riportate alcune statistiche sui revisori che hanno partecipato alla VQR. Esse fanno riferimento alla nazionalità italiana o “straniera”, intendendo con quest’ultimo termine l’affiliazione ad una istituzione estera e non la nazionalità del revisore.

Tabella 3.1. Numero di revisori per area distinti per nazionalità (italiana o straniera)

Tabella 3.2. Revisioni assegnate, effettuate e rifiutate per area e per nazionalità (italiana o straniera), con l’esclusione delle revisioni effettuate internamente dai membri GEV

Tabella 3.3. Numero e percentuale di prodotti totali e prodotti sottoposti a revisione *peer* per area

Figura 3.1. Numero di revisioni assegnate, effettuate e rifiutate per area e per nazionalità (italiana o straniera)

Complessivamente, la VQR2 ha impegnato 16.969 revisori di cui 13.546 italiani e 3.423 con affiliazione estera. In realtà, il numero di revisori esterni (escludendo cioè i membri GEV che hanno anche operato in qualità di revisori) intesi come persone fisiche distinte è inferiore, e pari a 12.731, in quanto i numeri della Tabella 3.1 sommano il numero dei revisori in ogni area, contando più volte i revisori che sono stati utilizzati in più di un SSD. Nell’Area 1 e nell’Area 9 i revisori stranieri sono circa il 60% del totale, mentre nelle altre aree prevalgono i revisori italiani. Come si evince dalla Tabella 3.2 e anche dalla Figura 3.1 i revisori con affiliazione italiana hanno dimostrato maggior disponibilità: il 78% dei prodotti loro assegnati è stato valutato, contro un valore equivalente del 66% per i revisori stranieri

Dalla Tabella 3.3 risulta che i GEV 8a, 10, 12 e 14 hanno valutato la totalità dei prodotti con modalità *peer* (nella tabella le percentuali in tali aree sono leggermente inferiori al 100% in quanto sono considerati anche i prodotti afferenti ad addetti dei GEV ma valutati da altri GEV). È importante sottolineare che un campione, pari a circa il 10% dei prodotti valutati bibliometricamente, è stato anche sottoposto a revisione *peer* al fine di misurare il grado di correlazione delle due metodologie di valutazione. Un’analisi dettagliata della metodologia di confronto e dei suoi risultati si trova nell’Appendice B.

Parte delle valutazioni *peer* sono state effettuate da membri GEV, con le stesse procedure delle valutazioni esterne. Complessivamente, la percentuale di valutazioni *peer* effettuate direttamente all’interno dei GEV è stata contenuta e pari al 13,6%. Ogni prodotto soggetto alla valutazione *peer* ha avuto almeno due revisioni. In qualche caso, per il ritardo nella consegna della valutazione da parte di alcuni revisori, e il conseguente invio a un terzo revisore, il numero di valutazioni è stato superiore a 2.



A ogni revisore è stato richiesto di valutare il prodotto sulla base di tre domande a risposta multipla¹⁰, una per ognuno dei criteri a, b, c della Sezione 2.5. A ogni risposta era associato un punteggio. La somma dei tre punteggi era confrontata con quattro soglie per generare una classificazione finale in cinque classi. La classificazione era proposta al revisore per consentirgli di confrontarla con la definizione delle classi 1, 2, 3 e 4 della Sezione 2.5 e, eventualmente, di modificare i punteggi. La valutazione *peer* richiedeva, oltre alla definizione dei punteggi assegnati sulla base dei tre criteri di merito, la formulazione di un parere scritto sul prodotto da valutare.

Ogni GEV ha costituito dei gruppi di consenso formati da due o tre membri, che, anche sulla base dei punteggi espressi dai due (o più) revisori e di un procedimento definito a priori, perveniva alla classificazione finale. Le valutazioni finali sono state approvate prima singolarmente dal Coordinatore GEV e poi complessivamente dal GEV in seduta plenaria, anche condotta per via telematica.

Al di là di piccole variazioni tra i GEV, il procedimento consentiva la richiesta di una terza revisione *peer* nel caso di valutazioni discordanti per due o più classi. Nella Tabella 3.4 si riportano per ogni GEV i numeri assoluti e le percentuali dei prodotti che hanno avuto revisioni discordanti per una, due, tre e quattro classi.

Mediando su tutte le aree, la percentuale di revisioni discordanti per almeno due classi è pari al 19,7%. L'argomento è anche ripreso nell'Appendice B.

Tabella 3.4. Numero e percentuali di revisioni *peer* discordanti per 1, 2, 3 e 4 classi per area

3.2 La valutazione bibliometrica

La valutazione bibliometrica dei GEV 2, 3, 4, 5, 6, 7, 8b, 9 e 11b ha riguardato gli articoli pubblicati su riviste indicizzate nelle basi di dati WoS e Scopus. L'ANVUR ha acquisito da Thomson-Reuters ed Elsevier, tramite il CINECA, le informazioni bibliometriche di tali archivi per gli anni 2011-2014 e per la produzione scientifica mondiale. Contrariamente alla scelta effettuata in altri paesi per esercizi di valutazione analoghi, l'ANVUR ha preferito utilizzare entrambe le basi di dati per evitare di legarsi a un solo gestore, e per sfruttare al meglio le caratteristiche di parziale complementarietà delle due basi di dati.

¹⁰ Per le domande e i punteggi si rimanda ai Rapporti finali di Area.

Rimandando ai Rapporti di Area per i dettagli sugli algoritmi bibliometrici utilizzati da ogni GEV, si descrivono brevemente nel seguito gli elementi principali.

L'algoritmo di valutazione dei GEV 2, 3, 4, 5, 6, 7, 8b, 9 e 11b si basa sul calcolo di due indicatori per ogni prodotto: il numero di citazioni del prodotto e il fattore di impatto della rivista di pubblicazione. Alle istituzioni è stato chiesto, nella fase di conferimento dei prodotti, di indicare nella scheda prodotto la base di dati (WoS oppure Scopus) e l'indicatore di impatto preferiti (IF5Y, Article Influence Score per WoS e IPP, SJR per Scopus). Nel caso di riviste appartenenti a più di una *Subject Category* le istituzioni hanno espresso una preferenza, sottoposta al giudizio dei GEV per conferma. Gli articoli pubblicati dalle riviste appartenenti unicamente alla categoria *multidisciplinary science*, che include riviste caratterizzate da una pluralità di argomenti scientifici, quali Nature, Science, ecc., sono stati riassegnati a un'altra *Subject Category* sulla base (i) delle citazioni contenute nell'articolo e (ii) delle citazioni fatte all'articolo. In particolare, per ognuna delle riviste citate si è individuata una (o più) *Subject Category* di appartenenza ed è stata scelta quella finale con una regola di decisione maggioritaria. Nell'assegnazione alla nuova *Subject Category*, il prodotto ha portato con sé il fattore di impatto della rivista di pubblicazione e il numero di citazioni ricevute, senza modificare le distribuzioni della *Subject Category* di destinazione.

Alla coppia di valori dei due indicatori caratteristici del prodotto considerato, con regole lievemente diverse da GEV a GEV (per le quali si rimanda ai Rapporti finali di Area), si è associata una fra sei classi: le cinque classi della VQR2 e una sesta classe (IR) ottenuta nel caso di indicazioni divergenti dei due indicatori (ad esempio, un prodotto con elevato numero di citazioni pubblicato su una rivista con impatto molto basso o viceversa). I prodotti IR sono stati sottoposti a valutazione *peer*.

Il GEV1 ha adottato un algoritmo di valutazione bibliometrico lievemente diverso, che non si basa direttamente sulle *Subject Categories* di ISI WoS e Scopus, ma che ha identificato delle categorie di riferimento, una per ogni SSD di competenza del GEV, che integrano le *Subject Categories* (SC) usate in WoS e la classificazione *All Science Journal Classification* (ASJC) usata in Scopus. Inoltre, Il GEV1 ha utilizzato, oltre alle basi di dati WoS e Scopus, e limitatamente all'indicatore d'impatto della rivista, MathSciNet dell'American Mathematical Society (MathSciNet). Per i dettagli si veda il Rapporto di Area del GEV1.

Il GEV13 ha utilizzato un algoritmo bibliometrico sensibilmente diverso dagli altri GEV bibliometrici, privilegiando la sede di pubblicazione e utilizzando il numero di citazioni per



“premiare” con un salto di classe i prodotti con un numero di citazioni significativo. Anche in questo caso, per i dettagli si rimanda al Rapporto di Area del GEV13.

Mentre nella VQR1 le autocitazioni erano incluse senza distinzioni nelle citazioni al fine di calcolare l'indicatore citazionale, nella VQR2 sono stati segnalati i casi in cui il numero di autocitazioni superava la soglia del 50% delle citazioni complessive, chiedendo in tali casi ai membri GEV responsabili della valutazione del prodotto una attenzione particolare.

Nella Tabella 3.5 si riportano i numeri assoluti e le percentuali di prodotti valutati bibliometricamente e di prodotti con classificazione IR per ogni GEV. L'attribuzione dei prodotti alle aree è fatta sulla base dell'addetto cui i prodotti sono stati associati. Ciò spiega come mai vi siano alcuni prodotti delle aree 8a, 10, 11a, e 14 che sono stati valutati bibliometricamente: per tali prodotti le istituzioni avevano suggerito per la valutazione un GEV bibliometrico diverso da quello di afferenza dell'addetto.

Tabella 3.5. Numero e percentuali di prodotti totali e prodotti valutati bibliometricamente e classificati IR per area

Come già anticipato, gli algoritmi dei GEV per la valutazione bibliometrica hanno utilizzato regole non identiche per l'attribuzione della classe a partire dai valori dei due indicatori. Una calibrazione accurata di questi algoritmi, in modo da rispettare le percentuali assegnate a ogni classe dal DM e dal Bando, è stata effettuata prima dell'approvazione e pubblicazione dei criteri, consentendo alle istituzioni di scegliere in maniera consapevole i prodotti da sottoporre a valutazione.

3.3 I prodotti “penalizzati”

Il DM e il Bando prevedevano l'assegnazione di punteggio zero ai prodotti “mancanti”, vale a dire prodotti attesi ma non conferiti dalle istituzioni, e ai prodotti “non valutabili”. Inoltre, erano da definire i casi di prodotti soggetti a penalizzazioni, quali ad esempio prodotti presentati due volte dallo stesso addetto. Le Tabella 3.3 e Tabella 3.5 mostrate in precedenza riportano per ogni area il numero di prodotti mancanti e di prodotti non valutabili. L'algoritmo deciso congiuntamente da tutti i GEV per attribuire le penalizzazioni prevede cinque casi distinti.

1. A ogni prodotto mancante, si applica al prodotto il punteggio di 0.
2. Ai prodotti non valutabili (le cause possono essere diverse e sono indicate dal Bando, ad esempio la mancanza del *file* pdf, o l'anno di pubblicazione al di fuori dei quattro anni della VQR2, ecc.) viene associato il punteggio di 0.



3. Se una istituzione presenta n volte lo stesso prodotto, il prodotto è valutato (ad esempio con Eccellente, punteggio 1), mentre gli altri $n-1$ ricevono valutazione nulla. Pertanto, a ciascuno è applicato un punteggio pari a $1/n$.
4. Se due istituzioni di tipologia diversa (ad esempio una università e un ente di ricerca vigilato dal MIUR) associano lo stesso prodotto allo stesso addetto, un prodotto è valutato (ad esempio con eccellente, punteggio 1), mentre l'altro è penalizzato con zero. Pertanto, a ciascuno dei due è applicato un punteggio pari a 0,5.

La mancanza del pdf o la presenza di un pdf incompleto o illeggibile non hanno comportato automaticamente la penalizzazione; in tali casi, infatti, l'ANVUR ha chiesto alle istituzioni coinvolte di inviare (o sostituire nel caso di *file* danneggiato) i pdf mancanti. Soltanto nel caso in cui la richiesta non abbia avuto seguito è scattata la penalizzazione.