Valutare l'eccellenza senza trascurare la differenza

Paolo Rossi – Dipartimento di Fisica dell'Università di Pisa e Consiglio Universitario Nazionale

Trento, 24 Novembre 2014

Premessa: è davvero possibile riconoscere "ex ante" l'eccellenza?

Vorrei iniziare questa relazione con un aneddoto che mi ha sempre impressionato non tanto per la qualità del candidato quanto per quella degli esaminatori.

Completato a Pisa il biennio d'Ingegneria Bruno Pontecorvo, seguendo il consiglio del fratello Guido, arriva a Roma nell'autunno del 1931 e chiede di essere immatricolato al terzo anno del corso di "Matematica e Fisica". Esaminato da Fermi e Rasetti., ottiene un giudizio positivo ma, al termine del colloquio, Fermi aggiunge un commento: "*La fisica è una sola ma disgraziatamente oggi i fisici sono divisi in due categorie: i teorici e gli sperimentatori. Se un teorico non possiede straordinarie capacità il suo lavoro non ha senso… Per quanto riguarda la sperimentazione invece anche una persona di medie capacità ha la possibilità di svolgere un lavoro utile*." Pontecorvo (che già in famiglia era considerato il meno intelligente) lo considera un avvertimento personale e decide che sarà un fisico sperimentale. "Ex post", i successivi contributi di Bruno Pontecorvo alla fisica teorica del XX secolo (in particolare per quanto riguarda la fisica dei neutrini) sono oggi considerati tra i più importanti in assoluto.

Altri aneddoti potrebbero aggiungersi a questo senza difficoltà, da Einstein a Flaubert, da Darwin a Giuseppe Verdi. Esempi che ovviamente non possono essere interpretati come un dato statistico, ma che ci obbligano a prestare la massima attenzione ai rischi insiti nella pretesa di identificare precocemente l'eccellenza, e questo indipendentemente dalla metodologia adottata ai fini di tale identificazione.

Che cos'è l'eccellenza?

Probabilmente un primo rischio di errore concettuale consiste nell'idea che esista una sola definizione di eccellenza scientifica. Questo errore è stato commesso per l'intelligenza, e ha portato alla definizione del concetto lievemente aberrante di IQ (Intelligence Quotient). Ora rischiamo di definire l'EQ, con risultati altrettanto perversi: ricordo che l'IQ è stato utilizzato per giustificare tesi razziste o sessiste.

In realtà esistono varie "eccellenze" possibili, anche restando strettamente nell'ambito scientifico. Certamente uno degli aspetti più interessanti del lavoro scientifico è la sua componente creativa, che peraltro è quella che si manifesta più sporadicamente e che non è possibile prevedere e/o controllare. Un'organizzazione scientifica ha tutto l'interesse a valorizzare la creatività, ma l'unico modo che si conosca per far ciò consiste nello "scommettere" su individui promettenti e poi lasciarli per quanto possibile liberi, sapendo perfettamente che si tratta di una scommessa che è assai più facile perdere che vincere. Si tratta di una politica che l'ERC ha dichiarato esplicitamente di voler praticare, almeno in una certa misura, ma bisogna evidentemente avere le spalle molto larghe, in termini di risorse, per poterla fare. Questo ruolo negli USA è svolto da importanti fondazioni, anche private (e.g. MacArthur).

Ma la gran parte della ricerca scientifica contemporanea si muove dentro i binari di quella che Kuhn ha definito "scienza normale", nella quale la componente creativa, pur importante, svolge un ruolo meno determinante in rapporto ad altre capacità, come quella di svolgere un lavoro sistematico, o di saper utilizzare strumenti (concettuali e materiali) molto sofisticati, la capacità di lavorare in una squadra o anche quella di saper coordinare una squadra.

Per non parlare della capacità di attrarre fondi, di trasformare risultati teorici in applicazioni pratiche. E da ultimo non credo che nell'attuale sociologia della ricerca si possa sottovalutare l'attitudine a disseminare i risultati del proprio lavoro, in quanto la credibilità scientifica è inevitabilmente misurata anche dal grado di risposta che la comunità di riferimento esprime in rapporto ai risultati che le vengono presentati. Si tenga conto del fatto che la quantità di articoli che viene giornalmente pubblicata in ciascun ambito disciplinare va oltre la capacità di lettura di qualunque studioso, e quindi la capacità di farsi leggere è in un certo senso essa stessa parte dell'"eccellenza" scientifica.

Questa diversità tra le componenti di un profilo "eccellente" è già una prima indicazione dell'importanza di introdurre una gamma di "parametri" valutativi sufficientemente ampia e capace di tener conto di capacità e attitudini anche molto differenziate tra loro. Sappiamo tutti che la diversità è un ingrediente essenziale per la sopravvivenza e per l'evoluzione degli ecosistemi: cerchiamo di non dimenticarcene quando ragioniamo di quel complesso ecosistema che è la comunità scientifica.

I criteri "oggettivi" di qualità scientifica

Premesso dunque che la qualità di un ricercatore non può mai essere ridotta alla valutazione numerica di un singolo parametro (*ranking)*, possiamo chiederci se sia possibile comunque individuare un certo numero di parametri "oggettivi" (in quanto misurabili) e di combinarli in modo tale da ottenere un giudizio (*rating*) per quanto possibile assoluto in quanto sciolto dalla soggettività del valutatore.

Non è difficile convincersi che il criterio più immediato, ovvero la quantità delle pubblicazioni scientifiche, è completamente privo di fondamento e di capacità selettiva. Esistono moltissimi ricercatori scientificamente prolifici che non hanno mai offerto in vita loro un contributo realmente originale, per converso esistono non pochi esempi di produzione scarsa ma qualitativamente eccellente. Si potrebbe obiettare che questi sono esempi estremi, ma tra questi estremi sta una varietà di casi che il dato numerico tende a mistificare. Ci sono autori che hanno sempre lavorato all'interno di collaborazioni (e oggi anche in grandissime collaborazioni) incrementando in tal modo il proprio *score* senza che sia possibile controllare il loro reale contributo; sono sempre più diffuse, purtroppo, le pratiche di plagio, di moltiplicazione degli stessi lavori su riviste diverse, di pubblicazione di risultati falsi o manipolati.

A questo punto è evidente che dovrebbe intervenire un criterio di merito, o di qualità, del contenuto delle pubblicazioni, ma ancora una volta si auspica che si tratti di un criterio non "soggettivo". Da tempo si considera che una *proxy* della qualità di un lavoro scientifico possa essere il numero delle citazioni che esso ha ricevuto, e che a sua volta l'Impact Factor (IF) di una rivista possa essere una *proxy* delle citazioni ricevute dagli articoli in essa pubblicati. Non vorrei spendere molte parole sull'IF, che è già stato così chiaramente demistificato da stupirsi che ci sia ancora qualcuno che gli dà qualche peso. Senza entrare in argomenti più sofisticati di teoria delle distribuzioni prive di scala e di calcolo delle probabilità, basti la semplice osservazione empirica che anche la maggior parte degli articoli pubblicati su riviste a elevato IF ricevono poche o nessuna citazione, e la "fortuna" di queste riviste è costituita dai pochi *bestseller* che ricevendo un numero enorme di citazioni innalzano la media complessiva.

Ma anche il conteggio delle citazioni individuali dipende da variabili delle quali è difficile produrre un'adeguata taratura: in primo luogo l'ambito disciplinare, che sfavorisce i settori di nicchia, anche all'interno di una specifica disciplina, per quanto grande possa essere la loro rilevanza concettuale o pratica; in secondo luogo l'originalità, che penalizza le ricerche svolte su tematiche poco frequentate o interdisciplinari in favore di una produzione *mainstream* che attira più facilmente l'attenzione di un gran numero di addetti ai lavori.

Se poi guardiamo alle citazioni cumulativamente ricevute da un determinato ricercatore dobbiamo riprendere anche alcune delle obiezioni già sollevate a proposito del numero delle pubblicazioni, ovvero l'effetto distorcente delle (grandi) collaborazioni (per il quale nessun correttivo tra quelli proposti ha ancora trovato un diffuso consenso) e delle pratiche opportunistiche. Non possiamo ignorare poi l'effetto delle autocitazioni, che in quanto tali si possono anche sottrarre, quello assai più subdolo delle sistematiche complicità (io cito te e tu citi me, a prescindere dai contenuti) e infine, ma non ultimo, l'effetto della deliberata manipolazione, la cui possibilità è stata più volte evidenziata da esplicite operazioni di demistificazione, ma che ha portato, salvo successive correzioni, persino a sostanziali "risalite" di Atenei non particolarmente illustri nelle famigerate "graduatorie" mondiali.

La letteratura scientometrica è ormai piena di generosi tentativi di aggirare queste difficoltà mediante combinazioni sempre più complesse e sofisticate dei parametri quantitativi disponibili, a partire dall'indice H (o di Hirsch) e dalle sue varianti. Quasi inutile dire che ognuno di questi tentativi soffre di due obiezioni fondamentali: la prima è la soggettività della combinazione prescelta, che si scontra con la soggettività delle altre proposte senza che esista alcun criterio "oggettivo" per effettuare una scelta (per non parlare della inutilità di definizioni spesso complicate che poi finiscono col produrre risultati fortemente correlati a quelli di altri parametri (basti guardare alla relazione tra H e la radice quadrata del numero delle citazioni), e la seconda è il carattere intrinsecamente statistico di ognuno di questi criteri, che li rende (parzialmente) utili per valutare una comunità (istituzionale o scientifica), ma ancora una volta del tutto arbitrari se usati per posizionare un singolo, soprattutto se in comparazione con altri singoli di qualità non troppo differente.

<u>Si può valutare l'eccellenza?</u>

La mia risposta a questa domanda, pur con tutti i *caveat* espressi in precedenza, è comunque positiva. Nel bel saggio di Sandro Graffi "*considerazioni sulla grandezza e decadenza dei concorsi universitari in Italia"*, pubblicato nel 2010 sulla rivista "*Quaderni di storia"* e purtroppo assai poco conosciuto, mediante un'analisi dei concorsi di Matematica svoltisi tra il 1885 e il 1940, confrontando il giudizio delle commissioni con quello della 'storia", si dimostra che una comunità di valutatori sufficientemente <u>esperta, coesa, indipendente e motivata</u> fu per molto tempo perfettamente in grado di identificare, tra soggetti assai più "giovani" di quanto non lo siano gli attuali candidati ai concorsi, coloro che con maggior probabilità avrebbero in seguito prodotto i risultati più rilevanti.

Oggi il compito è reso certamente più difficile dall'indebolimento di ciascuno degli aggettivi sopra elencati: la parcellizzazione delle discipline e delle competenze ha diminuito la competenza individuale e la coesione, mentre le trasformazioni sociali (università di massa, nuovi modelli di comunicazione, eccessivo interesse o peggio ancora disinteresse da parte delle classi dirigenti) hanno ridotto drasticamente l'indipendenza di giudizio e la motivazione.

Ma il compito è anche facilitato proprio dall'acquisizione di quel tipo di informazioni di cui poco fa abbiamo messo in discussione l'uso acritico, come i dati sulla produttività e l'impatto, che se usati con saggezza offrono elementi di giudizio certamente importanti sul profilo di candidati i quali, a causa della diffusa "neotenia" accademica (oggi pressoché chiunque abbia meno di trent'anni è considerato un "ragazzo" nel mondo universitario) si presentano alle valutazioni con *curricula* già assai articolati. Si parla in questo caso di *informed peer review*, che non è certo il rimedio per tutti i mali, ma che, congiunta a meccanismi regolativi e autoregolativi nella scelta dei *referees*, potrebbe assicurare una qualità delle valutazioni molto superiore a quella fornita dalle "mediane" e dagli "indicatori" applicati meccanicamente. Non si tratta di inventare nuovi parametri rispetto al numero dei lavori e delle citazioni, ma di usarli criticamente, tenendo conto del particolare contesto di ricerca e della storia individuale e di gruppo del candidato.

Rispettare la diversità

Resta comunque impregiudicata la petizione contenuta nelle premesse: ogni meccanismo di selezione qualitativa non può e non deve prescindere, se vuol conseguire risultati realmente positivi, dall'esigenza di assicurare uno spazio adeguato alla valorizzazione della diversità.

Abbiamo già sottolineato, e continueremo a farlo, l'importanza della diversità delle tematiche. La rincorsa alle citazioni e all'IF tende a far concentrare le ricerche su pochi filoni "di moda", ma questo pregiudica sia la nascita di nuovi filoni sia lo sviluppo di quegli ambiti, in particolare interdisciplinari, che pur avendo un valore strategico nel lungo periodo non riescono nel breve periodo a mettere in evidenza la propria potenziale rilevanza. Ricordiamo la risposta (probabilmente apocrifa, ma ben inventata) di Faraday a Gladstone: "*Non so a che cosa serva, ma è probabile che prima o poi ci metterete una tassa".*

Ma ci sono anche diversità più estrinseche rispetto ai temi specifici della ricerca, ma che non per questo risulta meno importante valorizzare. Pensiamo alle differenze culturali, che inducono nei ricercatori punti di vista molto differenti e quindi aprono la strada a soluzioni difficilmente concepibili all'interno di un quadro concettuale troppo rigido. La cultura ebraica ha condizionato fortemente la fisica teorica, così come il pragmatismo anglosassone ha favorito lo sviluppo di discipline con forti ricadute applicative, dalla genomica all'informatica. Esistono anche, e meritano di essere tenute in conto positivamente, importanti differenze territoriali, che oggi spesso invece agiscono da deterrente alla valorizzazione (esiste un "razzismo" accademico che si traduce in pre-giudizio negativo sulla base della provenienza).

Ma pensiamo in particolare alle differenze di genere, che condizionano non soltanto l'approccio ai problemi e la scelta del tipo di problemi da affrontare, ma anche la stessa organizzazione della ricerca, per la maggior capacità relazionale (non è questa la sede per discutere se di origine biologica o culturale) che favorisce il lavoro di gruppo e lo scambio di informazioni, in contrasto con l'individualismo e lo spirito di competizione che più spesso animano i ricercatori maschi.

Ma le differenze di genere, per essere valorizzate, richiedono uno sforzo particolare: se non ha senso pensare a"quote di genere" nella ricerca (e anzi ci sono indicazioni[1] che una presenza femminile nelle commissioni possa andare a scapito delle candidate), è tuttavia indispensabile che si crei una "cultura" della valutazione nella quale la taratura degli indicatori tenga adeguatamente conto dell'impatto che una serie facilmente identificabile di condizionamenti sociali (dalla maternità alla cura) esercita sulle ricercatrici, quando non accettino più o meno volontariamente di "alienarsi" rispetto ad alcune proprie vocazioni.

Anche qui, non si tratta di riconoscere una "produzione virtuale" al posto di quella reale, ma di valutare la produzione reale su una scala che può benissimo essere in qualche misura differenziata rispetto a quella usata per i colleghi maschi. Poi è evidente che il merito individuale (possibilmente non oscurato da pregiudizi di genere, come si verifica invece spesso in esperimenti psicologici) deve giocare un ruolo comunque preminente nella selezione: ma siamo sicuri che sia questo il criterio dominante che è stato applicato per decenni nella valutazione dei colleghi maschi che, in misura statisticamente non plausibile hanno finora trionfato nelle competizioni, soprattutto man mano che si saliva nella scala gerarchica del mondo accademico? (Le ordinarie sono oggi poco più del 21%, ma il dato più grave riguarda il fatto che il reclutamento femminile nella prima fascia è fermo da quasi 15 anni al 25%, per cui di questo passo non si arriverà mai nemmeno vicini alla parità.)

---

[1] M. Bagues, M. Sylos-Labini, N. Zinovyeva, *Do gender quotas pass the test? Evidence from academic evaluations in Italy,* LEM Working Paper Series 2014/14 (August 2014