

anno VI n. speciale

Direttore Responsabile

Daniela Cocchi

Direttore Scientifico

Marcello Chiodi

Co-direttori

Enrico Del Colle

Alessandra Righi

Segreteria di Redazione

Antonella Plaia

Dipartimento di Scienze
Statistiche e Matematiche "S.
Vianelli"

Università degli Studi di Palermo

Viale delle Scienze ed. 13 –

90128 Palermo

e-mail: plaia@unipa.it

on-line dal 9 maggio 2008

Reg. Trib. Di Roma n° 504/2002
2/9/02

Giornata su:

**La valutazione della ricerca in statistica:
esperienze, prospettive e confronti con altre
discipline**

organizzata dalla Società Italiana di Statistica

Roma, 21 gennaio 2008

Aula Gini, Facoltà di Scienze Statistiche,
Sapienza Università di Roma

In questo numero

<i>Daniela Cocchi:</i> La Società Italiana di Statistica e la valutazione della ricerca scientifica	2
<i>Fiorella Kostoris:</i> L'esperienza del CIVR e le prospettive dell'ANVUR nella valutazione della ricerca in Italia	5
<i>Alberto Zuliani:</i> La valutazione della ricerca: alcune riflessioni generali	15
<i>Alessandro Figà Talamanca:</i> Cosa possiamo imparare dalla VTR 2001-2003?	20
<i>Alessandro Carretta:</i> La ricerca scientifica in Economia Aziendale: primi risultati del processo di classificazione delle riviste scientifiche	30
<i>Pietro Reichlin:</i> La valutazione della ricerca in economia	33
<i>Benito Vittorio Frosini:</i> Valutazione della ricerca e valutazione delle riviste scientifiche in ambito statistico	39
<i>Franco Peracchi:</i> Valutazione della ricerca e reclutamento: quale ruolo per la SIS?	48
<i>Ugo Trivellato:</i> La valutazione della ricerca in statistica	50

La Società Italiana di Statistica e la valutazione della ricerca scientifica

Daniela Cocchi

Presidente della Società Italiana di Statistica

daniela.cocchi@unibo.it

Sommario

La Società Italiana di Statistica ha organizzato, il 21 gennaio 2008, un'iniziativa sul tema "La valutazione della ricerca in statistica: esperienze, prospettive e confronti con altre discipline", ospitata dalla Facoltà di Scienze Statistiche di Sapienza, Università di Roma. Questo numero monografico di Statistica & Società riporta tutti gli interventi dei relatori e due dei contributi alla tavola rotonda.

1 La valutazione della ricerca scientifica

Per la seconda volta, a distanza di oltre due anni e mezzo dalla prima occasione, la Società Italiana di Statistica ha organizzato, il 21 gennaio 2008, un'iniziativa riguardante la valutazione della ricerca sul tema "La valutazione della ricerca in statistica: esperienze, prospettive e confronti con altre discipline", ospitata dalla Facoltà di Scienze Statistiche della Sapienza. Il programma della giornata è il risultato dell'impegno e della competenza di Gianluca Cubadda e Nicola Torelli.

Ripensando al precedente workshop organizzato dalla Società Italiana di Statistica il 20 giugno 2005 su "Valutazione e criteri di finanziamento della ricerca: il punto di vista degli statistici", la situazione si è molto evoluta. In quel momento il dibattito sulla valutazione era agli inizi e si manifestava esprimendo diversi timori circa la sua corretta utilizzazione, verosimilmente dovuti alla mancanza di tradizione ed esperienza.

Dopo quella giornata, sono accaduti fatti importanti e significativi. Si è concluso l'esercizio CIVR VTR 2001-2003 (Comitato di Indirizzo per la Valutazione della Ricerca; Valutazione Triennale Ricerca). In seguito alla pubblicazione dei risultati, in diverse società scientifiche, tra cui la SIS, è continuata la discussione, anche avviando proposte per la graduazione delle riviste scientifiche di settore. Il CIVR e il CNVSU

(Comitato Nazionale per la Valutazione del Sistema Universitario) sono stati riconfermati, mentre da poche settimane è stata costituita l'ANVUR (Agenzia Nazionale di Valutazione del sistema Universitario e della Ricerca). Dovrebbe ripartire un esercizio di valutazione CIVR focalizzato sul triennio successivo a quello esaminato.

2 La valutazione della ricerca in Statistica

Nel 2005 il primo esercizio di valutazione triennale della ricerca da parte del CIVR era in pieno svolgimento. Non se ne conoscevano ancora gli esiti e nella nostra comunità scientifica serpeggiava il timore della scarsa visibilità della ricerca in statistica, unito alla generale diffidenza riguardo all'efficacia delle direttive ministeriali a proposito della valutazione della ricerca.

Si percepiva anche la preoccupazione riguardo alla classificazione delle materie che, nell'ambito dell'area 13 (economia e statistica, coincidente con il settore scientifico-disciplinare del CUN), attribuiva i risultati delle nostre ricerche a un sottogruppo denominato "matematica". A questo riguardo si è verificato un problema di definizione in quanto il termine "statistica", secondo gli standard ISI-CRUI, non era nemmeno incluso nelle parole chiave. C'era anche il timore della dispersione dei risultati delle ricerche in statistica al di fuori dell'area 13, alla quale facciamo riferimento riguardo ai settori scientifico

disciplinari: non eravamo sicuri che tutti i prodotti di ambito statistico fossero stati inseriti nell'area 13.

Nonostante queste difficoltà, i prodotti delle discipline statistiche hanno costituito oltre il 20% del totale dell'area. Non solo, ma l'esito della valutazione CIVR è risultato molto positivo per il sottogruppo degli statistici, che hanno ottenuto una valutazione più alta della media del gruppo.

L'area 13 si colloca in una posizione intermedia tra le discipline umanistiche e quelle scientifiche. Noi statistici siamo associati all'area 13, assieme ad aziendalisti, economisti, matematici finanziari, merceologi e storici dell'economia. La motivazione è di tipo storico e presenta come vantaggio l'aggancio con le discipline sostantive di riferimento più immediato. C'è tuttavia parecchia ricerca statistica anche al di fuori di quest'area e l'aumento delle sinergie è davvero auspicabile. La ridefinizione dei settori scientifico disciplinari, che potrebbe aiutare in questa direzione, è un argomento molto attuale, difficile e sofferto. I nostri rappresentanti al CUN si stanno adoperando per una riorganizzazione.

Nei mesi più recenti, la conclusione dell'esercizio CIVR e la presentazione dei suoi risultati hanno promosso discussioni approfondite, si sono sviluppati i confronti con analoghe esperienze internazionali, alcune associazioni delle varie discipline hanno organizzato iniziative di commento ai risultati (Società Italiana degli Storici dell'Economia, nel marzo 2007 a Roma, Società degli Economisti, ad Ancona nell'aprile dello stesso anno, Accademia Italiana degli Aziendalisti a Bologna in giugno).

Nelle società scientifiche sono state avviate anche iniziative per la graduazione delle riviste, almeno nell'AIDEA e nella SIS. Per la statistica questa iniziativa non è facile: la disciplina ha legami sia con quelle sociali e umanistiche sia con quelle scientifiche, ci sono differenze importanti a seconda che il canale di pubblicazione segua l'ambito delle riviste con *impact factor* o si collochi diversamente, anche se c'è consenso sul fatto che la considerazione esclusiva dell'*impact factor* sia fuorviante.

In seguito all'incontro del 2005 si è costituita, nell'ambito della SIS, una commissione, presieduta da Vittorio Frosini, con l'obiettivo di fornire indicazioni e linee guida per la graduazione delle riviste nazionali e internazionali, in cui gli

statistici pubblicano più frequentemente. L'iniziativa è importante anche perché i singoli atenei stanno sviluppando sistemi di valutazione autonomi al loro interno e una stessa pubblicazione può trovarsi in posizioni relativamente molto diverse nei sistemi valutativi locali. Nella SIS si stanno discutendo le modalità per proporre questa base informativa, punto di riferimento per la valutazione della ricerca scientifica svolta in ambito statistico.

3 Il ruolo della Statistica nella valutazione della ricerca

L'idea che i risultati della valutazione debbano essere utilizzati da parte degli organi di governo degli atenei si sta affermando, e si accetta che possano costituire uno degli elementi impiegati per finanziare la ricerca futura. Si sta riconoscendo che i criteri da definire possano essere flessibili e diversi, da usare in alternativa o in parallelo, a seconda dello scopo.

A questo proposito, la nostra disciplina ha strumenti propri per intervenire utilmente nella costruzione dei modelli per la valutazione della ricerca. A livello locale, però, il nostro ruolo non viene pienamente riconosciuto, in quanto non sempre siamo inseriti negli osservatori della ricerca e nei nuclei di valutazione.

Anche avvalendosi del contributo degli statistici è comunque difficile, all'interno di un ateneo, comparare i risultati delle valutazioni tra aree, ciascuna delle quali è stata effettuata separatamente, in base quindi a criteri che possono essere anche abbastanza diversi.

La graduazione dei risultati di una stessa disciplina tra le diverse sedi è più agevole, anche se, proprio nell'ambito dell'esercizio di valutazione CIVR, si è riscontrato l'inconveniente del diverso numero di prodotti selezionati per disciplina, che ha portato a non rendere sempre comparabili direttamente i dati tra le diverse sedi. La selezione delle migliori pubblicazioni è un criterio ormai condiviso, da associare comunque al numero dei potenziali autori. A questo riguardo si aprono due vie principali: l'utilizzo dello stesso metodo, per tutti gli atenei, come la proporzionalità del numero dei ricercatori della disciplina sul totale dell'area, o la libertà per gli istituti di ricerca di farsi valutare

soprattutto sulle linee di ricerca ritenute migliori.

Un conto è, poi, valutare l'eccellenza ed altro è valutare il livello medio della ricerca. Gli strumenti per valutare questo secondo aspetto devono essere approntati in modo diverso e si deve discutere se si tratta di un aspetto da approfondire a livello nazionale o a livello locale.

Il programma del nostro incontro ha previsto i contributi di studiosi di varie discipline: oltre allo statistico Alberto Zuliani, hanno accolto il nostro invito il matematico Alessandro Figà Talamanca, l'aziendalista Alessandro Carretta e l'economista Pietro Reichlin. L'intervento di Vittorio Frosini contiene anche la sintesi dei lavori della commissione per la classificazione delle riviste di ambito statistico istituita dalla SIS. Anche questa volta, inoltre, abbiamo il piacere e l'onore della

presenza di un componente CIVR, Fiorella Kostoris.

I nostri lavori si sono conclusi con una tavola rotonda su "La valutazione della ricerca in statistica e la statistica per la valutazione della ricerca" che ha toccato le due dimensioni del coinvolgimento della statistica riguardo alla valutazione. I partecipanti sono stati il presidente del CNVSU Luigi Biggeri, il demografo Francesco Billari, lo statistico economico Ugo Trivellato, che ha una lunga esperienza nei nuclei di valutazione, e il presidente del panel dell'area 13 nell'esercizio VTR del CIVR Franco Peracchi.

Questo numero monografico di Statistica & Società riporta tutti gli interventi dei relatori e due dei contributi alla tavola rotonda. Confidiamo che esso possa diventare uno dei riferimenti per il dibattito futuro.

L'esperienza del CIVR e le prospettive dell'ANVUR nella valutazione della ricerca in Italia

Fiorella Kostoris

Sapienza Università di Roma

fiorella.kostoris@tin.it

Sommario

Il processo di valutazione della ricerca italiana, iniziato dal CIVR, non deve arrestarsi. E' necessario che negli anni a venire prima il CIVR e poi l'ANVUR operino, riuscendo a combinare con criterio meritocratico e trasparente, dunque insieme equo ed efficiente, l'autonomia con la responsabilità delle nostre istituzioni scientifiche. In futuro è necessario che il confronto ordinamentale comporti conseguenze osservabili in termini di trasferimenti di risorse pubbliche per la ricerca; è consigliabile, inoltre, estendere la valutazione a tutte le istituzioni di ricerca destinatarie di fondi pubblici e non solo a quelle vigilate dal MUR, cui fin qui ci si è circoscritti.

1 Introduzione

Voglio ringraziare la Società Italiana di Statistica (SIS) per l'invito e per l'attenzione particolare che essa dedica al tema della valutazione; non si tratta di una formula di rito: non è certo un caso che l'attuale Presidente dell'ISTAT sia anche a capo dell'altro Comitato di Valutazione che, insieme al CIVR, opera nel MUR, e che il passato Presidente sia un noto esperto del campo. Del resto, io stessa sono già venuta in questa sede a parlarne il 20 giugno 2005 (in occasione dell'incontro su "Valutazione e criteri di finanziamento della ricerca: il punto di vista degli statistici"), in qualità di membro del *Board* del CIVR, e debbo concordare con la Presidente Cocchi nell'affermare che, rispetto ad allora, ci sono oggi per la SIS almeno due motivi aggiuntivi di soddisfazione. Innanzitutto perché, come è noto, proprio i risultati dell'esercizio CIVR sul primo triennio 2001-2003 indicano che, dei 4 settori che individuano l'Area 13 (detta di "economia e statistica") e cioè, nella dicitura ISI, l'*economics* propriamente detta, il *management*, la *mathematics* e la *sociology and anthropology*, quello della *mathematics* (dove prevalgono la statistica e l'econometria) appare al *top* per varie ragioni: per il tipo di produzione preferito (articoli), per la lingua adottata (inglese), per l'*impact factor* (IF) delle riviste utilizzate nella pubblicazione, per il

largo grado di consenso emergente in questo sottogruppo all'interno del Panel 13, e non ultimo per la quota di lavori eccellenti, pari al 28%, in linea con il 30% trovato nella totalità delle discipline scientifiche in Italia, ma molto superiore al meno del 20% di eccellenze nell'insieme dell'Area 13. Il vostro orgoglio dovrebbe, inoltre, crescere in ragione di un ulteriore elemento: nell'esercizio CIVR sul triennio 2004-2006, che sta cominciando ora, avete anche vinto una battaglia che, come ha ben chiarito la Presidente Salinetti, non è solo nominalistica, sicché il sottoinsieme che si chiamava *mathematics* diventerà *mathematics and statistics*. Perfino l'ISI Thomson, con i suoi 800 dipendenti disseminati in 7 Paesi del mondo, operanti in 35 lingue su 8.600 riviste (Bernini e Nepi, 2005) sembra pronto a piegarsi alla perseveranza della vostra Presidente! Nel poco spazio a mia disposizione non vorrei affrontare il tema, peraltro interessante, del *perché valutare* nell'ambito della ricerca: l'ho in parte già toccato con voi due anni e mezzo fa e ne ho scritto anche altrove (Kostoris Padoa Schioppa, 2006a, 2006b, 2006c). In Italia è ormai acquisita la consapevolezza che bisogna farlo per una molteplicità di considerazioni: innanzitutto per confrontare con parametri omogenei la nostra produzione scientifica dentro e fuori dai confini nazionali e per combinare con criteri meritocratici (Tognon, 2006), e dunque in termini insieme efficienti ed equi (Kostoris, 2007a), l'autonomia

con l'*accountability* delle istituzioni di ricerca destinatarie di fondi pubblici. Del resto, in una certa misura, a questo siamo impegnati anche dagli accordi europei, a cominciare da quelli dell'Agenda di Barcellona (Consiglio delle Comunità Europee, 2002; Commissione delle Comunità Europee, 1996 e 2002) e della solenne Dichiarazione di Bergen (Conferenza dei Ministri Europei Responsabili dell'Istruzione Superiore, 2005).

In aggiunta, non vorrei nemmeno approfondire sul puro terreno dei principi la questione, certo rilevante, del *come valutare*, con quali strumenti, con quale *governance*. Anche questi sono argomenti appassionanti, ma esiste in proposito una vasta letteratura, cui hanno contribuito alcune delle più autorevoli personalità della SIS. Vorrei, invece, circoscrivere il mio compito odierno alla trattazione di problemi più pragmatici, in stile più *british*, analizzando dapprima (nel paragrafo 2) i passi effettivamente compiuti dal CIVR in Italia nella valutazione *ex post* della ricerca con riguardo al periodo 2001-2003, verificandone poi (nel paragrafo 3) il grado di successo ed i principali limiti, esaminando infine i presumibili cambiamenti che si registreranno non tanto con riferimento al triennio 2004-2006 sotto esame nel nuovo esercizio del CIVR, quanto successivamente con la creazione dell'ANVUR, l'Agenzia di Valutazione dell'Università e della Ricerca (nel paragrafo 4), in cui confluirà il CIVR insieme al CNVSU. Alcune conclusioni sono riportate nel paragrafo 5.

2 L'esperienza del CIVR nel primo esercizio triennale 2001-2003

Non ho bisogno di dilungarmi troppo sul primo punto, perché molti hanno già letto la Relazione dei Panel pubblicata nel marzo 2006 (CIVR, 2006a, 2006b) e quella Finale del nostro *Board* (CIVR, 2006c) e perché la storia passata interessa solo in quanto *magistra vitae* per il futuro, secondo un meccanismo di *learning by doing*. Ricordo che, con riguardo all'intervallo 2001-2003, sono state valutate 102 Strutture, fra cui la totalità delle 77 Università statali o legalmente riconosciute, 12 Enti Pubblici di Ricerca, 13 Istituzioni Private. È stata sottoposta a giudizio la ricerca di ognuna delle 14 discipline scientifiche identificate dal CUN, oltre che di alcune Aree Speciali

interdisciplinari. Tutta la procedura è costata 3,5 milioni di euro, ha coinvolto direttamente o indirettamente 64.000 ricercatori, quasi 18.500 prodotti soggetti a valutazione, 150 panelisti, di cui $\frac{1}{4}$ stranieri, 6.000 esperti fra italiani (residenti nel Bel Paese o all'estero) e non, richiesti di redigere *referee reports*.

Tale processo è stato caratterizzato da trasparenza estrema, come ha evidenziato ad esempio Toniolo (2006), e si è articolato in 3 fasi.

La prima può definirsi di autovalutazione, in quanto le Strutture hanno dovuto scegliere i loro migliori prodotti, selezionandone un numero pari alla metà delle teste equivalenti di ricercatori (queste, chiamate ETP, a loro volta uguali al 50% di coloro che operano nelle Università, a causa del tempo dedicato anche alla didattica, ed al 100% di coloro che lavorano negli Enti di Ricerca). Si è così perduto molto, forse troppo tempo, in parte per l'inesperienza delle Strutture, in parte a causa della nostra richiesta di documenti cartacei, talora difficilmente reperibili. Nel prossimo CIVR tutte le procedure saranno velocizzate, anche perché basterà l'invio dei prodotti autoselezionati in forma telematica.

Invece è stata relativamente rapida la seconda fase, in quanto i Panel hanno completato le loro funzioni in circa un anno, concludendo la loro attività con la pubblicazione della loro Relazione. Come è noto, durante quel lasso temporale, ogni prodotto è stato valutato almeno da 2 *referees* indipendenti, il cui giudizio è stato poi portato all'approvazione dell'insieme dei membri dei sub-Panel e successivamente del corrispondente Panel. In certi casi, un terzo *referee* si è reso necessario, ad esempio quando i punti di vista dei primi due differivano troppo sensibilmente o quando essi non apparivano per nulla indipendenti l'uno dall'altro. Si è valutata secondo criteri meritocratici la qualità, con la rilevanza della ricerca, la sua originalità, il livello della sua internazionalizzazione, quale *proxy* dell'apprezzamento di cui godiamo nel resto del mondo e del nostro grado di competitività all'estero. In tale ultima fattispecie potevano dare un contributo anche gli indici bibliometrici, come l'*impact factor* della rivista utilizzata per la pubblicazione o il *quotation index* dell'autore su quel particolare *paper*. Ma il metodo precipuo del CIVR è stato e resta la *peer review*, sia perché esistono prodotti scientifici senza indici bibliometrici, ad esempio libri o capitoli o *papers* su riviste senza IF, sia perché mancano quasi del tutto indici bibliometrici

per alcune discipline, ad esempio quelle storico-filosofico-letterario-giuridiche. Il fatto che, nella nostra materia di “economia e statistica”, e solo in essa fra le 14 del CUN, sia emerso un acceso dibattito, con mozioni di maggioranza e minoranza esplicitate nella Relazione del Panel 13, circa i relativi pregi e difetti della *peer review* e dei suoi (imperfetti) sostituti o (ottimi) complementi, quali sono gli indici bibliometrici, dipende, a mio avviso, da due circostanze caratterizzanti nel merito e nel metodo l’Area 13: in questa, da un lato, sussiste una notevolissima varietà di contenuti, testimoniata dalla presenza per essa nel sistema universitario anglosassone non di 1 ma di ben 5 fra diversi *Departments* e *Schools*, i primi di natura più teorica (*Economics*, *Sociology*, *Anthropology*), le seconde a finalità più professionale (*Business*, *Management*), e, dall’altro lato, esiste nel nostro informe (o deforme?) italico aggregato di “Economia e Commercio”, una impostazione in cui l’*economics*, per alcuni, è legata alla storia, alla filosofia, al diritto, a carattere, dunque, qualitativo-istituzionale, mentre, per altri, è più vicina alle scienze *hard* che a quelle sociali, magari alla biologia, alla fisica, all’ingegneria, in un approccio, quindi, di tipo quantitativo. La prima ottica è quella di chi considera distorta la lente offerta dagli indici bibliometrici, la seconda è quella di chi, invece, se ne serve come di un utile strumento amplificante. Sia gli esperti nelle discipline sociali che non ne fanno mai uso, sia quelli operanti nelle scienze naturali che non ne fanno mai a meno risultano, in questo, coesi al loro interno. Gli economisti italiani sono, al contrario, a metà del guado.

Infine, la terza fase della valutazione del CIVR è stata quella in cui siamo intervenuti noi del *Board*, con la stesura della Relazione Finale. La Relazione dei Panel illustrava Area per Area il *ranking* delle varie Strutture. Noi abbiamo fornito il confronto ordinale delle Strutture, tenendo conto dell’insieme delle Aree in esse attive, e lo abbiamo fatto non solo con riferimento alla qualità della ricerca esaminata dai Panel (cui abbiamo dato, nella media ponderata, un peso di 4/9, come meglio risulta dalla Relazione Finale del Presidente del CIVR (Cuccurullo, 2007)), bensì anche con riguardo al grado di proprietà dei prodotti eccellenti (peso 2/9), alla tendenza alla mobilità internazionale dei ricercatori in entrata o in uscita (peso 1/9), alla propensione verso la formazione di dottorandi, assegnisti, ecc. (peso

0,5/9), alla capacità di valorizzazione nell’attrarre risorse esterne (peso 1/9) o nell’impegnarne di proprie non vincolate (peso 0,5/9).

3 Successi e limiti del primo esercizio triennale del CIVR e possibili cambiamenti nel secondo triennio

Passando a discutere del secondo punto, vi è un generale consenso nel sostenere che il CIVR ha finora ottenuto un certo successo: per la prima volta nella nostra storia siamo stati in grado di procedere ad una valutazione *ex post* generalizzata della ricerca italiana, che molti scettici nostrani ritenevano *ex ante* impossibile. L’*achievement* maggiore dell’esercizio del CIVR con riguardo al triennio 2001-2003 consiste proprio nell’aver mostrato che valutare si può anche in Italia, con rigore e serietà: nessuno ha sollevato lamentele su ingiustizie commesse, nemmeno chi risultava agli ultimi posti nei *ranking* pubblicati. Credo che questo sia dovuto, oltre che alla diffusa consapevolezza della necessità della valutazione con criteri trasparenti e condivisi ed alla maturità delle Strutture di ricerca del nostro Paese, alla professionalità e onestà dei valutatori del CIVR. Queste ultime sono state evidenziate dalla frequente concordanza piena fra i 2 *referee reports* indipendenti (nel 43-48% dei casi), mentre i giudizi sono risultati “contigui” con probabilità superiore al 40% e solo nel 9-13% dei casi si è manifestata discordanza totale fra i *referees*; in aggiunta, si è registrata l’unanimità all’interno dei Panel nel 87-96% dei casi (Reale *et al.*, 2006), ed è apparsa una correlazione positiva (pari allo 0,40) fra le valutazioni del CIVR e l’IF delle riviste in cui sono stati pubblicati i *papers* sottoposti a giudizio: Peracchi (2006a, 2006b) rifiuta l’ipotesi che tale fatto sia determinato dall’influenza degli indici bibliometrici sui valutatori, bensì supporta la tesi secondo cui la *peer review*, così come gli indici bibliometrici, sono entrambi generalmente capaci di cogliere abbastanza correttamente il valore di un prodotto scientifico.

Rimane, tuttavia, che molti commentatori hanno illustrato vari limiti del primo esercizio triennale del CIVR (alcuni dei quali in via di eliminazione con riguardo al secondo triennio), ed anch’io vorrei evocarne due ulteriori, che definirei

appartenere alla categoria (più larga) dell'incompletezza.

Infatti, da un lato, non si è proceduto ad una valutazione della ricerca individuale, bensì solo a quella delle Strutture, e quindi gli incentivi all'efficienza e all'equità, tendenzialmente raggiungibili attraverso un'allocazione delle risorse focalizzata sui meritevoli, sono risultati, financo *ex ante*, indeboliti: che sia rilevante arrivare ad un giudizio su ogni singolo ricercatore destinatario di fondi pubblici è opinione del Gruppo 2003 (2003), formato dagli scienziati italiani più quotati all'estero ed anche dell'attuale sottosegretario al MUR (Modica, 2006), ma, in proposito, numerose sono in Italia le obiezioni, spesso di natura autoreferenziale e corporativa (Lissoni, 2007). Proprio perciò temo che nemmeno nel nuovo CIVR tale limite sarà superato. Dall'altro lato, non si è giunti neppure ad un giudizio di tutte le Strutture che ricevono fondi pagati dal contribuente, bensì solo di quelle che vengono vigilate dal MUR, o di quelle che, ad esso estranee, hanno chiesto e pagato per essere vagliate, sapendo di risultare meritevoli. Che questa non sia una *best practice* è attualmente chiarito proprio dalla Relazione di accompagnamento al Regolamento dell'ANVUR (su cui cfr. paragrafo 4), quando essa ricorda che "la valutazione assurge oggi al ruolo di un diritto/dovere delle Università e delle Strutture di Ricerca: il diritto di essere valutate in relazione ai risultati conseguiti con le proprie attività; il dovere di sottoporsi in quanto istituzioni pubbliche o comunque sostenute da **trasferimenti pubblici** al giudizio valutativo" (il grassetto è mio). Così però non è avvenuto nel primo esercizio del CIVR, né presumibilmente accadrà nel secondo. Di conseguenza, per fare un esempio, Daniela Cocchi, se è autoselezionata dalla sua Struttura come creatrice di buoni prodotti scientifici, è sottoposta alla valutazione del CIVR se lavora all'Università, ma se fosse una statistica dell'ISTAT, dell'ISAE o dell'Istituto Superiore di Sanità sfuggirebbe a tale giudizio: il che comporta profonde distorsioni e inefficienti allocazioni delle risorse fra Enti per molti versi simili.

Quanto ai limiti segnalati da altri esperti, ecco quelli che mi paiono i principali: essi sono virgolettati, sebbene non sempre siano ripresi *verbatim*, perché tali commenti spesso derivano da molteplici fonti, anche se, per semplicità, qui sono riferiti ad una sola.

1) "Si sono valutati solo i *papers* migliori e non

quelli medi" (Consiglio Italiano per le Scienze Sociali, 2006). Ciò è vero, ma, a mio giudizio, è voluto perché in questo primo stadio, per un verso, si intendevano conoscere i punti di eccellenza (ed è bene cominciare da quelli, secondo Schiantarelli, 2006), per un altro, si desiderava contenere i prezzi e i tempi, dato che già così ad alcuni i costi sembravano eccessivi (ma i nostri 3,5 milioni di euro sono pochi, come ricorda Jappelli (2006), se confrontati con i 15 dell'equivalente esercizio inglese, il *RAE* (*Research Assessment Exercise*), il quale sottopone a valutazione fino a 4 *papers* per ogni ricercatore su un lasso di 7 anni).

2) "Si è ignorata la diffusione dei prodotti autoselezionati dalle Strutture fra i diversi ricercatori, sicché la variabilità fra essi è risultata ininfluente" (Tamborini, 2006). Questa considerazione è per me condivisibile, perché effettivamente, per dare un esempio, risultavano al CIVR equivalenti due situazioni molto eterogenee, pur a parità di qualità scientifica intrinseca e di grado di proprietà: l'una in cui N prodotti appartenevano tutti a 1 solo ricercatore ed un'altra in cui essi erano il contributo di N ricercatori di quella stessa Area e Struttura. Sarebbe invece opportuno distinguere i 2 casi suddetti e, più in generale, bisognerebbe tener conto, oltre che della qualità media, anche della sua dispersione fra ricercatori, o almeno si dovrebbe misurare il rapporto tra il numero dei ricercatori autoselezionati dalle Strutture per i loro prodotti e il numero dei prodotti (o alternativamente dei ricercatori equivalenti ETP) totali, come fosse un indice di concentrazione: in questa ultima eventualità, riprendendo il nostro esempio, si osserverebbe che, nel primo caso, il rapporto è $1/N$, mentre nel secondo è 1.

3) "Si sono valutate anche situazioni non rappresentative, ad esempio Strutture con Aree che sottoponevano pochi o perfino un solo prodotto" (Rubele, 2007). Commento *well taken*, direbbe il *Board* del CIVR, tant'è che nel secondo esercizio triennale 2004-2006 dovranno essere presentati almeno 3 lavori per Area allo scopo di acquisire un giudizio.

4) "Avrebbero dovuto essere impediti arbitraggi fra Aree, nel senso che la Struttura nel suo insieme doveva autoselezionare un numero di prodotti secondo un rapporto di 1 per ogni 2 ricercatori equivalenti a tempo pieno ($1/2$ ETP), ma le Aree più deboli di fatto hanno ceduto a quelle più forti spazi resi strategicamente liberi a causa dell'insufficienza qualitativa o quantitativa della

loro ricerca” (Arachi-Zanardi, 2006). Indubbiamente questo è talora accaduto, ed in tal modo le Strutture hanno eluso lo spirito se non la lettera delle Linee Guida (CIVR, 2006b). Perciò le regole cambieranno nel prossimo esercizio triennale del CIVR, nel senso che quel quoziente di 1/2 ETP dovrà essere soddisfatto a livello di ogni singola Area e non solo di Struttura. Inoltre, fra gli altri mutamenti previsti nella valutazione del periodo 2004-2006, è da segnalare quello concernente la sostituzione di circa il 50% dei panelisti del vecchio CIVR, al fine di combinare nel nuovo Comitato la continuità con l’innovazione.

5) “Non vi è stata abbastanza partecipazione dei soggetti valutati” (Consiglio Italiano per le Scienze Sociali, 2006). Questo commento lascia perplessi per vari motivi. Perché molti hanno presentato la critica esattamente opposta, sostenendo che non c’è stata abbastanza distanza fra valutatori e valutati (ad es. il Rettore Fabiani lo ha affermato proprio in questa sede, quando, poco meno di 3 anni fa, la SIS ne discusse, tanto che alcuni di noi del *Board* del CIVR decisero allora di mettersi in aspettativa dal ruolo universitario); perché la stragrande maggioranza dei commentatori ha lodato, a mio parere giustamente, il CIVR per aver largamente e trasparentemente coinvolto nel processo di valutazione l’intero mondo dei ricercatori italiani; e perché se anche questo presunto limite esistesse, la sua eliminazione comporterebbe numerosi effetti negativi e solo qualche potenziale beneficio. In particolare sarebbe costoso in termini di tempo e di denaro, ancorché forse non inutile, aggiungere alla fase documentale della valutazione CIVR colloqui e visite nelle Strutture *in loco*, ma con ciò aumenterebbe il rischio di potenziali commistioni tra vigilanti e vigilati, laddove invece dovrebbe sussistere piena separatezza.

6) “Non si è osservata alcuna conseguenza del I esercizio triennale di valutazione della ricerca sui finanziamenti destinati dal Ministero o sull’assegnazione dei posti alle Strutture, nonostante gli impegni presi in tal senso dalla Ministra Moratti al momento in cui l’attività del CIVR fu iniziata” (Guiso, 2006; Jappelli, 2008). Questo è un commento assolutamente corretto, ma non va rivolto al CIVR, bensì al Governo Prodi, che così ha purtroppo voluto che fosse. D’altra parte, va ricordato che perfino il *RAE*, nella sua prima edizione, non comportò implicazioni finanziarie immediate per le Strutture, quasi fosse

“un torneo di calcio estivo” (De Fraja, 2006). Oggi, tuttavia, alla vigilia del VI esercizio inglese (*RAE*, 2008), quasi il 20% delle risorse pubbliche trasferite alle Università della “perfida Albione” è allocato secondo i criteri di giudizio del *RAE* stesso ed emergono enormi differenze tra Strutture valutate a livelli pur contigui (ad esempio, considerando l’Area economica, il Dipartimento di Leicester ottiene un contributo di quasi 447mila sterline annue, mentre quello di Birmingham, pari in dimensione ma valutato al gradino subito sotto, riceve 138mila sterline, cioè meno di un terzo).

7) Alla critica espressa al punto 6) implicitamente risponde il Ministro in carica (Mussi 2007c), sottolineando un altro limite nell’impostazione del CIVR stabilita dal precedente Governo Berlusconi: “pur essendo molto apprezzabile la valutazione della ricerca realizzata dal CIVR, meglio sarebbe se la conseguente allocazione delle risorse pubbliche, in ragione del merito scientifico accertato, fosse regolata non solo sul livello della qualità delle Strutture, ma anche sul delta del loro miglioramento”. Poiché per un punto passano infinite rette, la costruzione del coefficiente Δ richiede informazioni su due punti. Questo avrebbe dovuto indurre l’attuale Ministro a far proseguire celermente e senza soluzione di continuità, dal primo al secondo triennio, l’attività del CIVR, magari rinnovandone in tutto o in parte la *governance*, con l’inserimento di persone di sua assoluta fiducia, al fine di ottenere al più presto le informazioni di statica e di dinamica desiderate. Al contrario, come ho già avuto modo di notare in una lettera a Augias su Repubblica, egli ha rallentato la conclusione del primo esercizio, poi, a partire dal suo completamento, ha lasciato passare inutilmente molti mesi fino all’autunno del 2007, quando ha costituito il *Board* del nuovo CIVR riconfermando tutti i precedenti membri, salvo infine, poche settimane dopo, non procedere rapidamente a emettere il bando relativo al secondo esercizio. Ancora oggi, mentre noi ci siamo impegnati a redigere la Relazione Finale a circa 1 anno e $\frac{1}{2}$ dall’inizio del processo, continuiamo a ignorare quando questo ufficialmente avverrà, con l’emanazione del Decreto del Ministro dell’Università e della Ricerca, da mesi pronto ma incomprensibilmente non firmato, come ha ancora recentemente ricordato Jappelli (2008). La caduta del Governo del gennaio 2008 ed il successivo, serio, intervento chirurgico cui è stato sottoposto Fabio Mussi (al

quale colgo l'occasione per inviare l'augurio di pronta guarigione) ulteriormente non aiutano.

8) "E' limitativa la procedura seguita dal CIVR di assegnare nella valutazione il prodotto di ricerca al posto dove è stato scritto anziché, come nel *RAE*, a quello dove attualmente si trova il corrispondente ricercatore. L'incentivo alla ricerca è maggiore nel sistema inglese, perché consente a un docente di portare con sé i suoi lavori e accrescere così il suo valore sul mercato" (De Fraja, 2006). Concorro con questa diagnosi, tanto più che, dati i ritardi esistenti tra l'eventuale assegnazione delle risorse addizionali e il giudizio sulla qualità della produzione scientifica, in Italia si rischia di premiare il merito del passato anziché quello attuale. Ma il disincentivo più grande nel nostro Paese dipende dalla circostanza che non si giudica né si incentiva ogni singolo ricercatore, bensì solo ogni singola Struttura. Con l'aggravante che, anche ammesso che alla fase della valutazione seguisse un correlato *ranking* dei finanziamenti pubblici per la ricerca, attualmente le Strutture monitorate dal CIVR mantengono una assoluta discrezionalità nel predisporre criteri premiali o all'opposto compensativi quando determinano al loro interno l'allocazione fra Aree delle risorse pubbliche ricevute: e questo, secondo alcuni (come Schiesaro (2007)), è davvero un grave limite rispetto al similare esercizio del *RAE* inglese, al punto da suggerire di "procedere a metter mano anche alla *governance* delle Strutture", destinatarie dei fondi differenziati in ragione della valutazione.

9) "Nel CIVR è emerso il rischio prospettico di appiattire, omologare, ridurre le differenze, in una parola di privilegiare il *mainstream*, perché, imponendo nella valutazione un *benchmarking* basato sul pensiero dominante, si induce la disciplina a privilegiare l'uniformità anziché la varietà e l'originalità" (Pasinetti, 2006). Al di là delle peculiari diatribe emerse in proposito nell'Area 13, personalmente ritengo che esisterebbe il pericolo denunciato solo se si utilizzassero unicamente indicatori ottenuti da forme ridotte, come in passato aveva proposto Roberto Perotti, per valutare il merito dei prodotti sulla base, ad esempio, di variabili esplicative quali l'IF della rivista su cui si è pubblicato: lo documentano anche uno studio internazionale di Genna e Martin (2001) e Turri (2005). Ma, dall'osservatorio del *Board* del CIVR, ci sembra che fortunatamente ciò non sia avvenuto nell'esperienza concreta dei panelisti e dei *discussants* coinvolti nel primo triennio di

valutazione, tanto meno nel caso specifico del Panel 13. Non abbiamo alcuna evidenza che di fatto si siano giudicati i prodotti della ricerca solo riferendosi al luogo della loro pubblicazione o ad altri indicatori bibliometrici, come il *quotation index*, che Breno *et al.* (2005) considerano migliore dell'IF; al contrario, è stato chiaro e documentato nel CIVR lo sforzo di "non scadere in criteri arbitrari e del tutto autoreferenziali" di giudizio (Tabellini, riportato in Locatelli, 2006). Vorrei aggiungere che non solo nel CIVR gli indicatori bibliometrici sono sempre risultati secondari rispetto alla *peer review*, ma in aggiunta quegli indicatori in Italia e nell'economia non privilegiano affatto in modo sistematico il *mainstream*: basti pensare che gli unici 2 economisti del nostro Paese che compaiono fra i 250 più citati del mondo nell'ultimo ventennio sono, secondo le classifiche dell'ISI Thomson, un neoclassico come Guido Tabellini ed un assoluto eterodosso come Giovanni Dosi.

10) "Nel disegno istituzionale del CIVR esistono pericoli di eccessiva vicinanza dei valutatori della ricerca non solo ai valutati, ma anche alle burocrazie ministeriali/accademiche ed alle stesse forze politico-sociali": da questa idea era nata, infatti, nella passata legislatura (gennaio 2006) una Proposta di Legge DS (Modica-Tocci, 2006), tendente alla creazione di una specifica Autorità per la valutazione della ricerca, a difesa della terzietà rispetto all'Esecutivo e della più generale autonomia del giudizio di merito scientifico. Personalmente ritengo che la condizione di separatezza di diritto e di fatto dal Governo, dai poteri forti, dalle burocrazie sia, per chi detiene la *governance*, fortemente auspicabile e dunque considero che, da tale punto di vista, emerga nel CIVR un'imperfezione da eliminare, ma so anche che, in ultima analisi, l'indipendenza è guadagnata (o perduta) tutta sul campo dagli uomini e dalle donne deputate al giudizio. Peraltro, da un lato, debbo dare atto sia a questo che al passato Ministro dell'Università e della Ricerca di non aver mai tentato di interferire nelle nostre attività (caso mai concordo con i molti che hanno osservato "la lentezza e la disattenzione del Ministro rispetto agli ottimi risultati del primo esercizio triennale di valutazione del CIVR", da Giavazzi (2006), a Perotti e Tabellini (2007), da Salvati (2007), a Kostoris (2006d, 2007b) ai quali due ultimi ha replicato Mussi (2007a, 2007b)); dall'altro lato, sono costretta a notare che il Governo Prodi non ha pienamente superato il

problema, avendo trasformato quella che avrebbe dovuto essere una Autorità indipendente per la valutazione della ricerca in una mera Agenzia ministeriale, l'ANVUR, dotata unicamente di personalità giuridica di diritto pubblico. Il sottosegretario al MUR, ancora il 17 gennaio 2007, ribadiva al Sole 24 Ore con rimpianto (Modica, 2007) il fatto che, in questa fase, la soluzione dell'*Authority* fosse divenuta utopica, mentre quella dell'Agenzia risultava "l'unica strada percorribile".

4 Prospettive dell'ANVUR

L'ANVUR ha già una vita normativa estesa e complessa, nonostante sia nata da poco con il Decreto-Legge n. 262 del 3 ottobre 2006, definitivamente convertito nella Legge 286 del 24 novembre 2006: basti leggere le Linee Guida del gennaio 2007 (ANVUR, 2007a), la bozza di Regolamento del 12 marzo 2007 (ANVUR, 2007b, 2007c), il suo corredo di commenti vari, dalla Commissione per l'Università dell'Accademia Nazionale dei Lincei (2007), al Consiglio di Stato (2007), dal CUN (2007), alla CRUI (2007) e infine lo schema di DPR del 18 dicembre 2007 (Presidenza della Repubblica, 2007). L'obiettivo dell'ANVUR è quello, già richiamato, di sottoporre a giudizio tutte le istituzioni di ricerca italiane destinatarie di fondi pubblici. Nonostante le ottime intenzioni dichiarate nelle norme e i presumibili buoni risultati di cui auspicabilmente sarà capace, non esistono concreti meccanismi giuridici volti ad assicurare che l'ANVUR sarà in grado di rendere operativo il criterio meritocratico meglio del CIVR, salvo nel desiderabile allargamento della platea dei valutati, come sopra menzionato. E questo per 3 principali motivi.

Innanzitutto, l'indipendenza dell'ANVUR è minata dalla mancanza di terzietà rispetto all'Esecutivo e dagli eccessi di controlli da parte dei vari *stakeholders*: tutti i 7 membri del suo Consiglio Direttivo sono, infatti, scelti direttamente o indirettamente dal Titolare del MUR e a lui o al suo Dicastero riportano, segnalano, propongono; inoltre essi sono tenuti a sottoporre i programmi e i criteri di selezione delle attività annuali al parere di un Comitato Consultivo, inizialmente non previsto nelle norme, formato da persone designate dai vertici delle Università, degli Enti di Ricerca, del CUN, della

CRUI, dell'Accademia dei Lincei, e perfino dai rappresentanti degli studenti e delle parti sociali! Non si capisce chi abbia introdotto nel testo definitivo del DPR sull'ANVUR i tanti paletti burocratico-sindacal-politici esistenti, dato che tutti ne sembrano, per ragioni opposte, scontenti: dalle associazioni sindacali che, come l'ANDU, gradirebbero presenze meno pesanti accademico-politiche, alla Conferenza dei Rettori che, con parere unanime del 20 marzo 2007, lamenta la subordinazione dell'ANVUR al Ministro, pur ricordando che le stesse attività di competenza del Ministero saranno oggetto di valutazione e che l'operato dell'Agenzia sarà sottoposto, dopo un triennio, al giudizio di una apposita commissione internazionale.

Secondariamente, non vi è sufficiente separatezza fra l'ANVUR e i nuclei di valutazione interna delle Università e degli Enti di Ricerca, in quanto le norme assegnano all'Agenzia il compito di svolgere su di essi funzioni non solo di "indirizzo e vigilanza", ma anche di "coordinamento" delle attività, specificatamente quando queste richiedano "confronti nazionali e internazionali sia quantitativi che qualitativi": l'Agenzia non si limita dunque, come vorrebbe il documento del 9 febbraio 2007 della Commissione Lincea, a predisporre anche criteri generali ai quali i nuclei interni di valutazione si debbano attenere; al contrario, essa si estende fino a creare interferenze fra vigilanti e vigilati, potenzialmente atte a condurre a collusioni dannose ad un giudizio meritocratico. Non basta, da questo lato, la pur lodevole attenzione che il Regolamento dell'ANVUR pone nell'evitare conflitti di interesse fra i membri del Consiglio Direttivo, chiedendo loro il tempo pieno incompatibile con qualsiasi rapporto di lavoro, anche a titolo gratuito, con Università, Enti di Ricerca, Ministeri, Autorità ed altre Agenzie governative.

In terzo luogo, le competenze dell'ANVUR sono oggi circoscritte sia nella valutazione dei singoli ricercatori, sia nell'assegnazione delle risorse alle Strutture in base a criteri meritocratici. Sul primo punto, il Regolamento assegna all'ANVUR solo il compito di giudicare *ex post* la qualità del reclutamento del personale delle Università e degli Enti di Ricerca, "contribuendo – come è scritto nelle Linee Guida – alla vigilanza del Ministero in tema di concorsi sulla base dei risultati e non delle procedure". Scompare, invece, qualunque riferimento al giudizio sulla ricerca individuale, pregnante al contrario nel Disegno di Legge

Modica-Tocci, laddove si stabiliva che “l’Autorità determina le procedure, le metodologie ed i tempi operativi per la valutazione periodica dell’attività di ricerca svolta dai singoli professori e ricercatori”. Sul secondo punto, il Regolamento stabilisce che l’Agenzia determini e proponga al Ministro parametri di ripartizione per l’allocazione dei finanziamenti pubblici, ma non disponga nulla, e, per un verso, segnali situazioni di istituzioni di elevata qualità, meritoria di quote aggiuntive premiali, per un altro, indichi condizioni di grandi scostamenti richiedenti appositi programmi di rientro. Al contrario, nel Disegno di Legge DS del gennaio 2006, era “riservata all’Autorità, sulla base dei risultati delle sue attività di valutazione, la ripartizione tra le Università e tra gli Enti di Ricerca di una quota del 2% del fondo annuale per il finanziamento ordinario”.

5 Conclusioni

Il processo di valutazione della ricerca italiana, iniziato dal CIVR, non deve arrestarsi: se ci si ferma, si indietreggia. E’ necessario che negli anni a venire prima il CIVR e poi l’ANVUR continuino a fornire quegli strumenti di giudizio statico e dinamico delle Strutture, che appaiono indispensabili a combinare con criterio meritocratico, e dunque insieme equo ed efficiente, l’autonomia con la responsabilità delle nostre istituzioni scientifiche. Il CIVR, pur con qualche limite parzialmente in corso di superamento, ha finora ben operato, pervenendo ad un *ranking* delle Università e degli Enti di ricerca accettato e trasparente. Non basta. Bisogna che tale confronto ordinamentale comporti conseguenze osservabili in termini di trasferimenti di risorse pubbliche, perché questo è un incentivo al miglioramento assai più credibile del mero “*naming and shaming*”, finora realizzato dal CIVR (Kostoris, 2007b). E’ consigliabile, inoltre, estendere la valutazione a tutte le istituzioni di ricerca destinatarie di fondi pubblici e non solo a quelle vigilate dal MUR, cui fin qui ci si è circoscritti. Da questi punti di vista, ma forse non da altri, l’ANVUR sarà probabilmente in grado di fare passi avanti rispetto al primo ed al secondo esercizio triennale del CIVR. Ma la strada è lunga e perigliosa. Speriamo almeno che sia una via senza ritorno!

Riferimenti bibliografici

- ANVUR (2007a), Linee guida per il regolamento ministeriale sulla struttura e sul funzionamento dell’Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca, gennaio, disponibile sul sito <http://aidea.economia.unibo.it>
- ANVUR (2007b), Bozza di regolamento recante disposizioni relative alla struttura e al funzionamento dell’Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca, marzo, disponibile sul sito <http://aidea.economia.unibo.it>
- ANVUR (2007c), Regolamento recante disposizioni relative alla struttura e al funzionamento dell’Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca, gennaio, disponibile sul sito <http://aidea.economia.unibo.it>
- Arachi G. e A. Zanardi (2006), “Un punto critico nella valutazione”, La voce.info, 8 febbraio.
- Bernini A. e V. Nepi (2005), La Qualità della Ricerca Scientifica in Italia, Firenze, Menarini Ricerca e Sviluppo.
- Breno E., Fava G.A., Guardabasso, V. e M. Stefanelli (2005), Un aggiornamento sull’impatto della ricerca scientifica e tecnologica italiana in ambito internazionale, 1981-2004. Analisi preliminare, Roma, CRUI.
- Comitato di Indirizzo per la Valutazione della Ricerca – CIVR (2006a), VTR 2001-2003. Risultati delle valutazioni di Panel di Area, Roma, Ministero dell’Istruzione, dell’Università e della Ricerca, 26 gennaio.
- Comitato di Indirizzo per la Valutazione della Ricerca – CIVR (2006b), Linee guida per la valutazione della ricerca. Pianificazione degli interventi, Roma, Ministero dell’Istruzione, dell’Università e della Ricerca.
- Comitato di Indirizzo per la Valutazione della Ricerca – CIVR (2006c), VTR 2001-2003. Relazione Finale e allegati, Roma, Ministero dell’Istruzione, dell’Università e della Ricerca.
- Commissione delle Comunità Europee (1996), Disciplina comunitaria in materia di Aiuti di Stato alla ricerca e sviluppo, Gazzetta Ufficiale delle Comunità Europee, n. C45/C, 17 febbraio.
- Commissione delle Comunità Europee (2002), “Più ricerca per l’Europa. Obiettivo: 3% del PIL”, Comunicazione della Commissione, COM(2002) 499 def.
- Commissione per l’Università dell’Accademia Nazionale dei Lincei (2007), Documento sull’ANVUR, febbraio, disponibile sul sito http://aidea.economia.unibo.it/index.php?option=com_docman&task=cat_view&gid=5&Itemid=4
- Conferenza dei Ministri Europei Responsabili dell’Istruzione Superiore (2005), Comunicato di Bergen, maggio.

- Consiglio delle Comunità Europee (2002), Conclusioni della Presidenza. Consiglio Europeo di Barcellona. 15-16 marzo.
- Consiglio di Stato (2007), Schema di decreto del Presidente della Repubblica recante regolamento concernente la struttura e il funzionamento dell'Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca, maggio, disponibile sul sito <http://aidea.economia.unibo.it>
- Consiglio Italiano per le Scienze Sociali (2006), La valutazione della ricerca. Libro bianco, Commissione coordinata da A. Zuliani, Venezia, Marsilio Editore.
- Consiglio Universitario Nazionale (2007), Parere sulla bozza di regolamento recante disposizioni relative alla struttura e al funzionamento dell'Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca, marzo, disponibile sul sito <http://aidea.economia.unibo.it>
- CRUI (2007), Parere sulla bozza di regolamento recante disposizioni relative alla struttura e al funzionamento dell'Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca, marzo, disponibile sul sito <http://aidea.economia.unibo.it>
- Cuccurullo F. (2007), Presentazione dell'8/2/2007, mimeo.
- De Fraja G. (2006), "Il campionato accademico", La voce.info, 8 febbraio.
- Genna A. e B. Martin (2001), "University Research Evaluation and Funding", Minerva, 41, pp. 277-304.
- Giavazzi F. (2006), "Errori e miti sull'Università", Corriere della Sera, 14 novembre, p. 1.
- Gruppo 2003 (2003), Manifesto, disponibile sul sito <http://gruppo2003.org/manifesto.php>
- Guiso L. (2006), "Un'Università allo stremo", La voce.info, 7 giugno.
- Jappelli T. (2006), "Una valutazione positiva", La voce.info, 8 febbraio.
- Jappelli T. (2008), "Università", La voce.info, 29 gennaio.
- Kostoris Padoa Schioppa F. (2006a), "I principi della valutazione della ricerca", Analysis, 3-4, pp. 8-12, ripubblicato in Fini R. e U. Margotta (a cura di), Il filosofo e il facchino. Università e ricerca nella formazione del capitale umano, Venezia, Mazzanti Editore, 2007.
- Kostoris Padoa Schioppa F. (2006b), "Tre consigli per valutare la ricerca italiana", Sissa News, 1, 5, febbraio, pp. 6-8.
- Kostoris Padoa Schioppa F. (2006c), "Risorse per la ricerca scarse e mal ripartite", Il Sole 24 Ore, 3 febbraio, p. 1.
- Kostoris F. (2006d), "Sull'Università il Governo affonda la meritocrazia", Il Riformista, 28 novembre, p. 3.
- Kostoris F. (2007a), "Questioni di merito", in A.A.V.V., Riforme. We Will, Venezia, Marsilio Editore, pp. 83-92.
- Kostoris F. (2007b), "Università, chi valuta che cosa", Il Sole 24 Ore, 14 aprile, p. 14.
- Lisconi F. (2007), "ANVUR, Agenzia per il reclutamento dei ricercatori", La voce.info, 23 aprile.
- Locatelli F. (2006), "Economisti in guerra sulla ricerca", Il Sole 24 Ore, 9 febbraio, p. 10.
- Modica L. (2006), "Università, ci vuole un rendiconto su prof e attività", Corriere del Mezzogiorno, 11 agosto, p. 1.
- Modica L. (2007), "Indipendenti e imparziali", Il Sole 24 Ore, 17 gennaio, p. 27.
- Modica L. e W. Tocci (a cura di) (2006), "Istituzione dell'Autorità per la Valutazione del Sistema delle Università e della Ricerca", Disegno/Proposta di Legge, gennaio.
- Mussi F. (2007a), "Valutazione zone franche", Il Sole 24 Ore, 14 aprile, p. 10.
- Mussi F. (2007b), "Ma io difendo la nuova Agenzia", Corriere della Sera, 15 agosto, p. 33.
- Mussi F. (2007c), "Università, si cambia: più fondi ai migliori. Intervista al Ministro Mussi", Europa, 15 febbraio, p. 4.
- Pasinetti L.L. (2006), in Comitato di Indirizzo per la Valutazione della Ricerca – CIVR (2006a), VTR 2001-2003. Risultati delle valutazioni di Panel di Area, Roma, Ministero dell'Istruzione, dell'Università e della Ricerca, 26 gennaio.
- Peracchi F. (2006a), "Un giudizio sul CIVR", La voce.info, 15 maggio.
- Peracchi F. (2006b) "Valutazione Triennale della Ricerca (VTR) 2001-2003. L'esperienza del Panel 13", Testo presentato alla CRUI, 28 marzo.
- Perotti R. E G. Tabellini (2007), "Perché si proteggono le Università peggiori", Il Sole 24 Ore, 11 marzo, p. 1.
- Presidenza della Repubblica (2007), Schema di Decreto recante regolamento concernente la struttura e il funzionamento dell'Agenzia Nazionale di Valutazione del Sistema Universitario e della Ricerca, dicembre, disponibile sul sito <http://aidea.economia.unibo.it>
- RAE 2008 (Research Assessment Exercise), Making a submission to the RAE 2008, disponibile sul sito <http://www.rae.ac.uk/aboutus/subs.asp>
- Reale E., Barbara A. e A. Costantini (2006), "Peer Review for the Evaluation of the Academic Research: the Italian Experience", Paper presentato alla Platform FTEval International Conference, Vienna, 24-25 aprile.
- Rubele R. (2007), "Siamo ancora troppo poco 'anglosassoni' per legare i fondi alla qualità degli atenei", Il Riformista, 10 maggio, p. 6.
- Salvati M. (2007), "Domande sul patto tra Governo e Atenei", Corriere della Sera, 14 agosto, p. 1.
- Schiantarelli F. (2006), "Un passo nella giusta direzione", La voce.info, 8 febbraio.
- Schiesaro A. (2007), "La governance non può attendere", Il Sole 24 Ore, 16 febbraio, p. 12.
- Tamborini R. (2006), "I pericoli dell'one-man-department", <http://www.lavoce.info>, 15 maggio.
- Tognon G. (a cura di) (2006), Una dote per il merito. Idee

per la ricerca e l'Università italiane, Bologna, Il Mulino.

Toniolo G. (2006), "Chi ricerca è premiato", Il Sole 24 Ore, 21 marzo, p. 8.

Turri M. (2005), La valutazione dell'università. Un'analisi dell'impatto istituzionale e organizzativo, Milano, Guerini e Associati.

La valutazione della ricerca: alcune riflessioni generali

Alberto Zuliani

Dipartimento di studi geoeconomici, linguistici, statistici, storici per l'analisi regionale
Sapienza Università di Roma
alberto.zuliani@uniroma1.it

Sommario

Gli investimenti in ricerca in Italia ristagnano da molti anni intorno all'1,1% del pil. Il contributo del settore privato è basso. Il confronto con gli altri paesi evoluti rende evidenti le nostre debolezze. In questo quadro, la valutazione diviene una pratica irrinunciabile. Sono disponibili strumenti e metodi collaudati. Occorre però una forte volontà politica. Bisogna assumere da parte di tutti un orientamento ad apprendere e i risultati della valutazione devono produrre effetti riguardo alla distribuzione delle risorse.

1 Il libro bianco del Consiglio italiano per le scienze sociali

Prendo le mosse dai contenuti del Libro bianco sulla valutazione della ricerca a cura del Consiglio italiano per le scienze sociali (2006) alla cui redazione ho preso parte, coordinando un'apposita commissione. Il volume si sofferma anzitutto su aspetti definitivi, rilevanti in un contesto i cui contorni sono sfumati; passa quindi in rassegna le principali esperienze di valutazione della ricerca effettuate nel paese, offre una rassegna ordinata sugli strumenti e metodi della valutazione e avanza alcune raccomandazioni. Al libro bianco rinvio senz'altro quanti volessero approfondire l'argomento.

Credo possa essere importante, in questa occasione, ricordare alcune principali dimensioni della ricerca e i maggiori problemi che essa incontra in Italia.

La spesa per ricerca e sviluppo è bassa. Essa ha costituito l'1,1% del PIL nel 2005, ultimo anno per il quale l'Istat ha reso disponibile il dato (Istat, 2007) ed è rimasta sostanzialmente intorno a questo livello ormai da molti anni. Nell'Unione europea a 15, nello stesso anno, il rapporto risultava pari a 1,91%. Si registra, inoltre, uno sbilanciamento fra investimenti effettuati dal settore pubblico (47,5% nel 2005) e da quello

privato (50,4% sempre nel 2005), rispetto a quanto avviene negli altri paesi evoluti, dove il contributo del secondo settore è assai più ampio. Il restante 2,1% è riconducibile al settore *no profit*.

Per dare conto della distanza dall'Europa, possono essere ricordati gli obiettivi di Lisbona che indicavano per il 2010, per l'Unione a 15, il 3% quale livello della spesa per ricerca e sviluppo rispetto al PIL e in 2/3 il contributo ad essa del settore privato.

Il confronto con un paese ben più dinamico, la Spagna, è istruttivo. Nel 1995 esibiva un rapporto fra spesa per ricerca e sviluppo e PIL pari a 0,79%, nel 2000 pari a 0,91% e nel 2005 a 1,12%. In Italia, l'andamento è risultato il seguente: 0,97% nel 1995; 1,05 nel 2000 e 1,10 nel 2005.

Come è noto, da noi operano tre circuiti di ricerca: università, enti pubblici di ricerca, ricerca privata. Ciascuno di essi presenta elementi di debolezza. Quanto alle università, le graduatorie internazionali relative sia alla ricerca sia alla formazione non ci vedono in buona posizione, naturalmente con eccezioni. Riguardo agli enti pubblici di ricerca, si sono susseguiti, nel corso degli ultimi quindici anni, molti interventi non coordinati. Ogni esecutivo ha ritenuto di avere la chiave di soluzione dei problemi, ma le misure prese di volta in volta hanno contribuito a rendere meno chiari (e durevoli) mandati e obiettivi. Infine, per il settore privato, la propensione

all'investimento in ricerca e sviluppo è risultata assai debole. La normalizzazione per distribuzione dimensionale delle imprese (da noi mediamente più piccole che negli altri paesi evoluti) e per composizione dei settori produttivi (da noi sono maggiormente presenti quelli "maturi") riequilibra un poco la quota di investimenti sul fatturato, ma dà conto, con differente evidenza, della fragilità del sistema produttivo del paese.

In generale, in tutti e tre i circuiti di ricerca, si riscontra un basso tasso di rinnovamento. L'Università è ripiegata su se stessa e riproduce le soluzioni organizzative di sempre. Faccio un esempio per tutti: in ogni provincia italiana, nessuna esclusa, è localizzato un presidio formativo universitario e molte delocalizzazioni (oltre 200 sono situate in comuni diversi da quelli che ospitano le sedi centrali degli atenei) rappresentano una replica in peggio delle proposte già presenti nella "casa madre". Si incontrano e colludono campanilismi minuti ed esigenze di riproduzione, non sempre per merito, della classe docente; con buona pace della qualità della formazione. Sono vistosi i fenomeni di "familismo", soprattutto in alcuni settori; le modalità concorsuali introdotte una decina di anni fa hanno accentuato un provincialismo già assai diffuso. La lettera aperta di undici ricercatori italiani, essenzialmente fisici e matematici, reclutati recentemente e con grande riconoscimento da organismi di ricerca francesi, rende evidente l'angustia nella quale l'università italiana è invischiata.

Gli enti pubblici di ricerca si sono molto burocratizzati ed è chiaro per tutti che burocrazia e ricerca mal si coniugano. Anche in queste istituzioni, i fenomeni di familismo sono estesi. Da ultimo, la contrattazione collettiva ha introdotto concorsi riservati, legati all'anzianità di servizio, una contraddizione profonda in un ambiente per definizione competitivo, che sta provocando arretramenti oltre che ingiustizie.

Nel settore privato l'attività di ricerca e sviluppo ristagna, l'ho già detto; sono un segnale evidente il *deficit* ormai cronico della nostra bilancia tecnologica e la lenta ma progressiva perdita di quote di mercato all'estero.

Ovviamente, ci sono dipartimenti universitari di eccellenza, da tutti riconosciuti; istituti di ricerca pubblici che vantano un'ottima produzione scientifica; imprese che innovano; ma complessivamente il sistema mostra una forte

difficoltà a cambiare marcia. Il rischio evidente è di "non farcela" a invertire le tendenze in atto.

Alcuni limitati segnali positivi si intravedono. Su di essi si deve lavorare. E' migliorato il dialogo fra scienza e società, anche a motivo dell'accresciuta sensibilità della popolazione ai problemi posti dagli avanzamenti scientifici: l'impatto ambientale; la bioetica, per fare soltanto due esempi. C'è forse qualche spazio perché la ricerca riguadagni posizioni nella scala delle priorità del paese. Si stanno sperimentando forme di collegamento fra ricerca pubblica e privata, attraverso gli *spin-off* accademici, gli *industrial liason offices* costituiti presso le università, i *joint labs* partecipati dalle università e dalle imprese. Ancora troppo poche iniziative, però, per produrre cambiamenti significativi. Comincia ad esprimersi la potenzialità delle fondazioni che, in misura ancora ridotta ma maggiore che nel passato, finanziano attività di ricerca. All'estero e specialmente nei paesi di cultura anglosassone, esse rappresentano un riferimento importante.

Dovrebbero riacquistare peso comportamenti orientati alla qualità; dovrebbe essere stimolata una competizione reale fra le istituzioni e fra i ricercatori; le iniziative di sostegno alle imprese dovrebbero essere affettivamente selettive e si dovrebbe chiedere conto dei risultati; occorrerebbe riaffermare il valore della mobilità del personale di ricerca, fra i tre circuiti nel paese e con l'estero, in quest'ultimo caso facendo sì che il flusso fosse bidirezionale.

2 Valutazione della ricerca

Fino a qualche anno fa, l'ambiente della ricerca, in particolare quello pubblico, è rimasto sostanzialmente autoreferenziale.

L'eterovalutazione è stata introdotta soltanto recentemente, anche perché i problemi di finanza pubblica hanno imposto una maggiore selettività degli interventi; essa ha suscitato resistenze diffuse, ma ha avviato anche processi di autovalutazione più rigorosi rispetto al passato. Nessuno, ritengo, metterebbe in dubbio, oggi, l'utilità della valutazione: essa stimola concorrenza emulativa; produce apprendimento e l'adozione di prassi migliori; contribuisce a diffondere tecniche di *R&D management* nel settore privato ed anche in quello pubblico.

La valutazione si avvale di metodi e strumenti diversi, seppure spesso integrati fra loro, a

seconda delle finalità che si propone: sostegno alla definizione di politiche in senso ampio; valutazione di leggi e misure di intervento; verifica del raggiungimento di obiettivi; valutazione e sostegno operativo delle istituzioni; valutazione di programmi e progetti, eventualmente in competizione; valutazione di gruppi e di individui.

Mi soffermo su due iniziative di valutazione che hanno interessato la ricerca pubblica in Italia: la valutazione dei progetti di interesse nazionale, PRIN, e la valutazione delle istituzioni effettuata con riferimento al triennio 2001-03 dal Comitato di indirizzo per la valutazione della ricerca, CIVR.

I meccanismi di cofinanziamento dei PRIN sono in generale indicati come un esempio di buona pratica. In effetti, essi hanno innovato profondamente rispetto alla prassi precedente di distribuzione delle risorse finalizzate alla ricerca, orientandosi a individuare e finanziare un numero ridotto di progetti meritevoli. L'impegno e la determinazione iniziali sono stati forti e inusuali. E' tuttavia possibile rinvenire alcuni limiti nella procedura e, recentemente, una tendenza involutiva.

Le critiche maggiori sono le seguenti: sono stati favoriti, almeno, nelle prime applicazioni, i grandi raggruppamenti, poiché sono state mutate acriticamente le logiche di finanziamento europee, le quali però tendevano dichiaratamente a suscitare collegamenti fra paesi e fra istituzioni; regole di rendicontazione non rigorose hanno fatto sì che il cofinanziamento da parte delle strutture di ricerca potesse attingere a fondi già destinati in precedenza a specifiche e differenti attività della stessa natura; le valutazioni sono risultate molto sensibili a giudizi di dettaglio; la limitata (o nulla) responsabilità dei valutatori riguardo ai giudizi espressi non ha favorito l'accuratezza; sono stati penalizzati, in qualche misura, i progetti interdisciplinari; la sostanziale assenza di valutazioni *in itinere* ed *ex post* riguardo ai risultati conseguiti non ha valorizzato il processo di valutazione. Tuttavia, il giudizio da dare è complessivamente positivo, con riferimento sia alla trasparenza sia all'incisività del meccanismo innescato. Per il futuro, occorrerà rimuovere i problemi segnalati.

Di estremo interesse risulta l'esperienza di valutazione delle istituzioni di ricerca effettuata dal CIVR per il triennio 2001-03. L'esercizio si è concluso nel dicembre 2006 e ha riguardato oltre

17.000 prodotti (libri, articoli, brevetti, progetti, disegni eccetera) e 102 strutture di ricerca (77 università statali, 12 enti pubblici di ricerca e 13 istituzioni di ricerca private) alle quali facevano capo circa 64.000 ricercatori. Un impegno importante, dunque, nel quale sono stati coinvolti 20 *panel* di esperti corrispondenti alle 20 aree disciplinari individuate dal CIVR.

Complessivamente, sono stati coinvolti nei *panel* 150 esperti i quali, a loro volta, si sono avvalsi di 6.500 valutatori. Il processo, sul quale si deve comunque esprimere un giudizio largamente positivo, presenta luci ed ombre. Si tratta del maggiore esercizio mai svolto precedentemente in Italia, realizzato attraverso una procedura interamente telematica, mediante il metodo internazionalmente condiviso del *peer reviewing*, coinvolgendo un congruo numero di esperti stranieri, più di un quinto del totale, con un costo di 3,5 milioni di euro, una cifra significativamente inferiore rispetto agli oltre 15 milioni spesi per l'analogo *Research Assessment Exercise (RAE)* nel Regno Unito.

Alcuni problemi andranno risolti nelle successive edizioni. La scelta di limitare l'analisi ad un prodotto ogni due ricercatori equivalenti a tempo pieno (*FTE*) nell'arco di un triennio (questo era il vincolo posto alle università, a motivo della produzione congiunta realizzata in tali istituzioni; dagli enti pubblici di ricerca doveva essere segnalato un prodotto per ogni ricercatore *FTE*) può aver provocato disomogeneità non controllabili. Nelle situazioni più attente, e si deve ritenere che questo sia normalmente avvenuto, le strutture di ricerca hanno selezionato i prodotti migliori, cosicché la valutazione ha finito per riguardare il livello della ricerca anche migliore e non di quella mediamente svolta. Il lavoro dei valutatori coinvolti è stato essenzialmente documentale; alcuni degli aspetti da considerare risultavano problematici e lasciavano ampia discrezionalità; l'algoritmo di valorizzazione/penalizzazione è stato esplicitato soltanto dopo che le istituzioni avevano selezionato i prodotti da sottoporre a valutazione. Tuttavia, come ho già detto, l'esperienza è stata molto positiva ed ha contribuito ad introdurre un orientamento alla valutazione nell'ambiente della ricerca italiana.

Alcune ulteriori considerazioni sono necessarie. Anzitutto, i risultati hanno avuto un impatto molto limitato sulla successiva distribuzione delle risorse per la ricerca e questo potrà nuocere alla credibilità delle replicazioni. La valutazione *deve*

avere conseguenze. In secondo luogo, al termine dell'esercizio - forse eccessivamente impegnativo, sia durante il suo svolgimento sia nella fase di gestione "politica" dei risultati - non è stata svolta un'analisi approfondita del cospicuo materiale raccolto, come invece sarebbe stato utile. Infine, le riflessioni innescate nei diversi settori disciplinari, in particolare in quelli risultati più deboli in sede di valutazione, sono ancora poche (Bartolini, 2007).

3 L'agenzia nazionale di valutazione del sistema universitario e della ricerca, Anvur

In questo contesto, è stata recentemente istituita l'Agenzia Nazionale di Valutazione del sistema Universitario e della Ricerca, ANVUR; essa si è impantanata negli adempimenti regolamentari, cosicché non è ancora operativa. L'Agenzia non nasce sotto i migliori auspici, a mio modo di vedere. Le sono stati assegnati troppi compiti, in qualche misura eterogenei, a fronte di risorse relativamente scarse. E' stata prevista l'istituzione di un Comitato consultivo, sostanzialmente inutile. Temo, date le premesse, una rapida burocratizzazione dell'organizzazione e della prassi della valutazione nel settore. Sarebbe stato necessario configurare una vera e propria autorità a presidio della valutazione della ricerca che, tutti concordano, rappresenta un presupposto essenziale per la crescita del paese. La partecipazione all'organo di governo dell'Agenzia sarebbe dovuta essere ben altrimenti attrattiva, per ruolo e per corrispettivo economico, rispetto alla previsione attuale. Le risorse dovrebbero essere commisurate ad una percentuale della spesa per la ricerca e non essere subordinate alle scelte congiunturali del bilancio pubblico o alla "sensibilità" del ministro in carica. Mi aspetto, dunque, che l'Agenzia abbia vita difficile, per quanto ho appena detto ed anche perché essa deve operare in un contesto nel quale, negli ultimi vent'anni, è sostanzialmente mancata competizione reale.

4 Difficoltà della valutazione

Prima di concludere, vorrei indicare due esempi dai quali emerge la difficoltà di valutare la

ricerca, il primo riferito ai prodotti, il secondo alla comparazione fra paesi. Per approfondimenti rinvio, ancora una volta, al Libro bianco del Consiglio italiano per le scienze sociali.

Un ricerca recente mostra l'insufficienza, la non robustezza, dell'*Impact Factor (IF)*. Esso risulta condizionato dalla popolarità delle riviste piuttosto che dalla loro rilevanza; privilegia le *Reviews* rispetto ai *Journals*. Un differente algoritmo (*PageRank algorithm, PR*), basato sulle modalità operative del motore di ricerca *Google* ed attento anche al secondo aspetto, produce graduatorie assai differenti. Ad esempio, con riferimento a 5.709 riviste scientifiche, il *Journal of Biological Chemistry* risulta primo per il *PR* e soltanto 180° per l'*IF*; l'*Annual Reviews of Biochemistry* è secondo per l'*IF*, ma 300° per il *PR* (Hascall et al., 2007). Aggiungo che è possibile influire sull'*IF* attraverso meccanismi di citazione incrociati. Mi viene in mente, a questo proposito, l'alterazione delle quote nel *parimutuel*, per le scommesse relative alle corse dei cavalli (Aczel, 2005).

Talvolta, le comparazioni fra paesi sono condotte in modo non corretto. Procedo con un esempio. In Italia, nel 1999, si contavano 11 ricercatori espressi in unità equivalenti a tempo pieno per 10.000 abitanti; negli Stati Uniti, nello stesso anno, il rapporto risultava di oltre quattro volte superiore. Supponendo, come è del tutto ragionevole, l'esistenza di una produttività marginale decrescente della ricerca in funzione del numero dei ricercatori, la produzione unitaria italiana andrebbe confrontata con quella dei migliori, più produttivi, ricercatori statunitensi, fino a raggiungere la quota di 11 ogni 10.000 abitanti; non è quello che si fa normalmente (sul punto, mi sono soffermato recentemente (Zuliani et al., 2007)).

5 Una breve conclusione

Di fronte alla valutazione, ed in particolare alla valutazione della ricerca, occorre avere un atteggiamento orientato all'apprendimento ed essere consapevoli che la materia non è semplice. L'articolo di Hascall e altri che ho già citato inizia con una frase di Albert Einstein: "*Everything in life should be as simple as possible, but non simpler*". Facciamone tesoro.

Riferimenti bibliografici

Aczel A. D. (2005), *Chance*, Raffaello Cortina Editore, Milano.

Bartolini S. (2007), “Come migliorare la qualità della ricerca in Italia. L’esperienza della valutazione triennale della ricerca nelle scienze politiche e sociali (2001-2003)”, www.sociologica.mulino.it.

Consiglio italiano per le scienze sociali (2006), *La valutazione della ricerca. Libro banco*, Marsilio, Venezia.

Hascall V. C. e altri (2007), “Impact Factor Page Rankled”, *ASBMB Today*.

Istat (2007), “La ricerca e sviluppo in Italia nel 2005”, *Statistiche in breve*.

Zuliani A. e altri (2007), “La misurazione dell’amministrazione pubblica italiana in un quadro comparativo”, *mimeo*.

Cosa possiamo imparare dalla VTR 2001-2003?

Alessandro Figà-Talamanca

Dipartimento di Matematica
Sapienza Università di Roma
sandroft@mat.uniroma1.it

Sommario.

Si ripercorre brevemente la procedura di Valutazione Triennale della Ricerca che si è conclusa nel gennaio 2006, indicando alcuni punti critici e suggerendo alcune revisioni della procedura.

1 I risultati del VTR 2001-2003

Ci sono molte cose che possiamo imparare dall'esercizio di Valutazione Triennale della Ricerca (VTR), conclusosi nel gennaio 2006 e promosso ed organizzato dal Comitato di Indirizzo per la Valutazione della Ricerca (CIVR).

Prima di tutto è stato dimostrato che è possibile, in Italia, una valutazione esauriente e completa della ricerca scientifica svolta dalle università e dagli enti di ricerca italiani, nell'arco di un triennio (2001-03).

Questa valutazione ha comportato una mole imponente di lavoro. Basta ricordare che sono stati esaminati da comitati di esperti 17.329 "prodotti", in massima parte pubblicazioni, proposti da 77 università e 12 enti di ricerca. I costi diretti di questa operazione sono stati all'incirca tre milioni e mezzo di euro, una cifra relativamente bassa. Non era affatto ovvio che una operazione di questo tipo, in tempi brevi, e a costi contenuti, fosse possibile. Molti scettici, compreso il sottoscritto, hanno dovuto, a questo proposito, ricredersi.

In secondo luogo si è prodotto un risultato che ha contraddetto in parte le lamentele usuali sul basso livello della ricerca scientifica italiana. Circa il 30% dei prodotti presentati è stato giudicato eccellente ed il 46% è stato giudicato buono. Questo significa che più di tre quarti dei prodotti presentati è stato giudicato buono o eccellente. A fronte di questi risultati forse dovremmo chiederci come mai l'impressione generale degli addetti ai lavori è che la ricerca scientifica italiana sia, in generale, inferiore a quella che si svolge negli altri paesi industriali. Una spiegazione possibile è il

fatto che fuori del proprio paese si conosce solo la ricerca di buon livello. Ciò che non raggiunge standard internazionali di qualità non varca le frontiere. Il confronto basato sulle esperienze personali dei ricercatori è quindi quello tra tutta la ricerca svolta in Italia, anche quella di basso livello, e la ricerca di maggior spicco che si svolge in altri paesi.

Dal punto di vista politico la conclusione più importante che si può ricavare dai risultati della VTR è che prodotti buoni o eccellenti sono stati presentati in tutte le aree da tutte le università grandi o medie (i dati riguardanti le "piccole" sedi, non sono, come vedremo, sufficientemente affidabili).

Questo risultato contrasta decisamente con i risultati dei *Research Assessment Exercises (RAE)* del 2001, svolti in Gran Bretagna, che evidenziano come 40 su 132 sedi universitarie hanno contribuito solo marginalmente all'attività di ricerca (Review of research assessment, 2003). A posteriori, la differenza tra il caso della Gran Bretagna e quello dell'Italia è spiegabile dalla diversa storia delle istituzioni universitarie. Sappiamo, infatti, che molte delle attuali università inglesi (corrispondenti a circa il 50% de laureati) hanno assunto questo nome, e la conseguente autonoma potestà di conferire le lauree, solo nel 1992. Precedentemente erano istituzioni parauniversitarie ai cui docenti, fortemente impegnati nella didattica, non era richiesto di condurre ricerca scientifica. Il passaggio di queste istituzioni al rango di università non ne ha immediatamente modificato la natura. Le valutazioni dei RAE, con poche eccezioni, hanno confermato le differenze tra le vecchie università e

quelle che fino al 1992 erano chiamate "Polytechnics". Al contrario il sistema italiano non ha mai previsto una diversificazione tra università di solo insegnamento e università di insegnamento e ricerca. In particolare, in linea di principio, a tutti i docenti è richiesto un impegno nella ricerca scientifica.

La prevedibilità teorica che ricerca scientifica buona ed eccellente fosse svolta in tutte le sedi, non rende meno importante la conferma empirica di questo risultato. Ad esempio, l'ipotesi di una facile, o almeno possibile, trasformazione del sistema universitario italiano in un sistema "dualistico" nel quale le sedi universitarie si distinguono in due o più classi sulla base dei livelli raggiunti dalla ricerca scientifica risulta indebolita (anche se certamente non esclusa) da questo risultato. Ricordiamo che la trasformazione del sistema universitario italiano in un sistema "dualistico" è stata negli ultimi anni fortemente caldeggiata da ambienti della Confindustria e da molti commentatori. Dopo le valutazioni del VTR 2001-2003 chi ancora sostiene l'opportunità di andare in questa direzione dovrebbe anche indicare gli strumenti amministrativi e soprattutto finanziari per attuare la trasformazione desiderata.

2 L'opportunità di un riesame critico della VTR

Fin qui i risultati macroscopici dell'esercizio di valutazione, cui si deve comunque riconoscere un grande successo sul piano organizzativo.

C'è però qualcosa di più che possiamo e quindi dovremmo imparare dalla VTR 2001-2003. Dovremmo esaminare con cura e nei dettagli la metodologia adottata e le difficoltà rilevate dai diversi comitati e dalle strutture di ricerca, per poter valutare proposte di cambiamenti allo scopo di migliorare l'esercizio nel futuro.

Ho già citato la "Review of research assessment" che ha seguito le analoghe valutazioni svolte in Gran Bretagna (RAE2001) a cura di una commissione presieduta da Sir Gareth Roberts. Questo riesame delle procedure e degli esiti è stato preceduto da un invito a presentare commenti e proposte sui RAE2001 rivolto a tutti i possibili interessati. L'invito ha dato luogo a ben 420 risposte che sono state attentamente vagliate. Sono stati inoltre utilizzati uno studio a cura della Science Policy Research Unit dell'Università del

Sussex, un programma di nove convegni organizzati da RAND Europe², 44 riunioni per consultare le istituzioni interessate ed infine riunioni pubbliche in diverse università.

La VTR 2001-2003 italiana, dopo la presentazione da parte del Ministro, nel gennaio 2006, è stata invece accolta da un generale disinteresse, a parte un dibattito intenso svoltosi nel mese di febbraio 2006 sul periodico elettronico lavoce.info e alcuni interventi che hanno chiesto l'immediata applicazione dei risultati nella distribuzione dei finanziamenti alle università e agli enti di ricerca. Non c'è stato invece alcun tentativo del Ministero di riesaminare criticamente le procedure ed i risultati di questa valutazione.

Non avrebbe alcun senso un vero riesame critico in un intervento breve, e sostanzialmente disinformato, come il mio. Vorrei però ripercorrere, anche se con inevitabile superficialità, la metodologia adottata nella valutazione, indicando alcuni, ma solo alcuni, punti critici che dovrebbero essere affrontati con maggiore attenzione e approfondimento. Mi limiterò ad esaminare brevemente il processo di valutazione dei "prodotti" (in grandissima parte pubblicazioni) delle università e degli enti di ricerca, quel processo che ha dato luogo alle classifiche tuttora direttamente consultabili sul sito <http://www.vtr.cineca.it>.

3 L'impostazione generale della VTR

Ricordiamo prima di tutto l'impostazione generale dell'esercizio di valutazione, che si è svolto secondo il percorso descritto qui di seguito.

Sono stati determinati 14 ambiti di ricerca corrispondenti alle 14 aree che raccolgono i settori scientifico disciplinari del sistema universitario italiano ai fini delle elezioni del Consiglio Universitario Nazionale, e cioè:

- 01 Scienze matematiche e informatiche
- 02 Scienze Fisiche
- 03 Scienze Chimiche

² Nel suo sito, RAND Europe si definisce come "*an independent not-for-profit research institute whose mission is to help improve policy and decision-making through research and analysis*".

- 04 Scienze della Terra
- 05 Scienze Biologiche
- 06 Scienze Mediche
- 07 Scienze Agrarie e Veterinarie
- 08 Ingegneria Civile ed Architettura
- 09 Ingegneria Industriale e Informatica
- 10 Scienze dell'Antichità, Filologico-Letterarie e Storico-Artistiche
- 11 Scienze Storiche, Filosofiche, Psicologiche e Pedagogiche
- 12 Scienze Giuridiche
- 13 Scienze Economiche e Statistiche
- 14 Scienze Politiche e Sociali.

Alle 14 aree sono state aggiunte 6 aree speciali corrispondenti, a:

- 15a Scienze e tecnologie per una società dell'informazione e della comunicazione.
- 15b Scienze e tecnologie per la qualità e la sicurezza degli alimenti
- 15c Scienze e tecnologie dei nano-microsistemi
- 15d Scienze e tecnologie aerospaziali
- 15e Scienze e tecnologie per lo sviluppo e la governance sostenibili: aspetti economici, sociali, energetici ed ambientali.
- 15f Scienze e tecnologie per la valutazione e valorizzazione dei beni culturali.

Per ognuna delle aree è stato nominato un Comitato d'Area (denominato "panel" in angloburocratese³) con il compito di valutare i "prodotti" appartenenti all'area.

E' stato chiesto alle università ed enti di ricerca

(denominati "strutture") che si sono sottoposti alla valutazione di scegliere un numero prestabilito di "prodotti" realizzati nel triennio 2001-2003. Il numero era commisurato al numero medio di ricercatori impiegati dalla struttura nel triennio. In pratica le università dovevano presentare, complessivamente, un numero di prodotti pari ad un quarto del loro personale docente e gli enti di ricerca un numero di prodotti pari alla metà dei loro ricercatori. Ogni prodotto selezionato doveva essere proposto dalla struttura per una delle 20 aree elencate sopra. Non era, ovviamente, necessario presentare prodotti per ognuna delle 20 aree.

Ai Comitati d'area è stato chiesto di valutare i prodotti presentati dalle strutture pervenendo a giudizi di merito complessivi. I Comitati dovevano avvalersi dei giudizi anonimi di esperti da loro stessi scelti, esterni ai comitati, e "anonimi" nel senso che i nomi di questi esperti non sarebbero stati divulgati al di fuori del comitato stesso, in analogia al sistema (denominato in angloburocratese "peer review"⁴) previsto per l'accettazione di articoli in una rivista scientifica. In effetti ogni "prodotto" avrebbe dovuto essere sottoposto al giudizio di almeno due esperti esterni. Ogni esperto era tenuto ad emettere un giudizio di merito complessivo classificando il "prodotto" in esame come "eccellente", "buono", "accettabile" o "limitato". Esamineremo nel seguito la definizione di queste classi di merito.

Sulla base dei giudizi espressi dagli esperti esterni il comitato d'area doveva classificare in via definitiva ogni prodotto, e, sulla base della classifica dei prodotti, doveva approvare una classifica (*ranking list* in angloburocratese) delle strutture in esame. I criteri inizialmente indicati dal CIVR per la formazione della classifica sono stati modificati a seguito delle osservazioni dei Comitati d'Area. I criteri finali prevedevano che si attribuisse ad ogni prodotto eccellente il coefficiente 1, ai prodotti buoni il coefficiente 0,8, a quelli accettabili il coefficiente 0,6 ed a quelli limitati il coefficiente 0,2. L'indice utilizzato per

³ I documenti del CIVR sono pieni di espressioni della lingua inglese, quasi sempre traducibili facilmente in italiano, ma a volte, come ad esempio "spin off", di incerto ed ambiguo significato. L'uso di parole straniere in documenti ufficiali dovrebbe essere evitato, non in nome del purismo linguistico, ma per un'esigenza di chiarezza. Molto spesso le parole inglesi utilizzate nei documenti italiani, private del loro naturale contesto linguistico, storico ed etimologico, risultano oscure anche a chi conosce la lingua inglese. L'uso di un'espressione inglese va quindi a scapito della trasparenza del processo amministrativo ed è censurabile alla luce dell'art. 97 della Costituzione italiana. Un caso paradigmatico è proprio quello della espressione "spin off" che ha un significato preciso nel contesto del diritto societario, mentre il verbo "to spin" ha diversi significati a seconda del contesto (filare, ruotare, ecc.).

⁴ Anche l'espressione "peer review" risulta parzialmente oscura se privata del suo contesto linguistico e culturale. Essa dovrebbe applicarsi quando l'esperto è un "pari", cioè un altro ricercatore che non è in alcun modo un superiore del ricercatore il cui lavoro è esaminato. Per questo alla fine, risulta più chiara e semplice l'espressione "esame da parte di esperti".

la graduatoria è la media così pesata delle valutazioni dei prodotti. Sono state anche previste, per ogni area, fino a quattro diverse classifiche e cioè classifiche diverse per le “megastrutture” che presentavano più di 75 prodotti, per le “grandi strutture” che presentavano da 50 a 74 prodotti, per le medie strutture che presentavano da 10 a 49 prodotti, ed infine per le “piccole strutture” che presentavano non più di 9 prodotti.

4 La scelta delle aree

Il primo passo di questa complessa procedura che sarebbe opportuno riesaminare, ed eventualmente modificare, è quello della scelta delle aree per la valutazione. In particolare colpisce la assoluta disomogeneità tra le prime 14 aree e le 6 aree speciali. Non è affatto chiaro su quali basi le diverse strutture, ed in particolare le università, avrebbero dovuto scegliere, per la presentazione di un prodotto, un'area tra le prime 14 o piuttosto una delle ultime sei. Si presume che per gli enti di ricerca titolari di progetti di ricerca che corrispondevano, in tutto o in parte, alle indicazioni contenute nel titolo delle ultime sei aree sia stato facile presentare i prodotti relativi alle ricerche inquadrati in quei progetti nelle aree speciali relative. Il problema appare di più difficile soluzione per i prodotti delle strutture universitarie dove la ricerca scientifica è prevalentemente “libera” e segue gli interessi, le inclinazioni, e la curiosità del ricercatore. Si è avuta pertanto una forte sovrapposibilità tra le tradizionali “aree CUN” e le aree speciali. Inoltre quasi tutte le sedi universitarie hanno presentato pochi “prodotti” attribuiti alle aree speciali. Nel complesso, ad esempio per l'Area 15a, 43 strutture hanno presentato prodotti, ma di queste una sola grande struttura ha presentato il 27% dei prodotti. Accanto a 7 medie strutture sono state contate 35 piccole strutture, cioè strutture che hanno presentato meno di 10 prodotti. Di queste ben 18 hanno presentato meno di 4 lavori. La Presidente del Comitato d'Area, per l'Area 15a, Bianca Falcidieno, giustamente osserva nella sua relazione che “è difficile pensare che sotto la soglia di dieci prodotti il giudizio sui singoli prodotti possa essere significativamente aggregato in un giudizio generale sulla struttura.” Gli stessi problemi si presentano per l'Area 15b alla cui valutazione ha partecipato una sola grande struttura, 3 medie

strutture e 25 piccole strutture alle quali si possono ben applicare le osservazioni espresse dal Comitato dell'area 15a. Per l'Area 15c le grandi e medie strutture sono un po' più presenti (1 megastruttura, una grande struttura e 9 strutture medie), e le piccole strutture si riducono a 29 su un totale di 40. Tuttavia, come evidenziato dalla relazione finale dell'Area 15c, c'è stata una forte sovrapposizione con le competenze delle altre aree, in particolare con l'Area 02 Scienze Fisiche, dove si ritiene siano stati presentati molti prodotti riferibili all'Area 15c. Anche nel caso dell'Area 15d il Presidente del comitato d'area ritiene che molti prodotti riferibili a quest'area siano stati presentati in altre aree (presumibilmente Area 02 Scienze Fisiche e 09 Ingegneria Industriale e Informatica), e conclude che in tal modo si è persa la possibilità di avere un quadro unitario delle ricerche in questa area. L'Area 15e è risultata molto eterogenea, al punto che, secondo il comitato d'area, le competenze dello stesso comitato non erano sufficientemente estese. Per l'Area 15f, dei 194 prodotti presentati 92 appartenevano in uguale misura ai settori di Archeologia e Arte ed Architettura, i restanti 102 erano divisi in 32 settori diversi. La maggior parte (56%) dei prodotti proveniva da 26 piccole strutture, i restanti da sei strutture medie. Anche in questo caso si applica l'osservazione del Comitato dell'Area 15a, sulla significatività di un giudizio generale su una piccola struttura sulla base di un campione di prodotti sotto la soglia di dieci.

In conclusione sembrerebbe opportuna una riflessione in merito al ruolo che le aree speciali hanno svolto nell'esercizio di valutazione. Questa riflessione dovrebbe innanzitutto prendere spunto dalle osservazioni dei comitati d'area. E' certamente ipotizzabile che una valutazione complessiva della ricerca svolta dagli enti pubblici su alcuni temi ritenuti importanti possa seguire una strada diversa dalla valutazione relativa al livello generale della ricerca scientifica nelle diverse sedi universitarie.

Per quanto riguarda le 14 aree tradizionali si deve rilevare che, storicamente, la formazione delle cosiddette “aree CUN” corrispondenti alle 14 aree è stata molto influenzata da divisioni, più o meno artificiali, che risalgono alle tradizionali divisioni delle “cattedre” nelle diverse “facoltà”. In particolare, come rilevato dal Presidente del Comitato dell'Area 01, non ha senso che l'Informatica risulti divisa in due aree diverse: la stessa Area 01 Matematica e Informatica e l'Area

09 Ingegneria Industriale e Informatica. Un confronto con le *unit of assessment* dei RAE inglesi rivela altre peculiarità del sistema italiano. Ad esempio, nella VTR italiana la Statistica, è valutata all'interno dell'Area 13 Scienze Economiche e Statistiche, che comprende anche la matematica per l'economia. Al contrario, nei RAE 2001, inglesi la Statistica apparteneva alla stessa *unit of assessment* della Ricerca Operativa. Lo stesso avviene nei RAE del 2008, per i quali è previsto un "panel" per i lavori in "Matematica Pura", "Matematica Applicata", "Statistica e Ricerca Operativa" e "Informatica e Scienza dei Calcolatori", cui fanno capo quattro *unit of assessment* una per ognuna di queste discipline.

Nel complesso le 14 aree italiane sono molto più eterogenee delle 68 *unit of assessment* degli esercizi inglesi del 2001, e delle 67 *unit of assessment*, previste dai RAE del 2008. Certamente aree non troppo frammentate consentono un confronto più ampio con discipline affini e consentono una migliore valutazione della ricerca interdisciplinare. Tuttavia, a questo proposito, dovrebbe essere presa in esame la soluzione adottata per i RAE del 2008, che è quella di un sistema gerarchico che comprende un numero più limitato di *panel* che coordinano i lavori di diversi *subpanel* (cfr. www.rae.ac.uk).

5 La scelta dei prodotti da parte delle strutture.

Osserviamo prima di tutto che le diverse strutture, in particolare le università, hanno dovuto, improvvisamente e celermente, operare scelte per le quali non erano preparate. E' quindi probabile che non siano sempre state operate le scelte più utili per una buona valutazione della sede. In particolare le 20 aree di riferimento non corrispondevano sempre a strutture universitarie abituate a scelte di merito scientifico (ad esempio nelle chiamate) come le facoltà ed i dipartimenti. E' probabile che non sia stato facile per una sede universitaria scegliere la produzione migliore per un'area facente capo a più dipartimenti e/o facoltà. E' probabile quindi che, in molti casi, si sia ricorso a criteri di proporzionalità. Un altro problema era probabilmente costituito dalle aspettative sulla natura del giudizio da parte dei comitati d'area. Ad esempio nella relazione dell'Area 01 "Matematica e informatica" viene

esplicitamente detto che i lavori presentati dal CNR non erano probabilmente i lavori migliori prodotti, in matematica ed informatica, da questo ente. Una spiegazione di questa possibile scelta sbagliata può essere dovuta al fatto che il CNR abbia scelto, per l'Area 01, i lavori pubblicati sulle riviste di più alto Impact Factor, nella convinzione che questo indice avrebbe guidato i giudizi del Comitato d'Area. Infatti l'Impact Factor medio dei lavori presentati dal CNR nell'Area 01, risulta il più alto, nel confronto con le "grandi strutture", mentre alla fine è il CNR a risultare l'ultima struttura nella graduatoria delle grandi strutture per questa area.

Queste difficoltà "interne" della scelta dovrebbero essere affrontate dalle diverse sedi, le quali indubbiamente rifletteranno sui risultati conseguiti in questa prima valutazione.

Ci sono però alcuni problemi tecnici che dovrebbero essere affrontati in sede nazionale, dall'organismo che organizzerà la prossima valutazione.

Osserviamo prima di tutto che la regola sul numero dei "prodotti" da presentare non riguardava le singole aree, ma l'intera struttura. Questo fatto avrebbe potuto consentire, a priori, la strategia di eccedere nel numero dei prodotti presentati dalle aree più forti a scapito del numero dei prodotti delle aree più deboli.

In realtà non è credibile che nel clima di confusione, mancanza di informazioni precise sui criteri di giudizi dei comitati d'area, e fretta imposta dalle scadenze, che ha caratterizzato la fase della scelta dei prodotti, ci siano state strutture in grado di scegliere consapevolmente questa strategia. Tuttavia, in un articolo dell'otto febbraio 2006, pubblicato da Lavoce.info, Giampaolo Arachi ed Alberto Zanardi analizzano i risultati della VTR per l'Area 13 Scienze Economiche e Statistiche, mostrando che, per questa area, le diverse università hanno presentato un numero di prodotti che, con poche eccezioni, differiva per eccesso o per difetto da quello strettamente legato al numero dei ricercatori. Osservano inoltre che se alle sedi che hanno presentato un numero di prodotti maggiore di quello atteso si togliessero, fino a raggiungere il numero atteso, i lavori risultati peggio classificati e alle sedi che hanno presentato meno lavori del numero atteso si aggiungessero ipotetici lavori fino a raggiungere il numero atteso classificandoli come il peggior classificato dei lavori presentati, la graduatoria delle sedi universitarie per l'Area 13

ne risulterebbe completamente stravolta. L'ipotesi sottintesa da quest'ultima osservazione è che il processo di selezione delle sedi fosse in grado di prevedere l'esito della valutazione dei lavori da parte dei diversi Comitati d'area. Si tratta di un'ipotesi non molto credibile. Tuttavia le osservazioni di Arachi e Zanardi mettono in luce un problema che non è opportuno trascurare.

Ricordiamo, a questo proposito che i RAE2001 della Gran Bretagna, cui si ispira la VTR italiana, non fissavano il numero dei prodotti presentabili dalla sede o da una particolare area presente nella sede. Infatti per ciascuna delle 68 *unit of assessment* non si presentavano "prodotti" ma piuttosto "docenti" per ognuno dei quali erano indicati non più di quattro lavori degli ultimi cinque anni (sette anni per le discipline umanistiche). A priori una sede poteva scegliere di presentare un solo docente, o tutti i docenti. Tuttavia il "premio", cioè il finanziamento ordinario distribuito sulla base dei risultati dei RAE, era determinato dal livello di qualità (approssimativamente mediano) raggiunto dai docenti presentati, ma anche dal numero dei docenti presentati. Questo significa che il numero dei docenti, e quindi dei prodotti presentati era limitato superiormente ed inferiormente da forti incentivi finanziari, anziché come nel caso italiano da una precisa prescrizione.

Vale la pena forse di osservare a questo punto che sarebbe stato opportuno prevedere un minimo di lavori presentati dalla sede, per ciascuna area. Nell'ipotesi di mantenere la regola che si presenta un lavoro per ogni quattro docenti universitari in ruolo, sarebbe però difficile fissare il minimo a dieci (come sembra proporre il Presidente del Comitato dell'area 15a). Tuttavia attribuire una valutazione, ed un posto in una graduatoria, in relazione ad una di 20 possibili aree di ricerca, ad una sede universitaria che presenta per quell'area solo un articolo scientifico, sembra abbastanza privo di senso. Un compromesso ragionevole potrebbe essere quello di non consentire per ciascuna area di ricerca la presentazione di un numero di lavori inferiore a quattro. Con le regole attuali quattro lavori corrispondono a 16 docenti presenti nell'area in questione. La questione del numero dei lavori presentati in un'area determinata verrà affrontata anche nel seguito quando discuteremo l'affidabilità della graduatoria per le piccole strutture.

6 Il processo di valutazione da parte dei Comitati d'Area

Secondo le prescrizioni del CIVR ogni prodotto presentato è stato assegnato dal Comitato d'Area a due relatori esterni che hanno avuto il compito di classificare il prodotto come "eccellente", "buono", "accettabile" o "limitato". Le definizioni di questi termini erano:

Eccellente: "Il prodotto si colloca nel 20% superiore in una scala di valori condivisa dalla comunità scientifica internazionale"

Buono: "Il prodotto si colloca nell'intervallo 60%-80% "

Accettabile: "Il prodotto si colloca nell'intervallo 40%-60% "

Limitato : "Il prodotto si colloca nel 40% inferiore"

Nonostante l'apparente precisione fornita dalle percentuali queste definizioni non sono univoche e possono portare a valutazioni diverse da parte di diversi relatori. Tra l'altro non è affatto chiaro quale è il denominatore cui si riferisce la percentuale. Facendo riferimento, ad esempio, agli articoli su riviste scientifiche, la percentuale si riferisce a tutti gli articoli pubblicati nell'area su tutte le riviste scientifiche? O si escludono dal denominatore di riferimento i lavori contenenti ricerche di routine? Con la prima interpretazione è probabile che in un'area ben rappresentata in Italia, come ad esempio la matematica, almeno la metà degli articoli pubblicati dai docenti italiani debbano essere considerati "eccellenti". Se invece ci si limita agli articoli pubblicati a livello internazionale ma che contengono risultati in qualche modo significativi, la proporzione di lavori "eccellenti" pubblicati dai docenti italiani dovrebbe essere minore. Insomma il riferimento alla comunità scientifica internazionale può costituire una valida indicazione, almeno per le discipline scientifiche, quando si parla della "scala di valori condivisa", ma è quanto meno incerto quando si introducono percentuali, almeno per le aree nelle quali la posizione dell'Italia non è lontana da quella degli altri paesi industriali.

Definizioni più precise avrebbero potuto fare riferimento per la classifica alla produzione scientifica effettivamente presentata dalle diverse strutture, limitando, ad esempio l'eccellenza al 5% migliore della produzione scientifica in esame. Il

Comitato d'Area avrebbe avuto in questo caso un ben definito universo di riferimento e sarebbe stato costretto ad un preciso confronto tra i diversi prodotti. Inoltre il Comitato non avrebbe avuto nemmeno la tentazione di gonfiare i risultati della propria area per "patriottismo d'area", come è quanto meno possibile che sia avvenuto, nelle aree scientifiche interessate a favorire un aumento complessivo dei finanziamenti destinati alla ricerca nella propria area⁵.

Il riferimento percentuale ai lavori effettivamente presentati, per la definizione della qualità, avrebbe l'effetto di rendere impossibili i confronti tra aree diverse, che però con il sistema che è stato adottato, se pur possibili, risultano abbastanza inaffidabili.

Ai relatori è stato chiesto di valutare i prodotti, fornendo un giudizio specifico, sulla base delle seguenti caratteristiche: qualità, importanza, originalità, internazionalizzazione. Quest'ultima caratteristica ha in qualche caso condotto ad interpretazioni diverse da parte dei relatori esterni, come ha osservato il Comitato dell'Area 01. Certamente ci sono aree all'interno delle quali qualsiasi contributo scientifico di qualche rilievo dovrebbe essere diretto ad una comunità internazionale, potenzialmente in grado di apprezzarlo ed utilizzarlo. In questi casi l'eventuale esistenza di collaborazioni internazionali per il conseguimento del risultato diviene irrilevante ai fini della valutazione.

Si deve comunque osservare che anche le definizioni adottate per la classifica dei prodotti dai RAE2008, non sono particolarmente chiare. Anche in questo caso si fa riferimento al livello "internazionale" della ricerca si parla infatti di *world class research*. Tuttavia per i RAE2008 i criteri generali sono precisati per ciascuna delle *unit of assessment*, in modo coerente con la ricerca scientifica che si svolge nelle diverse aree. Un possibile suggerimento per un prossimo esercizio di valutazione potrebbe essere proprio quello di chiedere ai comitati di area di precisare ulteriormente i criteri di valutazione che

adotteranno.

7 La graduatoria conclusiva

Dopo aver raccolto i giudizi dei relatori, e partendo ovviamente da questi giudizi, i Comitati d'Area hanno deliberato sul giudizio finale (eccellente, buono, accettabile e limitato) da attribuire a ciascun prodotto. Infine hanno stilato le graduatorie delle sedi (fino a quattro graduatorie diverse, come si è detto). Certamente ci saranno state discussioni ed in alcuni casi posizioni di minoranza. E' probabile tuttavia che in molti casi il giudizio dei relatori sia stato fatto proprio dal Comitato d'Area. Questa procedura, in apparenza del tutto ragionevole, nasconde però il fatto che i relatori esterni non hanno giudicato comparativamente i prodotti, mentre il risultato finale, quando si giunge alla classifica, è un giudizio comparativo. In altre parole è probabile che in molti casi il giudizio comparativo finale sia il frutto di giudizi espressi sulla base di scale di merito diverse da parte di diversi relatori esterni. E' anche da ritenere che in presenza di un numero sufficientemente ampio di prodotti da esaminare le differenze di giudizio dei diversi relatori si siano compensate e si sia pertanto arrivati ad un giudizio comparativo sostanzialmente valido per le diverse sedi.

Il problema è più complesso per le sedi che presentavano pochi lavori in un'area. Per illustrare questo problema mi riferisco alla graduatoria delle piccole strutture per un'area che mi è familiare, l'Area 01 "Matematica ed Informatica", che è riportata in calce a questo articolo. In questa graduatoria la Scuola Normale Superiore compare come la decima struttura per qualità della ricerca. La prima è la SISSA e cioè la Scuola Internazionale Superiore di Studi Avanzati, a pari merito (in termini di punteggio (*rating* in angloburocratese) con l'Università del Sannio. Il punteggio dell'Università del Sannio per la matematica risulta di 1,00 mentre quello della Scuola Normale, dove notoriamente operano diversi matematici di ottimo livello internazionale è di 0,87. Che credibilità hanno questi risultati? Chi può credere che in una qualsiasi graduatoria che rifletta la qualità, l'Università del Sannio dove opera un solo matematico attivo nella ricerca, a livello internazionale, superi di gran lunga la Scuola Normale?

⁵ Si deve osservare che in Italia una parte notevole dei finanziamenti destinati alle ricerche dei docenti universitari sono in realtà assegnati a grandi istituti nazionali. Questa circostanza rende meno plausibile la concorrenza tra "strutture" per ottenere maggiori finanziamenti e più comune la concorrenza tra aree, o sottoaree rappresentate dai grandi istituti nazionali.

Vediamo quindi come può essersi prodotto questo risultato paradossale. Seguendo le regole, l'Università del Sannio presenta un lavoro, dell'unico matematico attivo a livello internazionale tra i pochissimi docenti di matematica presenti nei suoi ruoli. Questo lavoro, in collaborazione con un autore straniero, viene giudicato eccellente da due relatori esterni. Ne risulta un punteggio pieno sul 100% dei lavori. La Scuola Normale, sempre secondo le regole, presenta tre lavori, anch'essi in collaborazione con autori esterni alla struttura. Uno di questi lavori viene giudicato eccellente e gli altri due buoni⁶. E' estremamente improbabile che i relatori che hanno giudicato il lavoro della università del Sannio abbiano anche giudicato uno o più lavori della Scuola Normale, si tratta infatti di ambiti subdisciplinari diversi. Sembra certo che nessun esperto abbia, in effetti, paragonato i lavori della Scuola Normale giudicati "buoni" con il lavoro presentato dalla Università del Sannio. In assenza di una valutazione comparativa dei lavori che senso ha pervenire ad un punteggio e a una graduatoria?

Si deve aggiungere che poiché non è precisato l'universo cui si riferiscono le percentuali usate nella definizione del punteggio attribuito ai lavori, ci si deve aspettare che queste percentuali siano state interpretate in modo diverso dai diversi relatori esterni.

Queste considerazioni ci dovrebbero quanto meno suggerire di non prendere in considerazione ai fini della stesura della graduatoria strutture che presentano un numero esiguo di "prodotti", nella speranza che con un alto numero di prodotti le idiosincrasie dei diversi relatori si compensino. Quanto alto dovrebbe essere questo numero? Certamente dovremmo escludere tutte le strutture che presentano un solo prodotto (nell'Area 01 ci sono cinque strutture che presentano un solo prodotto, nell'Area 15a ce ne sono sei su un totale di 43 strutture valutate). Forse dovremmo seguire le indicazioni della Presidente del Comitato d'Area 15a che suggeriva di prendere in considerazione solo le strutture che presentavano almeno dieci lavori. Si escluderebbero così dalla valutazione tutte le "piccole" strutture nella

definizione data dalle istruzioni del CIVR (per l'Area 01 le piccole strutture sono 27). Dieci prodotti corrispondono però ad un numero di docenti molto più alto, in effetti, secondo le regole adottate, corrispondono ad almeno 40 docenti. Ma forse bisognerebbe rivedere la regola che fissa così rigidamente il numero dei lavori presentati in relazione al numero dei docenti presenti.

Queste critiche non dovrebbero più applicarsi quando il numero dei prodotti presentati è alto e questi prodotti sono giudicati da un alto numero di relatori esterni. Uso però il condizionale perché è ben vero che, in presenza di un alto numero di prodotti esaminati, dovrebbe avvenire una compensazione tra interpretazioni diverse delle qualifiche di "eccellente", "buono", "accettabile" e "limitato", tuttavia è sempre possibile che, in qualche caso, la compensazione non abbia luogo e che, ad esempio, una struttura si trovi immeritatamente in cima o in fondo alla graduatoria per effetto della mancanza di un vero giudizio comparativo.

C'è da chiedersi a questo punto se valga la pena di insistere come ha fatto il CIVR perché la valutazione si concluda con un punteggio ed una graduatoria. Ricordiamo che la graduatoria e l'effetto "campionato" sono stati giudicati negativamente dalle autorità inglesi che, proprio per questo, hanno previsto per i RAE2008 un esito in termini vettoriali piuttosto che scalari, allo scopo di rendere difficile, se non impossibile, la formazione di una graduatoria univoca.

Infine dobbiamo osservare che un giudizio comparativo tra lavori presentati da sedi diverse potrebbe aver luogo in presenza di Comitati d'Area con competenze sufficientemente ampie da poter prescindere dal giudizio dei relatori esterni, e soprattutto in presenza di Comitati d'Area disposti ad assumersi responsabilità dirette, anziché nascondersi dietro il giudizio di relatori anonimi.

L'apparente burocratizzazione del processo di valutazione, che, ad esempio, impone di sollecitare il giudizio di due relatori esterni può aver favorito a questo proposito una eccessiva timidezza. Ricordiamo che, invece, i "panel" dei RAE2008 possono interpellare relatori esterni, ma non è obbligatorio farlo. Inoltre, con o senza relatori esterni è il "panel" che si assume la responsabilità delle decisioni.

⁶ Sembra che, secondo le regole interne del Comitato dell'Area 01, per ottenere il giudizio di "eccellente" fosse necessario che ambedue i relatori dichiarassero il prodotto eccellente.

8 Conclusioni

In conclusione, un riesame delle procedure e degli esiti del VTR 2001-2003 dovrebbe prendere in considerazione almeno i seguenti punti:

- 1) Una possibile ridefinizione delle aree che escluda prima di tutto le aree speciali e ridisegni le aree secondo criteri di affinità scientifica indipendenti dalla collocazione delle “cattedre” nelle facoltà.
- 2) Una più precisa definizione dei giudizi di merito (“eccellente”, “buono”, ecc.) che faccia eventualmente riferimento in termini percentuali ai prodotti presentati per ciascuna area e non ad una non meglio definita produzione internazionale.
- 3) Un chiarimento del criterio che fa riferimento all'internazionalizzazione.
- 4) L'indicazione di un minimo di prodotti presentati perché una struttura partecipi alla

valutazione per un'area determinata.

- 5) Una riflessione sull'opportunità di un esito che consiste in un punteggio “scalare” ed una graduatoria conseguente, anziché un punteggio “vettoriale”.

In aggiunta possiamo raccomandare di evitare nel futuro l'uso nei documenti ufficiali di termini inglesi anche se consacrati da tempo dall'angloburocratese. Questi termini corrono il rischio di svolgere la funzione del “latinorum” di don Abbondio, in un ambito, quello della valutazione, dove è particolarmente necessaria la chiarezza.

Riferimenti bibliografici

Review of research assessment (2003), Report by Sir Gareth Roberts to the UK funding bodies, May 2003.

	Struttura	Rating ¹	Prodotti pesati ²	Giudizi di merito ³						Prodotti	Ricercatori ETP	Grado di proprietà (medio) ⁵		IF ⁶ medio	prodotti con IF ⁶
				E%	E	B	A	L	NV ⁴			dei prodotti	dei prodotti eccellenti		
1	SISSA	1.00	3.00	100	3	0	0	0	0	3	6.83	0.83	0.83	1.40	2
2	Univ. SANNIO	1.00	1.00	100	1	0	0	0	0	1	2.83	0.50	0.50	0.66	1
3	Politecnico BARI	0.91	6.40	57	4	3	0	0	0	7	14.00	0.79	0.75	1.18	7
4	Univ. INSUBRIA	0.90	5.40	50	3	3	0	0	0	6	11.00	0.75	0.89	1.09	4
5	Univ. VERONA	0.90	3.60	50	2	2	0	0	0	4	10.17	0.58	0.42	1.09	4
6	Univ. ROMA TRE	0.89	8.00	44	4	5	0	0	0	9	16.50	0.74	0.79	1.18	9
7	Univ. SIENA	0.89	6.20	43	3	4	0	0	0	7	18.50	0.79	0.83	0.59	7
8	Univ. BASILICATA	0.88	4.40	40	2	3	0	0	0	5	13.17	0.60	0.75	1.21	5
9	Univ. BRESCIA	0.88	4.40	40	2	3	0	0	0	5	9.33	0.77	0.67	1.21	5
10	Scuola NORMALE Sup.	0.87	2.60	33	1	2	0	0	0	3	6.83	0.44	0.50	1.65	3
11	Univ. CATTOLICA	0.87	2.60	33	1	2	0	0	0	3	6.83	0.61	0.33	1.19	3
12	Univ. Politecnica MARCHE	0.85	3.40	25	1	3	0	0	0	4	8.00	1.00	1.00	1.00	2
13	Univ. CAMERINO	0.83	5.00	33	2	3	1	0	0	6	11.67	0.75	0.88	1.51	6
14	Univ. FERRARA	0.83	6.60	25	2	5	1	0	0	8	18.17	0.79	1.00	1.17	8
15	Univ. VENEZIA	0.80	5.60	29	2	3	2	0	0	7	11.00	0.47	0.42	1.01	7
16	Univ. MODENA e REGGIO EMILIA	0.80	7.20	22	2	5	2	0	0	9	23.50	0.60	0.75	1.05	7
17	Univ. CHIETI-PESCARA	0.80	4.00	20	1	3	1	0	0	5	8.00	0.60	0.50	0.85	4
18	Univ. PIEMONTE ORIENTALE	0.80	4.80	17	1	4	1	0	0	6	12.67	0.79	1.00	2.20	6
19	IUAV	0.80	0.80	0	0	1	0	0	0	1	3.83	1.00	0.00	0.00	0
20	Univ. SASSARI	0.80	0.80	0	0	1	0	0	0	1	1.33	0.50	0.00	0.68	1
21	Univ. URBINO	0.80	0.80	0	0	1	0	0	0	1	3.33	0.50	0.00	0.91	1
22	Univ. CALABRIA	0.76	6.80	33	3	3	2	1	0	9	17.00	0.53	0.57	1.23	9
23	Univ. REGGIO CALABRIA	0.70	2.80	25	1	2	0	1	0	4	4.00	0.88	1.00	0.39	3
24	Univ. MESSINA	0.69	6.20	11	1	4	3	1	0	9	20.50	0.71	0.50	0.86	9
25	Univ. CASSINO	0.60	0.60	0	0	0	1	0	0	1	3.33	0.67	0.00	0.38	1
26	Seconda Univ. NAPOLI	0.53	3.20	33	2	0	1	3	0	6	12.00	0.78	0.75	0.53	5
27	Univ. BERGAMO	0.40	0.80	0	0	0	1	1	0	2	3.00	0.59	0.00	0.54	2
Valori del raggruppamento ⁷		0.81	106.20	34	44	64	16	7	0	131	277.32	0.70	0.73	1.10	120

Tabella 1. Graduatoria delle piccole strutture per l'area 01 "Matematica e Informatica"

La ricerca scientifica in Economia Aziendale: primi risultati del processo di classificazione delle riviste scientifiche

Alessandro Carretta

Università di Roma "Tor Vergata"

carretta@uniroma2.it

Sommario

La presente nota illustra il progetto di classificazione delle riviste scientifiche realizzato dagli studiosi di Economia aziendale nell'ambito dell'Accademia Italiana Di Economia Aziendale (AIDEA).

Essa prende le mosse da una presentazione delle discipline riconducibili all'Economia aziendale, descrive presupposti, obiettivi, principali contenuti e fasi del progetto, presenta i primi risultati e prospetta i passi ancora da compiere per il completamento dell'iniziativa.

1 L'economia aziendale

L'economia aziendale comprende sei raggruppamenti disciplinari: economia aziendale, economia e gestione delle imprese, finanza aziendale, organizzazione aziendale, economia degli intermediari finanziari, scienze merceologiche, per un totale di oltre 1700 docenti.

I sei raggruppamenti presentano numerosi elementi comuni, ma anche alcune differenze significative, riconducibili almeno ai seguenti profili:

- i quadri scientifici di riferimento, focalizzati su misure e controllo, management, intermediari e mercati finanziari, ecc.;
- l'orientamento all'internazionalizzazione;
- l'ancoraggio alla "storia" della disciplina;
- la contaminazione con altre discipline;
- la propensione alle verifiche empiriche e il ricorso alle diverse metodologie di indagine;
- la dimensione, la struttura demografica e la distribuzione per Facoltà e Atenei.

2 Il progetto di classificazione delle riviste scientifiche

Il progetto viene avviato dall'Accademia Italiana di Economia Aziendale (AIDEA), sotto la

presidenza di Roberto Cafferata, nel quadro di una riflessione sulla valutazione della ricerca (post VTR 2001-03).

L'iniziativa si accompagna ad un forte cambiamento "culturale" delle discipline aziendali, che si trovano (area 13) in posizione di "confine" fra *hard science* ("naturali", aree 1-9) e *soft science* ("sociali", aree 10-12 e 14).

Il progetto può avere importanti ricadute anche sul funzionamento dei meccanismi di selezione e valutazione delle risorse, nella convinzione che la valutazione della ricerca sia utile, importante e centrale. In ogni caso, la riflessione in atto va inserita in un approccio sistemico di valutazione dell'attività accademica complessiva, con ricorso ad una pluralità di strumenti di indagine dei risultati, relativi non solamente alla produzione scientifica ma anche alle performance didattiche ed al presidio complessivo della relazione con gli studenti. Occorre inoltre assicurare una adeguata coerenza di "trattamento" delle specificità di contesto, che per alcune discipline sono irrinunciabili, nonché prestare attenzione alla articolazione ed alla varietà dei "prodotti" della ricerca, costituiti non solamente da articoli pubblicati sulle riviste, ma anche da libri e da altri output editoriali.

In particolare, AIDEA si è proposta, attraverso il progetto, di:

- aiutare i ricercatori nella selezione delle riviste su cui pubblicare i propri lavori di ricerca;

- supportare le strutture di ricerca nella selezione dei contributi da presentare nei prossimi esercizi di valutazione nazionale;
- supportare i soggetti attivi nella valutazione nei diversi momenti in cui si articola l'attività universitaria, ed anche nell'ambito della comparazione con i lavori prodotti nell'ambito di scienze sociali diverse da quella economico-aziendale;
- contribuire ad una maggiore valorizzazione e divulgazione della produzione scientifica italiana.

3 Le principali tappe del progetto

E' stato costituito un gruppo di lavoro AIDEA "misto" di colleghi senior e junior, composto da 35 persone, che ha tenuto otto riunioni fra l'ottobre 2005 e il giugno 2007, con il coordinamento di Angelo Riccaboni e Francesco Favotto.

Il progetto ha riguardato sia le riviste internazionali sia quelle italiane.

Per quanto concerne le riviste internazionali, è stata creata una "lista di riferimento AIDEA", articolata in cinque aree tematiche di classificazione (con individuazione di un referente per ogni area), non coincidenti con i sei raggruppamenti disciplinari "ufficiali".

Le aree tematiche sono:

- Banking & Finance
- Public Sector Management
- Accounting & Control
- Organization
- Management and Strategy (general management, marketing, innovation, operations, strategy, governance)

Per ciascuna area tematica la classificazione, su quattro livelli di qualità decrescente - da "A" a "D" -, si è basata su autonomi criteri di analisi e valutazione mediante combinazione di indici bibliometrici, studi e classificazioni presentate su riviste scientifiche, JQL e Journal ranking vari, meccanismi di selezione dei contributi utilizzati dalle riviste in esame (modalità di referaggio).

L'attività di classificazione è stata accompagnata da confronti puntuali con le comunità scientifiche di riferimento e con le sedi universitarie già in possesso di una qualche esperienza e sperimentazione, nonché da una gestione "flessibile" dei (pochi) casi di riviste presenti in più aree e valutate in modo difforme.

I risultati raggiunti sono stati oggetto di un

seminario di approfondimento, rivolto ai colleghi di AIDEA, che si è svolto a Bologna nel giugno 2007, approvati nel Consiglio direttivo di AIDEA del settembre 2007 e presentati ufficialmente nel Convegno annuale dell'Associazione, nell'ottobre 2007.

Per quanto riguarda le riviste italiane si è deciso di predisporre e proporre a queste ultime un processo di "accreditamento AIDEA", su base strettamente volontaria, basato sul soddisfacimento di alcuni criteri di "qualità", tra i quali:

- la presenza nei Comitati Editoriale e/o Scientifico di accademici autorevoli italiani e stranieri;
- l'attribuzione della direzione della rivista ad un accademico con elevato profilo scientifico;
- la presenza fra gli obiettivi della rivista di un chiaro riferimento all'avanzamento delle conoscenze scientifiche;
- il ricorso per la selezione degli articoli da pubblicare ad un doppio referaggio anonimo richiesto a studiosi esperti specializzati;
- una rendicontazione pubblica, su base periodica, dei nominativi dei referee e dei risultati dell'attività di selezione (articoli pervenuti, accettati, pubblicati, e relative tempistiche);
- il mantenimento di un archivio dei referaggi effettuati;
- adeguate modalità di diffusione e penetrazione della rivista;
- la presenza di un abstract in inglese per ogni articolo pubblicato;
- la disponibilità a fornire un rendiconto annuale su richiesta di AIDEA.

4 I primi risultati ed i prossimi passi

Il progetto ha consentito di pervenire a cinque liste di riferimento (una per area tematica), complessivamente con oltre 700 riviste, da un minimo di 70 a un massimo di 140 circa per area, condivise dalle comunità scientifiche di riferimento e da AIDEA (il documento è disponibile sul sito dell'Associazione <http://www.accademiaaidea.it>).

Esso ha inoltre dato un forte impulso alla "sensibilizzazione" sul tema di tutti i colleghi di Economia Aziendale. Diversi Dipartimenti hanno già "adottato" ufficialmente le liste di

riferimento AIDEA per una autovalutazione della ricerca e degli studiosi. Alcuni raggruppamenti disciplinari hanno avviato specifiche riflessioni ed azioni sul tema. Ad esempio, in Economia degli intermediari finanziari si sta attualmente lavorando per favorire l'adesione agli standard scientifici internazionali di alcune riviste italiane di prestigio e per lanciare iniziative di "allenamento" rivolte ai colleghi più giovani (una collana di working paper con criteri editoriali internazionali, diretta da Daniele Previati. Si veda il sito <http://www.adeimf.it>).

Le attività previste da AIDEA sono in prospettiva le seguenti:

- l'inserimento di nuove riviste, in un'ottica di manutenzione corrente delle liste;
- l'affinamento dei criteri di classificazione;
- il passaggio ad una lista unica, che accolga,

secondo criteri condivisi, le cinque liste di area;

- l'avvio e la gestione della lista di riviste italiane accreditate;
- l'avvio della valutazione dei libri;
- l'avvio di un monitoraggio sistematico sull'utilizzo delle liste da parte di gruppi di ricerca o Dipartimenti o Facoltà o Atenei per la distribuzione delle risorse, per le scelte di reclutamento, per le politiche di progressione delle carriere, ecc.;
- l'integrazione del ranking delle riviste scientifiche con lo sviluppo di un "data base" - sul tipo del progetto LUCA del Dipartimento di Studi Aziendali e Sociali dell'Università di Siena - in cui raccogliere con sistematicità le pubblicazioni internazionali degli aziendalisti italiani.

La valutazione della ricerca in economia

Pietro Reichlin

LUISS G. Carli
preichlin@luiss.it

Sommario

E' stato osservato che i criteri di valutazione della ricerca nelle scienze economiche tendono a scoraggiare le innovazioni e a privilegiare le linee di ricerca aderenti al "mainstream". In realtà, questa critica non sembra tenere conto degli sviluppi recenti delle discipline economiche, della diffusione di internet e della perdita di influenza delle università di maggiore prestigio.

1 Considerazioni generali

L'Italia spende per ricerca e sviluppo 15 miliardi di euro, l'1% del nostro PIL. Circa un terzo di questa cifra, costituita interamente da fondi pubblici, è assorbito dall'università. Una spesa non trascurabile (benché relativamente bassa nei confronti di molti paesi dell'OCSE) che sottraiamo a scopi sociali alternativi non meno importanti. Cinque miliardi di euro possono essere pochi o tanti, a seconda della qualità e della quantità del prodotto che generano. La valutazione della ricerca è, quindi, un obiettivo fondamentale delle politiche pubbliche. Questo è riconosciuto in tutto il mondo industrializzato, ove si è affermato il principio della *accountability* nei confronti degli *stakeholders*, cioè coloro che si aspettano, a diverso titolo, risultati concreti e ricadute positive. Ogni paese avanzato si è dotato, nel corso del dopoguerra, di organismi preposti alla valutazione della ricerca scientifica. L'esempio più noto è quello della NSF, una fondazione del governo federale, creata in USA nel 1947. La NSF ha avuto un ruolo fondamentale nella selezione dei progetti di ricerca portati avanti nelle università americane e nella promozione della ricerca di base, soprattutto quella che difficilmente può essere finanziata dall'industria privata. Nei paesi in cui il sistema universitario è interamente pubblico, l'importanza dei sistemi di valutazione è ancora maggiore. A questi, infatti, è attribuito il ruolo di distribuire fondi alle istituzioni, oltre che ai ricercatori, premiando le sedi universitarie più

produttive. Un sistema molto rigido basato su indici bibliometrici che determinano premi e penalizzazioni finanziarie per le università si è affermato recentemente nel Regno Unito.

Gli organi di valutazione operano in base al principio indispensabile che la qualità della ricerca possa essere misurata, anche se con un margine di errore. Per "qualità" si intende principalmente il fatto che la ricerca abbia "rilevanza". Tuttavia, la rilevanza non è un requisito facilmente verificabile, sia perché l'impatto di una ricerca può realizzarsi solo nel futuro, sia perché i suoi risultati possono essere complessi, non "tangibili", interpretati in modo difforme o, addirittura, manipolati. Queste difficoltà implicano che gli organismi di valutazione si basino sul giudizio degli esperti e, più in generale, su sistemi di *peer review*, cioè sul giudizio di *referee* provenienti dagli ambienti accademici.

Il sistema di *peer review* ha un raggio di applicazione più ampio di quello relativo alla valutazione. Nel mondo accademico si è affermato il principio che il lavoro scientifico acquisisce valore e credibilità se e solo se è pubblicato su un numero limitato di riviste accreditate presso il sistema accademico internazionale, dopo essere stato vagliato da *referee* anonimi. Le riviste hanno quindi il doppio ruolo della diffusione e della certificazione del lavoro scientifico. Quest'ultimo è tanto più importante quanto maggiore è la severità con cui è valutato dai *referee*, e quanto maggiore è l'impatto della rivista sulla comunità scientifica internazionale. Ne deriva che il curriculum scientifico di un ricercatore è un ottimo

segnale per inferire la qualità trascorsa della sua ricerca, e la lista delle pubblicazioni generate da un progetto di ricerca sono un buon modo per valutare l'output del progetto medesimo. Per questo motivo la *bibliometria*, cioè la definizione di una graduatoria delle riviste scientifiche in base a indici basati sull'*impact factor* o sul *citation index*, ha assunto una grande importanza nei sistemi di valutazione della ricerca.

Molti osservatori hanno espresso validi dubbi sull'affidabilità degli indici bibliometrici. E' vero che un prodotto pubblicato su una rivista ampiamente diffusa nella comunità scientifica è necessariamente un prodotto rilevante? La risposta è sicuramente negativa. Molti contributi pubblicati su riviste di fascia elevata (nella scala bibliometrica) negli anni passati sono poi stati dimenticati, mentre altri prodotti pubblicati su riviste di fascia media sono ancora ampiamente citati dalla letteratura corrente. Tuttavia, nell'ambito delle scienze economiche, è difficile trovare contributi che abbiano avuto un impatto duraturo e rilevante sulla comunità scientifica e che, nel contempo, non siano stati pubblicati sulle riviste internazionali di maggiore impatto.

Per evitare di commettere errori di valutazione è senz'altro opportuno non fidarsi degli indici d'impatto in modo cieco ed esclusivo. Si dovrebbe, in particolare, utilizzare la *peer review* senza farsi eccessivamente condizionare dall'*impact factor* di un prodotto o della rivista dove esso è stato pubblicato. Ma questo non è un consiglio privo di difetti. Ad esempio, supponiamo che, nell'ambito di una commissione di valutazione nazionale, un esperto A esprima un giudizio negativo su un prodotto pubblicato da una rivista di grande impatto. Questa stessa rivista ha pubblicato il prodotto in questione proprio perché ha ottenuto un parere positivo di un esperto B. Un osservatore terzo può ragionevolmente pensare che, in virtù della reputazione della rivista, l'esperto B sia altrettanto affidabile dell'esperto A. Ci troveremmo, quindi, in presenza di una difformità di vedute tra esperti, e non è chiaro cosa dovremmo concludere.

L'esempio serve per chiarire il seguente punto di vista. La critica ai criteri bibliometrici può essere di due tipi. Il primo riguarda *il modo in cui tali criteri sono costruiti*. Basta un'occhiata superficiale alla letteratura per rendersi conto del groviglio di problemi che incontra chiunque voglia stilare un ordinamento delle riviste scientifiche. Ad esempio: come tenere conto del fatto che

riviste più antiche tendono ad essere citate con maggiore frequenza delle riviste recenti? Il secondo tipo di critica riguarda *l'affidabilità del meccanismo di peer review* o di certificazione generato dalle pubblicazioni scientifiche.

In questo scritto non intendo discutere delle metodologie statistiche utilizzate per determinare gli indici bibliometrici. Viceversa, mi concentrerò sul problema generale dell'affidabilità della certificazione definita dalle riviste scientifiche, cioè sul problema dell'affidabilità dei sistemi di *peer review* internazionali e standardizzati, *versus* i sistemi di *peer review* nazionali o locali o segmentati per sottosettori e scuole di pensiero.

2 Critiche comuni ai criteri di peer review

La *peer review* ed i sistemi bibliometrici sono stati ampiamente discussi e criticati nella letteratura specializzata. Queste critiche hanno avuto una particolare enfasi in Italia, nell'ambito delle discipline economiche. L'efficacia di un sistema di valutazione presuppone una certa omogeneità di pensiero, cioè la condivisione dei criteri che definiscono la rilevanza delle linee di ricerca e la qualità dei risultati. Alcuni hanno messo in dubbio l'esistenza di questa omogeneità nell'ambito delle scienze economiche e hanno affermato la necessità di salvaguardare la sopravvivenza di un approccio specifico della tradizione italiana.

A tale proposito, vale la pena leggere la relazione finale del CIVR per l'area di Scienze Economiche e Statistiche, ove viene evidenziato un contrasto tra i componenti del *panel*. Il prof. Pasinetti si preoccupa di evitare che alcune aree di ricerca siano sottoposte ad una valutazione basata su criteri standardizzati e, in particolare, sul grado di diffusione internazionale delle riviste scientifiche di riferimento.

E' possibile condividere gli argomenti di Pasinetti e, nel contempo, accettare un sistema di *peer review* come quello adottato dal CIVR e da ogni altro organismo di valutazione della ricerca? Il quesito è lecito perché sembra contraddittorio affermare che la gran parte dell'accademia internazionale rifiuti una certa area di ricerca (come evidentemente accade se i suoi prodotti scientifici non trovano spazio sulle riviste più accreditate) e, nello stesso tempo, proporsi di sottoporre quella stessa area di ricerca al giudizio di una commissione di accademici

sufficientemente rappresentativa della comunità scientifica internazionale.

Sembra di capire che, secondo Pasinetti, esistono in Italia linee di ricerca rilevanti che non trovano spazio nelle riviste scientifiche internazionali più accreditate solo perché difformi o anomale rispetto al “*mainstream* accademico”. Pasinetti riconosce che il *mainstream* sia dominante a livello internazionale e che tale dominanza sia cresciuta negli ultimi decenni. Egli sviluppa anche un’interessante interpretazione di tale fenomeno. Secondo Pasinetti il *mainstream* non sarebbe altro che un prodotto della scuola di pensiero nata negli USA, e il suo successo sarebbe legato all’efficacia (e importanza) delle scuole di dottorato americane. Inoltre, Pasinetti osserva giustamente che, diverse volte nella storia, la maggioranza degli scienziati ha contrastato lo sviluppo di aree di ricerca che, successivamente, si sono dimostrate valide e tali da soppiantare le vecchie teorie. Si può dire, in sostanza, che affermare la validità generale della *peer review* e dei criteri bibliometrici tende a scoraggiare le rivoluzioni scientifiche.

Tuttavia, negare l’importanza del giudizio espresso da una maggioranza di scienziati ha anche conseguenze controverse. Naturalmente, qui non è in discussione la libertà di ricerca. Ma se accettassimo le tesi di Pasinetti, dovremmo forse rinunciare alla valutazione della qualità dei prodotti scientifici e accettare che qualsiasi linea di ricerca sia egualmente sovvenzionata? Potremmo cioè essere portati ad adottare una forma estrema di relativismo scientifico, secondo cui esistono solo opinioni o interpretazioni?

In realtà, questa non sembra essere una conclusione accettabile dallo stesso Pasinetti, il quale, avendo deciso di far parte di una commissione di valutazione insieme ad altri colleghi, è costretto ad accettare come evento normale e probabile l’esistenza di una maggioranza avversa alle aree di ricerca da lui preferite. Si noti, inoltre, che l’argomento principale usato da Pasinetti contro la *peer review* internazionale (il fatto che essa tende a scoraggiare le rivoluzioni scientifiche) non è affatto un argomento contro la valutabilità dei prodotti scientifici. Una rivoluzione scientifica è infatti tale se afferma un paradigma nuovo e superiore al paradigma antecedente.

Infine, si deve sottolineare che, in un sistema accademico che dipende unicamente da fondi pubblici, un approccio “relativista” (che rinuncia a valutare i prodotti scientifici su base possibilmente

oggettiva) ha l’effetto di frenare l’evoluzione ed il progresso degli approcci scientifici. Le università tendono a dividersi in scuole di pensiero e settori disciplinari immutabili, la cui consistenza è dettata dalla tradizione, anziché dal loro successo.

La critica alla *peer review* basata sulle pubblicazioni scientifiche ha tuttavia una sua legittimità ed autorevolezza.

Alcuni studi hanno esaminato la possibilità che si verifichino reciproci favoritismi tra i componenti di particolari gruppi accademici o che i componenti dei *panel* di valutazione cerchino di evitare il confronto aperto tra diversi punti di vista allo scopo di mantenere posizioni di potere. Altri studi hanno esaminato i difetti dei metodi quantitativi usati per la valutazione. In particolare, le principali osservazioni che sono state avanzate possono essere così riassunte.

- 1) La *peer review* tende a privilegiare i ricercatori più vicini al *mainstream* accademico e gli indici bibliometrici tendono ad esaltare la notorietà della rivista di riferimento, piuttosto che la qualità della ricerca (effetti di rete, *cross citation*).
- 2) Il fatto che l’inglese sia la lingua dominante nel sistema accademico internazionale tende a privilegiare la ricerca prodotta nei paesi anglosassoni.
- 3) La diffusione di una rivista scientifica è inversamente proporzionale al suo livello di specializzazione o tecnicismo.
- 4) Gli studi quantitativi ed empirici tendono ad essere citati più frequentemente degli studi a carattere qualitativo o teorico.
- 5) Gli indici di impatto sono poco sviluppati o del tutto assenti nel caso delle monografie.

3 I fatti

Ma veniamo ora ai fatti. Qui vorrei discutere l’allarme circa la presunta omogeneità di pensiero nell’ambito dell’economia e la mancanza di sufficienti incentivi alla revisione del pensiero dominante. Nella realtà, questo allarme sembra essere smentito da alcuni studi e osservazioni sull’evoluzione della disciplina.

Innanzitutto, chiunque abbia seguito la tipologia di contributi pubblicati sulle principali riviste scientifiche (le cosiddette riviste di *mainstream*), non può non essere colpito dalla crescita nel grado di frammentazione della disciplina, dalla crescita del numero di riviste e dall’esplosione di nuovi filoni di ricerca. Ad esempio, chi mai avrebbe

potuto prevedere nel 1970 l'affermazione della *Behavioral Economics*, dell'*Economia Sperimentale*, della *Neuro-Economics*, della *Teoria dei Giochi Evoluzionistica*, della *Prospect Theory*, dell'analisi del *Capitale Sociale*, della *New Institutional Economics* ecc. Il paradigma neoclassico basato sulla razionalità e sull'individualismo metodologico sembra fortemente discusso o contaminato.

Inoltre, si è assistito ad un consistente incremento di riviste nel corso degli ultimi trent'anni. Ad esempio, è impressionante quanto sia elevato il numero di riviste nate dopo il 1980. Tra queste il *Journal of Economic Behavior and Organization*, il *Journal of Economic Perspectives*, *Games and Economic Behavior*, il *Journal of the European Economic Association*, la *Review of Political Economy*, la *Review of Austrian Economics*, *Structural Change and Economic Dynamics*, il *Journal of Evolutionary Economics*, il *Journal of Public Economic Theory*. Alcune hanno rapidamente scalato le classifiche basate sugli indici bibliometrici. Gli stessi titoli mostrano una pluralità di approcci e di vedute alternative su come debba essere impostata la ricerca economica. Nell'ultimo decennio si è poi aggiunta una lunga lista di periodici elettronici. L'opportunità di pubblicare non si è certo ridotta: il settore delle scienze economiche è quello che conosce il maggior incremento annuale di riviste scientifiche. Questa diversificazione e frammentazione della disciplina non è facilmente dimostrabile con indici oggettivi. Kim, Morse e Zingales, in uno studio del 2006 classificano le pubblicazioni di maggiore impatto sulle riviste scientifiche prominenti dal punto di vista del soggetto (come definito dai codici EconLit), della metodologia e delle istituzioni di appartenenza, nell'intervallo di tempo 1970-2002. Ecco alcuni risultati interessanti di questa ricerca.

- 1) Il secondo articolo più citato è un contributo di Kahnemann e Tversky su *prospect theory*, cioè un approccio del tutto nuovo alla teoria delle decisioni in condizioni di incertezza, alternativo alla teoria classica dell'utilità attesa;
- 2) il numero dei contributi più citati classificabili nell'ambito della micro e della macroeconomia si dimezza, mentre sale enormemente quello riferibile alla teoria della crescita e dello sviluppo economico;
- 3) il numero di contributi di tipo teorico, tra quelli più citati, crolla dal 77 all'11%, mentre i

contributi di tipo empirico salgono dall'11 al 60%.

Qualunque sia lo scetticismo con cui si voglia accogliere le classificazioni degli autori, non si può evitare di concludere che la disciplina ha attraversato negli ultimi venti anni una fase di profondo mutamento, sia riguardo alle metodologie che all'oggetto dell'analisi.

Evidentemente, non è vero che la *peer review* ed il potere delle grandi istituzioni accademiche americane ha l'effetto di conservare la dominanza di un approccio scientifico.

Altri studi recenti hanno rilevato i seguenti fenomeni:

- a) la reputazione delle quattro riviste scientifiche più importanti, misurata in base ad un indice aggregato che tiene conto dell'impatto e delle citazioni, si è alquanto ridotta dal 1970 al 1990, passando dal 44,1% al 26,1% (Laband e Piette (1994));
- b) nell'insieme degli autori che pubblicano sulle riviste più importanti, la proporzione di coloro che provengono dalle università di élite statunitensi si è ridotto sostanzialmente dal 1970 al 2000. In particolare, l'appartenenza ad una di queste università non sembra fornire, oggi, un vantaggio comparativo degno di nota (Kim, Morse e Zingales (2006), Ellison (2007)).

Questo secondo gruppo di osservazioni ha spinto alcuni ad affermare che la *peer review* stia perdendo importanza o autorevolezza. In particolare, secondo Ellison (2007):

The necessity of going through the peer-review process has lessened for high status authors: in the old days peer-reviewed journals were by far the most effective means of reaching readers, whereas with the growth of the Internet high status authors can now post papers online and exploit their reputations to attract readers.

Sia Kim, Morse e Zingales (2006) che Ellison (2007) attribuiscono allo sviluppo di Internet e alla crescita dei mezzi di diffusione elettronica un ruolo chiave per spiegare la perdita di impatto delle istituzioni di élite sulle riviste di maggiore prestigio. Bisogna infatti ricordare che la ricerca scientifica si avvale in grande misura della qualità dell'ambiente in cui essa viene prodotta e, in particolare, dell'apprendimento reciproco e delle comunicazioni tra ricercatori. Nel passato ciò richiedeva prossimità "fisica" (cioè avere dei bravi colleghi nello stesso corridoio). Oggi questo è meno importante. Tutti possiamo accedere ai

lavori preliminari, o i *working papers* dei nostri colleghi, ovunque essi si trovino. Il costo di procurarsi ciò che di meglio è stato scritto su un qualunque argomento è enormemente diminuito in ogni parte del pianeta. In sostanza, le riviste scientifiche hanno oggi principalmente un ruolo di certificazione e assai meno il ruolo della diffusione del sapere.

Gli USA hanno senza dubbio una posizione dominante nelle scienze economiche. In gran parte questo si deve al fatto che la disciplina ha avuto, a partire dal dopo guerra, una svolta in senso quantitativo ed analitico, non familiare nella tradizione europea. Tuttavia, questa svolta è stata ormai condivisa ampiamente in ogni parte del mondo.

Si deve anche osservare che la distanza tra gli USA e l'Europa sta diminuendo. In uno studio accurato basato su un insieme di sistemi di ordinamento basati sul numero di pubblicazioni scientifiche e indici di impatto, Coupé (2007) ha mostrato che il numero di istituzioni accademiche europee incluse tra le prime 100 del mondo è raddoppiato dalla metà degli anni '80 al 2000, passando da circa 15 a circa 30 (il numero esatto dipende dai criteri di ordinamento). Purtroppo la situazione dell'Italia non è molto buona. Nessuna università del nostro paese è compresa nel gruppo di testa. A mio parere, ciò non dipende tanto dall'assenza di economisti di valore, quanto, piuttosto, dalla dispersione dei ricercatori di successo su un ampio numero di università. Queste ultime non riescono a fare politiche di reclutamento in funzione della massimizzazione dell'output scientifico. Si assiste, piuttosto, a criteri di reclutamento che mirano alla conservazione di un equilibrio tra sotto-settori scientifici e scuole di pensiero.

4 Osservazioni conclusive

Esistono metodologie alternative a quelle basate sugli indici bibliometrici e sulla *peer review* internazionale per valutare in modo sufficientemente trasparente e standardizzato la qualità e la rilevanza dei prodotti scientifici? Su questo esiste già un'ampia letteratura su cui non mi soffermo per difetto di spazio e di competenza.

Vorrei, tuttavia, riconsiderare la principale tra le osservazioni critiche da me citate in precedenza. Cioè che la *peer review* tende a privilegiare i

ricercatori più vicini al *mainstream* accademico e che gli indici bibliometrici tendono ad esaltare la notorietà della rivista di riferimento, piuttosto che la qualità della ricerca. Tale osservazione può essere definita come il rischio che una linea di ricerca "valida" o "rilevante" possa essere bocciata perché non gode di adeguate citazioni o non vanta lavori pubblicati su riviste di prestigio e di diffusione internazionale.

Come ho cercato di argomentare in precedenza, questo tipo di errore non ha carattere pervasivo nella situazione corrente. L'abbattimento dei costi di comunicazione e la facilità con cui si diffonde il pensiero scientifico hanno ridotto l'influenza dei luoghi di produzione scientifica che avevano, nel recente passato, una posizione dominante (ad esempio, le due Cambridge, Chicago, Oxford). Tuttavia, non possiamo escludere che questo errore possa essere commesso. Esso potrebbe essere ulteriormente limitato se riuscissimo ad affermare i seguenti principi.

Originalità. Tra i criteri di valutazione del CIVR o organizzazioni simili dovrebbe avere particolare peso il grado di originalità del prodotto.

Interdisciplinarietà. I panel di area potrebbero essere composti anche da accademici o studiosi esperti di discipline diverse. Ciò porterebbe maggiore enfasi sugli aspetti interdisciplinari dei prodotti e costringerebbe coloro che difendono la qualità di un lavoro a "spiegarne" la rilevanza generale.

Dibattito. Le istituzioni accademiche dovrebbero incoraggiare il confronto tra scuole di pensiero. In particolare, coloro che, pur operando nell'ambito di aree di ricerca marginali, sono convinti di meritare una maggiore attenzione da parte dell'accademia e delle riviste scientifiche di prestigio dovrebbero essere interessati al massimo confronto ed al dibattito più aperto. Tale confronto potrebbe essere incoraggiato, ad esempio, promuovendo riviste scientifiche specializzate nelle controversie, oltre che nel tentativo di comunicare con un pubblico non del tutto specializzato. Una cosa del genere è tra gli scopi del *Journal of Economic Perspectives*. La Società Italiana degli Economisti potrebbe sponsorizzare un'iniziativa simile.

Qualunque sia il valore dei prodotti scientifici che si pongono in alternativa al *mainstream* accademico, l'unico modo in cui essi possono emergere fuori dal "ghetto" in cui sono confinati è quello di cercare il massimo confronto con la comunità dei ricercatori. La storia del pensiero

economico dell'ultimo secolo dimostra che la disciplina ha subito rilevanti mutamenti e contaminazioni con discipline contigue. Tra queste ultime, la matematica, la statistica, la giurisprudenza, la sociologia, le scienze politiche e, da ultimo, le scienze cognitive e la neurologia. A tratti si ha l'impressione che le riviste più accreditate siano più accoglienti nei confronti di contributi che utilizzano metodologie e approcci non comuni, piuttosto che nei confronti di contributi tradizionali.

Riferimenti bibliografici

- Coupé, T. (2007), *Revealed Performances: Worldwide Ranking of Economists and Economic Departments 1969-2000*, ECARE, Université Libre de Bruxelles.
- Ellison, G. (2007), *Is Peer Review in Decline?*, NBER W.P. 13272.
- Kim, E. H., A. Morse e L. Zingales (2006), *What Has Mattered to Economics Since 1970*, CRSP Working Paper No. 612.
- Laband, D., e M. Piette. (1994), *The relative impacts of economics journals: 1970-1990*. *Journal of Economic Literature* 32: 640-666.

Valutazione della ricerca e valutazione delle riviste scientifiche in ambito statistico

Benito V. Frosini

Dipartimento di Scienze statistiche
Università Cattolica del Sacro Cuore - Milano
benito.frosini@unicatt.it

Sommario

La prima parte della relazione è dedicata a vari problemi concernenti la valutazione della ricerca in generale, e con riferimento a esperienze svolte in Italia, USA e Regno Unito. La seconda parte descrive una classificazione delle riviste di Statistica e Demografia, elaborata da una commissione nominata dalla Società Italiana di Statistica.

PRIMA PARTE

1 Valutazione della ricerca

Il tema della valutazione della ricerca è molto più ampio, e quasi indipendente dalla valutazione delle riviste scientifiche.

Molte pubblicazioni sono fuori dalle riviste: atti di convegni, lavori collettanei, monografie, rendiconti di accademie e simili, riviste culturali non strettamente scientifiche, quaderni di istituto o dipartimento (di regola in previsione di successiva pubblicazione su riviste).

Alcuni scopi della valutazione della ricerca:

(I) valutazione di un articolo sottoposto a una rivista per pubblicazione (o a convegno, o a opera collettanea);

(II) valutazione comparativa di due o più ricercatori: implica una valutazione sintetica dell'intera produzione scientifica;

(III) valutazione di istituzioni, tipicamente dipartimenti o istituti universitari.

Uno scopo abbastanza esplicito in recenti documenti ministeriali è quello di ricavare – dalla valutazione della ricerca e insieme ad altri “indicatori” – una graduatoria delle facoltà o dei corsi di laurea. Mi limito a dire che si tratta di una impostazione totalmente errata, per i seguenti motivi:

(a) la qualità della didattica – con esclusione dei corsi di dottorato – è scarsamente collegata con la qualità della ricerca;

(b) la “misurazione” della qualità della ricerca, basata su (1) valutazione positiva dei progetti presentati, e (2) valutazioni di istituzioni tramite l'esercizio VTR, è comunque scarsamente collegata con la reale “qualità della ricerca”, come si dirà meglio in seguito.

Il problema del collegamento della valutazione della ricerca con la graduatoria (o *rating*) delle riviste scientifiche, e più in generale con la sede di pubblicazione (es. rendiconti dei Lincei o dell'Istituto Lombardo di Scienze e Lettere, relazioni invitate negli Atti ISI ecc.), è sempre esistito. Tutti i valutatori hanno sempre attribuito qualche punto di merito al puro e semplice riconoscimento di una sede elevata di pubblicazione. E il motivo essenziale stava nella riconosciuta difficoltà a pubblicare in quelle sedi. Tale difficoltà dipendeva allora, come anche oggi, da due fattori collegati:

(a) il numero elevato di lavori sottoposti per la pubblicazione, che rende inevitabile un drastico sfortimento degli articoli arrivati, anche ove fossero tutti meritevoli;

(b) un rigoroso processo di revisione (*refereeing process*), che lascia filtrare solo lavori giudicati di buona qualità – oltre che adatti alla rivista.

Ne deriva che era ed è giustificata una preliminare classificazione degli articoli scientifici sulla base della sede di pubblicazione, assegnando a determinati articoli una valutazione almeno “buona”.

Ma si trattava e si tratta – per valutatori onesti –

solo della fase preliminare, o preparatoria, della valutazione. Infatti:

(a) in tutte le sedi di pubblicazione gli articoli presentano una notevole differenziazione qualitativa (la regola statistica è che “la variabilità nei gruppi è maggiore della variabilità fra i gruppi”), e non sarebbe corretto attribuire a tutti gli articoli di una rivista lo stesso *rating* (minimo, massimo, o medio di gruppo);

(b) anche per questo motivo non sembra facile graduare le stesse riviste fra loro, e soprattutto appare problematico passare da una eventuale graduazione (*scala ordinale*) a una *scala per intervallo*, o addirittura a una *scala di rapporto*;

(c) resterebbero comunque fuori – da questo *rating* preliminare – i moltissimi lavori non pubblicati su riviste scientifiche, che sarebbe ingiustificato e disonesto classificare come “inferiori”.

Prima conclusione: per una onesta e corretta valutazione della ricerca, le pubblicazioni devono essere lette, e valutate con sufficiente motivazione.

Negli ultimi vent’anni lo sviluppo e la diffusione degli *Impact Factors* (IF) e delle statistiche collegate (riviste citate e riviste che citano) ha messo a disposizione della comunità scientifica una quantità enorme di informazioni. Questo fatto ha spinto taluni:

(a) a utilizzare gli IF come indici della qualità di un articolo “medio” pubblicato da tali riviste;

(b) come ulteriore semplificazione, ad attribuire agli articoli pubblicati da tali riviste un “indice di qualità” proporzionale agli IF.

Non considerando le pubblicazioni prive di IF (che sono però la grande maggioranza), si sarebbe in tal modo raggiunto un risultato strabiliante:

(a) il giudizio sulla qualità dei lavori fatti da un individuo, o da un dipartimento (il massimo ammissibile di eterogeneità per l’impiego di IF), deriverebbe semplicemente da una somma di IF (eventualmente divisi per il numero degli autori);

(b) per lo stesso motivo, il giudizio da *soggettivo* (fatto a seguito di lettura da parte di un *referee*) diventa *oggettivo* (il numero dato dagli IF).

Se a *oggettivo* si sostituisce *formale*, si descrive meglio la procedura, che non ha nulla di *oggettivo*. Infatti, il minimo che si può dire è che un articolo su una rivista con IF = 0,8 può essere migliore di un articolo su una rivista con IF = 1,6; addirittura, un articolo su una rivista senza IF (vi sono molte riviste di qualità che non hanno uscite regolarissime, e pertanto non possono aspirare ad avere un IF) può essere migliore dei due tipi di

articoli citati. E non dimentichiamo che per argomenti “di frontiera”, studiati nel mondo da pochissimi ricercatori, è difficilissimo pubblicare su una rivista con IF elevato, indipendentemente dalla qualità del lavoro.

Due spunti da realtà di altri paesi.

a) U.S.A.: riferimento recentissimo: IMU-Net 24: July 2007.

Fra le *News*:

Committee on Quantitative Assessment of Research

The International Council of Industrial and Applied Mathematics (ICIAM), the Institute of Mathematical Statistics (IMS), and the International Mathematical Union (IMU) have formed a Committee of “Quantitative Assessment of Research” that will investigate various aspects of the quantitative assessment of mathematics. The Committee will, in particular, look into impact factors and similar ways to measure research output.

The Committee consists of:

Robert Adler (Haifa, Israel), appointed by IMS

Peter Taylor (Melbourne, Australia), appointed by ICIAM

John Ewing (Providence, USA), appointed by IMU

The Committee is expected to create a summary of its findings to be endorsed by the Executive Committees of ICIAM/IMS/IMU and to be published afterwards.

ICIAM, IMS, and IMU have formulated an aspirational charge to help set direction rather than prescribe the final outcome of the committee’s work.

Questo *aspirational charge* è stato firmato dalle tre organizzazioni, in vista della suddetta nomina, il 21 maggio 2007. Riprendo solo il primo e l’ultimo capoverso

(<http://www.ams.org/ams/impact-factor.pdf>).

The drive towards more transparency and accountability in the academic world has created a “culture of numbers” in which institutions and individuals believe that fair decisions can be reached by algorithmic evaluation of some statistical data; unable to measure quality (the ultimate goal), decision-makers replace quality by numbers that they can measure. This trend calls for comment from those who professionally “deal with numbers” – mathematicians and statisticians. Academic achievement is a complicated mix of contributions to research, teaching, and supervision of students, as well as

contributions to academic self organization/administration and to the scientific community. ICIAM, IMS, and IMU acknowledge that the evaluation of scientific quality is notoriously difficult, simplistic answers to complicated questions of judgment are unlikely. However, the committee is also asked to investigate whether it sees possible alternatives to measures based solely on citations that may help to evaluate research and academic achievement and indicate quality in a sensible way.

Questa impostazione mi trova totalmente d'accordo. Abbiamo una grande quantità di dati su molte riviste del nostro settore, che sarebbe assurdo respingere solo perchè taluni propongono l'IF come unico criterio – o come criterio principe – per la valutazione della ricerca, che è una proposta indifendibile. Il problema è come utilizzare e integrare questi dati, non per rendere automatica la valutazione (cosa da respingere con fermezza), ma come “aiuto sostanziale nella valutazione della qualità della ricerca”.

- b) United Kingdom: dato che il modello di riferimento per il primo VTR del CIVR è stato il RAE (*Research Assessment Exercise*) britannico, è importante ricavare alcune considerazioni comparative con tale esperienza. Il primo RAE fu tenuto nel 1986; l'ultimo si è concluso nel 2001; il prossimo – sesto della serie – è appena iniziato (la scadenza per le *submissions* era il 30 novembre 2007); i risultati dovrebbero essere noti entro la fine di quest'anno.

Osservazione 1

Una differenza fondamentale con la situazione italiana è che nel Regno Unito (e anche in USA) vi è una esplicita e voluta diversità di *mission* fra le 171 istituzioni universitarie: vi sono istituzioni votate prevalentemente alla ricerca, e istituzioni votate prevalentemente, o anche esclusivamente, alla didattica. Questa eterogeneità nota e dichiarata deve ovviamente trovare una relazione con la distribuzione delle risorse per la ricerca. In Italia, al contrario, tutte le università sono messe, almeno in linea di principio, sullo stesso piano, sia riguardo alla didattica sia riguardo alla ricerca.

Osservazione 2

Il RAE riguarda soltanto la ricerca (e anzi, solo una parte dell'attività di ricerca); dai suoi risultati dipende l'erogazione di una parte del finanziamento complessivo della ricerca.

Osservazione 3

L'ultimo RAE 2001, che si riferiva all'attività svolta nei 5 anni precedenti, ha costruito un indice di qualità (e quantità) sulla base di un massimo di 4 “prodotti” a persona riferiti al periodo quinquennale. Ci sono state proteste, e richieste di aumento di tale numero. Sempre mediando sulle persone, il VTR italiano ha ammesso un massimo di 1/2 prodotto a persona su un periodo di 3 anni. Il rapporto fra il numero di prodotti per persona per anno è 24/5, circa 5.

Attenzione: una differenza così grande non è solo quantitativa, essa non può non riflettersi *sugli stessi criteri di scelta* dei prodotti da sottoporre al RAE – o al VTR.

Osservazione 4

Vale la pena riportare le principali raccomandazioni contenute nella presentazione ufficiale del RAE 2001, e riprese quasi alla lettera nel RAE 2008:

(1) Judgments are made using the professional skills, expertise and experience of the experts; it is not a mechanistic process.

(2) All forms of research output (books, papers, journals, recordings, performances) are treated equally; panels are concerned only with the quality of the research. Similarly, all research (whether applied, basic or strategic) is treated equally.

(3) To assess submissions fairly and consistently within each UoA [Unit of Assessment], each panel draws up a statement describing its working methods and assessment criteria. These are published in advance of submissions being made.

Questo non è stato fatto nel caso VTR.

(4) 2.23) Evidence that research outputs have already been reviewed or refereed by peers may be used by panels as one measure of quality. However, the absence of such review may not, in itself, be taken to imply lower quality.

(5) Dalla declaratoria della UoA 24: Statistics and Operational Research:

3.17.12) . . . For all cited works ... the Panel will take note of the perceived editorial standards of journals, bearing in mind the character of research being described. The Panel will not use impact factors or citation indices. The Panel appreciates that the quality of material varies widely in some journals and edited books.

L'avvertenza è ancora più precisa nel RAE 2008: *The sub-panel will not use a rigid or formulaic method of assessing research quality. It will not use a formal ranked list of outlets, nor impact*

factors, nor will it use citation indices in a formulaic way.

3.17.17) In all cases quality will be measured by subjective assessment of actual intrinsic quality and potential impact.

2 Problema “costi-benefici”

Il RAE, ma anche il VTR (sia pure a un livello molto inferiore), richiedono un notevole impiego di risorse, soprattutto molte giornate di lavoro da parte di migliaia di ricercatori, nazionali e stranieri. Ne vale la pena?

Qualunque persona sensata si porrebbe un problema “costi-benefici”. Per quanto sostanzialmente giustificata dalla particolare struttura delle istituzioni universitarie britanniche, e dai buoni risultati ottenuti in passato riguardo agli incentivi dati allo sviluppo della ricerca, sulla stampa britannica sono apparse varie critiche. E in Italia?

Riguardo all’esperienza del VTR, sia per la struttura delle istituzioni universitarie italiane, sia per l’impostazione eccessivamente “al risparmio” di tutto l’esercizio di valutazione, sia per la mancanza di informazioni preliminari sui criteri da seguire, si deve affermare che è stato un grave errore e un grande spreco di risorse (proprio in riconoscimento del fatto che i componenti del CIVR e dei panel di area hanno lavorato con encomiabile impegno).

SECONDA PARTE

3 Valutazione delle riviste scientifiche in ambito statistico

Circa tre anni fa la Società Italiana di Statistica ha incaricato una commissione, da me presieduta, e formata da Graziella Caselli, Guido Consonni, Franco Peracchi, Maria Grazia Valsecchi, di preparare una “classificazione delle riviste di statistica”, con ampio mandato riguardo ai criteri di classificazione, che dovevano essere elaborati dalla stessa commissione.

La commissione si è riunita varie volte, ha elaborato in pieno accordo criteri operativi, e ha presentato alla SIS la sua relazione.

Come esposto più dettagliatamente nell’Allegato (riportato ai punti 5 e 6 nel seguito), la commissione ha fornito due classificazioni, una riguardante le riviste di Statistica in senso lato (Statistica, Calcolo delle probabilità, Econometria,

Statistica economica e sociale, Statistica medica, Statistiche applicate), e una seconda classificazione riguardante le riviste di Demografia. I criteri di classificazione che vengono esposti nell’Allegato solo con riferimento alle riviste di Statistica, sono stati recepiti anche per l’applicazione alle riviste di Demografia. Sostanzialmente, si è scelto di suddividere le riviste in tre classi di qualità decrescente:

AA: Riviste affermate, molto diffuse e di qualità mediamente eccellente;

A: Riviste con buona diffusione e di qualità mediamente buona/elevata;

B: Riviste con diffusione limitata e di qualità mediamente accettabile/discreta.

Non tutte le riviste considerate sono state inserite in una di queste classi, poiché comunque la qualità riconosciuta doveva essere *almeno* “accettabile/discreta”.

Si può osservare che, oltre alla qualità intrinseca delle riviste, hanno avuto un notevole rilievo la notorietà e la diffusione; fra la qualità intrinseca e la diffusione vi è naturalmente una certa connessione, ma certamente non una relazione biunivoca; non è stato però ritenuto opportuno individuare classi più dettagliate (ad esempio “riviste con limitata diffusione e di qualità elevata”); per quanto possibile, tale maggiore dettaglio classificatorio avrebbe richiesto uno sforzo operativo ben maggiore; si è preferito ricorrere a qualche approssimazione.

Alle tre classi ordinate **AA**, **A** e **B**, è stata aggiunta una classe **C**, che comprende “riviste che accolgono articoli di Statistica indirizzati a particolari ambiti applicativi”. Tale riferimento è ad articoli sostanzialmente metodologici (*Statistica applicata*, e non *Applicazioni della Statistica*), che non raramente trovano accoglienza su riviste scientifiche che hanno come specifica *mission* la presentazione di ricerche in un particolare ambito scientifico (ad esempio medico, informatico, sociologico, psicologico, economico, tecnologico ecc.). Una cernita di riviste di questo tipo, però non classificate a loro volta in classi ordinali, compare sotto la classe **C**. E’ del tutto ovvio che l’elenco proposto non può essere esaustivo, come certamente non esaustivi sono gli elenchi relativi alle classi **AA**, **A** e **B**; è anche ovvio che l’applicazione di criteri anche leggermente diversi avrebbero fatto includere alcune riviste della classe **C** in una delle precedenti classi ordinali.

E’ forse superfluo aggiungere che le scelte fatte

dipendono fortemente dalla composizione della commissione, che ha premiato soprattutto il contenuto metodologico delle ricerche, lasciando forse trasparire una non voluta preferenza per le ricerche di Statistica teorica rispetto alle ricerche di Statistica applicata. Appare invece del tutto ovvio che un confronto qualitativo fra una ricerca di Calcolo delle probabilità o di Statistica matematica con una ricerca di Statistica applicata può andare a favore di quest'ultimo tipo di ricerca. E' banale, ma vale la pena di ripetere la raccomandazione ricordata all'Osservazione 4, ripresa dal Research Assessment Exercise:

"All forms of research output ... are treated equally; ... panels are concerned only with the quality of the research. Similarly, all research (whether applied, basic or strategic) is treated equally".

Per i motivi appena ricordati, la commissione ritiene che la sua proposta non può che costituire la prima fase di una serie di successive verifiche, cui saranno chiamate altre commissioni negli anni a venire.

La commissione si è naturalmente documentata, nella fase iniziale dei suoi lavori, su ordinamenti e classificazioni già elaborati per discipline affini, sia apparsi su riviste, sia reperibili dai siti internet di alcune università americane. Gli articoli concernenti riviste economiche sono stati giudicati di scarsa utilità, dato che erano praticamente degli esercizi su dati IF e statistiche analoghe. Come è scritto nei "criteri di classificazione" della commissione, sono stati naturalmente utilizzati *anche* i dati IF tratti dalle elaborazioni di ISI-JCR, ove disponibili. Personalmente, ho stampato oltre 500 pagine (per una agevole consultazione e vari controlli) da ISI-JCR, sia per consultare i valori IF, sia per controllare le citazioni *da e per* altre riviste. Come ho scritto più sopra, sarebbe assurdo voler ignorare una tale quantità di dati, più o meno correlati alla qualità delle riviste scientifiche; come sarebbe però assurdo, ma solo estremamente comodo, ordinare le riviste solo sulla base di queste statistiche.

Che una classificazione qualitativa non possa farsi dipendere esclusivamente, e nemmeno prevalentemente, da indici IF e simili, è fin troppo ovvio, come è stato scritto da molti autori. Basterà in questa sede richiamare due informazioni tratte dal lavoro di Theoharakis-Skordia (T-S) "How do statisticians perceive statistics journals?", apparso su "The American Statistician", N. 2, 2003, pp. 115-123 (T-S hanno elaborato i risultati delle

2.190 risposte a un questionario sulla diffusione e l'utilizzazione delle riviste di Statistica, inviato a 12.053 statistici presenti in alcuni elenchi ufficiali di associazioni scientifiche). Anche se non si può negare un certo "effetto da auto-selezione", e si debba anche tener conto che il principale oggetto della ricerca non era la *qualità* delle riviste, alcuni risultati di questa ricerca sono tuttavia di qualche interesse. A p. 115 del lavoro T-S riconoscono (in accordo con la commissione) che i dati sulle citazioni "provide an aggregate measure of a journal's impact", ma sono solo parzialmente informativi riguardo alla qualità intrinseca. E ricordano, a proposito della loro utilizzazione per la valutazione di contributi scientifici, sia pure a livello *non* individuale *ma* di dipartimento o università, che nel celebrato Research Assessment Exercise del Regno Unito "impact factors or citation indices are not used to assess research output in journals". La seconda osservazione, sempre ripresa da T-S (p. 121), è che la correlazione tra IF e l'indice generale costruito da T-S (che grosso modo tende a misurare la "qualità percepita"), è 0,56, valore certamente non elevato (la quota di varianza spiegata è il quadrato di tale valore, cioè circa il 31%).

4 Fasi della classificazione delle riviste di statistica

4.1 Criteri di classificazione e fasce

In via preliminare la Commissione ha ripartito le riviste di Statistica in due categorie: (1) di contenuto statistico prevalentemente metodologico (in ambito propriamente statistico, probabilistico o econometrico), (2) di contenuto statistico indirizzato a particolari ambiti applicativi.

Successivamente ha predisposto per le riviste indicate al punto (1) una classificazione utilizzando i seguenti **criteri generali**:

1) Diffusione, notorietà, comitato editoriale, presenza in ISI-JCR, durata della rivista; livello qualitativo complessivo largamente condiviso dalla comunità scientifica internazionale.

2) Influenza sulle ricerche scientifiche degli statistici, testimoniata in particolare dai dati dell'Impact Factor (IF) in anni recenti. Va tuttavia rilevato che le valutazioni delle riviste non sono state basate solo - o principalmente - su IF, anche ove disponibile.

In relazione ai criteri generali sopra descritti, si propongono **tre fasce** ossia:

AA: Riviste affermate, molto diffuse e di qualità mediamente eccellente

A: Riviste con buona diffusione e di qualità mediamente buona/elevata

B: Riviste con diffusione limitata e di qualità mediamente accettabile/discreta.

Orientativamente, e seguendo – pur con qualche semplificazione – le indicazioni del CIVR, nella fascia **AA** si collocano le riviste il cui prodotto è generalmente *eccellente*, ossia si colloca nel 20% superiore nella scala di valore condivisa dalla comunità scientifica; nella fascia **A** quelle il cui prodotto è *buono/elevato* (si colloca nel segmento 50%-80%); nella fascia **B** quelle il cui prodotto è *accettabile/discreto* (si colloca nel segmento 30%-50%). Le riviste di livello inferiore non sono state inserite nella classificazione.

Si osservi che talune riviste non sono presenti nell'elenco allegato in quanto la loro durata è ancora insufficiente per esprimere un giudizio complessivo; anche per questo motivo, si auspica che il presente documento possa essere aggiornato periodicamente.

Per quanto riguarda le riviste che in generale sono indirizzate a particolari ambiti applicativi (Fascia **C**), la comparazione risulta alquanto più complessa a causa della eterogeneità e specializzazione delle riviste. In particolare, l'impiego dell'Impact Factor risulta inficiato dal contesto applicativo (ad esempio medico, informatico, sociologico, economico, tecnologico ecc.), così come i metodi impiegati sono di livello molto diverso in relazione alla disciplina di riferimento. Infine il numero di riviste di questo tipo è estremamente vasto, in rapida evoluzione, e meno controllabile.

4.2 Possibili utilizzazioni

Per gli organi di governo accademico centrali e periferici

1) Per orientare le politiche di distribuzione di risorse in relazione alle esigenze della ricerca scientifica

2) Per orientare le politiche di avanzamento di carriera, e più in generale le politiche di reclutamento del personale accademico.

3) Per orientare politiche di acquisto da parte delle biblioteche.

Per gli autori di articoli scientifici

E' utile che la comunità degli statistici italiani (e in particolare coloro che svolgono attività di ricerca scientifica) abbia a disposizione un elenco di riviste che segnali l'esistenza, largamente

percepita e condivisa dalla comunità scientifica internazionale, di una gerarchia qualitativa delle stesse. Questa informazione è particolarmente utile per i giovani che si avvicinano alla ricerca. Gli autori di articoli scientifici devono tendere a pubblicare su riviste di buona qualità e adeguata diffusione internazionale, ponendo particolare attenzione alle riviste dove più facilmente i lettori interessati possono presumere di trovare un certo tipo di ricerche.

E' utile impegnarsi seriamente per pubblicare su riviste molto diffuse e di elevata qualità per una serie di motivi collegati: 1) la nostra ricerca verrà maggiormente diffusa, conosciuta e dunque valorizzata nel mondo; 2) la valutazione sarà in generale più severa da un lato, ma più proficua in termini di riscontri e indicazioni dall'altro; 3) un articolo pubblicato su una rivista di alto livello sarà generalmente valutato molto positivamente nelle sedi appropriate (ad esempio concorsuali).

Per i valutatori

E' del tutto ovvio che ogni lavoro, specialmente in valutazioni comparative, deve essere esaminato nel merito specifico, indipendentemente dalla sede di pubblicazione. Tuttavia, nel caso di articoli pubblicati su riviste scientifiche, la particolare collocazione di un articolo può fornire una valida "informazione preliminare" sulla qualità dell'articolo, mutuata dalla qualità della rivista e dalla connessa serietà del processo di valutazione e revisione (*peer review*). Tale riferimento può essere particolarmente significativo quando lo scopo è la valutazione complessiva di un numero elevato di lavori, ad esempio pubblicati dai membri di un dipartimento in una serie di anni, e dovrebbe essere utilizzato nelle valutazioni comparative di procedure concorsuali. Ciò premesso, sembra opportuno rimarcare che per sovvertire una chiara gerarchia di merito fra più candidati, basata sulla predetta informazione preliminare, è necessario produrre una valutazione comparativa scrupolosamente argomentata e convincente.

5 Classificazione delle riviste di statistica

5.1 Fascia AA

Riviste affermate, molto diffuse e di qualità mediamente eccellente

Annals of Applied Probability

Annals of Probability
 Annals of Statistics
 Biometrika
 Econometrica
 Journal of Econometrics
 Journal of the American Statistical Association
 Journal of the Royal Statistical Society - B
 Probability Theory and Related Fields

5.2 Fascia A

*Riviste con buona diffusione e di qualità
 mediamente buona/elevata*

Advances in Applied Probability
 American Statistician
 Annales de l'Institut Henri Poincaré
 Bernoulli
 Biometrics
 Computational Statistics and Data Analysis
 Econometrics Journal
 Econometric Reviews
 Econometric Theory
 International Statistical Review
 Journal of Applied Econometrics
 Journal of Applied Probability
 Journal of Business and Economic Statistics
 Journal of Computational and Graphical Statistics
 Journal of Multivariate Analysis
 Journal of Statistical Planning and Inference
 Journal of the Royal Statistical Society - A
 Journal of the Royal Statistical Society – C
 (Applied Statistics)
 Journal of Theoretical Probability
 Journal of Time Series Analysis
 Oxford Bulletin of Economics and Statistics
 Review of Economics and Statistics
 Sankhya
 Scandinavian Journal of Statistics
 Statistica Sinica
 Statistical Science
 Statistics and Computing
 Statistics and Probability Letters
 Stochastic Processes and their Applications
 Stochastic Models
 Technometrics
 Test
 Theory of Probability and its Applications

5.3 Fascia B

*Riviste con limitata diffusione e di qualità
 mediamente accettabile/discreta*

Annals of the Institute of Statistical Mathematics

Australian and New Zealand Journal of Statistics
 Combinatorics Probability and Computing
 Communications in Statistics - A
 Communications in Statistics - B
 Communications in Statistics – Simulation and
 Computation
 Computational Statistics
 Estadística
 International Economic Review
 International Journal of Forecasting
 Journal of Financial Econometrics
 Journal of Forecasting
 Journal of Applied Statistics
 Journal of Nonparametric Statistics
 Journal of Statistical Computation and Simulation
 Journal of the Japan Statistical Society
 Journal of the Korean Statistical Society
 Korean Journal of Applied Statistics
 Methodology and Computing in Applied
 Probability
 Metrika
 Metron
 Model Assisted Statistics and Applications
 Revista Brasileira de Probabilidade e Estatística
 Revue de Statistique Appliquée
 Statistica
 Statistica Applicata (Italian Journal of Applied
 Statistics)
 Statistica & Applicazioni
 Statistical Methods and Applications (SMA-JISS)
 Statistical Modelling
 Statistical Papers
 Statistica Neerlandica
 Statistics
 Statistics Education Research Journal
 Stochastic Analysis and Applications
 Stochastics
 Structural Equation Modelling
 Student
 Studies in Nonlinear Dynamics and Econometrics
 Survey Methodology
 Teaching Statistics

5.4 Fascia C

*Riviste che accolgono articoli di Statistica
 indirizzati a particolari ambiti applicativi. Le
 riviste qui elencate sono chiaramente un
 sottoinsieme delle moltissime riviste scientifiche
 che accolgono articoli di Statistica Applicata,
 talvolta con apprezzabili contributi metodologici.*

Annals of Operations Research
 Applied Mathematics and Computation

Applied Psychological Measurement
 Applied Stochastic Models in Business and Industry
 Bioinformatics
 Biometrical Journal
 Biostatistics
 British Journal of Mathematical and Statistical Psychology
 Chemometrics and Intelligent Laboratory Systems
 Ecological Modelling
 Ecology
 Educational and Psychological Measurement
 Empirical Economics
 Environmental and Ecological Statistics
 Environmetrics
 European Journal of Operational Research
 Extremes
 Finance and Stochastics
 Fuzzy Sets and Systems
 Giornale degli Economisti e Annali di Economia
 IEEE (varie pubblicazioni)
 Infinite Dimensional Analysis Quantum Probability and Related Topics
 Insurance Mathematics and Economics
 International Journal of Game Theory
 Lifetime Data Analysis
 Journal of Agricultural, Biological, and Environmental Statistics
 Journal of Applied Mathematics and Decision Sciences
 Journal of Applied Mathematics and Stochastic Analysis
 Journal of Applied Meteorology
 Journal of Biopharmaceutical Statistics
 Journal of Chemometrics
 Journal of Computational Biology
 Journal of Educational and Behavioral Statistics
 Journal of Machine Learning Research
 Journal of Marketing Research
 Journal of Official Statistics
 Journal of Quality Technology
 Journal of Risk and Uncertainty
 Journal of Statistics and Management Systems
 Journal of Statistics Education
 Machine Learning
 Management Science
 Mathematical Biosciences
 Mathematics of Operations Research
 Multivariate Behavioral Research
 Open Systems & Information Dynamics
 Probabilistic Engineering Mechanics
 Probability in the Engineering and Informational Sciences

Psychometrika
 Quality and Quantity
 Quality Engineering
 Quality Progress
 Risk Analysis
 Scandinavian Actuarial Journal
 Statistical Applications in Genetics and Molecular Biology
 Sociological Methods and Research
 Statistical Methods in Medical Research
 Statistics in Medicine
 Stochastic Environmental Research and Risk Assessment
 Utilitas Mathematica

6 Classificazione delle riviste di demografia

6.1 Fascia AA

Riviste affermate, molto diffuse e di qualità mediamente eccellente

Applied Mathematical Demography
 Demographic Research
 Demography
 European Journal of Population
 Genus
 Mathematical Population Studies
 Population
 Population and Development Review
 Population Studies

6.2 Fascia A

Riviste con buona diffusione e di qualità mediamente buona/elevata

Annales de Demographie Historique
 Cahiers Québécois de Démographie
 Canadian Studies in Population
 International Migration Review
 Journal of Forecasting
 Journal of Reproduction & Fertility
 Popolazione e Storia (già Bollettino di demografia storica)
 Population et famille
 Population Research and Policy Review

6.3 Fascia B

Riviste con limitata diffusione e di qualità mediamente accettabile/discreta

American Demographics
 Boletín de la Asociación de Demografía Histórica

Boletín demográfico
 Bulletins de demographie et de planning familial
 Demos
 Demotrends
 Espace populations sociétés
 International Migration
 Polish Population Review
 Population et sociétés - bulletin mensuel
 d'informations démographiques, econ. et social
 Population Trends
 Revue Européenne des Migrations Internationales
 (REMI)
 Rivista italiana di economia demografia e statistica

6.4 Fascia C

Riviste che accolgono articoli di Demografia indirizzati a particolari ambiti applicativi. Le riviste qui elencate sono chiaramente un sottoinsieme delle moltissime riviste scientifiche che accolgono articoli di Demografia, talvolta con apprezzabili contributi metodologici.

Age and Ageing

Ageing and Society
 Annals of Epidemiology
 Annals of Human Genetics
 Environment and Planning – A Journal of
 Regional Science
 Epidemiology
 European Journal of Human Genetics
 Experimental Gerontology
 History of the Family
 Human Heredity
 International Journal of Population Geography
 Journal of Aging and Health
 Journal of the American Geriatrics Society
 Journals of Gerontology: Medical Sciences
 Lancet
 Mechanisms of Ageing and Development
 Nature
 North American Actuarial Journal
 Science
 Statistics in Medicine
 Theoretical Population Biology
 Twin Research and Human Genetics

Valutazione delle ricerca e reclutamento: quale ruolo per la SIS?

Franco Peracchi

Università di Roma "Tor Vergata"
franco.peracchi@uniroma2.it

Sommario

L'esperienza del CIVR mostra come un esercizio di valutazione della ricerca possa essere condotto in modo efficace attraverso un sistema di *peer review*. Le difficoltà incontrate da diversi Atenei italiani nel primo esercizio condotto dal CIVR sembrano trarre origine soprattutto dal mancato riconoscimento del legame che esiste tra reclutamento e promozioni da un lato ed esiti della valutazione della ricerca dall'altro. Un'utile funzione delle società scientifiche, quali la SIS, potrebbe quindi essere quella di contribuire a rendere più chiara ed evidente tale relazione.

Il messaggio principale che vorrei trasmettere a chi riprenderà, in un futuro che mi auguro non troppo lontano, l'esercizio di valutazione della ricerca è quello di evitare il ricorso a sistemi "meccanici" basati solo su indici bibliometrici, anche sofisticati. Non lo abbiamo fatto nel VTR 2001-2003, dove la valutazione si è basata invece su un sistema complesso di *peer review*. Tale sistema è, a mio avviso, più efficace di un sistema "meccanico" perché permette di utilizzare in modo flessibile tutta l'informazione disponibile, anzitutto quella fornita dalla lettura diretta del lavoro, ma anche quella derivante dal giudizio di esperti o di colleghi e dalla conoscenza della rivista dove il lavoro è stato pubblicato, senza escludere la considerazione di una serie di indicatori sintetici quali l'*Impact Factor* (IF) della rivista di pubblicazione, il numero di citazioni del lavoro, ecc. Trattandosi di valutazioni eminentemente soggettive, nasce il problema di come valutare i valutatori. Qui la statistica potrebbe aiutare molto. Si pensi alla possibilità di usare modelli e metodi con eterogeneità inosservata, effetti fissi, variabili latenti osservate solo in forma categoriale, ecc. Considerazioni di tempestività e di costo, anch'esse importanti, portano però a non escludere, quantomeno per fini particolari, sistemi "meccanici" basati su un numero limitato di indicatori sufficientemente condivisi. Da questo punto di vista andrebbe esplorato l'uso di

strumenti alternativi all'IF, come suggerito da Vittorio Frosini. Per esempio, considero molto interessante il lavoro svolto da Giovanni Cesareni, un collega di biologia molecolare a Tor Vergata, che ha presentato un *ranking* degli atenei italiani (un cui estratto è stato pubblicato nel dicembre scorso sul settimanale *L'Espresso*) utilizzando l'informazione contenuta in *Google Scholar*. Come notato anche da Francesco Billari, oltre alla sua notevole tempestività, questo sistema avrebbe il vantaggio di sfruttare una *open source* facilmente disponibile e continuamente aggiornata, evitando al tempo stesso i problemi di monopolio connessi all'informazione prodotta da *Thompson Scientific* o da *Scopus*. Anche in questo caso, gli statistici potrebbero giocare un ruolo importante nella costruzione di indicatori appropriati e nella loro valutazione.

È stato detto da taluni, sia oggi sia in altre occasioni, che le difficoltà che molti atenei hanno incontrato nell'esercizio VTR 2001-2003 traggono origine dal fatto che una parte importante del mondo universitario italiano non era preparata ad affrontare un esercizio di valutazione della ricerca. In realtà, mi sembra che le cose stiano diversamente. Come ricordato da Francesco Billari, la valutazione della ricerca viene infatti condotta di *routine* all'interno dell'università italiana, specie per quanto riguarda il reclutamento e la promozione del personale docente. Basti

pensare al sistema dei concorsi, al meccanismo della conferma in ruolo, e alla valutazione triennale dell'attività svolta dai ricercatori.

Perché dunque la valutazione esterna condotta con il VTR 2001-2003 è stata percepita come qualcosa di diverso e lontano da queste attività interne di valutazione? In parte vi è il fatto che, come ricordato da Luigi Biggeri, gli obiettivi e gli oggetti della valutazione sono diversi nei due casi. Ma la diversità degli obiettivi e degli oggetti può davvero giustificare una diversità così grande nelle modalità e, soprattutto, negli esiti dei due esercizi di valutazione, quello interno e quello esterno? Perché l'eccellenza tende a essere interpretata così diversamente nei due casi?

Una discontinuità così marcata può a prima vista apparire come mancanza di coerenza, oppure come problema di ordine morale. Non credo che sia questo il caso. Se anche lo fosse, credo di avere maturato una conoscenza sufficiente dell'accademia italiana da capire che non servirebbero i richiami a una maggiore coerenza o moralità dei comportamenti per cambiare le cose. Né credo che il problema stia semplicemente nella differenza che indubbiamente esiste tra valutazione interna di un singolo ricercatore e valutazione esterna di un'Ateneo o di un Dipartimento.

A mio avviso, il problema trae origine dal mancato riconoscimento del legame diretto tra reclutamento e promozioni da un lato ed esiti della valutazione esterna della ricerca dall'altro. Questo mancato riconoscimento ha due possibili spiegazioni, non mutuamente esclusive. La prima spiegazione è che

l'evidenziazione di tale legame venga percepita come un "rischio" basso che si può tranquillamente ignorare. Un'altra spiegazione è che le conseguenze della valutazione esterna, soprattutto dal punto di vista dell'allocazione delle risorse, vengano associate con un futuro così lontano da rendere l'esercizio di valutazione del tutto irrilevante. Tenuto conto della politica seguita finora dal MIUR, entrambe le spiegazioni appaiono plausibili.

Non si può però attribuire al MIUR la responsabilità esclusiva della situazione. Non solo perché è difficile pensare al MIUR come un organismo "esterno" all'Università italiana, ma anche perché altri fattori svolgono un ruolo importante. Mi limito a ricordare l'età media elevata del personale accademico italiano e la natura dei meccanismi decisionali e di allocazione delle risorse all'interno degli Atenei.

Vorrei concludere il mio intervento riprendendo la proposta avanzata da Cristina Marcuzzo di coinvolgere direttamente le associazioni scientifiche (la Società Italiana degli Economisti, la Società Italiana di Statistica, ecc.) nella valutazione della ricerca. Alla luce di quanto appena detto, un'utile funzione di queste associazioni potrebbe essere quella di contribuire a rendere più chiara ed evidente la relazione tra reclutamento universitario e valutazione della ricerca attraverso la pubblicazione di qualcosa di simile al Bollettino dei Concorsi prodotto per l'area economica da Roberto Perotti dell'Università Bocconi.

La valutazione della ricerca in statistica

Ugo Trivellato

Dipartimento di Scienze Statistiche, Università di Padova
trivell@stat.unipd.it

Sommario

L'intervento si snoda intorno a tre temi: il contesto istituzionale e le esperienze, per gran parte deludenti, in cui si è collocato l'esercizio di valutazione del CIVR; la forte discontinuità, rispetto alle precedenti esperienze, che connota la valutazione triennale della ricerca del CIVR, e il giudizio positivo che si può motivatamente dare sulla stessa; la presentazione di alcuni suggerimenti ritenuti utili per migliorare progressivamente il processo di valutazione della ricerca.

Sulla valutazione della ricerca in Italia, in particolare sull'esercizio di valutazione condotto dal CIVR, molto è già stato detto in questa giornata. In particolare, ho trovato persuasive parecchie delle riflessioni e delle proposte di Fiorella Kostoris, Alberto Zuliani, Alessandro Figà Talamanca e Pietro Reichlin. Così come condivido buona parte degli interventi dei colleghi che mi hanno preceduto in questa tavola rotonda conclusiva.

Ciò mi consente di restringere l'attenzione soltanto ad una parte – sia pure la più consistente – dell'esercizio effettuato dal CIVR: quella riguardante la ricerca svolta nelle Università nell'ambito delle 14 aree “tradizionali” del CUN. Trascuro, quindi, la ricerca svolta in strutture e centri non universitari, e in generale quella classificata nelle sei aree “speciali”.

Organizzerò le mie considerazioni intorno a tre interrogativi: (1) come si colloca l'esercizio di valutazione del CIVR nel contesto delle esperienze istituzionali di valutazione della ricerca in Italia? (2) qual è il significato dell'esperienza del CIVR, e quindi il giudizio che se ne può dare? (3) quali suggerimenti si possono prospettare per migliorare la valutazione della ricerca?

1 La valutazione triennale della ricerca del CIVR nel contesto istituzionale italiano

Per un giudizio meditato sull'esercizio del CIVR, è

utile collocarlo nel quadro degli ultimi 7-8 anni della vicenda universitaria italiana. Del suo svolgersi ordinario da un lato, scandito tra l'altro da un intenso processo di reclutamento di docenti. E dall'altro, e soprattutto, dell'istituzionalizzazione di procedure di valutazione: dell'attività universitaria in generale e della ricerca scientifica in particolare, a livello nazionale e dei singoli atenei – rispettivamente con il CNVSU e i Nuclei di valutazione.

In questi 7-8 anni non sono mancati contributi apprezzabili – a mo' di esempio, ricordo l'iniziale esperienza dei PRIN, ma l'esito a me appare complessivamente modesto, opaco. Detto in altre parole, il bicchiere è più vuoto che pieno. Quali, in sintesi, le ragioni di questo bilancio non soddisfacente?

Sono certamente chiamate in causa inadeguatezze del disegno e delle scelte a livello politico: del Parlamento, del Governo, del Ministero dell'Università – nelle diverse forme e nomi con cui si è venuto configurando, senza mai cambiare significativamente il *modus operandi*. Così si è definito un assetto istituzionale della valutazione non felice: con una autonomia a dir poco mediocre del CNVSU a livello centrale; con una intrinseca ambiguità dei Nuclei di valutazione degli atenei, strani “ircocervi”, a metà appendici periferiche della struttura centrale e a metà espressione dell'autonomia delle università. E, ancora, l'intero processo di allocazione e gestione delle risorse – dai finanziamenti ai meccanismi di reclutamento – è rimasto lontano da orientamenti che premiano il

merito, e che a questo fine siano *evidence-based*.

Ma ha concorso, alle perduranti inadeguatezze, il ruolo mediamente molto debole della comunità universitaria nel sollecitare e sostenere un'evoluzione in chiave meritocratica. Le stesse agenzie di valutazione, quella nazionale e quelle delle singole università, sono state largamente assorbite da un'attività faticosa di costruzione di ampie batterie di indicatori – di risorse fisiche e finanziarie, dello svolgersi del processo didattico, di operosità scientifica – che solo in alcuni casi risultano significativi, si prestano cioè a letture nitide. Una mole di lavoro – e di indicatori prodotti – notevole, ma non per ciò efficace, che richiama alla mente un'amara variante del titolo del basilare saggio di Piero Sraffa: "Produzione di indicatori a mezzo di indicatori".

Di queste inadeguatezze do tre scheletrici esempi: il primo mette in evidenza le responsabilità del contesto nazionale, e del Ministero in particolare; gli altri due segnalano responsabilità condivise, che in misura crescente coinvolgono i docenti universitari.

Il primo esempio riguarda il modo con cui in Italia si svolge il test di ammissione alle Facoltà di Medicina. C'è un numero chiuso, definito per singola sede. Il test si svolge sulla base di una prova comune, somministrata in una stessa data nelle varie sedi. Sin qui siamo nell'ovvio. Ma cadiamo immediatamente nel bizzarro – o forse sarebbe più appropriato dire nell'irragionevole, se aggiungiamo che non c'è una graduatoria unica nazionale, con gli ammessi che poi, in base alla loro posizione in graduatoria, scelgono la sede (con un evidente, duplice vantaggio: si iscrivono alla Facoltà di Medicina gli aspiranti che sono risultati migliori, e si innesta una positiva competizione fra le sedi). No. C'è invece una *graduatoria per sede*, con esiti paradossali. Ad esempio, il candidato che si è classificato quattrocentesimo a Padova – il centosessantunesimo degli esclusi – sarebbe stato tra gli ammessi a Ferrara. E il settecentesimo di Padova – il quattrocentosessantunesimo degli esclusi – sarebbe entrato a Campobasso (Della Zuanna, 2007).

Il secondo esempio, che ci porta a responsabilità condivise, attiene alla definizione di indicatori e parametri per la valutazione delle didattiche. Ora, tutti o quasi questi indicatori (laureati regolari su iscritti, e via dicendo), sono interni al processo di insegnamento, in definitiva endogeni. È immediato chiederci: che cosa ci dicono, anche indiretta-

mente, sull'output, cioè a dire sui livelli di apprendimento?

Se guardiamo all'esperienza di PISA – il *Project for International Student Achievement* promosso dall'OECD per misurare conoscenze e competenze degli studenti 15enni –, tra le molte ombre per l'Italia troviamo (i) un forte divario interno nei punteggi medi degli *achievement tests*, con il Mezzogiorno attestato su soglie molto più basse del Nord del paese, e contemporaneamente (ii) una correlazione negativa fra questi punteggi nei tests e i voti degli insegnanti, sicché se guardiamo a questi ultimi è nel Mezzogiorno che mediamente si registrano i risultati migliori. L'ovvia conclusione è che i voti degli insegnanti, più in generale i parametri interni, sono un termometro men che mediocre dello stato di salute della scuola italiana².

Ne discende immediatamente un interrogativo: in Italia l'Università è così diversa dalla scuola secondaria superiore da essere immune da questa vistosa deformazione dei parametri interni? Ne dubito. Sono, invece, propenso a ritenere che vi siano notevoli similarità fra il modo di valutare i risultati nella scuola secondaria e nell'Università. Ma se ciò è vero, è palese che i parametri interni servono a poco, e che moltiplicarli non serve a nulla. Occorre quindi mirare a costruire parametri esterni: pochi, ma significativi. Non è un'impresa facile. Ma più presto la si intraprende meglio è, se vogliamo disporre di indicatori che diano conto in maniera ragionevolmente affidabile delle conoscenze e competenze acquisite dai nostri studenti.

Il terzo esempio chiama in causa soprattutto le responsabilità nostre, di comunità di docenti, e riguarda i concorsi universitari, il loro svolgimento – come dire? – sovente non esemplare. I colleghi Billari e Peracchi si sono già fermati sull'argomento, e in maniera convincente. Vorrei aggiungere un'osservazione.

Alcuni interventi hanno sottolineato come gli strumenti di valutazione della ricerca debbano essere differenziati in relazione agli ambiti e agli obiettivi, che possono essere parecchio diversi. Come non convenire con una tale affermazione di

² Una recente, illuminante sintesi sullo stato della scuola italiana è il *Quaderno bianco sulla scuola*, curato dal Ministero della Pubblica Istruzione e dal Ministero dell'Economia e delle Finanze, Roma, agosto 2007.

metodo? Ma – aggiungo – come non vederne le insidie se, in nome della specificità dell’obiettivo valutativo, si costruiscono “torri d’avorio”, spazi sottratti a un rigoroso, onesto vaglio di merito? Certo, tra la valutazione di un articolo, la valutazione di un ricercatore e la valutazione di una struttura di ricerca – dipartimento o università che sia – vi sono differenze non trascurabili. Ma sottesa vi è anche una logica comune: quella della *peer review*. Che richiede di essere declinata con specificazioni diverse, le quali utilizzino in maniera variamente intensa parametri bibliometrici e giudizi informati sui singoli “prodotti”. Ma che non può non condividere due irrinunciabili capisaldi: il riferimento al merito scientifico e la *fairness*, pena cooptazioni arbitrarie, dannose per il futuro dell’università italiana.

2 Qual è il significato dell’esperienza del CIVR, e il giudizio che se ne può dare?

Nel quadro che ho delineato, con grossolane approssimazioni ma – penso – in maniera sostanzialmente veritiera, l’esercizio di valutazione triennale della ricerca svolto dal CIVR segna una netta discontinuità, un’innovazione decisamente positiva. Certo, si tratta di un primo passo, ancora lontano da un sistema collaudato e condiviso quale il *Research Assessment Exercise* (RAE) inglese³. Esso richiede quindi di essere consolidato e migliorato – e anch’io avanza qualche suggerimento in questa direzione. Ma ciò non può avvenire che muovendo dal riconoscimento dei meriti e della forte valenza innovativa dell’esperienza del CIVR.

Una prima riflessione che essa suggerisce riguarda la cornice e le forme in cui si è svolta. La valutazione della ricerca per il triennio 2001-2003 si è collocata *ai margini del quadro istituzionale ordinario*, a livello nazionale – col CIVR che si è sostituito al CNVSU – e dei singoli atenei – con le Commissioni scientifiche che si sono sostituite ai Nuclei di valutazione. È questo un sintomo inequivocabile dell’inadeguatezza mostrata dal quadro istituzionale e dai modi di operare degli

organismi ordinari.

Il segnale è rilevante non solo, e non tanto, per il giudizio sul passato, quanto soprattutto per le prospettive future, che dipenderanno largamente dall’efficacia dell’azione dell’ANVUR. Non dirò che l’esperienza del CIVR suoni come un campanello d’allarme, ma sicuramente è un *caveat* per il modo con cui, nei fatti, opererà l’ANVUR. Il disegno dell’Agenzia nazionale di valutazione presenta, infatti, ambiguità e rischi di appesantimento che ne possono attenuare, forse minare, la funzione innovativa. Riprendendo spunti già emersi nel dibattito, ritengo che il modo con cui l’ANVUR comincerà ad operare – e segnerà così una sorta di “interpretazione autentica” del proprio mandato – sarà decisivo soprattutto per due aspetti:

- (a) la focalizzazione della propria attività sul *core business* – la valutazione della ricerca e della didattica, mettendo la sordina a parecchi altri compiti (prossimi, a ben vedere, ad appesantimenti burocratici) che la normativa le assegna;
- (b) una chiara distinzione fra la valutazione della ricerca scientifica e quella della didattica, che, tra l’altro, lasci sullo sfondo lo sfuocato compito di “*valutazione complessiva di ciascuna università e delle sedi decentrate delle medesime*”, vistosamente esposto al rischio che “nella notte tutte le vacche sono nere”⁴.

⁴ L’art. 4, comma 6 dello schema di DPR istitutivo dell’ANVUR suona così: “È altresì compito dell’Agenzia la valutazione complessiva di ciascuna università e delle sedi decentrate delle medesime e di ciascun ente di ricerca, relativamente alla capacità di governare i processi e di darsi obiettivi strategici, al grado di raggiungimento di tali obiettivi, al livello raggiunto negli scambi internazionali, alla capacità di sostegno alla crescita del Paese e del proprio territorio di riferimento, ivi comprese le attività di formazione permanente al rapporto tra risorse e risultati, al grado di realizzazione dei principi della Carta europea dei ricercatori e del Codice di condotta per il loro reclutamento di cui alla raccomandazione della Commissione europea dell’11 marzo 2005, n. 2005/251/CE. È inoltre valutata la completezza e correttezza della comunicazione pubblica, soprattutto in materia di offerta formativa e di ricerche, nonché di servizi e strutture per gli studenti.” I rischi di elefantiasi, di indeterminatezza dei criteri di giudizio e di opacità dei risultati di una siffatta valutazione sono così palesi, che neppure vale la pena evidenziarli.

³ Sull’esperienza del RAE, alla quale mi riferirò anche nel seguito, vedi *Review of research assessment* (2003) e *RAE 2008* (2005).

La distinzione fra i due esercizi di valutazione – della ricerca e della didattica – è necessaria di per sé, per le marcate diversità fra le due attività (il che non significa negare le possibili, feconde interazioni fra le stesse). Lo è ancor più oggi, a fronte del processo – ormai in buona parte compiuto – verso la cosiddetta “università di massa”, con l’inevitabile distinzione fra *teaching universities* e università di alta formazione e di ricerca. Una distinzione inevitabile nei fatti, ma che è essenziale *rendere esplicita e governare bene*: il che significa innanzitutto riconoscere che i compiti delle diverse istituzioni sono differenti, e quindi che ciascuna di essa va valutata rispetto ai propri. Così, per le *teaching universities* l’obiettivo essenziale è di svolgere bene la formazione di 1° livello: e per non poche piccole università o sedi decentrate o facoltà/corsi di laurea in sedi decentrate questa è già una sfida ardua, se è vero – come sfortunatamente è vero – che hanno dotazioni, organizzazione e attività formative lontane dagli standard di un buon liceo o istituto tecnico. Così, per le università di alta formazione e di ricerca l’obiettivo è che la valutazione sia essenzialmente sul merito scientifico, con coerenti ricadute sui finanziamenti, e in tal modo stimoli una sana competizione verso il meglio.

3 Qualche suggerimento per procedere sulla strada intrapresa dal CIVR

Concludo con alcuni suggerimenti che potrebbero tornare utili per migliorare progressivamente il processo di valutazione della ricerca.

Il primo suggerimento è presto detto. Vi sono forti incertezze sui tempi di decollo dell’ANVUR. Ebbene, è di grande importanza evitare che l’esperienza della valutazione triennale del CIVR resti un episodio isolato, o comunque ripreso dopo un lungo intervallo. È dunque opportuno reiterare, subito, la valutazione della ricerca per il triennio 2004-2006. Detto con uno slogan, *chiedere che sia concesso al CIVR di fare il bis*, possibilmente con modeste varianti.

Serve poi *strutturare e rendere ancora più trasparente il processo di valutazione, innanzitutto nella fase preliminare*, in modo che (i) siano enunciati *ex-ante*, in maniera chiara e diffusa, i

criteri e i modi con cui si intende condurre la valutazione, e (ii) sia consentito alle Università/Facoltà/Dipartimenti/docenti di interloquire nel processo della loro definizione, quindi di “apprendere” e condividere la logica stessa della valutazione, e – a valle – le sue conseguenze su finanziamenti e reputazione. Su questo terreno, l’esperienza del RAE è esemplare.

Dall’esperienza del RAE, criticamente rivisitata in quest’ultimo quinquennio, vengono due altri insegnamenti. Essi riguardano l’articolazione flessibile con cui avviene il processo di valutazione, sotto due profili: quello della formazione dei panel di valutatori e quello dei criteri per la valutazione.

Ogni suddivisione disciplinare ha un certo grado di arbitrarietà, e vi è l’esigenza di trovare un *buon equilibrio fra una partizione in grandi aree – che eviti i rischi di autoreferenzialità – e l’attenzione alle specificità disciplinari*. La soluzione inglese del “two-tier model”, con 15 panel che definiscono i criteri e formulano i giudizi finali e 67 sub-panel che svolgono la concreta attività di valutazione, mi pare convincente. (Incidentalmente, tale modello potrebbe fornire una risposta soddisfacente anche al disagio patito nell’esercizio del CIVR dalla statistica, che si è trovata presa nel “tiro alla fune” fra aree scientifico-disciplinari e aree dell’ISI, al quale si è riferita la collega Salinetti nel saluto e nelle riflessioni di apertura.)

Quanto ai criteri di valutazione, l’empirismo anglosassone, combinato con un’altrettanto apprezzabile *fairness*, evita ogni aprioristica, rigida definizione di parametri – ad esempio, esclude esplicitamente il “formulaic use” di parametri bibliometrici quale l’Impact Factor o il Citation Index dell’ISI – e fa perno sulla *peer review*. Ciò non comporta, peraltro, che ai fini della *peer review* si costruiscano criteri di giudizio autarchici, con i panel o le società scientifiche nazionali che costruiscono proprie classificazioni delle riviste – una strada che giudico assai pericolosa, perché inevitabilmente esposta all’autoreferenzialità e alla mediazione al ribasso. Piuttosto, l’esercizio maturo, e *fair*, della *peer review* porta a *utilizzare in maniera saggia, abbastanza intensa ma non meccanica, una pluralità di parametri bibliometrici internazionalmente riconosciuti* (come ha segnalato Billari, non solo quelli dell’ISI, ma anche *Scopus-Elsevier*, *Google Scholar* e vari altri).

Inoltre – e qui delinea una prima proposta che potrebbe portare a modeste varianti all’iniziale

esperienza del CIVR, occorre tenere presente che i risultati della valutazione valgono essenzialmente per un confronto di strutture – dipartimenti e simili – entro un’area scientifica, non già fra aree. Il confronto è tanto più informativo quanto più esso poggia su una apprezzabile *variabilità dei punteggi* entro l’area. (E può essere informativo anche per vagliare comparativamente le *performances* di dipartimenti di aree diverse in uno stesso ateneo, quanto più esso poggia su una apprezzabile *omogeneità della variabilità* dei punteggi nelle varie aree.) Ora, l’esperienza del CIVR testimonia che ciò non è avvenuto. Restiamo alle strutture medie e grandi (le piccole sono infatti soggette a distorsioni, di cui dirò tra poco). Ebbene, il campo di variazione dei punteggi presenta marcati divari: è massimo, pari a 0,49, nell’area delle scienze economiche e statistiche; si colloca fra 0,20 e 0,30 per la gran parte delle aree; scende tra 0,12 e 0,18 per tre aree – le due umanistiche e l’ingegneria industriale e informatica. Si tratta di un fenomeno oggettivo, cioè di una differente dispersione dei migliori “prodotti” nelle diverse aree, misurata bene nell’esercizio di valutazione? O si tratta di una propensione dei diversi panel a usare metriche diverse, in particolare ad appiattire le valutazioni verso l’alto? Chiaramente, non sappiamo rispondere a questo interrogativo. E, per un certo verso, si potrebbe sostenere che rispondervi non è essenziale, perché possiamo trovare modi ragionevoli di normalizzare *ex-post* per la diversa variabilità entro le aree. Ma proprio a questo riguardo torna utile guardare al RAE. Anche l’esperienza, più che quindicennale, del RAE ha evidenziato la tendenza all’appiattimento delle valutazioni verso l’alto, verso il “*cliff edge*”. E ha indotto a una raccomandazione di rilievo: formulare non un punteggio di sintesi per una struttura, bensì un “*quality profile indicating the quantum of ‘one star’, ‘two star’ and ‘three star’ research in each submission. It will not be the role of the assessment to reduce this profile to summary metrics or grade*”⁵.

Assicurare un’*apprezzabile variabilità delle valutazioni entro ogni area* è un primo obiettivo che ritengo importante per l’auspicata reiterazione dell’esercizio di valutazione da parte del CIVR. Sempre nella prospettiva di proposte volte a

migliorare il processo di valutazione del CIVR, occorre poi riflettere attorno a un nodo cruciale: quali sono le strutture che debbono essere oggetto della valutazione? Le aree disciplinari dei singoli atenei, com’è stato nella prima esperienza? Propendo per il no, perché nella grande maggioranza dei casi non sono strutture che organizzano la ricerca; sono aggregazioni convenzionali utilizzate prevalentemente per lo specifico scopo della valutazione nazionale. Così come sarei drastico nell’escludere che la valutazione nazionale della ricerca debba riguardare i singoli ricercatori. Starà agli atenei, stimolati da un maturo processo di valutazione nazionale delle strutture seguito da coerenti ricadute sul fronte dei finanziamenti e della reputazione, mettere in atto modalità di valutazione/incentivazione dei docenti. *Le strutture oggetto di valutazione non possono che essere i dipartimenti*. Sono essi, infatti, le unità in cui si organizzano e si svolgono la formazione avanzata e la ricerca, e che sono destinatarie delle risorse per questi scopi.

Infine, è importante che *l’esercizio di valutazione della ricerca non sia circoscritto ai picchi di eccellenza*. Così è stato, con la regola di presentare un prodotto ogni quattro docenti, per l’arco di un triennio. Oltre, e più, che provocare distorsioni nella valutazione delle strutture piccole⁶, questa regola preclude un apprezzamento equilibrato della “qualità” della ricerca di una struttura – nei miei auspici, di un dipartimento. Per evitare questi inconvenienti si possono adottare varie soluzioni, che non appesantiscano eccessivamente il lavoro dei panel del CIVR. Una scelta ragionevole potrebbe consistere nel chiedere alla strutture di fornire, oltre che i prodotti “eccellenti” secondo le stesse regole del triennio 2001-03, due altre informazioni/prodotti:

- a) un prodotto per singolo “ricercatore attivo” nell’arco del triennio;
- b) l’indicatore suggerito da Fiorella Kostoris, cioè la frazione di “ricercatori attivi” rispetto al totale dei docenti.

⁵ Vedi Review of research assessment (2003), pag. 11.

⁶ La scelta di valutare soltanto l’eccellenza comporta, infatti, per le strutture piccole una vistosa distorsione verso la polarizzazione. Esse raggiungono il punteggio massimo, con la normalizzazione adottata dal CIVR pari a 1, quando vi afferisce anche un solo docente “bravo”, mentre hanno punteggi molto bassi quando ciò non accade.

Si tratta di una scelta parsimoniosa, che consentirebbe tuttavia una valutazione parecchio più informata e articolata della produttività e della qualità della produzione scientifica.

Riferimenti bibliografici

Della Zuanna G. (2007), “Quel test è uno specchio della scuola italiana”, <http://www.lavoce.info>, 31 luglio 2007.

Review of research assessment (2003), Report by Sir Gareth Roberts to the UK funding bodies, May 2003.

RAE 2008 (2005), *Research Assessment Exercise. Guidance on submissions*, June 2005, Ref RAE 03/2005, <http://www.rae.ac.uk/pubs/2005/03/rae0305.pdf>