

Valutazione della Qualità della Ricerca  
2011-2014 (VQR 2011-2014)

Bando di partecipazione

Versione provvisoria approvata dal Consiglio Direttivo ANVUR  
8 luglio 2015

**Proposte di integrazioni/revisioni avanzata  
dall'Associazione Italiana di Sociologia (AIS)**

L'Associazione Italiana di Sociologia, valuta positivamente la pubblicazione del Bando provvisorio per l'avvio dell'esercizio di Valutazione della Qualità della Ricerca, che sollecita esplicitamente le parti interessate ad esprimere pareri e formulare proposte tesi ad un miglioramento del processo di valutazione.

**La posizione dell'AIS circa la VQR pregressa: la lezione acquisita dagli esiti e usi della VQR**

L'AIS aveva già nel 2014 predisposto un documento analitico e circostanziato "Per una revisione degli indicatori di Valutazione della Qualità della ricerca in ambito non bibliometrico". In quel documento, di cui si allega il link al sito AIS, <http://tinyurl.com/odstzyg>, si sottolineava in premessa che: *La finalità emergente dal combinato-disposto del Decreto istitutivo della VQR e della legge 240/2010 assegna alla VQR una funzione esplicita di esercizio finalizzato al riparto delle risorse finanziarie tra le Università, assumendo implicitamente che meccanismi di premi/punizioni in base al "merito" scientifico siano strumenti efficaci di miglioramento del "Sistema universitario", senza entrare nel merito delle sue funzioni, delle priorità politiche da seguire, del funzionamento effettivo di tali. meccanismi.*

Si sottolineava che questa funzione attribuita alla valutazione avesse implicazioni rilevanti e presupponesse una serie di scelte preliminari. La prima implicazione è insita nel concetto stesso di "miglioramento" – che richiede l'esercizio di un giudizio di desiderabilità di certi risultati piuttosto che altri - presupponendo *che una valutazione orientata al miglioramento di un "Sistema Universitario" non sia possibile senza che si chiariscano profili e obiettivi del "Sistema universitario" che si vuole realizzare. Senza esplicitare quale Università vogliamo, non è possibile alcun responsabile esercizio di valutazione di quella che abbiamo, né è possibile distinguere tra risultati voluti e non voluti (perversi?) della valutazione e dei meccanismi che innesca nei comportamenti dei valutati.*

Si concludeva che: *La parabola della VQR si è così consumata nella ormai nota vicenda che ne ha "curvato" la funzione da esercizio di valutazione costruito sulle strutture e la qualità della loro produzione in "macchina" il cui prodotto finito sono alcuni parametri con cui si "pesano" le persone dei ricercatori/professori. Mentre nella proposta originaria alla VQR è stata assegnata la funzione di valutare le*

*strutture, ad essa si stanno di fatto attribuendo funzioni che finiscono per assumere come oggetto di valutazione i ricercatori/professori individualmente considerati – e non solo per iniziativa di singoli Atenei e Dipartimenti che arrivano a usi che potrebbero essere passibili di sanzione, ma in forma istituzionale, come ad esempio nelle valutazioni dei collegi docenti dei dottorati a fini dell'accREDITamento dei corsi stessi.*

In forza degli aspetti sopra evidenziati si sottolineava, infine *quanto possa cambiare il significato di un indicatore, se usato per valutare la struttura o l'individuo, il processo o il prodotto.*

Consapevole, tuttavia della centralità della valutazione quale meccanismo di regolazione e promozione della qualità della ricerca, oltre che delle strutture, l' AIS ha ritenuto di dare il suo contributo specifico ed operativo sugli indicatori di qualità utilizzati nella passata VQR, in particolare Rilevanza/impatto, Originalità/innovazione, Internazionalizzazione. Indicatori di cui erano stati sottolineati: la genericità delle dimensioni implicate, che spesso reciprocamente si sovrappongono; la non unicità (che si vuole misurare?), l'ampio margine di soggettività nell'interpretazione da parte dei valutatori.

Erano dunque state prodotte/proposte tre scale di valutazione (una per ogni criterio di valutazione), di cui si dava conto sia dal punto di vista metodologico che empirico (si veda <http://tinyurl.com/odstzyg> )

## **La valutazione dei prodotti di ricerca VQR 2011-2014:**

### **proposte AIS sui criteri di valutazione**

Nel bando provvisorio i criteri per esprimere il giudizio di qualità sono sempre tre, in particolare a) originalità, b) rigore metodologico; c) impatto attestato o potenziale nella comunità scientifica internazionale.

*1. elementi positivi:* i primi due criteri sono stati etichettati con una sola parola ciascuno, riducendo l'ambiguità dovuta al fatto che i due termini prima separati da "/" potessero essere variamente interpretati dai revisori;

*2. elementi critici:*

- mancanza di una scala di valutazione per i due criteri a) e b), che nella loro formulazione "colloquiale" non sono indicatori: si ritiene che a ciascun prodotto si debba associare alla fine un punteggio sintetico motivato e specificato dal "profilo" di qualità che risulta da indicatori di originalità e rigore metodologico ritenuti pertinenti ed esplicitati da ciascun *referee*, su ciascuna delle due dimensioni, dando così conto di quel punteggio. Si ritiene infatti – ed è bene ribadirlo - che i *reviewers* debbano dare conto dei punteggi assegnati ai prodotti riferendosi *esplicitamente* agli indicatori

pertinenti al prodotto valutato, che giustificano quei punteggi dando loro un significato chiaro e univoco;

- il terzo indicatore si riferisce all'impatto attestato o potenziale nella comunità scientifica internazionale: in termini di valutazione della qualità della ricerca l'impatto o è reale e misurabile (ad es. anche mediante qualche indicatore citazionale) o è una personale e soggettiva valutazione del *referee*, che può più o meno intercettare i desideri o la realtà del "produttore del prodotto". Cosa vuole dire che è potenziale? Lo è perché scritto nella lingua veicolare più diffusa nella comunità scientifica? perché pubblicato all'estero, anche se mai citato e sconosciuto? Come si valuta l'impatto negativo, vale a dire il caso in cui un prodotto sia abbondantemente citato nella comunità scientifica come esempio di ricerca mal condotta?;

- il criterio dell'internazionalizzazione, utilizzato nella precedente VQR, è stato in realtà riassorbito nell'impatto: anche in questo caso, tuttavia, manca una scala di valutazione tesa a mettere in evidenza quanto meno livelli diversi di internazionalizzazione e/o di impatto internazionale. Dando il suo punteggio, il valutatore di fatto si assume la responsabilità di qualificarsi come colui che possiede le conoscenze adeguate per decidere che un prodotto è eccellente o elevato: rientra cioè nel 10% dei prodotti eccellenti della comunità scientifica di riferimento o nel 20% di quelli elevati. Si potrebbe lavorare ex post sulle distribuzioni ottenute, standardizzando i punteggi per *referee* e per area;

- in modo un poco surrettizio viene poi introdotto un quarto criterio, apparentemente la sintesi dei tre, che riguarda il posizionamento del prodotto esaminato all'interno di una scala di qualità della produzione scientifica, esplicitamente precisata come internazionale nel caso delle due prime classi (eccellenza e qualità elevata) e genericamente indicata come "distribuzione della produzione scientifica dell'area a cui appartiene" nelle altre tre. Al riguardo due notazioni; in primo luogo appare arduo ritenere che il valutatore sia in grado di considerare l'intera produzione scientifica internazionale dell'area (non del campo particolare in cui opera, ma dell'area) strutturata secondo una scala articolata in decili; in secondo luogo non è chiaro se il riferimento sia sempre la produzione scientifica internazionale o meno (in questo secondo caso c'è evidentemente un "buco" relativo ai prodotti che si collocano al di sotto del primo 30% della produzione internazionale ma al di sopra dell'ultimo 70% della produzione nazionale, che si ipotizza non sia distribuita allo stesso modo). Complessivamente, il modo in cui è costruito l'indicatore sintetico sembra presupporre che tra i tre indicatori esista una relazione di covariazione praticamente perfetta. A) Non si tiene conto del fatto che un lavoro può replicare in altri contesti e altre condizioni un disegno di ricerca precedente (non essere quindi "originale", in quanto introduce un "nuovo modo di pensare": ma veramente l'ANVUR ritiene che il 10% delle pubblicazioni di ogni area introducano "un nuovo modo di pensare"? ) e pur tuttavia essere condotto con estremo rigore metodologico ed avere risultati significativi in termini scientifici. B) Si postula una relazione biunivoca che lega, da una parte, eccellenza e impatto internazionale (sia pur anche potenziale – e sin qui si potrebbe anche accettare l'idea che un prodotto eccellente prima o poi, direttamente o indirettamente, possa avere risonanza sulla comunità disciplinare ampiamente intesa – e, dall'altra – e qui emerge il pregiudizio negativo contro le comunità disciplinari nazionali – tra impatto nazionale e livelli appena accettabili di originalità e rigore

metodologico. Non si capisce perché un lavoro metodologicamente poco rigoroso debba avere impatto nazionale: si spera non ne abbia, se non per le critiche che si attirerà! (e qui si torna alla domanda sull'impatto negativo).

Non viene indicato in alcun modo l'obbligo per i *referee* di esplicitare le motivazioni dei giudizi assegnati e la possibilità di portarli a conoscenza, in forma anonima ovviamente, al soggetto valutato (e indirettamente alla comunità scientifica di riferimento);

- i tre indicatori hanno di fatto pesi diversi: sempre nel paragrafo 2.6.1. del Bando, nell'attribuzione del livello di qualità della pubblicazione l'impatto reale o presunto sulla comunità scientifica internazionale è l'unico criterio per accedere all'Eccellenza e al Livello elevato di qualità. Nulla osta alla scelta di dare pesi diversi - e motivati - ai criteri utilizzati, tenuto conto di quelle che sono le finalità del processi di valutazione, ma il meccanismo di attribuzione dei punteggi, così come formulato nel bando, in realtà valuta i prodotti utilizzando due scale, una delle quali, tuttavia, taglia drasticamente tra il 30% dei prodotti eccellenti o elevati solo in quanto 'internazionalizzati (realmente o presumibilmente!!!) e il rimanente 70% valutato per *default* mediocri o limitati. Con buona pace delle belle menti che purtroppo usano la lingua di Dante e a vantaggio di coloro che dicono banalità nella lingua veicolare per antonomasia della comunità scientifica.

Da ultimo, i punteggi assegnati alle varie classi cui sono riferibili i prodotti sembrano riprodurre l'usuale approccio secondo il quale il valore di un "prodotto" è sempre attribuito secondo una logica gerarchizzante, in forza della quale non vale l'eccellenza o la bontà di un prodotto in sé, ma il posizionamento in una fascia il più possibile elevata di una distribuzione. In altri termini, non possono esistere per definizione prodotti eccellenti in quota superiore al primo 10%, anche se il primo 20% fosse in ipotesi tutto di ex-aequo ...

- Come nella precedente VQR, la irrisoluzione del contenuto degli indicatori lascia ai GEV margini decisionali sui loro contenuti e sulle metriche stesse, che rendono non paragonabili tra loro i risultati, nonostante il DM imponga (?) all'ANVUR la produzione di *ranking*.

Ben strano appare quindi il compito c) del punto 3.3. assegnato ai GEV relativo alla illustrazione dei "punti di forza e di debolezza dell'Area a livello nazionale....." Che avrebbe un senso (forse, e non in relazione ai "suggerimenti") solo comparativamente rispetto alle altre aree, quindi NON compiuto dal singolo GEV.

- I "tre profili di qualità" (par. 2.6 e Appendice del bando) manifestano un problema grave.

Se si sta misurando la qualità della ricerca di un'Area o un settore il punto c) è eterogeneo (poiché riguarda capacità organizzative della struttura). Se si sta misurando la qualità della ricerca di un Ateneo o di un Dipartimento il punto c) deve giocare in modo del tutto diverso e autonomo: ad esempio come denominatore, oppure come un

indicatore autonomo (non sommato in percentuale agli altri) col quale “tarare” la combinazione di quanto in a) e b).

- Inoltre l'indicatore a) – ed è ipotizzabile che ciò si verifichi anche per il b) – soffre degli effetti di concentrazione dei ricercatori già evidenziati dalla precedente VQR e risolti dal metodo Poggi/Crui/Anvur, a cui qui non si fa minimamente riferimento.

### **La valutazione dei prodotti di ricerca VQR 2011-2014:**

#### **proposte AIS sulle procedure di lavoro e sulle pratiche**

Alcune raccomandazioni :

- nei settore non bibliometrici e non solo, i prodotti della ricerca devono essere letti dai *referee* che devono esprimere delle valutazioni sulla base di criteri chiari e condivisi nella più ampia comunità scientifica; l'attribuzione di punteggi deve essere motivata e non arbitraria. La motivazione deve sempre riferirsi ai criteri dati e non ad elementi extra non contemplati; il valore deve essere comparativo ma non necessariamente gerarchizzato in scale. Due prodotti possono essere eccellenti senza necessità di doverli collocare nel primo 10% di una distribuzione che sarebbe interessante chiedere sia descritta dal *referee* con un certo dettaglio, magari individuando i *milestone* di ogni decile;...

- il procedimento di attribuzione dei punteggi deve essere chiaro e il punteggio deve poter essere 'ricostruito': quando ci sono troppi passaggi e/o trasformazioni dal primo valore all'ultimo, si perde molto in chiarezza:

- i *referee* devono poter fare riferimento a criteri chiari e precisi, espressi con indicatori validi ed attendibili per evitare, in caso contrario, interpretazioni soggettive;

- deve essere data la possibilità a tutti i ricercatori e soprattutto a chi ha ottenuto un punteggio basso, di chiedere ai *referee* anonimi di specificare meglio e più analiticamente il perché della valutazione ;

- non si può inferire la qualità di un prodotto dal contenitore e viceversa. La passata VQR ha dimostrato in maniera chiara che la "correlazione" tra qualità del contenitore e qualità del contenuto non è stata molto elevata, dal momento che si sono rilevati prodotti di eccellenza in contenitori di prestigio basso, e prodotti mediocri in contenitori di alto prestigio;

- il tentativo di spingere verso pratiche di cui al punto precedente (valutare il contenuto solo in base al contenitore) deve essere arginato e osteggiato, per non creare oligopoli che selezionano l'accesso ai canali di pubblicazione, solitamente più sulla base di esigenze di mercato che non per un interesse per la qualità della ricerca.

**La valutazione dei prodotti di ricerca VQR 2011-2014:  
proposte AIS sull'uso della VQR per valutare le istituzioni**

L'uso dei risultati della VQR per valutare le istituzioni (ad es. Dipartimenti) , deve assolutamente tener conto anche di indicatori riferiti alle risorse disponibili (sia pubbliche che private) cui i singoli ricercatori hanno accesso - e deve quindi dare luogo a misure di valore aggiunto piuttosto che non a gerarchie assolute.

L'uso implicitamente punitivo delle valutazioni negative deve essere contrastato con la previsione di misure atte a favorire il miglioramento delle aree e sedi meno performanti, anche per controbilanciare l'impostazione della procedura valutativa in termini di darwinismo accademico che trasuda dal testo.

Infine, si propone che all'articolo 2, comma 4, si invertano i pesi assegnati ai profili di cui ai punti a) e b), dal momento che la distribuzione percentuale dei prodotti per area scientifica nelle cinque classi risente anche in modo rilevante delle diverse metriche adottate dai GEV, come messo in rilievo da molti commenti successivi alla VQR 2004-2010, mentre il posizionamento relativizzato come nel punto b) non ne risente.