

# SPAZIOFILOSOFICO

1/2015

**Numero 13**  
Valutazione



Fondatori

Enrico Guglielminetti  
Luciana Regina

Comitato scientifico

Enrico Guglielminetti (Direttore)  
Silvia Benso  
Gianfranco Dalmaso  
Ugo Perone  
Luciana Regina  
Brian Schroeder

© 2015 SpazioFilosofico  
Tutti i diritti riservati

ISSN: 2038-6788

Gli articoli filosofici della rivista sono sottoposti a blind review. La pubblicazione è subordinata per ogni articolo all'approvazione dei valutatori anonimi esterni alla direzione e all'accoglimento di eventuali richieste di revisione.



**SPAZIOFILOSOFICO**

**1/2015**

**VALUTAZIONE**

**a cura di Enrico Guglielminetti**



## INDICE

E. GUGLIELMINETTI-A. MIGLIETTA-P. MODERATO, <i>Valutazione? Ma non così! Editoriale</i>	11
E. GUGLIELMINETTI-A. MIGLIETTA-P. MODERATO, <i>Evaluation? Yes, but not This Way! Editorial</i>	15

## TEORIA

P. DAHLER-LARSEN, <i>The Evaluation Society: Critique, Contestability and Skepticism</i>	21
E. GUGLIELMINETTI, <i>Valutazione e novità</i>	37
R. MANZOTTI-P. MODERATO, <i>Neuroscienze e psicologia: valutazione e paradigmi</i>	55

## POLITICHE

D.M. MERTENS, <i>Philosophical Assumptions and Program Evaluation</i>	75
J.C. GREENE, <i>Evaluation as a Socio-Political Intervention</i>	87
G. SAPELLI, <i>Breve pensiero sulla valutazione di un economista roussoiano</i>	97

## PRATICHE

P. LEMMA, <i>Interventi di promozione della salute: una valutazione appropriata</i>	103
P. RICCHIARDI-E.M. TORRE, <i>Valutazione della scuola e del sistema scolastico: qualità formale e qualità effettiva</i>	113
M. CASSELLA-D. MUTTI, <i>Su alcuni aspetti della valutazione della ricerca nelle scienze umane. Bibliometria e dintorni</i>	123
A. MIGLIETTA, <i>La valutazione della qualità della ricerca nelle discipline manageriali: buoni propositi, cattive pratiche e urgenti cambiamenti</i>	139
E. FRANZINI, <i>Interrogare la valutazione nelle discipline umanistiche</i>	155

## STUDI

S. JACOB-W.S. AFFODEGON, <i>Conducting Quality Evaluations: Four Generations of Meta-Evaluation</i>	165
---	-----



## **VALUTAZIONE**



## VALUTAZIONE? MA NON COSÌ!

## EDITORIALE

La società dell'informazione e della conoscenza è anche una società della valutazione. I genitori dei ragazzi in età scolare hanno, per esempio, diritto a sapere in base a quali parametri i professori assegnano un voto. Nascono così un po' ovunque griglie di valutazione. Un liceo torinese prevede da 7 a 12 *items* per giustificare ciascun voto di condotta da 10 a 5. Consideriamo il primo *item*: per meritarsi un 10, lo studente dovrà mostrare, tra le altre cose, «pieno e consapevole rispetto degli altri e delle istituzioni scolastiche»; se invece dovesse mostrare solo «pieno rispetto degli altri e delle istituzioni scolastiche», senza consapevolezza, e se gli altri indicatori andassero di pari passo, meriterebbe il 9. Ci vuole, immagino, una raffinata ermeneutica per distinguere in concreto la fattispecie del pieno rispetto da quella del pieno e consapevole rispetto. La griglia quindi non è immediatamente applicabile. Qualcuno potrebbe anche obiettare che non sempre il rispetto pieno e consapevole è superiore al pieno rispetto. Immaginiamo un santo (o uno psicotico), ebbro di amore per gli altri e per tutte le forme di vita. Forse il suo grado di consapevolezza – perlomeno nel senso della lucidità razionale – potrebbe essere insufficiente. Magari ha tanto rispetto per le istituzioni dello Stato che piange di commozione ogni volta che vede il cancello della scuola, mentre il suo più consapevole compagno lo oltrepassa con distrazione, già pensando alla versione di latino che lo aspetta, anche se – ovviamente – non si sognerebbe mai di imbrattarlo o di prenderlo a calci. Siamo sicuri di voler dare il 9 al folle e il 10 al consapevole? Forse bisognerebbe allora discutere, complicare la griglia di valutazione, fare differenze ulteriori tra il pieno rispetto consapevole, non consapevole, infra-consapevole, sopra-consapevole (è difficile definire questi concetti, occorrerebbe molto tempo, ed è probabile che non si giungerebbe comunque a una soluzione condivisa). Oppure – visto che comunque l'esercizio di ermeneutica è inevitabile, griglia o non griglia – si potrebbe essere tentati di buttare a mare la griglia, e di affidarsi a concetti maggiormente sintetici, come quelli più tradizionali di comportamento *buono* o *cattivo* (e intermedi). Che se poi qualcuno chiedesse (con protervia?) che cosa vuol dire comportarsi *bene*, si potrebbe sempre rispondere elencando da 7 a 12 *items* (tra i quali un «pieno e consapevole rispetto degli altri e delle istituzioni scolastiche» non potrebbe mancare).

E tuttavia, alcune questioni si pongono qui: 1) se vi sia appunto un limite all'interpretazione, se cioè svolgendo la sintesi (comportarsi *bene*) in un'analisi di 7 o 12 *items*, abbiamo davvero fatto chiarezza, per noi e per l'utenza, o se non abbiamo piuttosto prodotto nuova ambiguità (che costituisce a sua volta un motivo di conflittualità), che richiederebbe di essere disambiguata; 2) se in generale non torniamo ad avere piuttosto bisogno di sintesi, che non (solo) di analisi ed elencazioni; 3) se l'opzione epistemica per un tendenziale azzeramento dei margini interpretativi, figlia della cultura analitica, non solo rischi di non mantenere che in parte le proprie promesse

di chiarezza e distinzione, ma possa addirittura rivelarsi controproducente: davvero vogliamo educare le nuove generazioni (e i genitori, i docenti, il personale tecnico-amministrativo...) a non sapere concepire il senso complessivo se non come una somma o una mera computazione di significati parziali?; 4) se educare al controllo passo per passo dei processi, nella (forse vana) speranza di eliminare approssimazione ed arbitrio, non costituisca un'ingegnerizzazione dell'atto educativo, che rischia di non trascurarne alcunché, eccetto l'essenza.

Questo numero intende discutere l'impatto delle tecniche e procedure di valutazione sulla produzione del *sapere* e della *realtà* sociale, nonché sulla definizione stessa di ciò che sono il *sapere* e la *realtà* in ambiti differenti (medicina, economia, management, psicologia, sociologia, filosofia, politiche pubbliche e non...).

È indubbio che tali procedure ingegnerizzate abbiano dato luogo a clamorosi infortuni, come la famosa A 1 con *outlook* positivo elargita da Moody's a Lehman Brothers nel marzo del 2008 (<https://www.moody.com/research/Moodys-affirms-Lehmans-A1-rating-outlook-now-stable--PR-151071>). Gli stessi sistemi di VQR (valutazione della qualità della ricerca) sono andati incontro a vibrante proteste, che non sembrano derivare dall'indisponibilità di professori e ricercatori universitari a essere valutati (si veda ad esempio, in Italia, l'appello per la filosofia di R. Esposito, A. Fabris, G. Reale: [http://www.lascuola.it/it/home/editrice\\_detail/un-appello-per-la-filosofia/tutte-le-news/](http://www.lascuola.it/it/home/editrice_detail/un-appello-per-la-filosofia/tutte-le-news/)).

La valutazione c'è sempre stata. L'esame di coscienza, per esempio, era una forma di vita esaminata, che si riteneva fosse indispensabile per acquisire l'abito della virtù. La catena delle riflessioni, ponendo ogni volta a oggetto lo stadio di consapevolezza appena raggiunto, costituisce, nell'idealismo trascendentale, il motore dialettico del progresso. In entrambi questi esempi la valutazione non serve a fotografare l'esistente ma a sollecitarlo verso un miglioramento che ha che fare con il perseguimento di una maggiore somiglianza con l'essenza e con la verità di ciò che viene esaminato. Sia nell'esame di coscienza sia nella riflessione trascendentale, però, possiamo riscontrare una tendenza egologica della valutazione, il rischio di mancare l'obiettivo del rinnovamento e di accontentarsi dell'esame, della fotografia: un eccesso di analisi e di riflessione può addirittura nuocere alla produzione di realtà e di novità.

La questione che si pone è quindi in primo luogo quale sia il rapporto tra la riflessione e la novità. In secondo luogo, si pone il problema di come le tecniche di valutazione condizionino le concrete modalità di produzione del sapere. "Valutare" non significa semplicemente fotografare quello che c'è, ma – in molti casi – costringe a una mutazione radicale dei metodi e degli oggetti stessi della ricerca. Il valutatore diventa quindi produttore, le tecniche di valutazione tecniche di produzione del sapere.

Se le agenzie di valutazione e *rating* sono agenzie di produzione della realtà, che non si limitano a selezionare i prodotti migliori disponibili sul mercato, ma impongono una linea di prodotti (una linea di politica economica, una linea politica *tout court*...) a esclusione di altre, la valutazione intrattiene un rapporto critico con la libertà e con la democrazia.

Una griglia di valutazione unilaterale può costituire un greve apparato categoriale, gettato sulla realtà per coartarla e per normalizzarla, impedendo anche solo che si affacci

quel *resto* di trascendenza, che pure nei differenti ambiti è quanto apprezziamo di più: il genio nell'arte, l'intuizione in economia, la visione in politica...

Nell'ambito specifico delle *humanities* e delle discipline sociali, si tratta di costruire una matrice gnoseologica della produzione & valutazione, di cui ancora si avverte la mancanza. Essa dovrebbe tenere conto di variabili specifiche, come – ad esempio – la potenza di un'idea. Come valutarla? In tale matrice, le ragioni dell'analisi non dovrebbero – crediamo – sopravanzare quelle della sintesi. Sebbene i valutatori (come molti filosofi) si concepiscano solo come analisti, andrebbe sempre ricordato che – come diceva Goethe – “la prova di una analisi ben riuscita sta nella sintesi”. E forse, proprio questa è la più immediata e naturale definizione della filosofia e delle scienze che vi si connettono: sapere sintetico.

Sulla base di tali considerazioni, si pongono (almeno) le seguenti domande:

- 1) in che modo, nei differenti ambiti, la valutazione condiziona la produzione?
- 2) in che modo la valutazione determina a priori ciò che deve (e in base a quale concezione di “dovere”) essere visto/riconosciuto, e per conseguenza potenziato?
- 3) qual è il rapporto tra analisi e sintesi nei processi di valutazione? Che rapporto c'è tra valutazione analitica e (iper-)specialismo? Possiamo andare verso nuove sintesi del sapere? Come?
- 4) c'è un nesso, e se sì quale, tra la pervasività degli attuali metodi di valutazione, l'aggressività delle nuove oligarchie (finanziarie, burocratiche, del sapere), l'arretramento della democrazia?
- 5) come è possibile valutare la *qualità*? Che cosa s'intende oggi, e che cosa si dovrebbe intendere per “impatto” di un programma di ricerca?
- 6) quale antropologia è sottesa ai processi dominanti di valutazione? In che modo questi lasciano spazio a considerazioni non meramente meccanicistiche o riduzionistiche dell'essere umano?
- 7) può darsi un conflitto tra la valutazione e la realtà da valutare, più o meno a quel modo in cui un eccesso di riflessione può risultare nocivo alla vita?
- 8) sono possibili *tecniche* alternative di valutazione in economia, in politica e nelle *humanities*? Se sì, quali?

La sfida non consiste tanto nel sottrarsi alla valutazione, ma nel criticarla e trasformarla. La qual cosa non può essere demandata solo ai tecnici della meta-valutazione, ma richiede un ampio coinvolgimento comunitario di tutti coloro che avrebbero titolo a prendere la parola, specialmente di coloro che di solito sono tenuti ai margini dei processi decisionali.

Enrico Guglielminetti  
Angelo Miglietta  
Paolo Moderato



## EVALUATION? YES, BUT NOT THIS WAY!

## EDITORIAL

Our society, based on the sharing of information and knowledge, is also an evaluation society. The parents of school age children have the right, for example, to know the criteria on the basis of which grades are given. Almost everywhere this generates evaluation grids. A Turin high school requires the consideration of between 7 and 12 items in order to justify a “behavior grade” (that is, a grade assessing general behavior at school) on a scale from 10 to 5. Let us consider the first of such items. In order to deserve a 10, students will have to show, among other things, “total and conscious respect for the others and the educational institutions.” If students showed “total respect for the others and the educational institutions” but without consciousness, all other markers being equal they would deserve a 9. Very refined hermeneutic skills, I would think, are undoubtedly required in order concretely to discriminate between “total respect” and “total and conscious respect.” The grid is thus not immediately applicable.

Someone could also object that it is not always the case that total and conscious respect is superior to total respect. Let us imagine a saint (or a psychotic), full of love for the others and for all forms of life. It could be that such an individual’s degree of consciousness—at least in the sense of rational lucidity—is insufficient. Such a student might perhaps have such a high respect for state institutions to be moved to tears every time the student sees the school gate. On the contrary, some other schoolmate might walk by the gate indifferently and absent-mindedly, already thinking of the Latin translation waiting for the students in class; even though, obviously, this schoolmate would never think of soiling or kicking the gate. Are we sure that we should give a 9 to the fool and a 10 to the one who is aware?

Perhaps one should debate this, complicate the evaluation grid, further differentiate between total and conscious, non-conscious, infra-conscious, super-conscious respect. But defining such concepts is difficult, would require a lot of time, and probably we would not reach a shared solution. Or, given that some hermeneutic exercise is inevitable whether we employ a grid or not, we might be tempted to throw away the grid and use more synthetic concepts such as the traditional notions of *good* and *bad* (and their intermediate). If one were then to ask (with some arrogance?) what behaving *well* means, one could always reply listing between 7 and 12 items (and among them, “total and conscious respect for others and the educational institutions” could not be absent).

Some questions emerge here, though. First, whether there is a limit to interpretation; that is, whether, once we have unfolded the synthesis (behaving *well*) into an analysis of 7 to 12 items, we have really reached some clarity for us and the grid users or whether we have instead produced some new ambiguity (which is in itself a new reason for conflicts), which would in turn require disambiguation. Second, whether in general we may be at a point where we again need more synthesis than simply analysis and lists. Third, whether the epistemic option for a progressive annulment of the interpretative margins, which is a legacy of the analytic tradition, not only risks succeeding only partly

in its promise for clarity and distinction, but may also disclose itself as counterproductive. Do we really want to educate the new generations (and parents, teachers, staff and support personnel, etc.) to conceive the total meaning merely as a sum or mere computation of partial meanings? Fourth, whether educating to a step-by-step process control in the (perhaps vain) hope of eliminating approximation and arbitrariness does not ultimately amount to engineering the educational practice, which ends up neglecting nothing except the essence.

The present issue of “SpazioFilosofico” aims at discussing the impact of techniques and procedures of evaluation on the production of *knowledge* and social *reality* as well as their effect on the very definition of what *knowledge* and *reality* are within the various fields (medicine, economics, management, psychology, sociology, philosophy, public and non-public policies, etc.).

Without doubt such engineered procedures of evaluations have produced clamorous accidents, such as the notorious A1+ rating assigned by Moody’s to Lehman Brothers in March 2008 ([https://www.moody.com/research/Moodys-affirms-Lehmans-A1-rating-outlook-now-stable--PR\\_151071](https://www.moody.com/research/Moodys-affirms-Lehmans-A1-rating-outlook-now-stable--PR_151071)). The systems of research quality evaluation (RQE) have been met with vigorous protests that do not originate from the unwillingness of university professors and researchers to be evaluated (see for example, in Italy, the plea for philosophy by Roberto Esposito, Adriano Fabris, and Giovanni Reale ([http://www.lascuola.it/it/home/editrice\\_detail/un-appello-per-la-filosofia/tutte\\_le\\_news/](http://www.lascuola.it/it/home/editrice_detail/un-appello-per-la-filosofia/tutte_le_news/))).

Evaluations have always occurred. Soul-searching, for example, was a form of examined life that was deemed indispensable for the acquisition of a habit of virtue. In transcendental idealism, the chain of reflections constitutes the dialectical engine of progress by posing as its object of reflection the stage of consciousness that has just been achieved. In both examples, evaluation works not in the sense of photographing reality but rather in the sense of spurring it toward an improvement that has to do with pursuing a closer resemblance with the essence and reality of that which is under examination. Both in soul-searching and in transcendental reflection, one can notice a self-oriented tendency, though; that is, the danger of missing the target of renewal and of instead making oneself happy with the examination, with the picture—an excess of analysis and reflection can be even detrimental to the production of reality and novelty.

The question to be addressed is, first of all, the relation between reflection and novelty. Secondly, there is the issue of how evaluation techniques condition concrete modalities of knowledge production. “To evaluate” does not simply mean to photograph that which exists; in many cases, evaluations force a radical change on the research methods and objects. The evaluator becomes producer, and the evaluation techniques become techniques of knowledge production.

Evaluation and rating agencies are agencies that produce reality; they do not confine themselves to selecting the best products available on the market but rather impose a line of products (a line of economic politics, a line of politics *tout court*, etc.) to the exclusion of others. Thus, evaluation stands in a critical relation with freedom and democracy.

A unilateral evaluation grid may represent a gloomy categorial apparatus that is cast on reality so as to coerce and normalize it and that thus prevents the even minimal

appearance of a *rest* of transcendence, which is conversely that which we appreciate the most—geniality in art, intuition in economics, vision in politics, etc.

In the specific field of the humanities and the social sciences, the issue is that of constructing an epistemological matrix of production and evaluation that we are conversely still missing. The matrix should take into account specific variables such as, for example, the power of an idea. How to evaluate it? We believe that, in the matrix, the reasons of the analysis should not overwhelm those of the synthesis. Although evaluators (as well as many philosophers) think of themselves only as analysts, one should always recall that, as Goethe used to say, “the proof of a successful analysis lies in its synthesis.” This is perhaps the most immediate and natural definition of philosophy and of the sciences related to it: synthetic knowledge.

On the ground of such considerations, at least the following questions may be raised:

1. In which ways, within the specific fields, does evaluation affect production?
2. In what ways does evaluation determine *a priori* that which *must* be seen/recognized and therefore empowered (that is, according to what conception of “must”)?
3. What is the relation between analysis and synthesis in evaluation processes? What is the relation between analytic evaluation and (hyper)specialization? Can we move toward new syntheses of knowledge? How?
4. Is there a connection, and which, between the pervasiveness of the current evaluation methods, the aggressiveness of the new (financial, bureaucratic, epistemic) oligarchies, and the regression of democracy?
5. How can one evaluate *quality*? What do we mean today, and what should we mean, with “impact” of a research program?
6. What anthropology underlies the dominating evaluation processes? In what ways do they allow for considerations of the human being that are not purely mechanistic or reductivistic?
7. Can there be a conflict between evaluation and reality to be evaluated, more or less similarly to the way in which an excess of reflection can be detrimental to life?
8. Is it possible to have alternative evaluation technics in economics, politics, and the humanities? If so, which?

The challenge lays not so much in avoiding as much as in critiquing and transforming evaluation. This cannot be delegated only to the meta-evaluation technicians; rather, it demands a wide community engagement on the side of all those who have some stake in the matter, especially those who are usually kept at the margins of decision-making processes.

*Enrico Guglielminetti*  
*Angelo Miglietta*  
*Paolo Moderato*

(Translated by Silvia Benso)



# TEORIA



Peter Dahler-Larsen

THE EVALUATION SOCIETY:  
CRITIQUE, CONTESTABILITY AND SKEPTICISM

**Abstract**

*The essay begins with the observation that evaluation relates to its object in a manner that is not only descriptive, but rather constitutive. Five domains where the constitutive effects of evaluation occur are presented and illustrated. Next, three kinds of social critique are discussed, and counter-critique is offered. Each of these critiques is shown to coincide with particular ideas and roles such as “the authentic life before evaluation,” “the rational architect of evaluation systems” or “the victim of evaluation.” Finally, by using the concept of “contestability differential” as a can-opener, all evaluations are shown to rest on a combination of something which is contested with something which is not contested. On that basis, the essay concludes with a discussion of how a critique of evaluation can be cultivated in a democratic society.*

1. *The evaluation wave*

Despite the inherent flaws in trying to reduce society to any *one* overarching dimension or principle (Morin 1988), the term *The Evaluation Society* does in fact capture many essential, frightening and awe-inspiring aspects of contemporary society. We do live in a society where evaluation, accreditation, auditing, benchmarking, performance management, quality assurance and similar documentation practices produce *datascares* as an important dimension in social life along with *idea-scapes*, *ethnoscares*, *technoscares* etc. (Appadurai 1996). The function of these *datascares* cannot be exhausted with reference to their descriptive aspects; instead it appears that they help define or constitute what they claim to measure. This observation raises the obvious question to which extent the contemporary evaluative grips on reality are conducive to structuring, regulating, and governing the social order in particular ways. Evaluating institutions may not be able to articulate how this takes place. So, the social, political and philosophical story-telling about evaluation should not be left to evaluators. What platform or position can be identified from which critique of the evaluation society can be articulated?

One reason why it is difficult to air critique is that evaluations are occupied with some large and positively sounding terms as *quality*, *sustainability*, *impact*, *equality*, *development*, *learning*, *transparency*, *innovation* etc. Since evaluators in their own view operate with indicators that approximately aspire to capture quality etc., evaluators often cannot understand why anybody would logically be against evaluation. Who are not in favor of quality? In contradistinction to earlier ideological tensions or class cleavages in society,

the tensions around, say, quality, appear to be non-existent, because according to those in favor of evaluation, everybody must be able to get on board the mission for quality. If evaluation successfully captures all positive concepts little space is left for alternative views. The lack of recognition of conflictual material in the very ambition to achieve quality makes it difficult to argue that there even exists the possibility of a critical position.

It should also be noted that it is not without personal risk to seek such position. A story illustrates this. At a conference, a new bibliometric evaluation system for researchers was debated. A university lecturer aired a harsh critique of the attempts to measure quality of research through bibliometric indicators. The presenter at the conference session replied that in his view, there is a strong correlation between a researcher's bibliometric score and his or her general reputation. In other words, researchers with a good reputation have nothing to fear and have no particular reason to be critical. The breathtaking implication, never articulated, is of course that the critical academic was critical because he was not a good researcher. To immunize oneself against that kind of tacit accusations it would be necessary first to score well on bibliometric indicators and then prepare one's critique. Critics would thus have to work hard to earn the right to air their views. However, if they achieved good bibliometric scores, their motivation to undermine the trustworthiness and social acceptability of the score would be reduced. Perhaps one of the most important social logics of performance indicators is exactly this *divide et impera* between high-scoring and low-scoring members of the same group, regardless of the validity of the indicator.

In other words, in our attempt to articulate a position of critique and study what happens, we can learn quite a bit about how smartly and cleverly the evaluation society functions. The strategies it adopts in incorporating critique and fending it off may be quite advanced.

There is another reason why we need to consider critique of the evaluation society carefully. Critique often has a tacit normative component ("it would be much better if..."). Every critique identifies with some agent or position. It is important to be clear about these normative frameworks. If not, we run the risk of airing a critique that is not in sync with its own time and not sufficiently respectful of the subtleties of what it is critiquing.

Let me put this problem in another way. I teach students in political science. I teach them evaluation because evaluation is one of their functions in their future jobs. I also teach them to be critical of evaluation because I think it is an important socio-political phenomenon that no one should be blind or ignorant about. What do I expect of the critical views? That they are so special or so naïve or so normatively self-enclosed that they must be put aside when real evaluation is to be done... as if critics and evaluators have nothing to say to each other? Or do I think that the critical view is so advanced and so relevant that it must be taken seriously by evaluators, too? Truly, critique finds itself in a very ambiguous and delicate situation when it turns out that critique becomes useful in order to improve the social systems it criticizes, but contemporary capitalist and bureaucratic organizations have already for some time cleverly integrated various forms of critique into more optimal forms of system operations (Boltanski and Chaipello 2007). However, alternatively I would also be concerned if my students were evaluators

until 5 p.m. and then critical after 5. p.m. The critical view should be aware of its own situatedness in society and acknowledge its own engagement in society.

My strategy in this paper is to begin, in an axiomatic way, with the illustration and depiction of an idea which I think is central to today's discussion about the evaluation society: The idea that evaluation stands not in a descriptive, but in a constitutive relation to what it claims to measure. Then I will subject this idea to three kinds of fairly conventional forms of critique, but what is new is that I will also be critical towards the critique. In the final section I will introduce the concept of "a contestability differential" as an ever-present element in evaluation and I will discuss, on this basis, whether it is possible to live with evaluation in a democratic way, or perhaps, even to cultivate the critique of evaluation and the democratic potential in evaluation at the same time.

## 2. *Constitutive consequences of evaluation*

Quantification begins with establishing the categories into which social phenomena are put so that they can be counted (Porter 1994). Statistical work not only reflects reality but establishes it by providing the players with a language to put reality on stage and act upon it (Desrosières 2002, p. 352).

Desrosières thus suggests that there is a constitutive element in the very act of measurement (putting reality on stage in a particular way) as well as in the subsequent act *upon* that construction of reality (which may make the construction more "real"). In social life, we can imagine that these two kinds of acts are difficult to separate (one is done with the other in mind), but we can also imagine that a particular evaluative staging of reality is not very successful if it does not lead to subsequent acts. If successful, however, evaluation can produce *constitutive consequences*.

To make this construct more visible and operational, I suggest we can observe these consequences in five domains. Under each domain I shall give examples of effects that are (perhaps) surprisingly concrete, while it should be remembered that this is exactly how such effects become embodied, in the concrete rather than in the abstract.

First, evaluation has an impact on the content on some work or practice. For example, testing in education is known to lead to "teaching-to-the-test". The test has a "wash-back" effect upon teaching itself, not only upon the organization of lessons, but also upon the curriculum and the choice of topics and materials (McNeil 2000).

Second, evaluation has an impact of timing of practices. Like budgets which are defined on a monthly or yearly basis and thus impact upon the timing of economic behavior, evaluation regimes impose their own rhythms on practices. For example, museums, schools, universities, hospitals and prisons have institutional dispositions for particular ways of anchoring themselves in time but their "goals" and "effects" become located in time in new ways according to how they are evaluated. Many evaluative issues are difficult to measure with validity, so time often becomes the universal currency in which quality and performance are expressed. How quickly? How often?

Third, evaluation has an impact on the (re-)configuration of social roles and identities. For example when students are asked to assess their degree of satisfaction with a teaching program, a new student role emulated on the basis of a consumerist approach

to education emerges. Next, teachers teach in particular ways when they are subject to student satisfaction surveys. Different degrees of popularity among teachers may challenge teacher-to-teacher relations and put different teachers in different relations with their school managers and principals.

Evaluation thus suggest a set of interrelated viewpoints quite similar to what Marx called *Charaktermasken*. There is a structural basis for the kinds of masks or roles which individuals take on in the evaluation society, such as “producer”, “consumer”, “manager” etc. At the same time, *Charaktermasken* are indicative of some level of ambiguity in relations between roles (can we unmask?), some question of the cleverness with which masks are carried (how cleverly do you perform with your mask?), and some tension regarding how each individual negotiates the relation between role-playing and personality (to what extent should I see my evaluation results as something that characterize me personally?).

These three kinds of constitutive effects tend to be interconnected like words, timing, and roles in a drama. They enroll each other in a larger evaluative assemblage.

Therefore, fourth, constitutive effects of evaluation (of the three kinds above) tend to coalesce into a larger world view that provides a sort of integrated or mythical image of what is going on. For example, with bibliometric indicators of research, it is suggested that what is interesting about research is only a particular kind of output called publications. The different kinds of products are allocated different kinds of statistical weight (depending on reputation, “impact factor”, etc.). On that basis synthesizing scores are developed. The overarching assumption that makes all this possible is that research should be understood as production. In a similar vein, an underlying assumption in PISA is that education is international competition. The meaningfulness of this idea is undergirded by an assisting myth which is that all countries have the same educational goals (Meyer 2008).

Fifth, the constitutive effects of evaluation extend to how we know, to our sources of knowledge. If an indicator has the implications suggested above, the meaning of an indicator changes when it is used as a part of an evaluation regime (Vulliamy and Webb 2001). When people change their interpretations and their actions as a result of the indicator, the indicator as a form of knowing is implied. I will now push this argument one step further and show that the same kind of argument also applies to other kinds of knowledge (officially regarded as knowledge or not) which are touched upon or enrolled by evaluation regimes. This is important because the richness of our insight into a particular phenomenon or practice under evaluation often depends on several kinds of knowledge.

Let me give two examples related to bibliometric indicators of research. A new bibliometric indicator has been defined in my country. All publications on a predefined list of journals and publishers (which does, regrettably, not include “Spazio Filosofico”) are allocated a certain amount of points. With the help of academic committees, all publications are divided into two groups depending on their reputation. Only the best 20 % is allowed into the privileged group that get more points than the rest. The purpose of this differentiation is to prevent a situation in which all researchers just produce more publications of bad quality in order to score more points, an effect known from a study in Australia (Butler 2003).

A comparative score of sums of points broken down by institution goes into an algorithm that determines the allocation of research funds across institutions every year.

In principle, one of the main assumed advantages in bibliometrics is the objectivity of counting publications as opposed to the alleged subjective assessment in peer review. This particular advantage, however, is illusionary, because the bibliometric indicator is not independent from peer review but rather dependent on it. All publications in the bibliometric system are subject to some form of peer view in order to be categorized as respectable enough for being counted. As a consequence, editors of books will contact me and say things like: “Professor Dahler-Larsen, we really invite you to contribute a chapter to a new anthology. It is subject to peer review and you will earn bibliometric points for your contribution. However, given your experience in the field, I am sure that the peer review will not lead to a rejection of your wonderful contribution. So please accept our invitation.”

Since the bibliometric indicator builds on peer review, it must be included by those who need to use bibliometric points in their negotiations with others. Whilst peer review in its classical meaning could assume both formative and summative functions, it is now deprived of the latter. The link or association between bibliometrics and peer review (or enrollment of the latter by the former, in Latour’s terms) does not leave the latter unchanged. In my analysis: As a function of the bibliometric indicator, we are now less sure of what the peer review might mean than we were before, but we do depend on it for our indicator system to function.

A second example: I asked one of the architects of the bibliometrical system how he wanted to make sure that I in fact planned to aim at the most prestigious publications just because they gave more points than the other ones. I could devise a cool strategy to make a high number of points just by producing many low-ranking publications. He answered that I was welcome to do so, but he believed that I would be so sensitive to my colleagues’ assessment of my work that it is in my own interest to make sure that the balance between high-ranking and low-ranking publications on my CV is not too skewed. I concluded that in order to not run amok, the bibliometric indicator still needed to be balanced with more conventional academic values. In other words, to work well, the indicator needs to prey on values that it does not itself embody. In a similar vein, we can imagine that other indicators in order to not produce totally anti-social behavior, still need that we know such things as norms, reputation, helpfulness, good practice etc. At the same time, it may also happen that a formal indicator tends to undermine or redefine the meaning of other forms of knowing that are embedded in other social norms and practices. We cannot just assume that collegial relations and professional conscience remain the same after a new evaluation regime is introduced which preys on but does not respect these other norms and forms of knowing.

*Critique one: Evaluation is antithetical to authentic life*

I shall now, as promised, discuss a number of critiques of the evaluation society that all respond, one way or another, to the observation that evaluation helps constitute something. The first of these forms of critique says that evaluation constructs artificial artefacts and is therefore antithetical to authentic life.

For example, using the practice of teaching and the categories of constitution mentioned above as an example, evaluation imposes a measurement regime which is against the very nature of teaching. Evaluation intervenes in the definition of content that would otherwise be chosen freely based on pedagogical considerations, evaluation intervenes in the spontaneous relations between teachers and students, evaluation imposes artificial time frames upon the teaching practice, evaluation confuses the reality of testing performance with real learning, and it undercuts the forms of knowing that springs from learning as an existential, relational and contextually embedded form of experience.

However, the problem with this line of reasoning is that it assumes an authentic and natural form of teaching and engagement with teaching that is ontologically prior to our knowledge-creation about teaching. If we argue that evaluation is against the very nature of teaching we tacitly assume that teaching springs out of nature and we thereby ignore the many investments human beings have made in different epochs in the phenomenon of teaching (socially constructed views of the human child; the role of authority; the role of education in relation to society; the changing visions of the good society to which good education is a preparation; etc.).

In a broader perspective, the critique that says evaluation is antithetical to authentic life tends to assume a certain pre-social destiny handed down to us. It is the identification with this pre-given order of things that allows the critique of evaluation to point to the *artificial* nature of data. This view risks lending itself to uncomfortable subscriptions to a metaphysical order of life. It too easily allies itself with a traditional, religious or even totalitarian undertone. We *know* what the authentic life is, and it commands us to live in a particular way.

However, if you listen to such commands, you can hear them in many variations, sending you off in different directions. How far should we go back? If books and newspapers and diaries are tools for systematic reflection, should they also be abandoned? Should we abandon thermometers and ask ourselves if we feel warm? Should we live like the Amish? Should we break all mirrors because they allow us to see ourselves from the outside? Or is it OK to make a systematic data-based inquiry into the effects of tobacco on lung cancer, but not OK if we call it an evaluation?

The command that sends us back to an “authentic” form of life must ignore, in Cornelius Castoriadis (1997) view, the responsibility we have as modern human beings to organize our own world and make our own laws. It also ignores, I believe, Gianni Vattimo’s (2004) observation that if we “take on” and “work through” the contemporary socio-philosophical condition, we cannot operate with a “handed-down” or metaphysical guarantee to support any argument.

However, a modified and humble or “weak” variation of the argument is possible. It goes like this. It is not possible to be reflexive about everything. No social system can question all its operations at the same time. There simply is no capacity for that (Bateson 1972). If the evaluation society promises endless development, endless change, endless accountability, and endless reflexivity, it is giving us illusions. In fact, quite a lot of social critique in recent years says that the ever-performing subject is in fact presently at war with itself (Han 2012). Too much flexibility can be destructive of the social fabric of

norms and of personal values (Sennett 2002), and it is possible to recommend a certain personal and social solidity that resists endless re-definition (Brinkmann 2014).

Evaluation transports a modern technical mentality according to which life consists of components that can be measured and replaced (Berger, Berger and Kellner 1973). Truly, there are forms of life that are embedded in frames of normativity that cannot be subject to any kind of evaluative perspective, any kind of componentiality. I am thinking of care, love, memory, pride, self-respect, geniality etc.

Modern existence seems to be caught in a paradox. Once we have discovered the reflexive standpoint, it is difficult to live as if the spontaneous form of life is the only one possible. We know it is not. We also know that there are existential “choices” or “ways of being” that lose their meaning if they are subjected to any kind of evaluative perspective. On a scale from one to ten, how do you assess the love of each of your children? There is great paradox in the fact that it requires (some kind of) reflexivity to even choose to protect such forms of life from (some kind of) reflexivity. It is difficult to *choose* to live spontaneously and authentically. It is like using one compartment of life to protect another compartment of life without succumbing to the compartmentalization of life.

Nevertheless, contemporary contributions to a critique of endless reflexivity seem to suggest that we doom ourselves if we have no “brakes” on mechanisms that enhance reflexivity. So, in a revised and moderated form, this first critique suggests that there is something which perhaps should be protected from evaluation not because it is authentic but because we care and find it wise to protect it.

*Critique two: Evaluation has counter-intentional side effects*

This critique often takes a starting point in the observation that measurement of complex phenomena is bound to be imperfect. Thus, indicators of the quality of public services, the impact of research, the innovation in the public sector, and sustainability of climate policies etc. are marred by flawed validity. Nevertheless, in a managerial context, these indicators are used for all sorts of purposes anyway (accountability purposes, steering purposes, information purposes etc.).

When imperfect measures are used, evaluation often has unintended consequences. So, according to this kind of critique, the problem with the constitutive consequences of evaluation is that they are *unintended* constitutive consequences. For example, if we measure the time from patients arrive at the emergency room until they encounter a nurse, some hospitals hire nurses to immediately say “hello” to each patient. The world is full of examples in which you can live up to what is being measured without living up to the intention behind the measurement. The discrepancy between the two is rooted in the validity problem described above.

For that reason, some advocates of evaluation spend quite a lot of time refining and cultivating the indicators used in evaluation, a process called purification by Latour (2004).

It can also be recommended to use a broader set of indicators (because there is an evaluation deficit in what is not being measured so far) or to use a more narrow set of indicators (because the general purpose has been lost in a jungle of measurements). There is a whole range of evaluative techniques concerning who gets measured how and

when all of which can be varied in order to improve validity. For example, the measurement of *effects* is almost like a whole discipline in itself that includes various schools of thought.

What remains, however, is that as long as measurements are *approximations* to the perfect measure, there will be unintended consequences of evaluation in practical use. This idea is not extremely radical, because it is accepted as a sensible middle ground between strong critics and strong believers in evaluation, performance measurement etc. (Norman 2002). The key point is perhaps only whether there are so many and so important unintended consequences of evaluation that they constitute a substantial reason for objection, and not least importantly, whether these unintended consequences can somehow be repaired.

What I would like to stress here, however, is the underlying identification of that kind of critique with the idea of *intentions* in evaluation. Logically, unintended presuppose an intention on the other side of the conceptual coin. However, a number of questions can be asked here (Dahler-Larsen 2014).

How can intentions be captured empirically if they are not stated? Do not tell me we can trust official political declarations of intentions! Which intentions count? Do we imagine an architect behind the evaluation whose intentions we share? Could other players have intentions, too, and what if all these intentions are not in alignment? If people invent new intentions, do we then go back to some “original” intentions or do we allow people to invent intentions along the way? Do we assume intentions behind a particular indicator, evaluation, or evaluation system? What if a network of evaluative phenomena amounts to a whole surveillant assemblage (Haggerty and Ericson 2000)? Is it not meaningless to assume one set of intentions behind such dynamic network?

The critique that claims that evaluation has unintended consequences more often than not *identifies with an icon of an evaluation architect* that rationally seeks to plan and control evaluation with the best of all intentions, but, alas, unfortunately, misses the target because the indicators fail to support him all the way. Would it really be better if evaluation was planned and controlled all the way? And perhaps even more importantly: Why should an analytical perspective identify with the so-called architect of evaluation when there are so many other perspectives in society one can identify with? If a scientific perspective is one that does not identify with any particular part in a political situation, why should evaluation research identify with this imaginary and overly rational evaluation architect?

Why miss the evolving and dynamic character of spontaneous evaluative initiatives? Some of these initiatives may have constitutive effects that are, in fact, not counter-intentional, but rather quite consistent with *some* political intentions (such as the redefinition of content, the reconfiguration of social relations in the direction of something more flexible, componential, and marketable). But my counter-critique goes one step further.

If we acknowledge that statistics are constitutive of what they claim to measure, and we apply the intended/unintended distinction thereto, perhaps we too early curse a measurement because it was not agreed upon or it was not collectively intended rather than in fact study and understand how it, for better or for worse, feeds into our collective sense-making and society-building. In a democratic deliberative perspective,

for instance, would we accept that some say “my values would support proposal A”, and others would say “I fear that proposal B would be disadvantageous for the weakest members of our society”, but not accept if one said “I have done a survey that leads to the conclusion that proposal C is the best proposal”?

In fact there exist some areas of political contention where the very ambition to do research or evaluation helps constitute that area as one that deserves attention. Some measurements of risks qualify here (Beck 1992). The same is true with the whole area of the “psycho-social work environment”.

One might argue that numbers are used strategically to make an argument more objective or technical than it deserves to be, because it is really just a statement from a particular viewpoint. However, if we insert into our common deliberations the no-longer-radical idea that statistics are social constructions, too, we can acknowledge their pragmatic and socially constructive qualities without succumbing to them as if they were cast in stone. Numbers can fool us in a thousand ways (!). But it is also part of the history of numbers (e.g., as embodied in the metric system) that they are supposed to help us agree to some common understanding of some aspect of something (Porter 1995).

The intended/unintended distinction is not one that deserves to be applied routinely to the constitutive aspects of evaluation, as if it leads to the highest wisdom of all to know whether a phenomenon that happens is or is not in alignment with some reconstruction of some alleged original intentions.

### *Critique three: Evaluation is power*

According to this third kind of critique, evaluation cannot be understood apart from its specific historical and institutional embeddedness. Many have found Orwell’s “Big Brother” and Foucault’s panopticon to be prime metaphors for understanding how the evaluation society combines surveillance with a structuration of the modern social order. Foucault’s contribution is to highlight how techniques for measurement, documentation and comparison become practices for governing at a distance in way that also involves discipline and self-scrutiny of subjects (in the interesting double meaning of “subject to” and “subject for”).

In education and in other fields there is a rich literature on colonizing evaluation practices that refer to Foucault (Shore and Wright 1999). Although it is probably correct in pointing to the link between evaluation practices and the larger institutional order, as well as to the production of monitorable subjects, this paradigm perhaps assumes too much of a centrally located point of observation, too much of a one-directional observation, and too much certainty about what is produced of what we have called “constitutive effects.” I am reminded of Žižek’s provocative warning that if we say that the outcome of totalitarianism is determinately known and nothing but tragic, we are almost giving the totalitarians too much. The best key to understanding these analyses, I think, is that they tacitly *identify with the victims of evaluation*. As if this category of victim is analytically easy to define, as if the members of this category are defined through and through by the “character mask” they wear, and as if the strategic move of victimization in itself supplies members of this category with some moral superiority. And as if the analysis of victimization takes place in a totally different world from the one in which

victimization takes place. If the analysis of victimization is correct, how is there even space for a critical analysis? Maybe these questions are not asked because if victims are morally superior, it would not be a good idea to search for alternative positions. It is better to remain a victim. The clearer the power structure, the easier the identification with victims.

An attempt to paradigmatically update the surveillant assemblage in a more “undeterminate” direction is provided by Haggerty and Ericson (2000). They assume, with Latour, that there are scattered centers of calculations that are not necessarily hierarchically related (sometimes police is filming a demonstration, but activists also film the police). Some of the new technologies of documentation and registration (cameras, survey software) are inexpensive and dispersed in ways that do not conform to authoritarian hierarchies. There is instead “a highly fractured rhizomatic criss-crossing of the gaze such that no major population groups stand irrefutably above or outside of the surveillant assemblage” (Haggerty and Ericson 2000, p. 618).

There is a potentiality in surveillant assemblages that becomes actualized only in particular ways when particular connections are made. There are constant negotiations going on, and new connections lead to the invention of new uses, and sometimes “endless redefinitions and reconfigurations” (Callon 2010, p. 165). (I am not sure that Foucault would object to observations like these; what I note, however, is that some of his epigons do not take up that research agenda).

When risk is imposed upon a part of a political steering system, that part is like to push back in order to avoid the risk, which leads to “spiraling regulatory logics” (Rothstein, Huber and Gaskell 2006).

Thévenot and the “pragmatic sociology” (Boltanski and Chiapello 2007) take the discussion of Foucault in a slightly different direction. They argue that a Foucauldian world is unlivable. They believe, with Durkheim, that any society needs some sort of moral fabric. In modernity, we have a high number of moral repertoires to draw from in our construction of institutionalized solutions to common problems that are seen to be more or less legitimate. In other words, as a corollary, evaluation practices need some form of justification which can, in principle, be interactively debated.

For example, in a case study in Denmark, I followed upper secondary school teachers who were discussing the meaning and consequences of student satisfaction surveys in their schools. When doing so, they drew on different repertoires. A part of the discussion had to do with whether student satisfaction data are truthful, valid and reliable. Another aspect dealt with fairness and justice, for example whether it is fair to compare schools in different socio-economic districts and whether it would have been more fair to include teachers in the planning of the survey at an earlier stage. It was also discussed, at the same time, whether the student satisfaction surveys could be useful for improving the student climate at the school. Truthfulness, fairness and utility all served as registers from which to draw arguments.

Such an analysis perhaps focuses too much at the micropolitics of evaluation inside the upper secondary school, but I admit that micropolitics are connected to macropolitics: the broader education policy, marketization of schools, increased competition among schools, etc. All I am suggesting is that in a particular case, it may be worth looking not only at how evaluation supports one-sided and hierarchical power

structures, but also how evaluation connects with actual arguments in a fragmented, diverse, and dynamic structure of power, including local negotiations.

Perhaps an a priori and general theoretical commitment to *either* a hierarchical power structure *or* a more flexible, diverse, fragmented and reflexive social order as conflicting paradigms is misplaced. Perhaps any particular socio-historical situation and any particular case study present us with a unique configuration that may draw differentially on the two paradigms, respectively.

### 3. *The contestability differential*

In a world handed to us by God or by tradition, evaluation cannot be carried out. Evaluation assumes that some aspect of social life is contingent. Evaluation is a planned inquiry deliberately designed to induce contingency. Evaluation assumes a set of expectations about potential social change, much like concepts in modernity open up a new horizon of expectations (Koselleck 2007).

Evaluation challenges a particular aspect of social life by saying: I will measure your quality, and maybe you need to change in order to improve what I define as your quality. It is the job of evaluation to make the evaluand contestable. It is easier to make sure evaluation is used if there is conflict and a pressure to act in the evaluation situation (Lederman 2012) which is equivalent to saying the evaluand is contested.

Evaluation is a special kind of social/institutional initiative because it is a practice that is deliberately organized in order to change another practice. To do so effectively, evaluation must protect itself from contestability. Evaluation needs to be backed up by, say, belief in methodology and data, in the credibility of the institution that carries out evaluation, and in the virtues related to using evaluation for good purposes such as learning or improvement. If evaluating institutions cannot count on such beliefs, they must have the power to carry evaluation anyhow. Without any of these social anchors, an evaluation would be futile.

A metaphor: Assume someone is using force to turn a screw with a screwdriver. Imagine that the screw is solidly anchored and the connection with the screwdriver is strong, and the person has no solid position on the ground, then the force exerted will in fact lead to a turning of the person in space instead of a turning of the screw. The person needs to make sure that his weight makes his feet stand solidly on the ground as he turns the screw. He or she also must make sure that the screwdriver has a solid grip on the screw. A child may not be able to do it. Perhaps it takes several attempts from a strong and heavy person with skills.

The same with evaluation. To function effectively, an evaluation must exploit the differential between the (relative) fluidity of the social material it seeks to change and the (relative) solidity of its own fixation in the world. I call this difference “the contestability differential.” All evaluation plays with the difference between what is solid and what is not solid.

Alternatively, when a contestability differential cannot be established, evaluation cannot take place. We may have so much strength and power in traditions and in institutions that they cannot be evaluated (that is why we do not evaluate flags or royal

families or the best and worst wars). OR: An evaluation is criticized so much for corrupt indicators, a filthy evaluation process, a manipulated result, or a lack of independence from political interest that perhaps there is not enough justification for using it. These observations, too, correspond to a failure in establishing a contestability differential: If the evaluation becomes more contested than what it seeks to evaluate, then it cannot operate.

Sociologically speaking, evaluation is a modern phenomenon that thrives on reflexivity and contingency. Evaluation makes its object soft and contestable and fluid. On the other hand, any particular evaluation itself needs to be relatively firmly anchored in something that is more solid and less contested. Evaluation can take on the “taken-for-granted” character that constructivists (Berger and Luckmann 1967) and institutionalists (Scott 1995) talk about. Evaluation can find support in normative, cognitive and regulatory institutional pillars such as belief in data-based decision-making, or incentives based on evaluation results. In Latourian language, we can talk about so many solid associations with various actants (people, inscription devices, resources, sanctions etc.) that it becomes possible for evaluation to operate as a “black box” that can be inserted as an operative element in large networks of activity.

The advantage of the contestability differential as a concept is that it allows us to see evaluation as a powerful force that (like the market) has the restructuring of social orders and relations as a primary function, without conceptually committing ourselves to always seeing evaluation fixed in the same way to any particular ideology or institution. Evaluation lends itself to more than one ideological agenda (Kipnis 2008). Neither are we committed to assuming that evaluation works deterministically in every instance. To work as a construction, it must first be constructed.

How does this take place in practice? It is necessary to ask this question because our belief in the value of the concept of contestability differential is sustained if the concept can be operationalized and used in empirical analysis.

Several options are available. For example, the evaluand (the object of evaluation) can be criticized for lack of effectiveness, quality etc. This seems to be one of the strategies that politicians use against public institutions such as schools. A softening of the object always makes evaluation easier. Next, evaluators and managers can talk smoothly about the many good consequences of evaluation (learning and development). They can align themselves with powerful institutional forces (expertise, manpower, management, financial incentives, legal consequences) and weave evaluation into organizational structures and processes through scripts and recipes such as “evaluation cultures”, “evaluation capacity”, “evaluation policies”, and a “general need to be learning-oriented and flexible”. They can also connect evaluation with evaluation imaginary (Schwandt 2009) in the larger social environment such as the myth of development or the myth of assurance (Dahler-Larsen 2012; Power 1997), the latter assuming that evaluation is the best response to a cultural anxiety about risk, crisis and potential disaster.

When evaluation is in a very powerful position, it does no longer need to justify itself (thus the deteriorating influence of “evaluability assessment”, an old-school procedure in evaluation which served to check whether a potential evaluand was in fact ready to be evaluated) (Dahler-Larsen 2014).

In that case, the contestability differential works very well. The abolishment of evaluability assessment indicates that the belief in systematic evaluation has become so strong that evaluation does not need to justify itself in each and every instance.

However, it may be costly to establish a strong contestability differential. It is always a delicate matter how much resources and how much institutional power should be invested in systematic evaluation. Evaluation may be expensive, and evaluation based on institutional force is only complied with as long as the subjects are faced with sanctions and a sense of necessity. It is difficult managerial balance to achieve the benefits of soft control while only reverting to harder forms when necessary. In some situations, there is a struggle between different elements of varying degrees of contestability, and evaluation has to fight from house to house.

In contemporary evaluation there is sometimes a structure, a function or an organization that serves as an “evaluation machine” without any subjective or human representation. When we are scared about or worried about “evaluation”, we are in fact faced with a large network of institutional elements, people, inscription devices, and resources that enroll us as “actants” with a particular “character mask” in relation to evaluation. We cannot always see the whole hinterland behind this construction. Nor are we interested. There is a “metadata” paradox here (Desrosières 2002). Although, from a technical or methodological perspective, we are interested in all the factors that influence on how evaluation results are produced, in a practical sense, we are not. We would be tired or die of boredom (Lindeberg 2007) or react much too slowly if we were to appreciate and understand all the details necessary to produce the large-scale evaluations; what we are faced with is that they are *actionable* already.

As analysts, however, it is our duty to tell a longer story.

#### 4. *Evaluation and democracy*

If we recognize that all evaluation is built on some manifestation of a contestability differential, there is no universal normative prescription that commands us to identify with a particular “character mask” a priori or with a particular foundational principle or myth that structures evaluation. Instead, we should be skeptical about general standpoints and general identifications. Personally, I am very skeptical of aligning evaluation with too much power, i.e. too solid a contestability differential. I am uncomfortable with the usurpation of political power and democratic roles by evaluating institutions (Neave 1997). In my personal view, there is more need than ever before to debunk the way that evaluating institutions build up their contestability differentials through a variety of means. I am particularly skeptical about the automatization, institutionalization and standardization of evaluation as it takes place in alliance with powerful organizations, and I am skeptical of the link between evaluation and ideologies such as a neo-liberal idea of all-encompassing productivity, marketization and competitiveness.

At the same time, I am also skeptical about generally not being willing to build any contestability differential that would make evaluation possible. We can learn from the sociology of knowledge that all knowledge is due to some element of social

construction, and some element of “black boxing” of what we think we know, but nevertheless we are doomed to live in a world where we must responsibly construct knowledge. Knowledge is a capacity to act (Stehr 2001). We are also doomed to find out how we can best handle our common social and political destiny through democracy. If politics and democracy are those domains where society works upon itself (Rosanvallon 2009), there is no need to totally abandon systematic and deliberate knowledge-production, although we have, of course, learned that knowledge production is not just descriptive but also constitutive. I do acknowledge that the “we” included in the previous sentence is also potentially contested. In Rosanvallon’s perspective, democracy is always historical and situational; we may later learn to discredit principles that served us well in an earlier epoch. Our democratic knowledge production by definition has a preliminary character.

Faced with that, it is democratically possible to ask for more evaluation in one area (wars) and less evaluation in others. I also think it would be fair to argue that some aspects of life are better protected without evaluation, although we may later change our priorities.

In a democratic context, it is a difficult task to build exactly the kind of contestability differential needed for a particular form of evaluation in a particular situation, not more, not less. We have to acknowledge the paradox inherent in this endeavor. We know we build that which we take for granted for a while. Everything can be contested, but not much at the same time. We know we have purposes, but our instruments do more than just help us with fulfilling these purposes in a transparent way. If evaluation is constitutive, it is by definition infused with ambiguity (Best 2008). There are no instruments that constitute things in a pure way. There is always an overflow. The good news is that once we have discovered the idea of the contestability differential as an ever-present ingredient in evaluation, we seek to provide any contestability differential with only the preliminary, temporary and fragile status it deserves.

I cannot be much more precise, but it is this kind of skeptical thinking about evaluation that I deem consistent with a democracy in which humans know that they set their own laws and live with the consequences. I think it would also be consistent with a kind of weak thinking (Vattimo 2004) according to which we construct our collective arguments paradoxically acknowledging that there are no firm foundations or guarantees undergirding our arguments about what we think we know. Evaluation can be based on no general and winning argument about truth, fairness or utility. We have to dare to make the humble and situated arguments as we go along.

### *References*

- A. APPADURAI (1996), *Modernity at Large. Cultural Dimensions of Globalization*, University of Minnesota Press, Minneapolis MI 1996.
- G. BATESON (1972), *Steps to an Ecology of Mind*, Ballantine Books, New York NY 1972.
- U. BECK (1992), *The Risk Society*, Sage, London 1992.

- P.L. BERGER-B. BERGER-H. KELLNER (1973), *The Homeless Mind. Modernization and Consciousness*, Vintage Books, New York NY 1973.
- P.L. BERGER-T. LUCKMANN (1967), *The Social Construction of Reality*, Doubleday, New York NY 1967.
- J. BEST (2008), *Ambiguity, Uncertainty, and Risk: Rethinking Indeterminacy*, in “International Political Sociology”, 2 (2008), pp. 355-374.
- L. BOLTANSKI-E. CHIAPELLO (2007), *The New Spirit of Capitalism*, Verso, London 2007.
- S. BRINKMANN (2014), *Stå fast - et opgør med tidens udviklingstrang*, Gyldendal, København 2014.
- L. BUTLER (2003), *Explaining Australia’s Increased Share of ISI Publications. The Effects of a Funding Formular Based on Publication Counts*, in “Research Policy”, 34 (2003), pp. 565-574.
- M. CALLON (2010), *Performativity, Misfires, and Politics*, in “Journal of Cultural Economy”, 3 (2/2010), pp. 163-169.
- C. CASTORIADIS (1997), *World In Fragments. Writings on Politics, Society, Psychoanalysis, and the Imagination*, Stanford University Press, Palo Alto CA 1997.
- P. DAHLER-LARSEN (2012), *The Evaluation Society*, Stanford University Press, Palo Alto CA 2012.
- P. DAHLER-LARSEN (2014), *Constitutive Effects of Performance Indicators: Getting Beyond Unintended Consequences*, in “Public Management Review”, 16 (7/2014), pp. 969-986.
- A. DESROSIÈRES (2001), *How Real Are Statistics? Four Possible Attitudes*, in “Social Research”, 68 (2/2001), pp. 339-355.
- K.D. HAGGERTY-R.V. ERICSON (2000), *The Surveillant Assemblage*, in “British Journal of Sociology”, 51 (4/2000), pp. 605-622.
- B.-C. HAN (2012), *Træthedssamfundet*, Møller Forlag, København 2012.
- A.B. KIPNIS (2008), *Audit Cultures: Neoliberal Governmentality, Socialist Legacy, or Technologies of Governing?*, in “American Ethnologist”, 35 (2/2008), pp. 275-289.
- R. KOSELLECK (2007), *Begreber, tid og erfaring*, Hans Reitzels Forlag, København 2007.
- B. LATOUR (2004), *Why Has Critique Run Out of Steam? From Matters of Fact to Matters of Concern*, in “Critical Inquiry”, 30 (2/2004).
- S. LEDERMAN (2012), *Exploring the Necessary Conditions for Use in Program Change*, in “American Journal of Evaluation”, 33 (2/2012), pp. 159-178.
- T. LINDEBERG (2007), *Evaluative Technologies: Quality and the Multiplicity of Performance*, Copenhagen Business School, København 2007.
- G NEAVE (1998), *The Evaluative State Reconsidered*, in “European Journal of Education”, 33 (3/1998), pp. 265-284.
- R. NORMAN (2002), *Managing Through Measurement or Meaning? Lessons from Experience with New Zealand’s Public Sector Performance System*, in “International Review of Administrative Sciences”, 68 (2002), pp. 619-628.
- T.M. PORTER (1994), *Making Things Quantitative*, in “Science in Context”, 7 (3/1994), pp. 389-407.
- M. POWER (1997), *From Risk Society to Audit Society*, in “Soziale Systeme”, 3 (1997), pp. 3-21.
- H. ROTHSTEIN-M. HUBER-G. GASKELL (2006), *A Theory of Risk Colonization: The Spiralling Regulatory Logics of Societal and Institutional Risk*, in “Economy and Society”, 35 (1/2006), pp. 91-112.

- W.R. SCOTT (1995), *Institutions and Organizations*, Sage, Thousand Oaks CA 1995.
- R. SENNETT (2002), *The Corrosion of Character. The Personal Consequences of Work in the New Capitalism*, W. W. Norton & Company, New York NY 2002.
- C. SHORE-S. WRIGHT (1999), *Audit Culture and Anthropology: Neo-Liberalism in British Higher Education*, in "The Journal of the Royal Anthropological Institute", 5 (4/1999), pp. 557-575.
- N. STEHR (2001), *The Fragility of Modern Societies. Knowledge and Risk in the Information Age*, Sage, London 2001.
- G. VATTIMO (2004), *Nililism and Emancipation: Ethics, Politics, & Law*, Columbia University Press, New York NY 2004.

Enrico Guglielminetti

## VALUTAZIONE E NOVITÀ

### Abstract

*The thesis of this essay is that the current research quality evaluation system is allergic to greatness. The mantra of innovation seems intentionally made to caution us against novelty. Through an excursus on Foucault's trans-subjectivization and Benjamin's immanent critique of the work of thought, the essay contrasts the assessment of greatness with process and product evaluation, and provides recommendations aimed at identifying synthetic indicators of novelty and greatness within the humanities.*

«È da tenere presente che quando uno capiti a meditare su un comandamento al cui riguardo constati non avere abitudine alcuna di peccare, non occorre che vi si fermi tanto tempo; ma, secondo che trovi se stesso più o meno in fallo rispetto a quel comandamento, più o meno dovrà soffermarsi nella considerazione e verifica di esso, e lo stesso ci si regoli per i peccati mortali».

(IGNACIO DE LOYOLA, *Esercizi Spirituali*, § 242)

### 1. *Trans-soggettivazione e auto-soggettivazione*

La valutazione c'è sempre stata. Costituisce – potremmo dire – un universale antropologico<sup>1</sup>. Individui e organizzazioni si prendono cura di sé, riflettono sui risultati

---

<sup>1</sup> Proprio per questo, essa va incontro anche ad aberrazioni, tra le quali una delle più perniciose consiste nel valutare una cosa irrilevante in luogo di una rilevante: «Nei paesi comunisti la valutazione (*bodnocení*) e il controllo dei cittadini è l'attività sociale principale e costante. Se un pittore deve ricevere il permesso di esporre, se un cittadino deve ottenere il visto per andare in vacanza al mare, se un calciatore deve entrare nella nazionale, per prima cosa si devono raccogliere tutti i giudizi e le informazioni su di lui [...]. Questi giudizi però non hanno mai a che fare con la capacità (*schopnosti*) del cittadino di dipingere, con la sua capacità di giocare a pallone o con la sua salute che necessita di un soggiorno al mare. Hanno a che fare semplicemente con quello che viene chiamato il “profilo politico del cittadino” [...]. Dal momento che ogni cosa (la vita di ogni giorno, l'avanzamento nel lavoro, anche le vacanze) dipende da come il cittadino sarà valutato (*bodnocen*), chiunque (se vuole giocare a pallone nella nazionale, oppure organizzare una mostra o passare le vacanze al mare) deve comportarsi in modo tale da ricevere un

raggiunti, sugli ostacoli che incontrano o che hanno incontrato, sulla distanza che ancora li separa dalla meta finale (o intermedia), sulla definizione stessa di questa meta.

In una lezione del 10 febbraio 1982, Foucault – riferendosi in specie ad Hadot – distingue tre forme della cura di sé, o della conversione: l'*epistrophē* platonica, la *metanoia* cristiana e – appunto – la cura di sé in epoca ellenistica e romana. Secondo Foucault, questa costituisce una terza via rispetto alle prime due. È forse proprio quest'ultima – aggiungiamo noi – che costituisce il modello della “valutazione”, come oggi è concepita.

Secondo Foucault, «mentre l'*epistrophē* platonica consisteva in un movimento che poteva condurci da questo mondo all'altro – dal mondo di quaggiù, al mondo di lassù – la conversione di cui ora ci stiamo occupando, nella cultura ellenistica e romana, ci induce a spostarci da quanto non dipende da noi, a ciò che invece dipende da noi»<sup>2</sup>. Ugualmente netta è la differenza con la *metanoia* cristiana: «Innanzitutto, la conversione cristiana implica un mutamento improvviso. Ovviamente, quando lo definisco improvviso, non voglio affatto dire che tale mutamento non possa, e che addirittura non debba, essere stato preparato, e preparato persino da molto tempo, da tutto un itinerario. Ma, preparazione o meno, sforzo oppure no, ascesi o assenza di ascesi, resta nondimeno il fatto che, in ogni caso, affinché vi sia conversione, è necessario un evento unico, improvviso, al contempo storico e metastorico, capace di sconvolgere e trasformare, d'un colpo, il modo d'essere del soggetto»<sup>3</sup>. In altre parole, «può esservi conversione solo a condizione che vi sia, all'interno dello stesso soggetto, una rottura. Il sé che si converte è un sé che ha rinunciato a se stesso»<sup>4</sup>.

Tutt'altra grammatica regola la *conversio ad se* ellenistica: «In primo luogo, nella conversione ellenistica e romana non esiste propriamente una rottura»<sup>5</sup>. Più precisamente, abbiamo una rottura del sé rispetto al mondo (bisogna distogliersi dal mondo, per concentrare l'attenzione esclusivamente sul sé), «che non è, però, una rottura di sé rispetto a sé»<sup>6</sup>. «Se dunque la conversione (nel senso della *metanoia*) cristiana o postcristiana» – continua Foucault –, realizzandosi come rottura, «rappresenta una sorta

---

giudizio favorevole (*zbohdnocen příznivě*)» (M. KUNDERA, *Nesnesitelná lehkost bytí*, Sixty-Eight Publishers, Toronto 1985, p. 90; trad. it. G. Dierna e A. Barbatò, *L'insostenibile leggerezza dell'essere*, Adelphi, Milano 1994, p. 102). Se definiamo *para-valutazione* la valutazione di una cosa per l'altra, la questione che oggi infuria nelle università italiane è se – fatte le debite proporzioni – il giudizio, per esempio, circa il grado di internazionalizzazione e diffusione o circa la collocazione editoriale di una ricerca non costituisca appunto una *para-valutazione* rispetto alla ponderazione della capacità dello studioso di produrre ricerca di qualità. Il dibattito perderebbe gran parte della sua virulenza, se – anziché parlare di valutazione della *qualità* della ricerca – si discutesse più sobriamente di valutazione del grado di internazionalizzazione o dell'impatto attuale di una ricerca. Se, in luogo di una VQR (valutazione della qualità della ricerca), avessimo una VIR (valutazione dell'impatto della ricerca), dove per “impatto” si intendesse una quantità (come tale misurabile, per esempio attraverso indici citazionali), e *non* una qualità, molte delle osservazioni avanzate in questo contributo sarebbero superate. Il problema è dunque solo nominalistico? Forse, ma talvolta in un nome c'è più di quanto non si sospetti.

<sup>2</sup> M. FOUCAULT, *L'herméneutique du sujet. Cours au Collège de France. 1981-1982* (d'ora in poi = HS), Seuil/Gallimard, Paris 2001, p. 202; trad. it. M. Bertani, *L'ermeneutica del soggetto. Corso al Collège de France (1981-1982)*, Feltrinelli, Milano 2004<sup>2</sup>, p. 186.

<sup>3</sup> HS, pp. 202-203; it., p. 187.

<sup>4</sup> HS, p. 203; it., p. 188.

<sup>5</sup> *Ibidem*, *ivi*.

<sup>6</sup> HS, p. 204; it., p. 189.

di trans-soggettivazione», la conversione ellenistico-romana «non costituisce [...] un modo per introdurre nel soggetto, lasciandovene il segno, una cesura essenziale. La conversione è qui, piuttosto, un processo lungo e permanente che vorrei definire non tanto di trans-soggettivazione, bensì di auto-soggettivazione», nel quale, «dopo aver fissato se stessi come obiettivo e meta», si cerca di stabilire «un rapporto pieno e adeguato di sé con se stessi»<sup>7</sup>.

Nei testi di Plutarco, Epitteto, Seneca o Marco Aurelio, il *se convertere ad se* (convertirsi a sé) ha quindi un significato ben distinto da quello platonico, da un lato, e da quello cristiano, dall'altro. La consegna di volgere lo sguardo verso se stessi ha qui di mira una concentrazione: come suggerisce Plutarco, occorre «essere come un cane che si tiene al guinzaglio, avere lo sguardo ben fisso dinanzi a sé, non pensare ad altro che non sia un obiettivo e una meta»: «L'esercizio di concentrazione del soggetto è un esercizio per mezzo del quale tutta l'attività, e tutta l'attenzione, del soggetto dovranno essere riportate verso quella tensione che lo conduce al suo scopo finale»<sup>8</sup>. Si tratta di una «concentrazione di tipo teleologico [...]. Si tratta di avere costantemente davanti agli occhi, nel modo più chiaro possibile, ciò verso cui si tende, e di avere in un qualche modo una coscienza chiara di tale scopo, di quello che è necessario fare per arrivare a esso, delle possibilità effettive che si hanno di raggiungerlo»<sup>9</sup>. «A dover diventare oggetto di coscienza, di vigilanza, di attenzione [...] è dunque quel che ci separa dalla meta, è la distanza tra noi stessi e lo scopo finale. Come potete vedere – continua Foucault – per intendere tutto ciò dobbiamo pensare, di conseguenza, a una concentrazione di tipo atletico, a qualcosa come la preparazione alla corsa o alla lotta»<sup>10</sup>. Come nel tiro con l'arco, «si tratta di pensare alla traiettoria che ci separa da ciò verso cui vogliamo muovere, o da ciò che vogliamo raggiungere. E pertanto, ciò su cui dovremo concentrare tutta la nostra attenzione dovrà essere proprio questa traiettoria, che va da sé a se stessi»<sup>11</sup>. «Così, credo sia proprio la presenza di sé a sé – a causa della distanza che ancora sussiste tra sé e sé – ovvero la presenza di sé a sé che si ha nella distanza di sé da se stessi, a dover diventare l'oggetto e il tema del rovesciamento dello sguardo. Lo sguardo, che un tempo era orientato sugli altri, dovrà ora essere riportato non tanto a sé come oggetto di conoscenza, ma per l'appunto alla distanza rispetto a se stessi. Il soggetto dell'azione ha certo a disposizione, per raggiungere il proprio scopo, degli strumenti, ma soprattutto ha come imperativo quello di raggiungerlo. E la cosa, che egli deve raggiungere, è il proprio sé»<sup>12</sup>.

<sup>7</sup> *HS*, p. 206; it., p. 191.

<sup>8</sup> *HS*, pp. 212-213; it., p. 198.

<sup>9</sup> *HS*, p. 213; it., p. 199.

<sup>10</sup> *Ibidem*, *ivi* (trad. it. leggermente modificata).

<sup>11</sup> *HS*, p. 214; it., p. 199.

<sup>12</sup> *HS*, p. 214; it., pp. 199-200 (trad. it. leggermente modificata). Nell'epoca ellenistica, secondo Foucault, non risulta peraltro «mai del tutto chiaro», né viene «mai definitivamente stabilito, se il sé rappresenti qualche cosa a cui si fa ritorno in quanto già dato in partenza, o se il sé sia invece una meta che ci si deve prefiggere e a cui si potrà eventualmente, nel caso si pervenga alla saggezza, avere alla fine accesso» (*HS*, p. 205; it., p. 190).

## 2. Valutazione e assoluto

Nel suo scritto giovanile su *Il concetto di critica d'arte nel romanticismo tedesco*, Walter Benjamin stabilisce un nesso strettissimo tra valutazione e assoluto. Secondo Benjamin, solo le vere opere d'arte (ma il discorso si estende a tutte le opere di pensiero) sono criticabili. Se qualcosa è criticabile, dunque valutabile, allora è un'opera autentica. Non solo: se qualcosa è un'opera autentica, allora la valutazione non interviene da fuori, non è una critica soggettiva, che stia in capo cioè a un soggetto terzo della valutazione, ma è immanente all'opera stessa. Le opere, che sole sono valutabili, sono per essenza autovalutazioni. Non c'è prima l'opera e poi la valutazione, ma l'opera stessa è valutazione, a quel modo in cui – secondo la teoria fichtiana – non c'è prima il pensiero e poi la riflessione, ma il pensiero è la riflessione. Nel medium dell'arte, come in quello del pensiero, la valutazione immanente è da un lato l'autodistruzione della forma di esposizione dell'opera singola, dall'altro la sua salvazione, la sua elevazione all'infinito e all'assoluto. Nel concetto romantico di critica d'arte, qual è stato sviluppato specialmente da Friedrich Schlegel e da Novalis, si ritrova quindi indubbiamente una tendenza mistica. Ma – per Benjamin – si tratta, qui come altrove, di una *profane Erleuchtung*, di una illuminazione profana. Intrecciando il concetto romantico di critica d'arte con quello holderliniano di “sobrietà” (*Nüchternheit*) dell'arte, Benjamin giunge alla conclusione che la criticabilità/valutabilità immanente dell'opera faccia corpo con la sua razionalità, con la sua strutturale ripulsa dell'irrazionale e del mito (cui invece, almeno in qualche misura, soggiace Goethe), con il carattere prosaico (romantico, nel senso di romanzesco) della poesia in quanto tale. La poesia, se è vera, è assoluta; ma questo infinito poetico non ha nulla di ebbro o di folle, piuttosto qualcosa di meccanico, esatto: l'infinito deve potersi insegnare. L'umile mestiere del recensore, o del critico (e Benjamin non ha mai voluto essere altro che un critico), non fa altro che proseguire la tendenza all'autovalutazione che ogni opera è. Con ciò, la critica – come suona la chiusa della dissertazione benjaminiana – procede all'«assolutizzazione dell'opera»; ché anzi il processo critico-valutativo «può essere espresso, con un'immagine, come la generazione dell'abbagliamento nell'opera. Tale abbagliamento – la luce sobria – spegne la molteplicità delle opere. È l'idea»<sup>13</sup>.

Vi sono certo elementi aporetici nella concezione romantica, che Benjamin già rileva in questa sede, e su cui non mancherà di lavorare in seguito. L'opera successiva di Benjamin si può anzi intendere complessivamente come un'emendazione della concezione romantica, cui pure Benjamin continuerà a restare legato in punti essenziali. Il primo elemento aporetico è quello che si potrebbe definire un eccesso di positività. Appena nata, l'opera – come autovalutazione – è già dissolta (nel che va ravvisato l'elemento propriamente critico della critica), ma è dissolta nell'assoluto. Ogni opera

---

<sup>13</sup> W. BENJAMIN, *Der Begriff der Kunstkritik in der deutschen Romantik* (vol. 3 di ID., *Kritische Gesamtausgabe*, a cura di C. Gödde e H. Lonitz in collaborazione con il Walter Benjamin Archiv; d'ora in poi = *GA* 3), a cura di U. Steiner, Suhrkamp, Frankfurt a.M. 2008, p. 131 (indico tra parentesi anche la paginazione della precedente edizione in ID., *Gesammelte Schriften*, vol. I: *Abhandlungen* [d'ora in poi = *GS* I], a cura di R. Tiedemann e H. Schweppenhäuser, Suhrkamp, Frankfurt a.M. 1980, p. 119); trad. it. C. Colaiacomo, *Il concetto di critica nel romanticismo tedesco*, in ID., *Opere Complete*, vol. I: *Scritti 1906-1922*, a cura di E. Ganni, Einaudi, Torino 2008, p. 449.

autentica coincide con l'assoluto, e – tramite questo – con ogni altra opera. Nel medium dell'arte, l'opera appare come una monade: riflette dentro di sé ogni altra opera, e tutte le opere la riflettono. In forza di questa specularità, l'arte – come i romantici la concepiscono – non è in fondo altro che il *Nous* di Plotino<sup>14</sup>. La critica è dunque un trarre in salvo ciò che è già da sempre salvo, valutare è valorizzare. In questo modo si lascia però senza risposta il problema della valorizzazione di ciò che *non* vale, che – sotto forma di apocatastasi – diventerà uno dei nuclei propositivi del pensiero di Benjamin. Con un discorso temerario, Benjamin si chiederà se non sia possibile una redenzione, se cioè non si possa in qualche modo allargare le maglie del *Nous*, costringendo a entrarvi anche un materiale refrattario. Per i romantici, invece, ciò che è dentro (poetico), è dentro, e ciò che è fuori (im-poetico), è fuori: l'unica decisione possibile – come suggeriva Novalis – è se qualcosa sia poesia oppure no<sup>15</sup>, all'interno di questa decisione non ve ne sono altre: non si può decidere se una poesia valga di più o di meno, perché ogni vera poesia è già l'assoluto; e non si può decidere se ciò che non è poesia si avvicini di più o di meno alla poesia, appunto perché si tratta di ambiti incomunicanti, separati da un intervallo infinito.

Il secondo elemento aporetico, strettamente connesso con il primo, è l'eccesso di continuità. Tutte le opere costituiscono un'opera sola, che è poi l'infinito. Non esiste separazione possibile né tra l'opera e l'infinito, né delle opere tra loro. Con sempre maggiore vigore, fino alle tesi sul concetto di storia, Benjamin sottolineerà invece il significato cruciale dell'interruzione. Non solo c'è ovviamente interruzione tra il contesto colpevole di ciò che vive e l'avvento del tempo messianico, ma la stessa verità – cioè appunto l'Intelligenza – ha per Benjamin, come appare chiaro dalla *Vorrede all'Origine del dramma barocco tedesco*, una struttura discontinua.

In questo senso, il confronto tra i romantici e Goethe, che occupa l'ultima sezione dell'opera sul *Concetto di critica d'arte nel romanticismo tedesco*, ha un significato cruciale, individua esattamente «il problema sistematico fondamentale della filosofia dell'arte [...] come il problema del rapporto di idea e ideale nell'arte»<sup>16</sup>. Da un lato i romantici, l'idea dell'arte; dall'altro Goethe, l'ideale dell'arte. Il punto debole di Goethe, se così è lecito esprimersi, è che per lui, «di fatto, la critica dell'opera d'arte non è né possibile, né necessaria [...]. Goethe rifiuta di riconoscere la criticabilità come momento essenziale dell'opera d'arte»<sup>17</sup>. Il prezzo di questo rifiuto è però il mito<sup>18</sup>, cioè – come si potrebbe

<sup>14</sup> Nell'Intelligenza dell'arte, questo *mirroring* è portato a perfezione: «Lassù tutto è trasparente, nulla è tenebroso e impenetrabile, ognuno è manifesto ad ogni altro nel suo intimo e in ogni dove, poiché la luce è manifesta alla luce. E infatti ognuno porta in sé tutto e in ogni altro vede tutto: perciò ogni cosa è dappertutto, ogni cosa è tutto e ciascuno è tutto e lo splendore è infinito. Ciascuno di quegli esseri è grande, poiché lassù anche il piccolo è grande; lassù il sole è tutti gli astri, e ogni stella è un sole e tutti gli astri insieme. Eppure, ogni singolo essere è diverso, e, nello stesso tempo, tutte le cose appaiono in esso» (PLOTINO, *Enneadi*, V 8, 4, 4-11; cito seguendo la trad. it. a cura di G. Faggini, Bompiani, Milano 2000, p. 911). Ogni opera è luce, ma la luce è manifesta alla luce.

<sup>15</sup> Cfr. *GA* 3, p. 85 [= *GS* I, p. 79]; it., p. 414.

<sup>16</sup> *GA* 3, p. 128 [= *GS* I, p. 117]; it., p. 447.

<sup>17</sup> *GA* 3, p. 130 [= *GS* I, p. 119]; it., p. 448.

<sup>18</sup> «Anche il concetto goethiano di stile, in ultima analisi, racconta un mito» (*GA* 3, p. 130 [= *GS* I, p. 118]; it., p. 448).

dire – una salvezza “parmenidea”<sup>19</sup>, una salvezza senza salvazione delle opere stesse, perché ciò che davvero vale, nell’arte, non è ciò che è creato (dunque, l’opera), ma ciò che è increato, dunque gli archetipi, o gli *Urphänomene*<sup>20</sup>. L’ideale goethiano, come si potrebbe dire, salva l’uomo ma senza l’uomo, salva la produzione, ma senza i prodotti. Per questo, per lui, le opere non sono “frammenti”, che concregono nell’unità dell’assoluto, come per i romantici, ma semplici “torsioni”: «Ma ciò che vietava a Schlegel questa soluzione [...] era il fatto che essa conduce a una valutazione estremamente condizionata della singola opera (*zu einer höchst bedingten Einschätzung des einzelnen Werkes führt*)»<sup>21</sup>.

Da un lato dunque una salvezza astratta, che salva tutto eccetto ciò che v’è da salvare, e che per Benjamin, teorico della redenzione, costituisce una forma del mito e della follia (c’è mito, ogni volta che la regola tende a fare piazza pulita del regolato). Dall’altro però un eccesso di positività e di continuità, una felicità a prezzi di saldo, che i principi goethiani della discontinuità e della insuperabile differenza tra ciò che è prodotto e ciò che non lo è aiutano a emendare. La sintesi tra queste due posizioni, l’idea cioè di una felicità *a caro prezzo*, costituirà la risposta benjaminiana al «problema sistematico fondamentale della filosofia dell’arte».

Consideriamo ora in maggiore dettaglio l’idea romantica della “valutazione”: «In quanto è conoscenza dell’opera d’arte, la critica è autoconoscenza dell’opera stessa; nella misura in cui la giudica (*beurteilt*), questo avviene come autovalutazione (*Selbstbeurteilung*) nell’opera»<sup>22</sup>. Questa autovalutazione è però solo impropriamente un giudizio: «Del tutto atrofizzato è infatti, in essa, un momento necessario di ogni giudizio (*Beurteilung*), quello negativo [...] il momento positivo di questo potenziamento della coscienza (*Bewußtseinssteigerung*) supera di gran lunga quello negativo»<sup>23</sup>. Ecco dunque emergere con nettezza il tratto peculiare della critica romantica, e cioè «la piena positività di questa critica, per cui essa si differenzia radicalmente dal concetto moderno»<sup>24</sup>. «È chiaro: per i romantici la critica è molto meno il giudizio (*Beurteilung*) su un’opera, che non il metodo del suo compimento»<sup>25</sup>. In questo senso, la critica è una funzione estatica: «La critica adempie il suo compito se, quanto più conclusa è la riflessione e più rigorosa la forma dell’opera, tanto più variamente e intensamente le porta fuori di sé (*heraustreibt*), risolvendo in una riflessione più alta la riflessione originaria, e così di seguito»<sup>26</sup>.

I romantici sollecitano dunque una «critica immanente (*immanente Kritik*)» dell’opera<sup>27</sup>. È questo «il fondamento di un genere di critica completamente diverso, che non si atteggia in modo giudicante (*nicht beurteilend eingestellten*), e la cui importanza maggiore non

<sup>19</sup> L’ideale goethiano dell’arte esprime «ciò che nell’arte è immobile in senso eleatico» (*GA 3*, p. 124 [= *GS I*, p. 114]; it., p. 443).

<sup>20</sup> In questo senso, l’arte «non è creazione, ma natura (*nicht Schöpfung, sondern Natur*)» (*GA 3*, p. 123 [= *GS I*, p. 112]; it., p. 442).

<sup>21</sup> *GA 3*, p. 125 [= *GS I*, p. 114]; it., p. 444.

<sup>22</sup> *GA 3*, p. 71 [= *GS I*, p. 66]; it., p. 402.

<sup>23</sup> *GA 3*, pp. 71-72 [= *GS I*, p. 66]; it., p. 403.

<sup>24</sup> *GA 3*, p. 72 [= *GS I*, p. 67]; it., p. 404.

<sup>25</sup> *GA 3*, p. 74 [= *GS I*, p. 69]; it., p. 405.

<sup>26</sup> *GA 3*, p. 79 [= *GS I*, p. 73]; it., p. 409.

<sup>27</sup> *GA 3*, p. 83 [= *GS I*, p. 77]; it., p. 412.

consiste nella valutazione (*Einschätzung*) della singola opera, bensì nell'esposizione delle sue relazioni con tutte le altre opere e, infine, con l'idea dell'arte»<sup>28</sup>. «La critica, dunque, esattamente al contrario della concezione attuale della sua essenza, non è, nella sua intenzione centrale, giudizio (*Beurteilung*), ma, da un lato, compimento, completamento, sistematizzazione dell'opera; dall'altro, la sua dissoluzione nell'assoluto [...] i due processi, in ultima analisi, coincidono»<sup>29</sup>.

Ne discendono, secondo Benjamin, tre conseguenze strettamente connesse: la critica è un «fatto», non è possibile istruire un'«indagine a parte» per stabilire se qualcosa abbia valore oppure no, perché anzi «la pura e semplice criticabilità di un'opera si configura come giudizio di valore positivo sopra di essa»; «se un'opera è criticabile, essa è un'opera d'arte»; il brutto non può essere criticato<sup>30</sup>.

L'idea romantica della critica è dunque pervasa da un «“rivoluzionario furore di oggettività”»<sup>31</sup>: «La critica, che nella concezione odierna è il massimo della soggettività, fu per i romantici l'elemento regolatore di ogni soggettività, casualità e arbitrarietà nella nascita dell'opera. Mentre, secondo le idee attuali, essa si compone della conoscenza oggettiva e della valutazione (*Wertung*) dell'opera, l'elemento distintivo del concetto romantico di critica sta nel non conoscere una speciale valutazione soggettiva (*subjektive Einschätzung*) dell'opera nel giudizio di gusto. La valutazione (*Wertung*) è immanente alla ricerca oggettiva e alla conoscenza dell'opera. Non è il critico che dà su questa il giudizio: è l'arte stessa, nel momento in cui o accetta in sé l'opera nel medium della critica, oppure la respinge e, proprio con ciò, la valuta al di sotto di ogni critica (*unter aller Kritik schätzt*). La critica dovrebbe porre in atto, con ciò che essa tratta, la selezione fra le opere»<sup>32</sup>.

### 3. Valutazione di processo e di grandezza

I concetti foucaultiano di trans-soggettivazione e benjaminiano di critica immanente dell'opera di pensiero ci aiutano a vedere quello che non va nell'attuale sistema di valutazione.

Ci sono – mi pare – due forme di valutazione: una è la valutazione di *processo*, l'altra la valutazione di *grandezza*. Quest'ultima è – certamente – una valutazione di *prodotto*; non però una valutazione di conformità del prodotto a standard prefissati (come quando, per esempio, dobbiamo valutare se ci sia del polistirolo nell'argine del Carrione a Carrara), ma una valutazione – appunto – della sua *forza e novità*, che è anche la capacità di determinare nuovi standard<sup>33</sup>.

<sup>28</sup> GA 3, p. 84 [= GS I, pp. 77-78]; it., p. 413.

<sup>29</sup> GA 3, p. 84 [= GS I, p. 78]; it., p. 413.

<sup>30</sup> GA 3, p. 85 [= GS I, pp. 78-79]; it., pp. 413-414.

<sup>31</sup> GA 3, p. 88 [= GS I, p. 81]; it., p. 415.

<sup>32</sup> GA 3, p. 87 [= GS I, p. 80]; it., p. 415.

<sup>33</sup> Una buona traduzione analitica del termine “grandezza” (che però non ne esaurisce la pregnanza) può spiegare altresì che cosa intendiamo per “leggi di natura”: «Ciò che noi valutiamo (*What we value*) in un sistema deduttivo è una combinazione appropriatamente bilanciata di semplicità e forza (*a properly balanced combination of simplicity and strength*) [...]. Nella scienza vi sono standard – vaghi, a dire il vero – per valutare (*for assessing*) la combinazione di forza e semplicità offerta da sistemi deduttivi» (D. LEWIS, *Counterfactuals*, Basil Blackwell, Oxford 1973, pp. 73-74).

Supponiamo di dovere valutare un atleta, per esempio Usain Bolt. La valutazione di grandezza si affida a un gesto sintetico<sup>34</sup>: avere corso i 100 mt piani in 9"58 il 16 agosto 2009 a Berlino. La valutazione di processo indica invece i mezzi, o i passaggi, che si sono resi necessari per raggiungere il risultato (o, eventualmente, che sarebbero necessari per migliorarlo): avere utilizzato, e in quale misura, una combinazione di allenamento pliometrico, allenamento con i pesi, allenamento sulla flessibilità ed esercitazioni sugli sprint per circa 3 ore al giorno; consumare 6 pasti al giorno contenenti il 60% di proteine, il 30% di carboidrati, il 10% di grassi; partecipare a un certo numero di gare internazionali, eccetera.

Supponiamo ora di dovere valutare un santo cattolico. La valutazione di processo indicherà i mezzi, o i passaggi, che si sono resi necessari per raggiungere il risultato: essersi confessato almeno una volta alla settimana, avere partecipato alla funzione eucaristica tutti i giorni, avere regolarmente recitato la liturgia delle ore, aver fatto l'esame di coscienza quotidiano, eccetera. La valutazione sintetica o di grandezza, di nuovo, farà riferimento a un gesto perfetto: per esempio, essersi avvicinato da solo al lupo di Gubbio, avere fatto su di lui il segno della croce chiamandolo "fratello lupo", e avergli comandato in nome di Cristo di non fare più male a nessuno.

Supponiamo infine di dover valutare un filosofo. La valutazione sintetica farà riferimento a un'opera: per esempio, avere scritto *Essere e tempo*; quella di processo insisterà sugli *step* necessari per raggiungere il risultato: avere studiato filosofia, avere conseguito la libera docenza, essere stato assistente di un professore di filosofia, avere studiato per decenni per un certo numero consistente di ore al giorno, e via discorrendo (oggi aggiungeremmo: avere pubblicato l'opera in questione su una rivista indicizzata ISI/Scopus, avere fatto esperienza di ricerca all'estero, e simili).

Il sistema universitario di valutazione si basa sulla valutazione analitica, o di processo. Questo è del tutto legittimo. Certamente Usain Bolt non avrebbe corso i 100 piani in 9"58, se non si fosse allenato per tre ore al giorno, non si fosse esercitato in un certo modo, non avesse seguito una certa dieta; né San Francesco avrebbe potuto ammansire il lupo di Gubbio, se non fosse innanzitutto stato un santo nel senso lato della festa di Ognissanti: se cioè non avesse creduto in Dio, non avesse studiato il catechismo, non si fosse confessato, non avesse pregato; né Heidegger avrebbe scritto *Essere e Tempo*, se non si fosse laureato in filosofia, non avesse frequentato altri filosofi, non avesse mai studiato o avesse studiato poco.

Il problema è che non basta. A parità di allenamento, qualcuno è Usain Bolt, qualcun altro un ottimo sportivo che corre, supponiamo, i 100 piani in 11"58; a parità di preghiere, qualcuno è San Francesco, qualcun altro un buon cristiano (oppure un sepolcro imbiancato); a parità di studio, qualcuno è Martin Heidegger, qualcun altro un onesto professionista della filosofia. Non c'è niente di male. Anzi, è da una comunità di cristiani che nasce Francesco; da una comunità di filosofi che vien fuori Heidegger; da una comunità di atleti che emerge Bolt.

---

<sup>34</sup> Sul gesto sintetico, o "completo", cfr. G. MADDALENA, *Gesto completo: uno strumento pragmatista per l'educazione*, in "Spazio Filosofico", 10 (1/2014 [numero monografico sul tema *Educazione*, a cura di E. Guglielminetti e L. Regina]), pp. 31-41, <http://www.spaziofilosofico.it/numero-10/4487/gesto-completo-uno-strumento-pragmatista-per-leducazione/#more-4487>.

La valutazione analitica di processo indica la presenza dei prerequisiti individuali e istituzionali perché la grandezza possa essere raggiunta, ma non dice se sia stata raggiunta. Credo sia questa, in fondo, la ragione della resistenza della comunità universitaria alla valutazione.

Per un verso, tale resistenza è miope, perché sembra non ammettere che la forma è sempre anche formula: il gesto perfetto, unico e irripetibile, è sempre anche analizzabile in una serie di micro-gesti fungibili, che la valutazione di processo può controllare. Così, un amore esemplare non sarà mai la sommatoria di averle regato dei fiori, avere contribuito al mantenimento della famiglia, averla portata in vacanza, avere sopportato la suocera, essere stato gentile, ma è altresì vero che se non le hai mai regalato dei fiori, non l'hai mai portata in vacanza, non hai mai dato un soldo in famiglia, l'hai sempre presa a male parole, è un po' difficile che il tuo sia stato un grande amore.

Per l'altro verso però, la resistenza si basa sul fatto che la forma non è mai – appunto – riducibile alla semplice formula. La forma *avere corso i 100 mt piani in 9"58* non è la conseguenza di un tipo di dieta. La dieta – o l'allenamento – può essere per tutti (è una formula), la forma è di uno solo, o – al limite (come nel caso dei record sportivi) – di pochissimi.

In fondo, la collettività universitaria assomiglia a una comunità di samurai. Il valore principale è il prestigio (o l'onore). C'è una differenza abissale tra correre in 9"58 e correre in 10 netti. C'è una differenza abissale tra scrivere *Essere e Tempo* e scrivere – poniamo – un'eccellente introduzione a Heidegger. Una valutazione che non faccia *questa* differenza, è giustamente avvertita come disonorevole.

Si potrebbe obiettare che la valutazione della qualità della ricerca è una valutazione di prodotto, non di processo. Non abbiamo forse un catalogo nazionale dei *prodotti* della ricerca? Il fatto è però che questa valutazione di prodotto viene operazionalizzata appunto come una valutazione di processo: il prodotto è un processo. Avere pubblicato un certo numero di libri, saggi in volume, articoli in rivista indica il grado di *attività* o di *inattività* di un ricercatore, e l'essere-attivo è appunto un processo. Se dico che Bolt ha partecipato come finalista a un certo numero di meeting di atletica nell'anno in corso, registro la sua attività, non la sua bravura.

A ciò i sostenitori del sistema attuale di valutazione possono obiettare che la valutazione non tiene conto solo del *grado* di attività, ma anche del *luogo*; non distingue solo tra chi corre e chi no, ma fa differenza anche tra chi corre *qui* (per esempio, alle olimpiadi) e chi corre *là* (per esempio, alla *Turin Marathon*).

Se uno è finalista in un meeting internazionale, non sarà forse bravo? Ne convengo. E, tuttavia, si può sempre chiedere: bravo *quanto*? Supponiamo che, da domani, l'intera umanità inizi a soffrire di un fastidioso male all'anca sinistra. Stante la capacità di soffrire degli atleti, i meeting si tengono lo stesso. Tuttavia, il dolore fisico, che determina un vistoso zoppicamento, impedisce al più bravo dei bravi di correre i 100 mt piani, supponiamo, in meno di 100 secondi. In questo caso, una valutazione di grandezza stabilirebbe che Bolt (cioè Bolt con il male all'anca) corre i 100 mt alla stessa velocità di un anziano sano *prima* della epidemia di male all'anca. Viceversa, la valutazione di prodotto, operazionalizzata come valutazione di processo, stabilirebbe *tout court* l'eccellenza di Bolt, che entrerebbe dunque a far parte con tutti gli onori dei programmi di finanziamento connessi all'*Exzellenzinitiative*. È come dire che, supponendo che non

disponessimo più del cemento armato, l'argine al polistirolo di Carrara sarebbe eccellente. Forse sarebbe il migliore degli argini esistenti, ma continuerebbe a essere leggero, e non meritevole di premi e riconoscimenti.

Per converso, se anche Bolt si rifiutasse, per motivi politici o religiosi, di correre a meeting internazionali, se non vincesses mai nessuna gara e si presentasse alle olimpiadi regolarmente ubriaco, ma continuasse a correre i 100 piani in 9"58 in allenamento (e se il controllo di questa prestazione fosse affidabile), resterebbe un grandissimo atleta, il più grande velocista nella storia dell'umanità fino a oggi. Quello che conta, in ultima analisi, è solo il tempo. Se questo tempo venga ottenuto nella finale olimpica, o nel campetto dietro casa il giorno dell'onomastico di Bolt, conta come avere pubblicato *Essere e tempo* in una rivista indicizzata Scopus (una banca dati privata) o nel giornale della parrocchia di Meßkirch (un'istituzione pubblica): cioè niente.

La valutazione di processo (o di prodotto & processo) assomiglia alla cura di sé foucaultiana. È un programma di allenamento, come tale indispensabile, che però esclude quella forma di rottura, che caratterizza invece, secondo Foucault, la *metanoia* cristiana. Ciò che le manca, è la trans-soggettivazione, cioè ogni forma di sensibilità al salto, alla scarto qualitativo, alla trascendenza. Anche quando questo scarto viene in qualche modo considerato – come nell'aporetica classificazione delle riviste di fascia A, B o C –, è di nuovo nel senso della formula, dell'analisi senza sintesi, che esso viene interpretato. Se due articoli sono pubblicati in riviste di fascia A, è possibile che vi sia tra essi uno scarto qualitativo essenziale (come, ovviamente, è possibile che articoli in fascia C siano migliori di articoli in fascia A), scarto che la valutazione non vede. Sostenere che un articolo in fascia A valga di più di un articolo in fascia C, è come dire che vincere al meeting di Berlino è più importante che vincere al meeting di Cesena. E se a Cesena venisse stabilito il record del mondo<sup>35</sup>?

Si pongono qui alcune questioni distinte. La prima: circa la differente grandezza di contributi ospitati sulla stessa rivista. Consideriamo una rivista esemplare, lo "Jahrbuch für Philosophie und phänomenologische Forschung di Husserl": non so se si possa dire, per esempio, che i due contributi del 1927 – M. HEIDEGGER, *Sein und Zeit* I e O. BECKER, *Mathematische Existenz* – abbiano la stessa potenza; o se la *Philosophie des Grafen Paul Yorck von Wartenburg* pubblicata da Fritz Kaufmann nel volume del 1928 valga quanto le *Vorlesungen zur Phänomenologie des inneren Zeitbewusstseins* pubblicate da Husserl nel medesimo vol. IX. Non sarà che qualcuno corre i 100 mt piani in 9"58 e qualcun altro in 10"58? La seconda questione è: quante, e quali, riviste hanno la stessa potenza/magnitudine delle 11 annate dello "Jahrbuch"? A giudicare dal numero di riviste in fascia A, non c'è mai stata un'epoca così grande in filosofia come l'attuale. Sarebbe interessante selezionare 10 riviste internazionali in fascia A e mettere a confronto, annata per annata, i singoli contributi dello "Jahrbuch" con i contributi di ciascuna rivista, assegnando a ciascuno un gradiente di magnitudine. Forse scopriremmo

---

<sup>35</sup> A chi rispondesse che, in questo caso, il meeting di Cesena potrebbe sempre essere riposizionato in fascia A per l'anno successivo, obietto che – l'anno successivo – il nuovo record del mondo potrebbe essere stabilito a Senigallia, e così via. È difficile *programmare* la novità.

che lo “Jahrbuch” merita la tripla A, e qualche altra rivista la tripla C, sebbene il sistema di valutazione le consideri equivalenti<sup>36</sup>.

Il concetto di critica del romanticismo tedesco va nella stessa direzione della trans-soggettivazione foucaultiana. L'idea romantica di valutazione come autovalutazione ha di mira l'assoluto. L'opera è criticabile, dunque superabile, per definizione, ma – paradossalmente – solo l'insuperabile (l'opera autentica, che è un possesso per sempre) può essere oggetto di un tale superamento. Viceversa, se si esclude l'assoluto (la prestazione assoluta, come avere scritto la *Divina Commedia*), non si esercitano le nuove leve all'amore per la grandezza; ma se non si esercitano le nuove leve all'amore per la grandezza, non c'è insegnamento.

Entrambe le forme di valutazione sono indispensabili<sup>37</sup>. Ma senza trans-soggettivazione, cesura, aspirazione all'assoluto, la valutazione è falsata. Solo la grandezza costituisce un'autentica novità, e un sistema, che si vuole innovativo, difficilmente può ignorare la novità.

In realtà, non è proprio così. L'innovazione non è la novità, ma sta alla novità come l'auto-soggettivazione sta alla trans-soggettivazione foucaultiana. In-novare è dirigersi verso il nuovo, con un programma di allenamento. La novità, invece, è un evento unico, che, programma o no, preparazione o no, determina una rottura. Ne discende la conseguenza che può esserci tensione, e perfino contraddizione, tra innovazione e novità. Noi, oggi, cerchiamo l'innovazione, non la novità, anche perché questa non tanto si cerca, quanto piuttosto si trova.

Qui è dato vedere in opera il conflitto tra metafisiche concorrenti. Per i teorici dell'auto-soggettivazione, quanto più di cura di sé, tanto più di miglioramento individuale e collettivo, tanto maggiore progresso. È uno schema ben noto, di matrice illuministica, che ha dalla sua buoni elementi di plausibilità, specie in ambito tecnico. E tuttavia: la differenza tra la *metanoia* e la *conversio ad se* si basa sul fatto che lo scarto è appunto uno scarto, che *non* si ottiene per accumulo. Per i teorici della trans-soggettivazione la novità è dunque innanzitutto una sorpresa, non l'esito di una pianificazione. Pianificare la sorpresa potrebbe essere anzi il modo più sicuro per vanificarla. Per stare ancora alla categorizzazione foucaultiana: non è la nostra volontà di potenza di diventare santi, che ci fa diventare santi. Il primo passo sulla via della santità

---

<sup>36</sup> Ma, terza considerazione, la valutazione dovrebbe essere annuale o biennale, proprio come si fa con la crescita, il PIL o il debito sovrano. Se lo “Jahrbuch” avesse pubblicato le *Ideen I* nell'annata del 1913 e non anche, per esempio, *Der Formalismus in der Ethik und die materiale Wertethik II* di Max Scheler nel II volume, del 1916, la sua valutazione avrebbe dovuto essere diminuita. E questo a prescindere dal fatto che, essendo passati 3 anni dal primo al secondo volume, lo “Jahrbuch” sarebbe probabilmente considerato oggi una rivista poco affidabile, perché non abbastanza regolare nelle sue uscite. Va poi considerato che lo “Jahrbuch” era un prodotto di mera rilevanza nazionale (c'erano solo tedeschi), che non si era dotato nemmeno di un comitato scientifico internazionale e che, pubblicando solo in tedesco, risulterebbe oggi illeggibile da, e di conseguenza irrilevante per, una percentuale esorbitante di addetti ai lavori. Uno degli effetti perversi della valutazione/internazionalizzazione è, infatti, che il numero di giovani ricercatori di filosofia che fanno il tedesco sta fortemente diminuendo. E se sapere il tedesco fosse (pressoché) indispensabile per fare filosofia?

<sup>37</sup> Non intendo qui associarmi alle geremiadi anti-valutazione, ma fare emergere alcune aporie: più o meno le stesse di cui ci lamentiamo, non sempre a torto, quando scopriamo che Standard & Poor's assegna il rating BBB- all'Italia e BBB+ alla Colombia: forse c'è qualcosa che non va

potrebbe invece essere accorgersi che essa, come tutto ciò che vale veramente, accade per grazia, che cioè *non* dipende da me o da quanto mi alleno. Ché anzi l'auto-soggettivazione manca strutturalmente la santità, e può – al limite – proporsi la saggezza. Allo stesso modo, la valutazione della qualità della ricerca, così come viene attualmente implementata, manca strutturalmente la grandezza, e può – al limite – proporsi l'efficacia e l'efficienza<sup>38</sup>.

#### 4. *Parti contendibili*

È noto che la *Habilitationschrift* di Benjamin venne bocciata con l'argomento che «*Geist kann man nicht habilitieren* (non si può dare la libera docenza allo spirito)». Molti filosofi analitici sarebbero oggi sicuramente ben disposti a ripetere con protervia questo giudizio. Al di là di questa valutazione, la frase (attribuita a Erich Rothacker) rischia di diventare il simbolo di un'università governata in pompa magna, oggi come allora, dai vari Franz Schultz e Hans Cornelius (i professori di Benjamin a Francoforte), dove Benjamin non trova posto. La standardizzazione del sapere, mossa forse da uno spirito di risentimento, sembra temere l'originalità. Si produce così uno scollamento tra la società, nel cui corpo vengono iniettate dosi sempre più massicce di stranezza, e le istituzioni del sapere, che si votano al conformismo.

Ogni sistema di valutazione produce *insider* e *outsider*, e una serie di casi dubbi. Ma il criterio, in base al quale sei dentro o sei fuori (o sei sulla linea), può essere più o meno virtuoso, ed è appunto su questo che la comunità degli *stakeholders* vorrebbe dire la sua.

Qui può venire di qualche utilità il concetto – formulato originariamente da van Fraassen e poi sviluppato da Lewis – di “supervalutazione”, che, pur non essendo direttamente collegato con il tema del nostro numero, può interagire con esso circa la questione delle decisioni non prese.

Vi sono casi – osserva Lewis – in cui l'indecisione semantica è inevitabile, e tuttavia vogliamo poter continuare a parlare. Per fortuna, spesso le decisioni semantiche che *non* abbiamo prese non hanno importanza. Così, se dico che un famoso architetto ha progettato la casa di Fred, non mi è in realtà mai venuto in mente di disambiguare il mio pensiero, chiarendo se per “casa” intendessi qualcosa che includesse o meno l'annesso garage. Ma non importa: la mia affermazione sarà vera “in entrambi i casi”. «Spesso – osserva Lewis – ciò che uno vuol dire sarà vero in tutti i modi differenti di prendere la decisione non presa». Si tratta appunto del metodo delle «*superevaluations*. Una

---

<sup>38</sup> È forse di un qualche interesse notare qui che, delle tre cantiche dantesche, solo il *Purgatorio* assomiglia a una palestra in cui, come un cane al guinzaglio, si presta ogni attenzione allo svolgimento del programma penitenziale individuale, fino al punto di non avere quasi più tempo per altro. La valutazione, che cerca l'innovazione, ci fa passare – giustamente – per il purgatorio della pianificazione, ma non dovrebbe mai dimenticare che la meta è ben altra: qualcosa di assoluto, il paradiso dell'arte o della filosofia, non qualcosa di relativo, e destinato a finire non appena si passi all'assoluto.

proposizione è *super-vera* se e solo se è vera in tutti i modi di prendere le decisioni semantiche non prese»<sup>39</sup>.

Il concetto di supervalutazione è connesso a quello di vaghezza: quella gocciolina d'acqua ai margini della nuvola, fa ancora parte della nuvola, o è solo vicino a essa? Tutte le cose, per Lewis, sono in qualche modo vaghe, perché hanno “*questionable parts*”, parti di cui si può dubitare se vi appartengano o meno<sup>40</sup>.

Analogamente, ogni processo valutativo presenta una serie di decisioni *non* prese. Per esempio, la valutazione romantica decide circa la grandezza, ma non decide circa i livelli di grandezza. Si valutano solo i capolavori, il che non significa che tutti i capolavori (tutte le opere valutabili) siano su un piano di parità. Possono sempre insorgere questioni su casi-limite.

Nella sua configurazione attuale, la valutazione della qualità della ricerca lascia invece indeciso se i membri di una classe condividano il carattere (o l'etichetta) attribuito alla classe. Se dico che i filosofi, le cui monografie sono pubblicate da Oxford University Press, sono eccellenti, non sto veramente *decidendo* che la Prof.ssa X sia una studiosa eccellente, né – tanto meno – sto dicendo che, per “eccellente”, intendo davvero “eccellente”, per esempio eccellente come Wittgenstein o come Frege. Non prendo nessuna decisione su nessun individuo (su nessuna opera), ma decido su una classe di individui a esclusione di altre. Se dovessi controllare la mia decisione caso per caso sulla lista dei “passeggeri”, potrei avere brutte sorprese, ma il sistema non lo richiede, perché è formulato in quello che, con Lewis, potremmo definire un «linguaggio interpretato in un modo imperfettamente decisivo (*an imperfectly decisive way*)»<sup>41</sup>.

Nel caso della valutazione romantica, la vaghezza riguarda le goccioline d'acqua ai margini della nuvola del capolavoro: la *Commedia* lo è certamente, ma *Il Fiore* (che alcuni attribuiscono a Dante)? *Essere e Tempo* lo è di sicuro, ma *La dottrina delle categorie e del significato in Duns Scoto* (la dissertazione per l'abilitazione di Heidegger)? L'aspirazione dell'evaluando neoromantico è dunque: fare parte della nuvola, scrivere un capolavoro, stare abbastanza dentro la nuvola da non costituirne una “*questionable part*”.

Nel caso della valutazione come oggi è concepita, la vaghezza riguarda invece, da un lato, la trasmissibilità o la portabilità di un'etichetta da una classe ai suoi membri, dall'altro, e più decisamente, il significato stesso dell'etichetta: che cosa vuol dire “eccellente”? Per un verso, quindi, la vaghezza diminuisce: non esistono casi dubbi, o marginali, circa il fatto se un ricercatore abbia pubblicato o no presso Oxford University Press: basta guardare il catalogo. Per l'altro, però, aumenta: è l'etichetta stessa, che potrebbe rivelarsi, a una più attenta considerazione, come una gocciolina ai margini della nuvola. Essere nel gruppo degli “eccellenti”, non è forse una “*questionable part*” dell'essere nel gruppo dei grandi? L'eccellenza è la nuvola, o non piuttosto la gocciolina? Stiamo guardando il dito o il bersaglio? Non c'è nessuna contraddizione a pensare un'università di eccellenti, che non produca alcun capolavoro: molte monografie (un po' involute) su

<sup>39</sup> D. LEWIS, *Many, but Almost One*, in J. BACON-K. CAMPBELL-L. REINHARDT (a cura di), *Ontology, Causality and Mind. Essays in Honour of D.M. Armstrong*, Cambridge University Press, Cambridge-New York 1993, pp. 28-29.

<sup>40</sup> *Ibidem*, p. 23.

<sup>41</sup> *Ibidem*, p. 29.

*Duns Scoto*, nessun *Essere e Tempo*. La vera decisione *non* presa del nostro sistema di valutazione è quindi se l'eccellenza sia la grandezza.

Contestualmente, cambiano le aspirazioni: l'evaluando non cerca più la riuscita assoluta, ma quella relativa. L'aspirazione a scrivere un capolavoro viene via via sostituita dalla più realistica aspirazione a essere pubblicati presso Oxford University Press, o almeno da Einaudi. Un romantico la troverebbe una piccola aspirazione. Ma non è tempo di romanticismo.

Certamente, si dirà, una cosa è la valutazione sistemica, affidata all'Anvur, un'altra è la valutazione individuale, affidata alle commissioni giudicatrici. E tuttavia, è molto difficile pensare che i due processi possano essere scollegati. Se la valutazione sistemica ignora la grandezza, la valutazione individuale rischia di tenerle dietro. Opportunamente, nei nostri concorsi a cattedra, la valutazione delle pubblicazioni scientifiche mette al primo punto la voce "originalità, innovatività, rigore metodologico e rilevanza di ciascuna pubblicazione". Qui, almeno in parte, è lasciato effettivamente uno spazio aperto per una valutazione di merito. E tuttavia, sarebbe interessante vedere che cosa succederebbe, se sostituissimo la voce "originalità" con quella "originale-come". Se dico che le pubblicazioni di un candidato sono originali (e tutti i commissari lo dicono, perlomeno del vincitore di un concorso), voglio dire con ciò che esse sono originali-come *Essere e tempo*? Come credo si evinca da questo esempio, anche la valutazione individuale e di merito è fatta in modo da cautelarsi contro la novità, la grandezza e l'assoluto.

Hanno ragione i romantici: non ha senso, perlomeno ai fini della valutazione, introdurre livelli di grandezza tra opere grandi. Sarebbe una complicazione inutile stare a disquisire se originale-come Heidegger sia di più, o di meno, di originale-come Foucault, o Benjamin, o David Lewis, o Luigi Pareyson. È probabile che i filosofi abbiano idee piuttosto precise al riguardo, e che differenze ci siano. Così come è probabile che ci siano differenze tra Dante, Manzoni e Leopardi, ma è opportuno e necessario che un sistema di valutazione resti cieco a queste differenze. Quello che non è opportuno è che resti cieco alla grandezza, sebbene si tratti di un valore desueto, e che qualcuno potrebbe considerare di destra (in realtà, niente è più democratico della grandezza, che è un bene per tutti).

In altri termini: per quanto possa essere difficile stabilire degli indicatori al riguardo, bisognerebbe sempre anche valutare la potenza di un pensiero. Ciascun lettore si accorge della (differente) potenza di un romanzo, o di un libro di filosofia. Ma il sistema di valutazione universitario non dice nulla, in nessun momento, al riguardo, cioè al riguardo della cosa stessa. Forse perché non è pensato per i lettori: sono talmente numerose le pubblicazioni scientifiche, che ai valutatori non può essere chiesto davvero di leggerle.

## 5. *Il bambino e l'oculista*

Quando vanno dall'oculista, i bambini – non sapendo ancora leggere – provano la propria vista su una tavola ottometrica pensata apposta per loro, come questa:



*Originale-come (oppure: potente-come):*

**FASCIA A:**



Leibniz



Kant

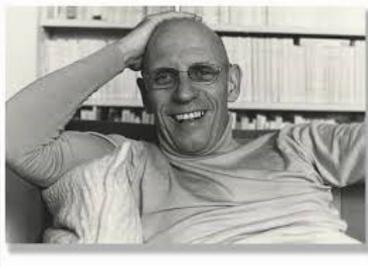


Hegel

**FASCIA B:**



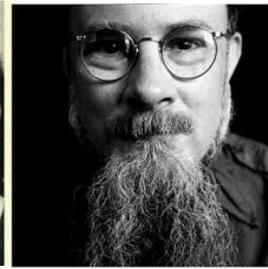
Benjamin



Foucault



Pareyson



Lewis

**FASCIA C:**



Guglielminetti

**FASCIA D:**



*Absit iniuria imaginibus...*

Non che io creda che il Miur possa proporci veramente una tavola del genere: per la carità. Vi sarebbe, oltretutto, una valanga di ricorsi. E tuttavia, a chi dubitasse che possano esistere indicatori sintetici di potenza, la nostra tavola dovrebbe fare sorgere qualche dubbio.

Le fasce A e B esprimono la grandezza (cosa che le riviste di fascia A e B non esprimono). A proposito di responsabilità civile degli impiegati pubblici, si potrebbe immaginare che, qualora uno o più commissari dichiarino che una candidata è *originale-come*, supponiamo, Leibniz o Lewis, il concorso (in ipotesi, da ricercatore universitario a tempo determinato) venga temporaneamente sospeso. Una commissione internazionale di esperti sarebbe chiamata a validare tale giudizio. Se il giudizio venisse confermato, la candidata dovrebbe essere chiamata direttamente dal Miur come professore ordinario. Anzi no, data l'eccezionalità della cosa, dovrebbe essere chiamata in una fascia apposita, sovraordinata rispetto a quella degli ordinari: una fascia di eccellenza, cui si accederebbe solo per chiara fama. Nel caso in cui il giudizio dovesse invece purtroppo non venire validato, i commissari che lo hanno espresso potrebbero, per esempio, saltare un giro, venendo esclusi per 5 o per 10 anni (a seconda dell'entità dell'errore) dalla partecipazione alle commissioni giudicatrici su tutto il territorio nazionale.

Se invece il candidato fosse ritenuto *originale-come* una delle figure di fascia C, la commissione giudicatrice avrebbe due strade: o semplicemente assegnare il posto a concorso al vincitore, oppure – qualora ritenesse tale posizione inadeguata alla bravura del concorrente – proporlo direttamente per un posto di professore associato o di professore ordinario. Di nuovo, il concorso verrebbe sospeso. Una commissione nazionale di esperti valuterrebbe la proposta. Se fosse respinta, il candidato sarebbe confermato nel ruolo di ricercatore a tempo determinato (o magari indeterminato), senza ovviamente alcuna sanzione nei confronti dei commissari, che hanno fatto solo il proprio dovere. Se invece la proposta fosse accettata, la commissione di esperti avrebbe facoltà di decidere se chiamare il vincitore su un posto di prima o di seconda fascia.

I candidati inseriti nella fascia D (che andrebbe modulata in modo più rispettoso di quanto non abbia fatto io nel contesto di questo piccolo carnevale docimologico), risulterebbero ovviamente non vincitori (o non idonei), e potrebbero sempre riprovarci la volta successiva.

È solo uno scherzo. Che – come tutte le cose ridicole – vorrebbe però dare a pensare.



Riccardo Manzotti  
Paolo Moderato

## NEUROSCIENZE E PSICOLOGIA: VALUTAZIONE E PARADIGMI

### Abstract

*The evaluation of the quality of scientific research is a thorny issue, particularly in those fields that require a revolutionary approach rather than an incremental progress. This is the case of the sciences of the mind that struggle to find a shared theoretical landscape. Currently, psychology dwells on the explanatory promissory notes made by neuroscience. Yet, it is unclear whether neuroscience will deliver what it is currently promising. In this paper, we outline the dangerous loop between society's reaction and research assessment regarding neuroscience's pretence of becoming the new science of the mind.*

«GALILEO: Una delle principali ragioni della povertà della scienza, è la pretesa di essere così ricca».

(BERTOLT BRECHT, 1939)

Come valutare la ricerca se non si conoscono ancora quelle che saranno le vie che porteranno a effettivi risultati? Come definire criteri di valutazione se la scienza stessa non sa quali saranno i modelli scientifici che avranno successo nel futuro più o meno immediato? Se si guarda alla storia della scienza, è evidente che non è una strada rettilinea, ma un percorso tortuoso e incerto. Brevi tratti possono dare l'impressione di una progressione costante o di criteri noti a priori, ma si tratta di un'illusione. A lungo termine la direzione da prendere è ignota.

Se ci si rifà alla contrapposizione tra scienza ordinaria e scienza straordinaria proposta da Thomas Kuhn, si hanno due momenti. Nel primo, si accettano i paradigmi e gli assunti di base, e si raccolgono dati sperimentali. In questa fase, il progresso è incrementale e si possono definire criteri di valutazione che ripropongono conservativamente i modelli di successo utilizzati in precedenza. Nel secondo momento, i paradigmi vengono posti in discussione e si propone un paradigma nuovo che è, in larga misura, incommensurabile rispetto agli assunti precedenti. La parola chiave, ai fini della valutazione della valutazione, è proprio questa: "incommensurabile", ovvero non misurabile sulla base dei criteri precedenti.

Gli esempi sono innumerevoli, da Boltzmann che si suicida perché la sua depressione è aggravata dall'incomprensione dei colleghi per la sua teoria statistica della termodinamica, fino a Wegener che morì cercando ulteriori prove per la deriva dei continenti nella generale incredulità dei geologi. Per non parlare del triste caso di Ignác Fülöp Semmelweis, che finì in manicomio, dove morì, per aver osato sfidare le opinioni del tempo in materia di igiene e antisepsi dei reparti di ostetricia, e, nonostante avesse provato con evidenze che aveva ragione, fu schernito e trattato da folle.

In tutti questi casi si è assistito a un fenomeno ricorrente, sia a livello sociologico sia a livello di ricerca scientifica: un paradigma viene scelto da un gruppo di ricercatori che diventa maggioritario e che investe su di esso credibilità, sforzi e risorse. Quando tale paradigma comincia a diventare – per vari motivi spesso non di natura strettamente scientifica – auto-rinforzante, si determina un processo virtuoso di rafforzamento di un'ortodossia dominante nel panorama scientifico.

In questo contesto, un temibile orizzonte epistemico di non ritorno è rappresentato dal successo mediatico di una linea di ricerca. Il rischio è che si determini un *feedback* tra ricerca e pubblico che diventa rapidamente autoreferenziale – una certa linea di ricerca è accolta con interesse dal grande pubblico perché conferma qualche pregiudizio radicato nel fantomatico “uomo della strada”. La comunità accademica riceve un ritorno positivo dal fatto di seguire tali studi e quindi si orienta in tale direzione. Ulteriori studi suscitano ulteriore entusiasmo nel pubblico e così via, in una spirale che diventa sempre più prepotente nel dirigere i processi di ricerca e nel fornire i principi di valutazione. Come si legge nei *Promessi Sposi* a proposito del dotto aristotelico Don Ferrante: «Fin che non faceva che dare addosso all'opinione del contagio, trovava per tutto orecchi attenti e ben disposti: perché non si può spiegare quanto sia grande l'autorità d'un dotto di professione, allorché vuol dimostrare agli altri le cose di cui sono già persuasi». Tutto ciò senza considerare eventuali meccanismi opportunistici che particolari lobby o comunità di ricercatori possono scientemente mettere in atto.

Questo meccanismo può instaurarsi facilmente in quei settori in cui ci si trova di fronte a un problema straordinario in senso kuhniano, ovvero un problema i cui confini e la cui soluzione non sono definibili sulla base delle ricerche passate. La mente è un caso da manuale: per il suo studio non esistono facili soluzioni né chiare linee di valutazione, e quindi i criteri adottati finiscono con l'essere la conferma dell'ortodossia dominante piuttosto che effettivi giudizi oggettivi. Lo studio della mente è un cimitero di posizioni dominanti che, con il tempo, si sono rivelate completamente ingiustificate. Un esempio classico è rappresentato dal dibattito sulla localizzazione delle funzioni e dei contenuti mentali che, periodicamente, oscilla tra globalizzazione e localizzazione puntuale (Zeki 2001; Uttal 2001). Il caso della mente è particolarmente sensibile all'opinione pubblica perché coinvolge la concezione popolare del sé e quindi la natura dell'uomo. Al contrario, altri settori di ricerca scientifica – quali la parallelizzazione dei processi computazionali – sono molto meno suscettibili di essere influenzati dalla *vox populi*. Nel caso della mente, al contrario, ogni affermazione degli scienziati si traduce invariabilmente in una conferma o in una negazione dei pre-giudizi che ognuno ha circa la propria esistenza.

È quindi facile lasciare che i criteri di valutazione diventino prigionieri delle metafore e degli slogan che, al pari di una campagna di propaganda politica, hanno queste

caratteristiche fondamentali: confermano i pregiudizi diffusi, non richiedono particolari sforzi concettuali, vestono con parole nuove idee vecchie, promettono di risolvere tutto in un futuro più o meno vicino. Un grande psicologo come Gaetano Kanizsa ammoniva di sorvegliare con grande attenzione l'uso delle metafore. Tutte queste scorciatoie, note come euristiche, e i loro caratteristici effetti di condizionamento sui processi decisionali sono stati studiati e ampiamente analizzati da due psicologi, Amos Tversky e Daniel Kahneman fin dalla metà degli anni '70<sup>1</sup>.

Esiste poi un'ulteriore caratteristica di grande importanza perché la comunità scientifica abbracci incondizionatamente ed entusiasticamente un programma di ricerca: deve richiedere ingenti e continui investimenti finanziari. Nel 1998, al convegno *Toward a Science of Consciousness III* (svoltosi a Tucson, in Arizona), uno degli autori di questo articolo si trovò, insieme a molti altri, a porre domande sulle politiche della ricerca al premio Nobel Gerald Edelman. Alla domanda circa quali fattori garantissero la buona accoglienza di una linea di ricerca da parte della comunità scientifica, il famoso scienziato rispose, con un certo cinismo: «deve dare molto da fare agli scienziati, cioè deve creare dei posti di lavoro». Tutto questo marchingegno mediatico-scientifico-economico-valutativo funziona – è chiaro – per orizzonti temporali non illimitati perché anche la pazienza del pubblico non è infinita. Sono così possibili due esiti. Nel primo caso un certo paradigma di ricerca viene progressivamente screditato, spesso in cambio di un nuovo paradigma identico al primo ma mascherato da una nuova terminologia e da qualche innovazione tecnologica di sicuro impatto mediatico. Nel secondo caso, emerge una effettiva soluzione spesso *incommensurabile e del tutto imprevedibile e imprevedibile* da parte dei criteri di valutazione adottati che, come nel caso della relatività speciale di Einstein, consente di mettere da parte il quadro di riferimento concettuale precedente (Lavazza-Manzotti 2011). Questo secondo esito, per quanto riguarda la mente, non è ancora avvenuto. Stiamo ancora aspettando la rivelazione.

La ricerca scientifica è spesso più conservatrice di quanto non si pensi. In ogni epoca, molte teorie e ipotesi sono guidate da casi passati. Il caso della mente è esemplare – ogni epoca cerca di addomesticarne la natura cercando di ridurla a qualcosa di familiare. Nel '500, prevalse la metafora pneumatica – la mente era una specie di pneuma catturato nei ventricoli cerebrali. Nel '600 dominò il modello idraulico, in seguito quello meccanico. Nell'800, fu la volta dell'elettromagnetismo – ancora oggi si dice comunemente che alcune persone sono “esaurite”, quasi fossero batterie elettriche. Nel '900 prevalsero le teorie dell'informazione che identificavano la mente con il *software* che permette il funzionamento dell'*hardware* neurale. Ogni epoca propone una metafora di successo che gli scienziati utilizzano per divulgare le loro scoperte. Il rischio è diventare prigionieri di queste metafore, di reificarle, e di scambiarle per risultati consolidati.

Il fatto è che il mito della misurazione oggettiva è, appunto, un mito. Si vede quello che ci si aspetta di vedere, e, spesso, le valutazioni servono solo a confermare quello che le commissioni si aspettano di vedere. Come l'aiutante di Newton si acconciò a dichiarare di vedere sette colori primari per soddisfare le pretese del suo illustre mentore, come il servo del re dei Gepidi confermò la vittoria (fasulla) contro i Longobardi di

---

<sup>1</sup> Per queste ricerche Kahneman ha vinto il premio Nobel 2002 per l'economia. Tversky, purtroppo, era morto prematuramente.

Alboino, come lo stesso Cartesio dichiarò che una palla di fucile a grande distanza fa più danno che a corta distanza, così, spesso, la ricerca dei dati empirici cerca ciò che – si sa fin troppo bene – troverà una benevole accondiscendenza nei criteri dei valutatori.

In questo nostro intervento vorremmo mostrare come, nel caso della mente, ci troviamo in una situazione in cui il ricorso alle neuroscienze, come futura scienza della mente, ha molte delle caratteristiche di un paradigma sostenuto da un intreccio mediatico piuttosto che da effettivi risultati empirici e concettuali. Con questo non vogliamo affatto sminuire i brillanti risultati che le neuroscienze stanno ottenendo nella descrizione dell'attività del sistema nervoso e delle sue basi biologiche. Quello che ci appare preoccupante è che siano state avanzate ipotesi finora mai dimostrate circa i fondamenti neurali della mente e che l'adesione a tali ipotesi, che dovrebbero essere oggetto di verifica e non criteri di scelta, diventi il parametro di valutazione dei progetti di ricerca. Pensiamo se la stessa cosa fosse avvenuta alla fine dell'800: interi campi di ricerca si sarebbero fossilizzati cercando di avvalorare con nuovi dati punti di vista considerati oggi irrimediabilmente datati.

### 1. *Le neuroscienze come futura scienza della mente?*

Prendiamo in considerazione il caso particolare della ricerca sulla mente come intreccio tra le neuroscienze e la psicologia. È un dato di fatto che, anche per una certa latitanza ontologica della psicologia (Manzotti-Moderato 2011), le neuroscienze si siano imposte come scienza forte per giungere a una spiegazione della mente. L'approccio delle neuroscienze, inevitabilmente, propone il cervello come organo deputato a produrre la mente – è così, sempre di più, sempre più tempo di cervello. In realtà, l'interesse per quest'organo non passa mai di moda (anche se non sempre all'interesse ne corrisponde l'uso). Nelle ultime tre decadi, complice la possibilità di visualizzare l'attività cerebrale in soggetti umani coscienti, il cervello è diventato oggetto di un irresistibile e crescente interesse. L'idea di poter trovare l'anima dentro il corpo, dentro la testa, piace un po' a tutti e, tutto sommato, è facile da capire. È come la vecchia idea dell'anima, solo che le neuroscienze ci forniscono un alibi per crederci in modo nuovo. In fondo, come l'anima era per lo più invisibile ma diventava visibile in casi particolari (i fantasmi), così la mente è invisibile ma grazie ai potenti mezzi della tecnologia (*brain imaging*) ci viene promesso che, un giorno non lontano, potremo vederla.

Così come dentro il DNA si è trovato l'alfabeto della vita, così nei neuroni si cerca la chiave per la mente. Il cervello è diventato così il principe delle iperboli. È quotidianamente citato come il sistema più complesso dell'universo (chissà cosa esiste in giro poi...), come un sistema che ha più neuroni delle stelle della galassia, come qualcosa che è "più grande del cielo" e così via, di metafora in metafora. Si potrebbe sospettare che una delle motivazioni alla base di questo entusiasmo sia il desiderio, molto umano, di considerarsi in qualche modo speciali. Sociologicamente, si può coltivare il ragionevole dubbio che il cervello stia al mondo naturale come la terra geo-centrica di Tolomeo stava al cosmo – è il perno intorno al quale si cerca di far girare il rapporto tra natura e conoscenza, tra universo e uomo, tra mondo e materia. Tale perno potrebbe rivelarsi molto più debole del previsto.

Complice una certa divulgazione, il cervello è diventato il nostro alter-ego. Spesso, per la divulgazione scientifica, noi siamo i nostri cervelli. L'anima è stata sostituita dal cervello. Già nel 1983, l'attore Steve Martin si era innamorato di un cervello (si presume femminile...). Due anni fa, sul "Sole 24 Ore" del 20 Maggio 2012, Motterlini e Monti dichiaravano senza imbarazzo che è meglio interrogare i cervelli piuttosto che le persone, che i cervelli predicono quale spot pubblicitario sia migliore, e che il nostro cervello sa delle cose di noi stessi che noi stessi non conosciamo (Motterlini-Monti 2012). Si tratta di affermazioni ancora più stupefacenti proprio perché gran parte dei lettori sembrano accettarle quasi come fossero ovvietà. Sono l'espressione di una tendenza sempre più forte: identificare la nostra mente – e tutto quello che ne discende – con quello che fanno i nostri neuroni. Si tratta di un'ipotesi che, nel 1994, il premio Nobel Francis Crick aveva definito sorprendente (*astonishing*) e che il pubblico trova sempre più convincente. Si assiste così al tentativo, un poco da prestidigitatori, di trasformare le neuroscienze in una vera e propria scienza della mente (Manzotti-Moderato 2011). Le neuroscienze, come il mago di Oz, promettono di portarci dal Kansas dei neuroni al paese della mente.

Questa marcia trionfale, accompagnata dagli squilli di tromba della cassa mediatica, si traduce in una serie di presunte discipline scientifiche che vorrebbero tradurre in termini neurali quasi tutti gli aspetti della mente e del comportamento umani: neuro-marketing, neuro-etica, neuro-teologia, neuro-filosofia, neuro-estetica, neuro-linguistica, neuro-psicologia, e così via. Come sostengono Carlo A. Umiltà e Paolo Legrenzi nel loro libro eponimo, ci troviamo di fronte a una vera e propria neuro-mania (Legrenzi e Umiltà 2009), che tuttavia non contribuisce, per ora, alla spiegazione del "chi siamo".

L'influsso delle neuroscienze non si limita al laboratorio, ma si estende fino a toccare il cuore dell'esistenza umana. Proprio per questo, bisognerebbe sorvegliare molto attentamente le ipotesi non dimostrate che ne determinano – come per ogni ricerca scientifica – metodi e obiettivi. L'ambizioso tentativo di delineare i fondamenti fisici della mente umana, riducendoli ad attività neurale, potrebbe essere minato alla base da numerosi e gravi motivi di scetticismo. Forse, prima di unirsi alla marcia trionfale e salire sul carro dei vincitori, varrebbe la pena di considerare criticamente queste voci contrarie. Stupisce che, al momento, sia così raro sentire voci dissonanti rispetto a quella che è stata chiamata la nuova ortodossia neurale. È innegabile che, in questo momento, la sensibilità collettiva e mediatica determini un vantaggio in chi si associa alla fiducia incondizionata nei confronti delle promesse delle neuroscienze. Il pubblico è sempre molto ben disposto a spiegazioni basate sul funzionamento dei neuroni, anche (o forse soprattutto) quando non è ovvio il legame tra il dato microscopico e gli eventi quotidiani. Scriveva il filosofo e matematico Alfred N. Whitehead che «ogni epoca ha ipotesi nascoste che tutti accettano e nessuno critica apertamente: sono proprio tali ipotesi che definiscono i limiti della comprensione» (Whitehead 1920, p. 127).

Premettiamo che non ci riferiamo, in quanto segue, al successo che le neuroscienze, in tutte le loro articolate declinazioni, hanno riscosso a livello del loro obiettivo proprio di ricerca (ovvero l'attività neurale e il sistema nervoso) fin dai tempi di Golgi e Cajal. Questo è un successo scientificamente solido e senza precedenti. La nostra critica, se valida, si rivolge soltanto al tentativo di fare il grande passo e, per usare le parole di Daniel Dennett, trasformare l'acqua del cervello nel vino della mente. In particolare, vogliamo criticare la tendenza di considerare i metodi e gli obiettivi delle neuroscienze

quali principi per valutare la ricerca sulla mente in altri settori, *in primis* nella psicologia o nella filosofia della mente.

Sia nell'opinione pubblica che nella comunità scientifica ha preso consistenza la convinzione secondo cui presto le neuroscienze potranno dirci tutto sulla mente: che cosa pensiamo, perché pensiamo qualcosa, che cosa ci fa innamorare, perché odiamo qualcuno, perché siamo disonesti, persino perché compriamo i prodotti di una certa marca. Eppure, se consideriamo i quattro classici successivi passaggi della ricerca scientifica – descrizione, spiegazione, previsione, controllo – scopriamo che, nei confronti della mente, le neuroscienze non sono riuscite a guadagnare neppure il primo passo. Infatti, il vocabolario delle neuroscienze non ha gli strumenti per affrontare le proprietà del mentale: intenzionalità, qualità, unità, significato, prima persona, *l'hic et nunc*. Non potendo descrivere i termini della mente non può nemmeno avanzare delle spiegazioni che ne giustifichino la comparsa, e, parallelamente, non può prevederne caratteristiche e contenuti. Per quanto riguarda la mente, la credibilità delle neuroscienze è fondata su una promessa più che su risultati concreti – come un creditore che paga con delle cambiali, le neuroscienze oggi acquistano credito scientifico con delle cambiali epistemiche, con “spiegherò”.

Di fronte a questi – e ad altri – ostacoli pratici e concettuali, nella comunità scientifica, per la prima volta si vedono accenni di perplessità relativamente alle neuroscienze quali disciplina della mente. Nel 2010 lo psicologo Ranier Mausfeld sosteneva che «sebbene da oltre 200 anni molti reputino che la mente sia una funzione del cervello, si dimentica che, nonostante i considerevoli risultati sperimentali delle neuroscienze, la nostra comprensione teorica del rapporto tra mente e cervello è vicina allo zero; non abbiamo nessuna idea di come e perché l'attività neurale produca la mente» (Mausfeld 2012, p. 66). In modo analogo, contro la corrente di pensiero prevalente, negli ultimi anni diversi autori hanno iniziato a sollevare obiezioni contro l'idea che le neuroscienze siano la disciplina destinata a spiegare la mente umana (Bennett e Hacker 2003; Illes e Bird 2006; Manzotti e Tagliasco 2008; Legrenzi e Umiltà 2009; Illes e Moser *et al.* 2010; Manzotti e Moderato 2010; Malafouris 2012; O'Connor, Rees *et al.* 2012). Le critiche sono molte e articolate, ma si possono ricondurre a cinque aree prevalenti che qui vogliamo esporre e riunire: 1) i metodi delle neuroscienze non sono adatti a spiegare la mente, 2) i fondamenti ontologici delle neuroscienze non sono compatibili con la mente, 3) si fa confusione tra localizzazione e spiegazione, 4) esiste un complesso intreccio tra neuroscienze, società e media, 5) i fattori culturali e politici rendono autoreferenziali i criteri di valutazione.

## 2. Problemi tecnici

Per quanto riguarda i metodi, senza perderci in interminabili dettagli, possiamo limitarci a sottolineare che, al momento, le tecniche disponibili sono affette da problemi quantitativi e qualitativi. In senso quantitativo, si deve tenere presente che i segnali registrati hanno un dettaglio molto basso. Per esempio, l'EEG (elettroencefalogramma) registra soltanto l'effetto cumulato di miliardi di processi neurali. Anche nella versione più semplice è paragonabile (per difetto) a qualcuno che registrasse il rumore prodotto

dalla città di Manhattan sedendosi sulla riva opposta del fiume Hudson. Al contrario, le registrazioni condotte con microelettrodi (molto invasive e quindi possibili solo in particolari casi clinici o sugli animali) riescono a cogliere soltanto il segnale di un numero ridotto di neuroni e quindi è molto difficile porli in relazione con il cervello nel suo complesso.

Per quanto riguarda la tecnica che più ha riscosso l'interesse dei mezzi di comunicazione, ovvero la fMRI, non si deve dimenticare che, al di là della bellezza delle immagini, la sua risoluzione spaziale è ancora molto bassa. Ogni punto colorato in una fMRI corrisponde a milioni di neuroni. Inoltre la fMRI è molto lenta: ogni processo più breve di qualche secondo è integrato e sommato a tutti gli altri, e quindi l'immagine che si ottiene è una media nel tempo. Potremmo dire che se osservassimo un danzatore con questa tecnica alla fine sapremmo solo quali muscoli ha usato con maggior frequenza e intensità, ma non conosceremmo i movimenti o il momento in cui li ha compiuti. Oppure, estremizzando molto (ma lo fanno anche molti altri), come gli astrofisici ci ricordano, quando noi guardiamo la costellazione di Andromeda stiamo guardandola com'era circa 3,5 milioni di anni fa.

La fMRI non misura direttamente l'attività neurale, ma altri fenomeni biologici (come la quantità di ossigeno nei vasi sanguigni) che si suppone siano in relazione con essa. Tale quantità non ha un significato assoluto, ma viene confrontata con una presunta attività media stimata in base a varie altre ipotesi. La conseguenza di tutti questi passaggi è che ciò che viene proposto come risultato finale è completamente diverso da una foto ottenuta con una lastra a raggi X, ma è una ricostruzione grafica di una serie di elaborazioni statistico-matematiche. Non è una vera foto, come quella celebre dell'anello al dito della moglie di Röntgen, ma piuttosto un'elaborazione ottenuta dal calcolatore sulla base di correlazioni con altri fenomeni biologici.

Ovviamente i neuroscienziati conoscono benissimo i limiti dei loro strumenti e sanno che ciò che misurano non è la mente, ma qualcosa che ha una relazione misteriosa con essa. Il pubblico però non è così attento e spesso ritiene che quello che viene mostrato sotto forma di punti colorati dentro il nostro cranio sia proprio la traccia della nostra mente, l'ombra dell'anima. Da tempo, neuroscienziati come Nikos Logothetis mettono in guardia contro i limiti metodologici di queste tecniche. In un articolo su "Nature" del 2008 dal titolo significativo di *What We Can Do and What We Cannot Do with fMRI*, Logothetis sottolinea il fatto che, molto spesso, si ignorano domande fondamentali circa il significato dei dati raccolti con questi strumenti e si traggono conclusioni errate basate su premesse non supportate dai dati scientifici (Racine, Bar-Ilan *et al.* 2005; Logothetis 2008; Garnett, Whiteley *et al.* 2011). In poche parole, la messa a fuoco nell'osservare i processi neurali non è ancora sufficiente per comprendere i dettagli di quello che avviene, ma consente solo di localizzare in modo approssimativo il luogo di attività cerebrali che sono correlate con certi processi mentali.

Le neuroscienze, per loro natura, tendono a concentrarsi sulla parte (il neurone e le sue combinazioni) più che sull'intero (la persona umana nel suo complesso). Un neuroscienziato e un filosofo, Maxwell Bennett e Peter Hacker, hanno etichettato questa riduzione del tutto alla parte, della persona ai suoi neuroni, come la *fallacia mereologica* (Bennett e Hacker 2003): si scambia il fenomeno nel suo complesso (la persona umana) con una sua parte (il cervello) e si attribuiscono a questa parte le proprietà del fenomeno

più generale. E quindi si dice che il cervello “crede”, “pensa”, “vede”, “sente”, “interpreta”, “vuole”. Mentre, per quanto ne sappiamo, è l’essere umano nel suo complesso che crede, pensa, vede, sente, interpreta e vuole. La riduzione dell’essere umano alla parte “cervello” non è dimostrata ma postulata. Questa riduzione dell’essere umano ai suoi neuroni è stata proposta più volte. La convinzione secondo cui la scienza dei neuroni è destinata magicamente a divenire la nuova scienza della mente potrebbe nascondere un errore non dissimile da quello degli epigoni di Tolomeo: porre il cervello al centro della spiegazione della mente potrebbe essere una specie di epiciclo mentale (Manzotti e Moderato 2011). Forse una nuova scienza della mente non coinciderà necessariamente con l’evoluzione delle neuroscienze, così come la nuova astronomia copernicana non è stata un’evoluzione degli epicicli. La mente potrebbe richiedere un cambiamento di prospettiva radicale che potrebbe non essere tra le possibilità delle neuroscienze.

Banalizzando un po’, un diffuso stile di spiegazione del comportamento umano a partire dal dato neuroscientifico è il seguente. Sulla base di uno studio condotto in laboratorio si trova che, in corrispondenza di un certo tratto cognitivo (per esempio l’amore romantico) che si determina chiedendo ai soggetti se lo possiedono, una certa area cerebrale di qualche millimetro (spesso queste aree non sono affatto isolate ma connesse ad altre aree opportunamente trascurate) è più attiva del resto del cervello (per esempio, Bartels e Zeki 2000, p. 3829). Invariabilmente la conclusione è che finalmente si è spiegato “con tanto di immagini del cervello” in che cosa consiste l’amore romantico. È un modo di procedere bizzarro e criticabile. Ma, si sarebbe tentati di chiedersi, perché questo risultato dovrebbe dirci qualcosa sulla natura dell’amore romantico? Che cosa sappiamo in più? Ancora una volta distinguiamo il dato sperimentale, indubitabilmente interessante, dalla sua importanza per la comprensione dei processi mentali. Questi studi creano un’illusione che contrasta con la loro effettiva valenza esplicativa (Weisberg, Keil *et al.* 2008). Su questo punto Hardcastle e Stewart concludono che la maggior parte degli studi dei neuroscienziati, per quanto riguarda la mente, non dice nulla che non fosse già noto agli psicologi, che non sono proprio da considerare la punta di diamante della ricerca scientifica! Al massimo, le immagini del cervello ci hanno dato una coreografia per trattare di fenomeni mentali che erano già perfettamente noti in ambito psicologico (Hardcastle e Stewart 2009; Alvarez 2011).

Le neuroscienze sono inesorabilmente attratte dal principio di località, ovvero l’idea che i fenomeni siano spiegabili a partire da principi localizzati. La tendenza a cercare di localizzare parti della mente in corrispondenti parti del cranio raggiunse il suo apice nell’800 con la famosa frenologia di Gallo e Lombroso – una pseudoscienza in seguito screditata dopo aver discriminato migliaia di persone sulla base di evidenze sperimentali del tutto infondate. Eppure, secondo i criteri di valutazione dell’epoca, la frenologia poteva essere considerata un approccio empiricamente accettabile.

Come ripetutamente osservano un neuroscienziato come Carlo Umiltà e uno psicologo come Paolo Legrenzi, confondere la spiegazione di un fenomeno (il suo “perché” e il suo “che cosa”) con la sua localizzazione (il “dove”) ha una facile presa sul pubblico. Tutte le volte che si suggerisce di avere individuato il luogo di un certo fenomeno, ecco che il pubblico ha la sensazione di avere capito qualcosa. Dov’è la felicità? In una certa area cerebrale. Dov’è l’infedeltà? È in una proteina. Oltretutto, gli

studi recenti non dimostrano neanche dove vengono portate a termine determinate attività mentali. Al massimo, gli studi di *brain imaging* (dalla PET alla fMRI, dall'EEG alla MEG) mostrano grossolanamente quali aree sono coinvolte nell'esecuzione di certi compiti. Si capisce subito che "essere coinvolti" non è proprio la stessa cosa che essere "responsabili" di un certo fenomeno. L'interruttore della luce è sicuramente coinvolto nell'accensione delle lampadine, ma non è sufficiente a far circolare la corrente elettrica. Una spiegazione "localistica" della mente umana sarebbe evidentemente insufficiente. Eppure, nel caso della mente, ci comportiamo proprio in questo modo, continuando a scambiare il dove con il perché dei fenomeni.

### 3. Scontro di ontologie

Per quanto siano difficili da risolvere, i problemi tecnici potrebbero essere risolti nel futuro più o meno vicino. Lo strumento si può perfezionare ed evolvere, ma il problema risale al modello che ne guida l'uso. I dotti sono troppo spesso innamorati dei loro strumenti per interrogarsi sulla loro interpretazione. Quando i dotti di Padova guardarono dentro il cannocchiale di Galileo non videro altro che immagini colorate. Quando si misurò per la prima volta la radiazione cosmica di fondo, si pensò a un'interferenza. Il fatto è che la differenza tra i fondamenti delle neuroscienze e i fondamenti della mente rimane irrisolta. I mattoni con cui hanno a che fare le neuroscienze sono fatti di attività neurale distribuita nel tempo e nello spazio: un fenomeno fisico di grande complessità. I mattoni della mente sono qualcosa di completamente diverso: emozioni, sensazioni, idee, concetti. Con questo non si vuol affatto sostenere che la mente sia fuori dalla fisica, come sostengono i dualisti, ma sembra plausibile che la mente non possa essere identica all'attività chimica dentro i neuroni. Insomma, i nostri pensieri, che pure sono probabilmente fisici, producendo effetti (a volte felici a volte meno), sembrano molto diversi da serie di impulsi elettrici, onde di potenziale, concentrazioni di sostanze chimiche.

I filosofi, gli psicologi e i neuroscienziati lottano da tempo con questo problema – tradizionalmente etichettato come il problema mente-corpo – che già affliggeva Cartesio, Darwin ed Einstein. In tempi recenti, con una brillante mossa di *marketing* filosofico, il filosofo David Chalmers ha ribattezzato la questione, con un gioco di parole consentito dal duplice significato di "*hard*" nella lingua inglese, con un nuovo nome che, almeno, ha il pregio di sottolinearne l'estrema difficoltà concettuale: l'*hard problem* (Chalmers 1996).

Finora, nonostante l'impegno profuso dalle neuroscienze – e da alcuni suoi esponenti di spicco quali i premi Nobel Francis Crick e Gerald Edelman – per spiegare la mente a partire dall'attività neurale non vi sono stati risultati di rilievo. Anzi, potremmo dire che non vi è stato alcun risultato. La comparsa della mente cosciente è tanto misteriosa oggi quanto ai tempi di Galileo. E, senza la coscienza, la mente è un po' come una democrazia senza libertà di pensiero, qualcosa che si fatica a definire. Come ammette il neurologo Christof Koch, al momento non c'è alcuna ipotesi scientifica che spieghi perché l'attività di milioni (o miliardi) di neuroni debba produrre qualcosa di simile alla nostra mente cosciente (Koch 2004; Koch 2012). Per quanto se ne sa, tutti questi

neuroni potrebbero semplicemente *funzionare*, come fanno le cellule del fegato o quelle del sistema immunitario, senza che il soggetto abbia la minima sensazione cosciente. Lo psicologo cognitivo ungherese Steven Harnad sottolinea che le neuroscienze finora spiegano il funzionamento dei neuroni, ma non dicono nulla né sul funzionamento né sulla natura della mente (Harnad 2000).

Anche tralasciando la coscienza, rimane il dato di fatto che il linguaggio della mente non è traducibile in quello dei neuroni e viceversa. Ogni traduzione, finora, è metaforica e arbitraria, come dimostrato dal fatto che qualsiasi attività neurale, per quanto se ne sa, potrebbe essere legata a qualsiasi contenuto mentale. In modo analogo, lo stesso bit all'interno di un computer può rappresentare un carattere, un punto di luce, una nota, un colore. Non esiste alcun legame necessario tra l'informazione nei computer e nel nostro cervello e il contenuto della mente. Le neuroscienze, come si è detto prima, hanno offerto solo degli autorevoli "spiegherò".

Molto spesso nella scienza si assiste a questa situazione: un'ipotesi non dimostrata (per esempio la terra al centro del sistema solare) viene incondizionatamente assunta come vera, e, di conseguenza, si devono aggiungere ipotesi accessorie per giustificare il punto di partenza (errato ma prestigioso). Ci sono innumerevoli esempi di questo tipo: gli esperimenti di Cartesio (mai eseguiti) per negare il principio di inerzia, i vortici per giustificare l'attrazione gravitazionale attraverso un contatto meccanico, l'etere per giustificare l'assolutezza dello spazio newtoniano, e così via. La strategia seguita è sempre la stessa: la scienza impiega un principio esplicativo che ha avuto successo in un campo e lo reitera in un altro. Il trucco, spesso, funziona. Ma non sempre. Prima o poi, arriva il momento in cui l'astuzia della ragione o l'imprevedibilità della natura rompono i vecchi schemi esplicativi. E quasi sempre gli studiosi se ne accorgono solo dopo ripetuti fallimenti. Chi ricorda più la clamorosa cantonata del grande fisico Lord Kelvin quando contestò Darwin sulla base di una stima dell'età della terra di pochi milioni di anni? Kelvin non aveva previsto la radioattività. Il grande fisico inglese voleva imporre i metodi della propria disciplina (i fenomeni di ossidazione) a un fenomeno (il sole) che richiedeva un salto concettuale. Non sarà che le neuroscienze, in questa fase storica, stanno cercando di imporre il proprio modello (di indubbio successo) a un fenomeno che non rientra nella loro stretta competenza? Forse è necessario un salto di paradigma scientifico: il futuro della scienza non è scritto dentro i metodi dei predecessori delle neuroscienze.

La fisica, per esempio, ha costantemente allargato le proprie radici ontologiche. Molti fenomeni che non sarebbero stati accettati al tempo di Galileo, sono stati progressivamente integrati: forze che agiscono a distanza, spazio e tempo integrati, dualismo onda-particella, buchi neri, equazioni d'onda, etc. Tali nuovi fenomeni hanno richiesto modifiche radicali sia nei contenuti sia nei criteri di valutazione. Il caso della mente è ancora aperto. Nel passato recente molti studiosi della mente (da Brentano a Mach, da Whitehead a James) hanno preso in considerazione ambiziosi schemi ontologici che offrirono una possibilità di comprensione della mente. Questi tentativi, forse incompleti ma promettenti, sono stati abbandonati. Le neuroscienze sembrano prigioniere di un'ortodossia nella descrizione della natura che non permette elasticità teorica e che impone meccanismi di valutazione rigidi. Si assiste così a un atteggiamento di sostanziale conservazione – le neuroscienze sembrano negare che la mente possa

richiedere una revisione delle basi fisiche. I neuroni sono descrivibili utilizzando un'ontologia fisica relativamente meccanicistica e semplice. La mente rientrerà in questi confini, o richiederà un allargamento della nostra nozione di *physis*? Non possiamo stabilirlo a priori, e solo teorie adeguate – e oggi al di fuori della portata e della capacità previsionale delle neuroscienze – potranno darci una risposta. Oggi queste teorie, semplicemente, non esistono. Einstein diceva che, quando un problema si presenta ripetutamente con caratteristiche di insolubilità, la soluzione consiste in un cambio delle premesse, piuttosto che nell'accanirsi nel proporre spiegazioni sempre più complesse.

#### 4. *Neuroni, società e media*

Esistono altri motivi di preoccupazione che riguardano la valenza sociale delle neuroscienze. In un recente articolo intitolato *Neuroscience in the Public Sphere* sulla rivista "Neuron", O'Connor, Rees e Joffe propongono varie cause che possono spiegare questa rapida identificazione tra mente e cervello (O'Connor, Rees *et al.* 2012). In particolare, questi autori sostengono che tre fattori condizionano il modo in cui le neuroscienze presentano i propri risultati: 1) considerare il cervello un fattore competitivo, 2) usare presunte differenze a livello cerebrale per valutare le persone, infine 3) trovare nel cervello la giustificazione per i comportamenti umani. Quello che stupisce è che ognuno di questi punti non sia di natura scientifica, ma piuttosto derivi dall'autorità sociale che le neuroscienze hanno ottenuto attraverso i media.

Il primo punto è facilmente comprensibile. In una società altamente competitiva come la nostra, il cervello diventa il terreno sul quale si gioca il riconoscimento del merito e delle capacità. Tutto quello che permette di quantificare il livello di prestazioni cognitive diventa oggetto di grande interesse. Come gli atleti sono tentati dal doping per raggiungere prestazioni eccellenti, così molti si chiedono se non sia possibile potenziare le capacità cognitive attraverso il cervello. In modo analogo a quanto era avvenuto con la frenologia di Gallo e Lombroso, il cervello è diventato un elemento di discriminazione e differenziazione. La presenza di presunte strutture cerebrali condizionerebbe le capacità e le potenzialità degli individui. Anche senza scomodare inquietanti distopie, queste presunte differenze cognitive rischiano di riproporre antiche barriere e privilegi.

Il cervello è sempre più spesso proposto per giustificare su basi biologiche alcuni comportamenti e rifiutare altri. In questo senso vanno letti alcuni lavori che cercano di trovare le radici di particolari stili di vita in propensioni innate a livello neurale. Il cervello diventerebbe così la misura di tutte le cose, o almeno di tutte le regole della morale umana. In ambito sociale, le neuroscienze danno l'illusione di poter deresponsabilizzare la mente. Se io trovo qualcosa dentro una certa area corticale e posso dire che i miei processi mentali derivano da quella certa attività, la responsabilità delle mie azioni e intenzioni non è più mia: non sono io che rubo, ma il mio sistema talamo-corticale. Non a caso, oggi, molti autori riflettono sulla possibilità di definire sia la morale che l'etica a partire da studi neuroscientifici dando così luogo a discipline quali la neuroetica.

Le neuroscienze possono così dare vita a un metacriterio per stabilire la validità eziologica di comportamenti e valori etici. Inevitabilmente tale ruolo si traduce in un loro

valore aggiunto che ne condiziona l'applicazione e inquina i criteri per valutarne la correttezza scientifica. Al pari della citata frenologia, le neuroscienze acquistano un valore sociale che si presta a una loro strumentalizzazione. In merito, Mausfeld ha recentemente affermato che, alla base di alcune posizioni delle neuroscienze rispetto alla mente, si trovano motivazioni al confine tra le finalità politiche e gli influssi sociali (Mausfeld 2012). Secondo Mausfeld, le neuroscienze non offrono una vera spiegazione. Al contrario si limitano a illudere di averlo fatto proponendo processi materiali quali presunte cause dei processi mentali. È questa concretezza che conferisce forza alle spiegazioni delle neuroscienze. Tuttavia sarebbe un grave errore non accorgersi che tra i dati neurali e i fatti mentali esiste uno iato incolmabile. I primi non descrivono, non spiegano e non permettono di prevedere i secondi. La loro contemporanea presenza è un fatto che si constata, ma non si spiega.

Non si tratta di una domanda solo filosofica o scientifica, in quanto il meccanismo dei finanziamenti scientifici e il rapporto tra scienza e società sta facendo guadagnare alle neuroscienze una notevole autorità politica e normativa. Secondo Jan Slaby, in una società basata sul valore individuale (almeno in teoria) delle singole menti, ogni ipotesi forte su quello che la mente è e fa si traduce invariabilmente in norme e giustificazioni per gruppi sociali (Slaby 2010). Le neuroscienze non possono fingere di ignorare l'intreccio centrifugo di scienze umane, pregiudizi psicologici, interessi finanziari e scommesse scientifico-tecnologiche che, come un uragano, si trascinano dietro. Nessuna scienza oggi è innocente. Sui mezzi di comunicazione siamo inseguiti quotidianamente da dichiarazioni roboanti circa la scoperta delle basi neurali di molti tratti umani: dalla propensione all'acquisto alla fede politica, dall'esperienza religiosa alla propensione alla pedofilia, dall'arte all'etica. Oltretutto, con il rischio che si prendano decisioni sulla base di risultati preliminari in settori a rischio quali determinate sindromi o condizioni (autismo, iperattività, stato vegetativo, dislessia e simili, si veda Racine, Waldman *et al.* 2010; Choudhury e Slaby 2011; Gonon, Bezard *et al.* 2011; Burnett 2012).

La comunità scientifica non è del tutto esente da responsabilità per quanto riguarda il modo errato con il quale i risultati delle neuroscienze sono presentati e assorbiti dalla società. In una recente analisi, Gonon *et al.* evidenziano che «c'è una enorme differenza tra i dati sperimentali e le conclusioni comunicate dai media. Questa differenza è una conseguenza sia del processo di comunicazione che dei modi con i quali gli autori di pubblicazioni scientifiche presentano il proprio lavoro» (Gonon, Bezard *et al.* 2011, p. 1). Questi modi fuorvianti dipendono dalle aspettative che gli autori hanno circa gli aspetti più appetibili dei propri studi, spesso confondendo premesse e risultati.

## 5. *Valutazione come competizione*

Alla fine, però, la scienza trova il modo di progredire, per lo meno nel lungo termine. Si dice che il grande valore del metodo scientifico è di essere autocorreggente, perché basato sull'intersoggettività e sulla competizione. C'è un esempio che vorremmo citare, quello relativo a una ricerca, poi dimostratasi fraudolenta, eseguita da un gastroenterologo inglese, Wakefield, che sosteneva di aver trovato la causa scatenante di una forma di autismo in un effetto collaterale patogeno del vaccino trivalente. La storia è

lunga e complessa, e ha effetti ancor oggi, poiché molti genitori sull'onda emotiva di quei dati, poi rivelatisi falsi, rifiutano di vaccinare i propri figli, con esiti epidemiologici e personali gravissimi. Ma la domanda che ci dobbiamo porre è: perché una rivista come “Lancet”, una bibbia della medicina, ha pubblicato i dati di una ricerca svolta su soli 12 bambini? Tutti sanno che la numerosità del campione è un elemento critico per valutare (ritorna il nostro termine) la validità di una ricerca. La risposta che abbiamo ipotizzato è che la redazione della rivista, certamente non composta da incompetenti, voleva probabilmente lasciare aperta la porta a un'ipotesi originale, per quanto debole statisticamente, in un campo drammatico come quello del disturbo autistico, per evitare di arroccarsi nel conservatorismo. In altre parole, la rivista “Lancet” sapeva di non essere il giudice ultimo e quindi di poter contare sul fatto che l'ipotesi sarebbe stata valutata da futuri altri ricercatori che, seguendo la prassi protocollare, l'avrebbero sottoposta a verifica empirica. Così infatti è stato, anche se, in questo caso, fu un'indagine giornalistica a scoprire che Wakefield aveva intenzionalmente compiuto una frode a scopo di lucro. Ma questo dettaglio non ha importanza per il processo di autocorrezione scientifica, i dati potevano anche essere stati raccolti onestamente, la valutazione intersoggettiva li avrebbe sottoposti a verifica – nella comunità scientifica vale il criterio della competizione che impedisce a qualsiasi agenzia di diventare dominante troppo a lungo. Anche se le comunità accademiche sono per loro natura conservatrici e dominanti, non esiste una struttura centralizzata che si possa imporre per sempre e, sia pure con qualche ritardo, i criteri vengono aggiornati e migliorati. La competizione garantisce l'intersoggettività della valutazione.

Il punto è che, soprattutto nelle fasi di ricerca straordinaria, quando paradigmi e criteri devono essere aggiornati, non ci può essere un giudizio *super partes* che non sia la riproposizione di criteri obsoleti e superati (se non fossero stati obsoleti, avrebbero già portato alla soluzione dei problemi). Quindi la competizione diventa la vera chiave di valutazione, ma una competizione che non può essere frutto dei soli giudizi interni alla comunità e che deve realizzarsi darwinianamente<sup>2</sup> attraverso la selezione che risulta dal confronto con il mondo esterno. Altrimenti, senza competizione, la valutazione è soltanto la giustificazione di un sistema feudale interno alla classe accademica, con i suoi conti, vassalli, e patti più o meno segreti.

Infine vorremmo ricordare un aspetto che tocca da vicino la vita accademica: i criteri per l'allocazione dei fondi di finanziamento. È ben noto che la vita accademica non è esente da bisogni concreti. Nel migliore dei casi tale approvvigionamento finanziario dovrebbe avvenire in modo da valorizzare le ricerche più significative. Ma chi decide quali siano queste ricerche? Altri ricercatori che, inevitabilmente, saranno condizionati dalla visione scientifica prevalente. Se una certa visione supera una massa critica tenderà a produrre una certa polarizzazione a suo vantaggio. Molti scienziati e ricercatori tenderanno a privilegiare particolari orientamenti per paura di essere esclusi dalla comunità che decide che cosa è “scienza” e che cosa non lo è.

L'introduzione di indicatori “oggettivi” e “quantitativi” per valutare il successo dei singoli ricercatori tende inevitabilmente a riprodurre la direzione più ortodossa di ricerca.

---

<sup>2</sup> Nel senso vero della selezione, per pressione ambientale, non nella versione del darwinismo sociale, o legge del più forte, con cui viene da molti confusa.

Se i criteri sono stabiliti sulla base dell'aderenza dei risultati a una serie di metodi precedentemente adottati, c'è il rischio del diffondersi di strategie opportunistiche (anche in buona fede). La ricerca veramente innovativa non può essere pianificata. Nella sua raccolta di consigli ai giovani ricercatori, il grande neurofisiologo Ramon y Cajal dichiarava che «non ci sono regole per fare importanti scoperte [...] la creatività non segue alcuna regola, ma le forgia autonomamente». Su questo tema, Condorcet aveva sintetizzato con la massima: «Il mediocre può essere educato, il genio educa se stesso». E, per quanto umile, per essere produttivo ogni ricercatore deve avere una scintilla di creatività.

Prima abbiamo citato il caso del cambio dei paradigmi scientifici. L'uso di indici di misurazione oggettiva tende a non incoraggiare la scoperta rivoluzionaria (se non a posteriori, una volta che la scoperta abbia superato ogni scetticismo). I parametri "oggettivi", per lo meno sul breve periodo, sono adatti soprattutto alla scienza "incrementale", ovvero a quel tipo di ricerca che reitera i metodi considerati oggettivi e li applica a un dominio crescente di casi sperimentali. Il caso della mente potrebbe però esulare della scienza ordinaria che si esprime attraverso progressi incrementali e potrebbe richiedere un cambiamento rivoluzionario. D'altronde, neuroscienziati di fama hanno esplicitamente denunciato la natura rivoluzionaria del problema della mente. Il recentemente scomparso premio Nobel Edelman scriveva che «per comprendere la mente, sono necessari nuovi modi di concepire l'attività neurale» (Edelman e Tononi 2000, p. 132). Lo stesso campione del riduzionismo neurale e premio Nobel Crick riconosceva che «il problema della mente non è risolvibile utilizzando i metodi e i concetti scientifici attualmente disponibili: sono indispensabili concetti radicalmente nuovi» (Crick e Koch 2002, p. 11).

Come scrive Schopenhauer, il talento è la capacità di colpire il bersaglio che nessuno riesce a colpire, ma il genio è la capacità di vedere il bersaglio che nessuno sa che esiste. Un po' come la moderna distinzione tra l'abilità di chi risolve i problemi (*problem solver*) e l'abilità di chi li inventa (*problem finder*). I metodi oggettivi di misurazione della ricerca scientifica misurano lo scostamento da un bersaglio noto (o in gran parte anticipato), ma non possono misurare la distanza da bersagli ignoti. Anzi, per quanto riguarda scostamenti radicali dalle conoscenze sedimentate, tali metodi potrebbero persino fornire una misura negativa per approcci veramente rivoluzionari (Gamow 1961). Ancora una volta la creatività non è valutabile ma trova nella competizione selezionata dal mondo esterno la sua chiave di volta.

A questo proposito, vale la pena di ricordare che molti neuroscienziati hanno abbracciato una visione incrementale del loro lavoro. Per esempio, molti ricercatori sostengono esplicitamente che il progresso delle neuroscienze sarà possibile solo attraverso l'accumulo incessante di nuovi dati circa il funzionamento del sistema nervoso (Koch 2004; Sporns 2011; Yarkoni, Poldrack *et al.* 2010). Quest'accumulo è sicuramente prezioso, e il suo valore, per lo studio dell'attività nervosa in quanto tale, è innegabile, ma la domanda che ci si deve porre è: quanto di quello che viene scoperto circa il sistema nervoso è importante per la mente? Non sappiamo ancora quale sia il livello critico per la mente... le sinapsi? I neurotrasmettitori? I segnali nervosi?

Finora le neuroscienze non sono riuscite a spiegare molto di quello che stava fuori dai loro confini (sono state bravissime invece, e non bisogna stupirsi, a spiegare i

meccanismi neurali). Di fronte ai tanti aspetti della mente per i quali le neuroscienze non dispongono di strumenti adeguati, la comunità scientifica ha spesso messo in atto quella curiosa strategia epistemica per risolvere gli enigmi fastidiosi così ben descritta da Robert Musil: dimenticarsene. Problemi quali quello del senso e del significato, del libero arbitrio, dell'intenzionalità e, naturalmente, della mente cosciente sono alle radici della persona umana e non possono essere elusi solo perché non si traducono facilmente in attività neurale. Come racconta la barzelletta dell'ubriaco che cercava le chiavi solo sotto la luce del lampione (dove non c'erano) perché almeno lì ci vedeva bene, così le neuroscienze rischiano di guardare solo laddove nel passato hanno trovato qualcosa. A volte si ha come l'impressione che le neuroscienze assomiglino un po' alla favola dell'imperatore nudo. Ci mostrano un cervello nudo e ci chiedono di valutare i suoi meravigliosi vestiti mentali che noi non vediamo perché, forse, non c'è nulla da vedere.

Secondo Mausfeld, queste tendenze autoreferenziali nel processo di valutazione corrono il rischio di indebolire la spinta a sviluppare quelle profonde revisioni concettuali che, alla lunga, sono la radice del progresso scientifico. Purtroppo, nel breve periodo, è evidente a tutti che il modo migliore per guadagnarsi una valutazione positiva da parte dei propri colleghi consiste in una produzione aderente all'ortodossia prevalente basata sull'applicazione meccanica di metodi di ricerca consolidati. Le neuroscienze sono particolarmente sensibili a questa tendenza. Una volta che si sia acquistata dimestichezza con la terminologia e i metodi, e una volta che si disponga di adeguate attrezzature, è possibile reiterare tali approcci su questioni di carattere incrementale (Mausfeld 2012). Approcci diversi sono sistematicamente rifiutati (Chemero 2009; Manzotti 2012) perché incoerenti con le idee dominanti (ma insoddisfacenti).

L'intreccio tra neuroscienze e psicologia diventa così particolarmente critico proprio sul piano della valutazione perché una disciplina diventa fondante dei criteri di valutazione di un'altra disciplina senza che la prima abbia, almeno finora, dimostrato di poter risolvere gli interrogativi di fondo della seconda, né sul piano ontologico né sul piano epistemico. In questo processo è stata cruciale la mancanza, da parte della psicologia, di un programma di ricerca preciso circa la naturalizzazione della mente (Manzotti e Moderato 2011).

Rifacendoci alle considerazioni iniziali, la mente non si pone come un problema ordinario, ma piuttosto come un momento di scienza straordinario che, per sua natura, richiede un cambiamento radicale di paradigma che non può essere valutato con criteri tradizionali. In questo senso, i criteri proposti dalle neuroscienze non peccano per innovatività, ma per mancanza di originalità: l'idea che la mente non sia altro che una proprietà invisibile che viene secreta dai neuroni è troppo facile e troppo simile ad antiche nozioni animiste per non essere sospetta, anche se la si (tra)veste con parole e termini apparentemente nuovi.

## Riferimenti

- M.P. ALVAREZ (2011), *The Magnetism of Neuroimaging: Fashion, Myth and Ideology of the Brain*, in “Papeles del Psicologo”, 32 (2/2011), pp. 98-112.
- A. BARTELS-S. ZEKI (2000), *Neural Basis of Romantic Love*, in “Neuroreport”, 11 (2000), pp. 3829-3834.
- M. BENNETT-P. HACKER (2003), *Philosophical Foundations of Neuroscience*, Blackwell, Malden MA 2003.
- D. BURNETT (2012), *Neuroscience Fiction in Newspaper*, in “The Guardian”, 1° maggio 2012 (<http://www.theguardian.com/science/blog/2012/may/01/neuroscience-fiction>).
- D.J. CHALMERS (1996), *The Conscious Mind: In Search of a Fundamental Theory*, Oxford University Press, New York NY 1996.
- A. CHEMERO (2009), *Radical Embodied Cognitive Science*, MIT Press, Cambridge MA 2009.
- S. CHOUDHURY-J. SLABY (2011), *Critical Neuroscience: A Handbook of the Social and Cultural Contexts of Neuroscience*, Wiley-Blackwell, New York NY 2011.
- F. CRICK (1994), *The Astonishing Hypothesis: the Scientific Search for the Soul*, Touchstone, New York NY 1994.
- F. CRICK-C. KOCH (2002), *The Problem of Consciousness*, in “Scientific American: Special Edition”, 12 (1/2002), pp. 10–17.
- G. EDELMANN-G. TONONI (2000), *A Universe of Consciousness*, Allen Lane, London 2000.
- G. GAMOW (1961), *The Great Physicists, from Galileo to Einstein*, Dover, New York NY 1961.
- A. GARNETT-L. WHITELEY-H. PIWOWAR-E. RASMUSSEN-J. ILLES (2011), *Neuroethics and fMRI: Mapping a Fledgling Relationship*, in “PLOS One”, 6(4/2011), pp. 1-7.
- F. GONON- E. BEZARD-T. BORAUD (2011), *Misrepresentation of Neuroscience Data Might Give Rise to Misleading Conclusions in the Media: The Case of Attention Deficit Hyperactivity Disorder*, in “PLOS One”, 6 (1/2011), pp. 1-8.
- V.G. HARDCASTLE-C.M. STEWART (2009), *fMRI: A Modern Cerebroscope? The Case of Pain*, in J. BICKLE (a cura di), *The Oxford Handbook of Philosophy and Neuroscience*, Oxford University Press, Oxford, pp. 200-225.
- S. HARNAD (2000), *Correlation vs. Causality: How/Why the Mind/Body Problem is Hard*, in “Journal of Consciousness Studies”, 7 (2000), pp. 54-61.
- G. HICKOK (2014), *The Myth of Mirror Neurons*, Norton & Company, New York NY 2014.
- J. ILLES-S.J. BIRD (2006), *Neuroethics: a Modern Context for Ethics in Neuroscience*, in “Trends in Neurosciences”, 29 (9/2006), pp. 511-517.
- J. ILLES-M.A. MOSER-J.B. MCCORMICK-E. RACINE-S. BLAKESLEE-A. CAPLAN-E.C. HAYDEN-J. INGRAM-T. LOHWATER-P. MCKNIGHT-C. NICHOLSON-A. PHILLIPS-K.D. SAUVE-E. SNELL-S. WEISS (2010), *Neurotalk: Improving the Communication of Neuroscience Research*, in “Nature Reviews Neuroscience”, 11(1/2010), pp. 61-9.
- C. KOCH (2004), *The Quest for Consciousness: A Neurobiological Approach*, Roberts & Company Publishers, Englewood CO 2004.
- C. KOCH (2012), *Consciousness. Confessions of a Romantic Reductionist*, MIT Press, Cambridge MA 2012.

- A. LAVAZZA-R. MANZOTTI (2011), *Modelli di creatività: dall'elaborazione inconscia e implicita al fringe jamesiano*, in “Giornale italiano di psicologia”, 38 (1/2011), pp. 47–76.
- P. LEGRENZI-C. UMITÀ (2009), *Neuro-mania. Il cervello non spiega chi siamo*, Il Mulino, Bologna 2009.
- N.K. LOGOTHETIS (2008), *What We Can Do and What We Cannot Do with fMRI*, in “Nature”, 453 (2008), pp. 869-78.
- R. MANZOTTI-P. MODERATO (2010), *Is Neuroscience the Forthcoming ‘Mindscience’?*, in “Behaviour and Philosophy”, 38 (1/2010), pp. 1-28.
- R. MANZOTTI-P. MODERATO (2011), *I confini della mente. Verso una nuova ontologia per la psicologia?*, in “Giornale Di Psicologia”, 5(1-2/2011), pp. 19–39.
- R. MANZOTTI-R. PEPPERELL (2012), *The New Mind: Thinking Beyond the Head*, in “AI & Society. Knowledge, Culture and Communication”, 24 (1/2012), pp. 1–12.
- R. MANZOTTI-V. TAGLIASCO (2008), *L'esperienza. Perché i neuroni non spiegano tutto*, Codice, Milano 2008.
- G. MARCUS (2012), *Neuroscience Fiction*, in “The New Yorker”, 30 novembre 2012 (<http://www.newyorker.com/news/news-desk/neuroscience-fiction>).
- R. MAUSFELD (2012), *On Some Unwarranted Tacit Assumptions in Cognitive Neuroscience*, in “Frontiers in Psychology”, 3 (67/2012), pp. 1-13.
- M. MOTTERLINI-M. MONTI (2012), *Il cervello sa se lo spot funziona*, in “Il Sole 24 Ore”, 20 maggio 2012 ([http://www.ilsole24ore.com/art/cultura/2012-05-20/cervello-spot-funziona-081510\\_PRN.shtml](http://www.ilsole24ore.com/art/cultura/2012-05-20/cervello-spot-funziona-081510_PRN.shtml)).
- C. O'CONNOR-G. REES-H. JOFFE (2012), *Neuroscience in the Public Sphere*, in “Neuron”, 74 (2012), pp. 220-226.
- E. RACINE-O. BAR-ILAN-J. ILLES (2005), *fMRI in the Public Eye*, in “Nature Reviews Neuroscience”, 6 (2005), pp. 159-166.
- E. RACINE-S. WALDMAN-J. ROSENBERG-J. ILLES (2010), *Contemporary Neuroscience in the Media*, in “Social Science & Medicine”, 71(4/2010), p. 725.
- J. SLABY (2010), *Steps towards a Critical Neuroscience*, in “Phenomenology and the Cognitive Sciences”, 9 (3/2010), pp. 397-416.
- O. SPORNS (2011), *Networks of the Brain*, MIT Press, Cambridge MA 2011.
- G. TONONI (2004), *An Information Integration Theory of Consciousness*, in “BMC Neuroscience”, 5 (2004), pp. 1-22.
- W.R. UTTAL (2001), *The New Phrenology: The Limits of Localizing Cognitive Processes in the Brain*, MIT Press, Boston MA 2001.
- D.S. WEISBERG-F.C. KEIL-J. GOODSTEIN-E. RAWSON-J.R. GRAY (2008), *The Seductive Allure of Neuroscience Explanations*, in “Journal of Cognitive Neuroscience”, 20 (3/2008), pp. 470-477.
- A.N. WHITEHEAD (1920), *Concept of Nature*, Cambridge University Press, Cambridge MA 1920.
- T. YARKONI-R.A. POLDRACK-D.C. VAN ESSEN (2010), *Cognitive neuroscience 2.0: Building a Cumulative Science of Human Brain Function*, in “Trends in Cognitive Sciences”, 14 (11/2010), pp. 489-496.
- S. ZEKI (2001), *Localization and Globalization in Conscious Vision*, in “Annual Review of Neuroscience”, 24 (2001), pp. 57–86.



# **POLITICHE**



Donna M. Mertens

## PHILOSOPHICAL ASSUMPTIONS AND PROGRAM EVALUATION

### **Abstract**

*The transdisciplinary nature of evaluation allows its application in diverse contexts, with diverse stakeholder groups, to address diverse social problems, through the use of diverse methodologies. Given these multiple dimensions of diversity, there are also diverse sets of philosophical assumptions that underlie the choices that evaluators make regarding their methodologies. This essay explores the different philosophical framings that are guiding thinking in the evaluation community through the lens of four paradigms: postpositivist, constructivist, pragmatic, and transformative.*

Program evaluation is described by Michael Scriven (2003) as a transdiscipline that is characterized as a discipline that supplies “essential tools for other disciplines, while retaining an autonomous structure and research effort of [its] own” (p. 19). Evaluation is not merely the application of social science methods to solve social problems; rather, evaluators use social science methods to examine the merit, worth and significance of a program or project or policy for the purposes of describing values associated with different stakeholder groups, as well as reaching evaluative conclusions “about good and bad solutions to social problems” (p. 21). The transdisciplinary nature of evaluation allows its application in diverse contexts, with diverse stakeholder groups, to address diverse social problems, through the use of diverse methodologies. With these multiple dimensions of diversity, it should come as no surprise that there are also diverse sets of philosophical assumptions that underlie the choices that evaluators make regarding their methodologies.

“Evaluation is situated in a broad landscape in terms of its diverse meanings in different disciplines, sectors, nations, and venues. The hallmarks of the evaluation field are its interdisciplinary roots and the ways in which the resultant conversations around the meaning of evaluation have benefited from this diversity of perspectives” (Mertens & Wilson 2012, p. 1). The evaluation field has experienced many decades of differences of opinions about which methodologies are best; at times these differences have been acrimonious. However, Shadish (1998) claims that differences about methodologies are not based on arguments about methods choices, but they are reflective of the different philosophical assumptions that guide methodological choices. He wrote that most debates in the evaluation field are “about epistemology and ontology, about what assumptions we make when we construct knowledge, about the nature of many fundamental concepts that we use in our work like causation, generalization and truth” (p. 3).

Mertens (2009; 2015) and Mertens and Wilson (2012) built on the work of Guba and Lincoln’s (1989; 2005) concept of paradigms in research and evaluation; they described four sets of philosophical assumptions that constitute a paradigm: axiology, ontology, epistemology and methodology. Mertens and Wilson identified four major paradigms that are operating in the world of evaluation: postpositivist, constructivist, transformative, and pragmatic. These paradigms are associated with four branches of evaluation that reflect the diverse perspectives in the field. Christie and Alkin (2013) identified three branches of evaluation: Methods, Use and Values. Mertens and Wilson (2014) added the fourth branch of Social Justice. The paradigms and branches of evaluation align in the following way: The Methods Branch maps onto the postpositivist paradigm, the Use Branch onto the pragmatic paradigm, the Values Branch onto the constructivist paradigm, and the Social Justice Branch onto the transformative paradigm. The following figure depicts the primary focus of each paradigm and its associated evaluation branch.

Figure 1. Evaluation Paradigms and Branches (adapted from Mertens & Wilson 2012, p. 56)

Paradigm	Branch	Description
Postpositivist	Methods	Focuses primarily on quantitative designs and data
Pragmatic	Use	Focuses primarily on data that are found to be useful by stakeholders; advocates for the use of mixed methods
Constructivist	Values	Focuses primarily on identifying multiple values and perspectives through qualitative methods
Transformative	Social Justice	Focuses primarily on viewpoints of marginalized groups and interrogating systemic power structure through mixed methods to further social justice and human rights

At first glance, it should be apparent that there is a possibility of overlap between the various paradigmatic positions and evaluation branches. For example, constructivists or proponents of other paradigms and branches can work to advance social justice and human rights; this is not the sole territory of the transformative social justice evaluator. However, the transformative paradigm arose out of the expressed dissatisfaction of members of marginalized communities about the assumptions that were made in evaluations conducted in their communities and the recognized need for more culturally responsive thinking (Mertens 2009; Hood, Hopson & Frierson 2015). The four paradigms and their associated philosophical assumptions are next discussed in reference to the evaluation branches.

### 1. *Postpositivism and the Methods Branch*

The philosophical origins of the postpositivist paradigm can be traced back to Sir Francis Bacon (1561-1626) in his articulation of the principles of the scientific method that included the ontological assumption that one reality exists and it is independent of the observer (Howell 2013; Turner 2001). This leads to an epistemological assumption

that tasks researchers with the adoption of a distanced manner in order to capture the knowledge or reality that exists outside of the individual. The axiological assumptions associated with the postpositivist paradigm align with those included in the National Commission for the Protection of Human Subjects of Biomedical and Behavioral Research (1979) in its Belmont Report. The ethical principles of beneficence, respect and justice from the Belmont Report provide the framework for most ethical review boards in the United States. The interpretation of these principles lead to ethical procedures such as confidentiality, informed consent, and avoidance of coercion (Mertens & Wilson 2012). The methodological assumption reflects the use of scientific methods that allow the evaluator to discover laws about human behavior through empirical observations, prioritizing the use of experimental designs in the form of randomized control trials that require random selection of subjects and random assignment to intervention conditions (Mertens 2015).

White (2013) explains that the use of randomized control trials in evaluation increased in the late 2000's because of the frustration expressed by donors that there was a "lack of rigorous evidence as to which development programs are effective" (p. 62). Although White was writing about international development evaluations, a similar frustration was also expressed by many domestic funders of social programs. Thus, a focus on impact evaluations occurred in order to get an answer to the question: What difference did an intervention make? In keeping with the assumptions of the postpositivist paradigm and the Methods Branch of evaluation, White writes: "For interventions with a large number of units of assignment, this question is best answered with a quantitative experimental or quasi-experimental design. And for prospective, or *ex ante*, evaluation designs a randomized control trial (RCT) is very likely to be the best available method for addressing this attribution question if it is feasible" (p. 61). White goes on to acknowledge that such a design only answers one narrow question: Did it work? In order to obtain answers to other questions, such as how well was the intervention implemented, was the targeted population reached, and what barriers to participation were encountered, White recommends the use of mixed methods designs.

The stated benefits of the randomized control trial design is that it can answer the question about impact whilst requiring no understanding of the complex causal chain associated with the effect of the intervention (White 2013). This is because of the assumption that randomization of the control and treatment groups controls for any differences in baseline characteristics of the participants. Everything is presumed to be equal, except for the administration of the intervention to one group and not to the other. Mixed methods approaches can be added to an RCT by collecting qualitative and quantitative data to answer questions about the causal chain and to interpret results from the statistical analysis of the RCT. Given the real-world context of evaluation, the conditions necessary for RCTs can be difficult to meet. If the intervention is not well-designed and culturally appropriate, then the study will only confirm its failure. Other paradigmatic stances and evaluation branches place greater emphasis on determining what is needed and provision of data to make adjustments throughout the course of the program so that the evaluation is dynamic and responsive to the culture and the community's changing context. The constructivist paradigm and the Values Branch

reflect one of the paradigms that assumes that evaluators need to be more personally involved with the targeted communities.

## *2. Constructivism and the Values Branch*

House (1990) described the movement of evaluation in the direction of the Values Branch as follows: “Philosophically, evaluators ceased to believe their discipline was value-free and realized their practice entailed promoting the values and interests of some groups over others, though they were by no means clear on what to do about this discovery...If diverse groups wanted different things, then collecting the views of people in and around the programs themselves seemed to make sense. Qualitative methodology useful for obtaining the views of participants came into vogue” (p. 25). The arguments that ensued in the world of evaluation as to whether quantitative or qualitative methods were better were based on differences in philosophical assumptions associated with the postpositivist and constructivist paradigms.

The historical, philosophical roots of the constructivist paradigm are found in the late 1700s work of Immanuel Kant (1781/1966) and Husserl (1936/1970). The constructivist ontological assumption holds that humans create knowledge based on processing their experiences through interaction with external stimuli. Epistemologically, the evaluator needs to interact with participants and to engage in meaningful dialogue and reflection to create knowledge (Guba & Lincoln 2005). Schwandt (2000) adds to these assumptions in a methodological sense by noting that a constructivist attempts to reach an understanding of meaning from the perspective of the persons who have the experiences. It is possible that the persons themselves do not understand the experience fully. Thus, an evaluator can use methods that help make visible understandings for diverse stakeholders through the use of multiple methods.

The ontological, epistemological, and methodological assumptions of the constructivist paradigm and the Values Branch serve to expand the understanding of ethics in evaluation. The assumption of diverse values at play in evaluations brings up the issue of the evaluator’s own values and those of the various stakeholder groups. Constructivists in the world of research recognize that because the researcher is the instrument, the researcher needs to make a careful inventory of their own values and how those values color their perceptions in the research context. The same holds true for evaluators. How do they make visible their own values and the values of the various stakeholder groups? How do they insure that the results of their evaluation are accurately reflective of the different values, beliefs, and interests of the different constituencies? Methodologically, this means that evaluators need to develop a relationship with the stakeholders and immerse themselves in the community sufficiently to engage in meaningful reflective dialogue with participants. Constructivists have a strong tendency to use qualitative methods. When mixed methods research emerged as a growing phenomenon in the United States in the late 1990s, some constructivist researchers rejected the possibility of combining qualitative and quantitative methods on the grounds that the assumptions of the postpositivist and constructivist paradigms were

incompatible. As the evaluation world explored how to address this conundrum, adherents of the pragmatic paradigm appeared.

### 3. Pragmatic paradigm and the Use Branch

With the passage of legislation in the United States in the 1960s under the Great Society initiative, evaluators realized that their work had the potential to inform policy decisions at the highest level. Several evaluation scholars, such as Daniel Stufflebeam (1980), Carol Weiss (1998), and Michael Patton (2010), raised the consciousness of the evaluation community regarding the use (or nonuse) of their findings. The pragmatic paradigm aligns closely with the Use Branch in the sense that the focus is on the conduct of evaluations that can provide information in a way that the intended stakeholders can use the results as a basis for informing decision making.

The pragmatic paradigm began in the second half of the 19<sup>th</sup> century with the contributions of William James, John Dewey, George Herbert Mead and Arthur F. Bentley (Mertens & Wilson 2012). These scholars rejected the idea that truth could be discovered by scientific methods. Neopragmatism emerged in the 1960s, with scholars such as Abraham Kaplan, Richard Rorty, and Cornel West (Maxcy 2003). “These philosophers have distinguished themselves from the early pragmatists by their emphasis on common sense and practical thinking” (Mertens & Wilson 2012, p. 89).

Hall (2013), Greene (2007), and Denzin (2012) argue that the invocation of pragmatism as a philosophical base for the Use Branch of evaluation is misleading. Rather, many evaluators in the Use Branch do not consciously act from the philosophical assumptions associated with pragmatism. Rather, they adopt a utilitarian, “what works” approach that has been criticized as an a-paradigmatic stance that emphasizes convenience instead of engaging with philosophical underpinnings.

The tension between pragmatism as a philosophical frame and utilitarianism as a practical frame for evaluation is one that is generating a great deal of thought in the evaluation world. At present, the Use Branch operates with an axiological assumption that aligns with the utilitarian theory of ethics, which holds that the value of something is a function of its consequences (Christians 2005). Morgan (2007) describes the ethical stance of pragmatism as gaining knowledge in pursuit of desired ends. Rather than doing an evaluation for the sake of an evaluation, pragmatists see the value of the evaluation *as how it is used and the results of that use*” (Mertens & Wilson 2012, p. 90, italics in the original).

In ontological terms, Tashakkori and Teddlie (2003) state that pragmatists avoid arguing about metaphysical terms such as truth and reality. They argue that the value of evaluation is not based on whether they discover the truth, but on the demonstration that the results work with respect to the problem that is being studied (Mertens & Wilson 2012). Epistemologically, the evaluator is free to develop whatever type of relationships with stakeholders is appropriate for the matter under investigation. The nature of the relationship is judged in terms of its ability to get the results of the evaluation used by the intended stakeholders. Methodologically, this philosophical stance has been used to justify the use of mixed methods in evaluation (Morgan 2007;

Tashakkori & Teddlie 2003). The underlying methodological assumption of the Use Branch is that the method should match the purpose of the evaluation (Patton 2010).

Hall (2013) argues that evaluators would benefit by paying closer attention to the tenets of pragmatism, especially as it was conceptualized by Dewey (1923/1998). Dewey contributes the following ideas that are relevant to evaluators: First, his thoughts on intelligent action allows the evaluator to increase their contextual sensitivity and examine “the tangible processes for how inquiry and credible evidence are achieved. Second, his views on intelligent action advance reflection, ethics, and social justice. And third, Dewey’s pragmatism is relevant because, like many evaluators, his main objective is to address societal problems by taking action in an intelligent way” (Hall 2013, p. 17).

Dewey’s reflection on the meaning of Truth include a rejection of the idea of an absolute truth in favor of a transactional realist perspective that sees truth and knowledge as being “temporal and embedded in and generated through our experiential transactions. Truth is linked to action, and has to be tested continuously and substantiated. It is in this way transactional realism supports an experimental inquiry approach in which verification plays a significant role to determine future actions” (Hall 2013, p. 17). Hence, evaluators would adopt a critically reflective stance in a dynamic system to understand the complexity of their work in order to advocate for the use of the results of experimentation for intelligent action.

Dewey emphasized the importance of social inquiry in the advancement of social justice (Hall 2013). However, because of the utilitarian nature of much of the scholarship in the Use Branch, social justice issues have not been given priority. The transformative paradigm emerged as a response to the need for an explicit philosophical framing for evaluations that prioritize human rights and social justice (Mertens 2009; 2015).

#### *4. Transformative paradigm and the Social Justice Branch*

The philosophical roots of the transformative paradigm and the Social Justice Branch are eclectic, reflecting early work by Kant (1781/1966) and Hegel (1812/1929) regarding the importance of critically interrogating subjugation and the master-slave relationship in order to address issues of power and inequities (Mertens & Wilson 2012). Later, philosophers such as Marcus (1998), Habermas (1971), and Horkheimer (1972) extended thinking about value-laden perspectives in inquiry and the prioritization of social justice as the starting principle for research and evaluation. A multitude of theoretical perspectives contribute to expanded understandings of the transformative paradigm. Kincheloe and McLaren (2005) elucidate the contribution of critical theory as follows: “A critical social theory is concerned in particular with issues of power and justice and the ways that the economy; matters of race, class, and gender; ideologies, discourses; education; religion and other social institutions; and cultural dynamics interact to construct a social system” (p. 92). While this statement reflects some of the diversity in terms of dimensions that are used as a basis for discrimination and oppression, additional theoretical perspectives also contribute to the transformative paradigm.

Additional theoretical perspectives include feminists such as Irigaray, Kristeva, and Cixous (see Kincheloe & McLaren 2005); indigenous and postcolonial philosophers such as Asante (1992) and Chilisa (2011) from Africa, Cram (2009) from New Zealand's Maori community, Freire (1970) from Latin America and LaFrance and Crazy Bull (2009) from the American Indian community; and disability and deafness rights theorists (Mertens, Holmes & Harris 2009; Sullivan 2009).

These philosophical roots and theoretical perspectives contribute to the transformative paradigm because they all address issues of power inequities, privilege, and the consequences of these for achieving social justice. The transformative paradigm can be described as follows:

The transformative paradigm offers a meta-physical umbrella that brings together these various philosophical strands. It is applicable to people who experience discrimination and oppression on whatever basis, including (but not limited to) race/ethnicity, disability, immigrant status, political conflicts, sexual orientation, poverty, gender, age, or the multitude of other characteristics that are associated with less access to social justice. In addition the transformative paradigm is applicable to the study of the power structures that perpetuate social inequities (Mertens 2009, p. 4).

The transformative axiological assumption reflects an awareness of the pervasiveness of discrimination that occurs in many communities and the ethical responsibility of the evaluator to understand critical dimensions of diversity in order to challenge societal processes that perpetuate an oppressive status quo (Mertens & Wilson 2012). Thus, the transformative ethical assumptions extend on the ethical principles explicated in the Belmont Report (discussed earlier in this article).

Respect is critically examined in terms of the cultural norms of interaction in diverse communities and across cultural groups. Beneficence is defined in terms of the promotion of human rights and an increase in social justice. An explicit connection is made between the process and outcomes of evaluation studies and the furtherance of a social justice agenda (Mertens 2009, p. 49-50).

Constructivists also place emphasis on evaluators understanding their own positionality and values. However, transformative evaluators have an explicit mandate to take this a step further by working to transform the status quo (Ponterotto 2005). The American Evaluation Association (AEA) revised its guiding principles in 2004 to include an explicit statement about the importance of recognizing diversity and acting in an ethically responsible and culturally competent manner. In 2011, AEA published a *Statement on Cultural Competence* that calls upon evaluators to engage in a constant state of learning in order to guard against being blinded by their own assumptions that differ from those of the stakeholders whose backgrounds are different from their own. "Cultural competence requires awareness of self, reflection on one's own cultural position, awareness of others' positions, and the ability to interact genuinely and respectfully with others" (<http://www.eval.org/p/cm/ld/fid=92>).

The transformative ontological assumption recognizes the multi-faceted nature of reality. Human beings often believe that they know what is real, but each concept of what is real is influenced by the positionality of the person. A person who is in a position of unearned privilege by virtue of skin color, gender, or lack of a disability might hold one version of reality. However, a person who is not in that privileged

position may hold quite a different version of reality. The evaluator's responsibility is to design studies in ways that make visible the differences in perspectives about what is real, the factors that influence those perceptions (e.g., poverty, education, gender, race/ethnicity, religion), and then critically examine the consequences of accepting one version of reality over another. History is replete with examples of the acceptance of the privileged views of reality and the harmful consequences of that action. Native American Indians were taken from their families and forced to relinquish all aspects of their culture in the US government's attempt to "civilize" them, resulting in high suicide rates and drug abuse. Similarly, Aboriginal Australians were also taken from their homes and forced to live in boarding schools with disastrous results that are still being felt decades later in the form of cultural disintegration. This concept of ontology comes into play in evaluation work when evaluators encourage stakeholders to critically examine their own assumptions about the target population and the interventions and to obtain data from the targeted population on these topics as well.

Epistemologically, knowledge is not viewed as absolute nor relative; it is created within a context of power and privilege. Evaluators need to develop respectful and collaborative relationships that are culturally responsive to the needs of the various stakeholder groups in order to establish conditions conducive to revealing knowledge from different positions. Tensions can arise because of the power differences and the challenge of working through sensitive issues related to discrimination and oppression. The evaluator needs to develop effective communication strategies in order to navigate the inherently political terrain of an evaluation study.

The transformative methodological assumption does not dictate any particular approach to evaluation. "Rather, methodological decisions are aimed at determining the approach that will best facilitate use of the process and findings to enhance social justice; identify the systemic forces that support the status quo and those that will allow change to happen; and acknowledge the need for a critical and reflexive relationship between the evaluator and the stakeholders" (Mertens & Wilson 2012, p. 172). Mixed methods are often used in transformative evaluations because of the need to establish a dialogic relationship and a deep contextual understanding. Qualitative and quantitative methods can be used together because they reveal different aspects of the phenomenon under study and are responsive to different information needs of the various stakeholder groups. "The methods used need to capture the contextual complexity and be appropriate to the cultural groups in the evaluation. A cyclical design can be used to make use of interim findings throughout the evaluation study. And follow-up is needed to facilitate use to enhance the potential for the program evaluation findings to achieve the strengthening of human rights" (Mertens 2013, p. 33). This supports the credibility of findings because the stakeholders are engaged throughout the process and their perspectives are reflected in respectful ways.

### *Conclusions*

As a transdiscipline, evaluation's pathway is complex and fraught with challenges. Evaluation's inherent political nature means that the assumptions for research need to

be critically examined in order to understand their applicability to the evaluation context. The plurality of philosophical paradigms offers opportunity for exploration of ways to understand the assumptions that guide evaluators and the consequences of accepting one set of assumptions over another.

Increased attention to mixed methods has led to increased discussion of how adherents of each paradigm could incorporate mixed methods into their practice (Mertens & Hesse Biber 2013). Claims about program effectiveness can be bolstered by having multiple forms of evidence. However, a definitive claim of causality in the social world is not possible because there is always a margin of error and competing explanations. There are also concerns about who used their power to decide what the intervention should be, who should be included in the program, how they will be recruited and supported, who will implement the program, what kind of changes are made throughout the course of the program, and what kinds of data were collected by what methods. A final important element is who interprets the data and who has the power to use the data to make changes. These elements that are the heart of evaluation raise issues of power, representation, and interpretation and thus inherently lead to questions about ethics in evaluation. Continued exploration of the philosophical underpinnings of evaluation can be fruitfully conducted by integrating consideration of axiological, ontological, epistemological and methodological assumptions. Hopefully, these discussions will advance understandings of how evaluators can work with diverse stakeholders in a variety of contexts, build productive and positive linkages with policy makers, and develop and refine new strategies for planning, implementing, and using evaluation.

### References

- AMERICAN EVALUATION ASSOCIATION (2004), *Guiding Principles for Evaluators*, <http://www.eval.org/p/cm/ld/fid=51>.
- AMERICAN EVALUATION ASSOCIATION (2011), *Public Statement on Cultural Competence in evaluation*, <http://www.eval.org/p/cm/ld/fid=92>.
- M.K. ASANTE (1992), *Kemet, Afrocentricity and Knowledge*, Africa World Press, Trenton NJ 1992.
- B. CHILISA (2011), *Indigenous Methodologies*, Sage, Thousand Oaks CA 2011.
- C.G. CHRISTIANS (2005), *Ethics and Politics in Qualitative Research*, in N. DENZIN-Y.S. LINCOLN (eds.), *Handbook of Qualitative Research Methods*, Sage, Thousand Oaks CA 2005<sup>3</sup>, pp. 139-164.
- C.A. CHRISTIE & M. ALKIN (2013), *An Evaluation Theory Tree*, in M. ALKIN (ed.), *Evaluation Roots*, Sage, Thousand Oaks CA 2013, pp. 11-58.
- F. CRAM (2009), *Maintaining Indigenous Voices*, in D.M. MERTENS-P.E. GINSBERG (eds.), *Handbook of social research ethics*, Sage, Thousand Oaks CA 2009, pp. 308-322.
- N. DENZIN (2012), *Triangulation 2.0*, in “Journal of Mixed Methods Research”, 6 (2/2012), pp. 80-88.

- J. DEWEY (1923/1998), *The Development of American Pragmatism*, in L.A. HICKMAN-T.M. ALEXANDER (eds.), *The Essential Dewey: vol. 1. Pragmatism, education, democracy*, Indiana University Press, Bloomington IN 1998, pp. 3-13 (original work published in 1923).
- P. FREIRE (1970), *Pedagogy of the Oppressed*, Seabury, New York NY 1970.
  - J. GREENE (2007), *Mixed Methods in Social Inquiry*, Wiley, San Francisco CA 2007.
  - E. GUBA-Y.S. LINCOLN (1989), *Fourth Generation Evaluation*, Sage, Newbury Park CA 1989.
  - E. GUBA-Y.S. LINCOLN (2005), *Paradigmatic Controversies, Contradictions, and Emerging Confluences*, in N. DENZIN-Y.S. LINCOLN (eds.), *Handbook of Qualitative Research Methods*, Sage, Thousand Oaks CA 2005<sup>3</sup>, pp. 191-216.
  - J. HABERMAS (1971), *Knowledge and Human Interests*, trans. J. Shapiro, Beacon Press, Boston MA 1971.
  - J.N. HALL (2013), *Pragmatism, Evidence, and Mixed Methods Evaluation*, in D.M. MERTENS-S. HESSE BIBER (eds.), *Mixed Methods and Credibility of Evidence in Evaluation* (“New Directions for Evaluation, 138 (2013)), pp. 15-26.
  - G.W.F. HEGEL (1812/1929), *Science of Logic*, trans. W.H. Johnston and L.G. Struthers, George Allen & Unwin, London 1929 (original work published in 1812).
  - M. HORKHEIMER (1972), *Critical Theory*, ed. and trans. M.J. O’Connell, Seabury Press, New York NY 1972.
  - S. HOOD-R. HOPSON-H. FRIERSON (eds.) (2015), *Continuing the Journey to Reposition Culture and Cultural Context in Evaluation*, Information Age Publishing, Charlotte NC 2015.
  - E. HOUSE (1990), *Research News and Comment: Trends in Evaluation*, in “Educational Researcher”, 19 (1990), pp. 24-28.
  - K.E. HOWELL (2013), *An Introduction to the Philosophy of Methodology*, Sage, London 2013.
  - E. HUSSERL (1936/1970), *The crisis of European sciences and transcendental phenomenology*, trans. D. Carr, Northwestern University Press, Evanston IL 1970 (original work published 1936).
  - I. KANT (1781/1966), *Critique of Pure Reason*, trans. M. Müller, Doubleday, Garden City NY 1966 (original work published 1781).
  - J.L. KINCHELOE-P.MCLAREN (2005), *Rethinking Critical Theory and Qualitative Research*, in N. DENZIN-Y.S. LINCOLN (eds.), *Handbook of Qualitative Research Methods*, Sage, Thousand Oaks CA 2005<sup>3</sup>, pp. 303-342.
  - J. LAFRANCE- C. CRAZY BULL (2009), *Researching Ourselves Back to Life: Taking Control of the Research Agenda in Indian Country*, in D.M. MERTENS-P. GINSBERG (eds.), *Handbook of Social Research Ethics*, Sage, Thousand Oaks CA 2009, pp. 135-149.
  - G.E. MARCUS (1998), *Ethnography Through Thick and Thin*, Princeton University Press, Princeton NJ 1998.
  - D.M. MERTENS (2015). *Research and Evaluation in Education and Psychology*, Sage, Thousand Oaks CA 2015.
  - D.M. MERTENS (2013), *What Does a Transformative Lens Bring to Credible Evidence in Mixed Methods Evaluations?*, in D.M. MERTENS-S. HESSE BIBER (eds.), *Mixed Methods and Credibility of Evidence in Evaluation* (“New Directions for Evaluation, 138 (2013)), pp. 27-35.
  - D.M. MERTENS (2009). *Transformative Research and Evaluation*. NY: Guilford.

- D.M. MERTENS-S. HESSE BIBER (2013), *Mixed Methods and Credibility of Evidence in Evaluation*, in D.M. MERTENS-S. HESSE BIBER (eds.), *Mixed Methods and Credibility of Evidence in Evaluation* (“New Directions for Evaluation, 138 (2013)), pp. 5-13.
- D.M. MERTENS-H. HOLMES-R. HARRIS (2009), *Transformative Research and Ethics*, in D.M. MERTENS-P.E. GINSBERG (eds.), *Handbook of social research ethics*, Sage, Thousand Oaks CA 2009, pp. 85-102.
- D.M. MERTENS-A.T. WILSON (2012), *Program Evaluation Theory and Practice*, Guilford, New York NY 2012.
- D. MORGAN (2007), *Paradigm Lost and Paradigm Regained: Methodological Implications of Combining Quantitative and Qualitative Methods*, in “Journal of Mixed Methods Research”, 1 (2007), pp. 48-76.
- M.Q. PATTON (2010), *Developmental Evaluation*, Guilford, New York NY 2010.
- J.G. PONTEROTTO (2005), *Qualitative Research in Counseling Psychology: A Primer on Research Paradigms and Philosophy of Science*, in “Journal of Counseling Psychology”, 52 (2/2005), pp. 126-136.
- T. SCHWANDT (2000), *Three Epistemological Stances for Qualitative Inquiry: Interpretivism, Hermeneutics, and social Constructionism*, in N. DENZIN-Y.S. LINCOLN (eds.), *Handbook of Qualitative Research Methods*, Sage, Thousand Oaks CA 2000<sup>2</sup>, pp. 189-213.
- M. SCRIVEN (2003), *Evaluation in the New Millennium: The Transdisciplinary Vision*, in S.I. DONALDSON-M. SCRIVEN (eds.), *Evaluating Social Programs and Problems*, Erlbaum, Mahwah NY 2003, pp. 19-41
- W. SHADISH (1998), *Evaluation Theory is Who We Are*, in “American Journal of Evaluation”, 19 (1998), pp. 1-19.
- D. STUFFLEBEAM (1980), *Interview: An EEPA interview with Daniel L. Stufflebeam*, “Educational Evaluation and Policy Analysis”, 2 (1980), pp. 85-90.
- M. SULLIVAN (2009), *Philosophy, Ethics, and the Disability Community*, in D.M. MERTENS-P.E. GINSBERG (eds.), *Handbook of social research ethics*, Sage, Thousand Oaks CA 2009, pp. 69-74.
- A. TASHAKKORI-C. TEDDLIE (2003), *Major Issues and Controversies in the Use of Mixed Methods in the Social and Behavioral Sciences*, in A. TASHAKKORI-C. TEDDLIE (eds.), *Handbook of mixed methods in social and behavioral research*, Sage, Thousand Oaks CA 2003, pp. 3-50.
- J.H. TURNER (2001), *The Origins of Positivism: The Contributions of Auguste Comte and Herbert Spencer*, in G. RITZER-B. SMART (eds.), *Handbook of Social Theory*, Sage, Thousand Oaks CA 2001.
- C WEISS (1998), *Evaluation Research*, Prentice Hall, Upper Saddle River NJ 1998.
- H. WHITE (2013), *The Use of Mixed Methods in Randomized Control Trials*, in D.M. MERTENS-S. HESSE BIBER (eds.), *Mixed Methods and Credibility of Evidence in Evaluation* (“New Directions for Evaluation, 138 (2013)), pp. 61-73.



Jennifer C. Greene

## EVALUATION AS A SOCIO-POLITICAL INTERVENTION

### Abstract

*The logic of evaluation is presented and examined, resulting in the claim that evaluation (of public policies and programs) itself is an intervention in the public sphere. Evaluation accomplishes this primarily via the advancement of selected stakeholder interests and values, represented most evidently in the evaluation's purpose, audience, and key questions, which in turn, structure the quality criteria upon which judgments of program quality and effectiveness are made. Illustrative examples are offered.*

In societies all around the globe, the contemporary practice of evaluation powerfully influences media headlines, political debates, and, most consequentially, actual social and economic policies and programs. Evaluation data on politically-contested programs are used, often by all sides in the debate, as fodder for public arguments about policy directions and program parameters. For programs of significant public interest or consequence (for example, an experimental program in health care subsidies for the poor), evaluation results capture national headlines. Interestingly, challenges to and debates about evaluation itself – its questions, designs, methodologies, and especially its values – rarely enter the public debate<sup>1</sup>. This is so, even though evaluation is far from a homogenous social practice, *and* the character of any particular evaluation study significantly shapes, even constitutes in part, the results obtained.

In this essay, I will take up the challenges of evaluation itself as an intervention in public policy and program decision making<sup>2</sup>. I will argue that the presence of an evaluation in the spaces occupied by a public program influences that program in particular ways, depending on the character of the evaluation. So, what is evaluated is a program-with-an-evaluation, rather than a program pursuing a course void of evaluative influences. For some evaluation approaches, the influences on the program are intentional and for others, more of an unintended and unobserved side-effect. The argument will begin with a general discussion of the logic of evaluation and how this logic constitutes an intervention, followed by a discussion of how the character of this

---

<sup>1</sup> The National Education Policy Center at the University of Colorado at Boulder (<http://nepc.colorado.edu/>) is a rare exception, although the public reach of the work of this center is not known.

<sup>2</sup> This essay concentrates on social-economic programs, mostly in the public sphere, that are designed to support people in need. Programs designed for the well-off or the wealthy, for example tax codes, are much more rarely evaluated (Datta-Grasso 1998) and their evaluations may have different kinds of influences.

intervention is connected to the evaluation approach being used. These discussions will focus on perhaps the most powerful lever of influence in evaluation, that of the values being advanced by the inquiry. In discussing evaluation's values, I will somewhat summarily assert that democratic values are the most defensible of evaluation's varied portfolio of value stances. Finally, I will illustrate the arguments made by offering examples of evaluation in action.

I offer these thoughts as a lifetime "theorist" and practitioner of program evaluation in the United States. My evaluation practice has focused largely on formal educational programs for children and youth, and has also included evaluations of non-formal educational programs (for example, summer camps or after-school programs) and a smattering of evaluations of programs in other domains.

### 1. *On Evaluation as a Values-Engaged Intervention*

This part of the argument first establishes the values parameters of evaluation and then turns to how these values parameters influentially serve to position evaluation itself as an intervention.

#### *Quality criteria in evaluation as conveyers of values*

Evaluation is the systematic assessment of the operations and/or outcomes of a program or policy, compared to a set of explicit or implicit standards, as a means of contributing to the improvement of the program or policy. (Weiss 1998, p. 4)

This definition of evaluation by the late Carol Weiss is commonly cited, both because it is relatively broad and inclusive *and* because it explicitly states that evaluation involves standards and thereby value judgments. The core of the evaluative enterprise, captured in this definition, is the comparison of empirical data collected about a given program (its underlying theory, context, operations, and/or outcomes) to established standards or criteria that define what constitutes program quality in that evaluation setting. Interestingly, few evaluation proposals or reports explicitly articulate and justify the quality criteria being used. Especially so, a key question becomes, where do these program standards or quality criteria come from, and just how are they determined?

In brief, standards or criteria for judging program quality are embedded in the logic of the particular evaluation approach being employed in that context. Table 1 outlines the generic logic of evaluation. This outline presents core elements of evaluation (in theory and in practice), beginning with the program/policy and evaluation contexts; followed by the evaluation's primary purposes, intended audiences, key questions, and the evaluation approach to be used<sup>3</sup>; and then the criteria to be used to judge program quality; the evaluation methodology; plans for communication and utilization; and finally, meta-evaluative criteria. Whether explicitly stated or not, these elements describe evaluation's basic logic. And, as presented in the next section, different constellations of

---

<sup>3</sup> Different evaluation approaches are differentially well suited to address different evaluation purposes, audiences, and key questions.

these elements describe different approaches to evaluation, and approaches gain coherence with strong and logical connections among their distinct elements.

Table 1  
A Framework for Evaluative Logic

<b>Evaluation element</b>	<b>Description</b>
Program and context	Description of the program to be evaluated (needs, goals/objectives, activities, staffing) and of the organization(s) involved; the settings in which the program is being implemented; the program's policy context; and particular politics of relevance.
Evaluation context	The impetus for the evaluation – who wants it and why? Who are the “automatic” evaluation audiences? What politics of importance are related to the request for the evaluation?
Purpose	The reasons the evaluation is being conducted. Stated reasons include to improve the program, to contribute to policy decisions, to generate knowledge, to provide accountability for public funding, to promote organizational growth or change, to educate consumers, to empower participants, to catalyze social change. Unstated reasons include to politically support or undermine the program, to stall for time, to damage a political opponent.
Audience	Which stakeholders' needs for information and evaluation questions are being addressed in the evaluation? Stakeholders include funders, policy and other decision makers, program developers, program managers, onsite administrators, staff, program participants and their families and communities, interested citizenry, the media.
Questions	<p>The particular questions the evaluation will address; evaluation questions are generally framed by purpose and audience, but require further specification.</p> <p>Sample evaluation questions for a <i>process evaluation</i> of a science education program are:</p> <ul style="list-style-type: none"> <li>• How well does the curriculum engage students of varying learning styles and achievement histories?</li> <li>• How well does the professional development training prepare teachers to implement the program effectively?</li> </ul>
Approach	The evaluation approach (or approaches) to be used in the evaluation and the rationale for this approach (or approaches). Extant evaluation approaches include decision-oriented, responsive, educative, utilization-focused, accountability-oriented, democratic, and praxis-oriented evaluation.
Judging program quality	The criteria or standards to be used to make judgments of program quality, the justification for these criteria, and the process to be used for making such judgments, including who is to be involved

Design and methods	The overall methodological design to be used in the evaluation, and the specific data collection and analysis methods that will be used within this design. The criteria or standards to be used to make judgments of program quality, the justification for these criteria, and the process to be used for making such judgments, including who is to be involved.
Communication and reporting	How the evaluation team will communicate and report to key clients and other audiences, both during the evaluation and at the end of the study.
Utilization	Intended uses of the evaluation process and results. Categories of common evaluation uses are instrumental, conceptual, symbolic, and political use.
Meta-evaluation	The criteria or standards to be used to make judgments of evaluation quality, and the process for making such judgments, including who is to be involved.

Collectively, the other elements in this evaluative logic serve to determine the element representing the standards or criteria for judging program quality. Most centrally, these criteria are determined by the evaluative elements that directly engage the political and values-laden evaluative issues of whose interests are being addressed by the evaluation, namely, what is the purpose of the evaluation, who is it for, and what specific questions will the evaluation address<sup>4</sup>. Different combinations of these elements invoke different criteria for judging quality. For example, an evaluation conducted to inform a policy decision about whether to reauthorize and refund a given program clearly addresses the outcomes-oriented information needs of policy makers. In this context, a good or high quality program is one that meets its intended outcomes. As another example, an evaluation conducted to better understand how students from different socio-demographic groups respond to an innovative technology curriculum addresses the educational information needs of program designers, and likely teachers as well. In this context, a good program is one that serves the distinct learning profiles of students from multiple socio-demographic groups equally well.

So, the values advanced by a given evaluation are most readily apparent in the criteria developed to judge program quality, which, in turn, are most directly influenced by evaluation purpose, audience, and key questions. Specific quality criteria in evaluation are drawn from varied sources, which include: (a) stated policy goals or program objectives; (b) facets of the program's theory; (c) implicit goals and objectives, unstated assumptions; (d) relevant theory and research; (e) salient dimensions of the context (for example, political or cultural factors); and (f) key stakeholder or evaluator commitments. Again, because quality criteria themselves are typically not stated, neither are the sources from which they are drawn.

<sup>4</sup> Renowned American evaluation theorists Shadish, Cook, and Leviton (2001) include valuing as one of the five core components of evaluation. The other four components are knowledge, use, practice, and social programming. And international evaluation theorist Michael Scriven presents the core logic of evaluation as centered on judgments of quality based on comparisons of empirical results to established standards (<http://michaelscriven.info/>).

*Evaluation as an intervention*

So, in what ways is evaluation itself an intervention and what role do the value dimensions of evaluation play in this conceptualization of the evaluation enterprise?

Broadly, *evaluation perturbs the program and context being evaluated by infusing (some form of) evaluative thinking and perspectives into ongoing program conversations and reflections*. And a competent and thoughtful evaluator, of whatever theoretical persuasion, will do this with purpose and intention, aiming for influences that enhance the aims and objectives of the particular evaluation approach being implemented and that advance the values of his/her preferred approach. These conversations and reflections, then, afford ongoing opportunities, and in some cases internal or external pressures, for stakeholders to rethink selected program aspirations, to reconsider who should be eligible for program participation, or even to revise a troublesome component of the program as it is being implemented. In these ways, the evaluation can function as an intervention in program design, implementation, and aspiration.

While there are likely multiple specific ways in which evaluation influences the program being evaluated and its context(s), I offer the following three, all fully interrelated, and all of which directly engage evaluation's values dimensions. As elaborated in the next section, the particular character of these three evaluative influences depends on the evaluation approach being implemented. First, evaluation influences *who participates* in ongoing program conversations. Representatives of the targeted evaluation audiences are usually the stakeholders identified for ongoing consultation and conversation about the program and its evaluation. And such audiences can range from policymakers to advocacy groups to intended program beneficiaries. In some evaluations, one or more stakeholder advisory boards are convened for just this purpose. Second, evaluation influences *the substance or content* of the ongoing conversations. That is, the ongoing evaluative conversations can draw attention to particular issues of importance, including issues overlooked in the program's vision. One example here is how well the program is reaching all types of eligible beneficiaries. Another concerns the cultural and linguistic appropriateness of the program's design and implementation for diverse kinds of participants. Third, evaluation influences the *relational fabric* of the program's administration, oversight, and implementation. All evaluators interact in some ways and establish some kind of relationships with some stakeholders. The valence and character of these interactions, also viewed as the social relations of evaluation (Abma 2006), matter. They can modestly influence particular program components, as well as more significantly shape the overall ambience of the program's context of oversight and delivery, and thereby the consequent relevance and acceptance of the evaluation results.

A brief example can illustrate the evaluator's power and obligation to exert these influences responsibly. The example comes from an award-winning evaluation (Brandon, Smith, Trenholm, & Devaney 2010), commended for its high methodological quality, its high utility in influencing policy, *and* its purposeful, fair-minded engagement with stakeholder concerns and values, on all sides of a highly emotional and contested issue. This US evaluation was of four promising abstinence education programs, all designed to reduce teen pregnancy by teaching youth about the benefits of abstinence

(“just say no”) and the risks of sexual activity before marriage. Politically, abstinence is promoted by conservatives while liberals favor birth control alternatives or abortion. At times, this debate can get very heated, as beliefs about these issues run deep.

The evaluation, funded by the US government, was intended to address policymakers’ questions about the impact of abstinence education on sexual activity to aid future policy making. As preferred by the policy audience, the evaluators designed an experimental impact study to assess the effects of the four programs on key behavioral outcomes, including sexual abstinence, risks of pregnancy, and incidence of sexually-transmitted diseases. Survey data were collected in four waves over a 10-year period. The results found no evidence that abstinence education had increased rates of sexual abstinence or in any way affected sexual risks among participating youth.

In the context of the present argument about evaluation as a values-engaged intervention, what was most significant about this evaluation was how the evaluation team intentionally worked to foster ongoing discussion among holders of competing values and standpoints in this highly charged political context of abstinence education, keeping both detractors and advocates respectfully engaged in the evaluation throughout the long 10-year period. The evaluators chose not to ignore the politics of the evaluation – as is customary in experimental work – but rather to engage them head-on. Specifically, the evaluators inclusively chose to allocate time and resources to listening and dialoguing with stakeholders who held opposing views on abstinence. The evaluators established a national technical advisory group, comprised of technical experts with diverse stances on abstinence, and consulted with them on multiple technical issues throughout the evaluation. And they used a parallel process to engage diverse program stakeholders at each local level, again demonstrating respect for program staff. In a report on the evaluation process the evaluators said, “You need to understand the program from the local stakeholders’ perspective and why they think the program is worthy and how they implement it. And you need to get their buy-in so they believe the study is credible.”

In this study, the evaluators primarily addressed the first evaluative influence presented above, that of fostering reflective and critical conversation among an intentionally diverse set of stakeholders, instead of just with designated policymakers and/or key program staff. This ongoing conversation influenced some measurement decisions, enhanced participant understanding of the specific nature of the four abstinence programs, and significantly increased the credibility of the evaluation results. The evaluation, that is, constituted a substantial intervention in the policy and program context that surrounded the abstinence debate at that time.

## *2. The Multiple Countenances of Evaluation as an Intervention*

In this section, I will present brief intervention-and-values profiles for two contrasting approaches to evaluation from among five broad families of evaluation approaches<sup>5</sup>.

---

<sup>5</sup> These families are differentiated by their major foci: (1) outcomes and policy, (2) performance and accountability, (3) understanding and learning about the social problem at hand; (4) deep contextual

These profiles follow the general evaluation logic of Table 1, and are intended to further illustrate the argument made above regarding the values-based character of evaluation as an intervention. The two approaches are (a) the outcomes and policy-oriented evaluation approach, arguably the most common evaluation in our public sectors, and (b) the democratic evaluation approach, which most explicitly promotes particular values – specifically, equity, fairness, and justice – throughout its practical enactment. The context for both profiles will be an evaluation of an innovative web-based curricula and resources for students and teachers in middle school science (for children ages 11-14). The program, *Science for All*, has been thoroughly field-tested and revised and is now being implemented on a pilot basis for two years in three school districts (urban, suburban, and rural) in one state. The evaluation is being conducted by a well-established evaluation center at a state university. Both the program and the evaluation are funded by state education dollars.

*Policy-oriented evaluation: Intervention as “business as usual”*

The policy-oriented evaluation team would view state education policy makers as the primary audience for the evaluation, as they are the funders and decision makers in this context. Likely decisions resulting from this evaluation (which also constitute the primary intended evaluation uses) include continuation of the program and possible expansion to other schools and districts. Secondary audiences would be the administrators and teachers in the pilot schools. The team would view the main evaluation purpose as assessing how well the program “works” or how well it accomplishes its intended outcomes of science learning. Though not likely stated as such, attainment of intended outcomes would also constitute the primary criteria for judging program quality and effectiveness. In turn, key evaluation questions would focus on outcome attainment, likely in terms of average scores. Resources permitting, evaluation questions about teacher (and student) experiences with the *Science for All* program may also be posed. Consonant with these evaluative decisions, an experimental or quasi-experimental methodology would likely be selected, as the experimental counterfactual (what happens in similar contexts without the program) can provide strong evidence on the attribution of observed outcomes to the program being evaluated. An advisory board of science education and experimental evaluation experts may be convened. The evaluators would likely provide regular progress reports (written and oral) to this board and to the key evaluation audiences.

In terms of the three evaluation-as-intervention influences on the program and its context discussed above, in this policy-oriented evaluation, (1) participants in evaluative conversations about the program are primarily existing decision makers, (2) these conversations are focused on the program as designed and especially intended outcomes, and (3) the relationships established via the evaluation do not likely challenge extant hierarchies, boundaries, or norms. Therefore, the values advanced in this evaluation include support for the decision making status quo, a valuing of ends (learning outcomes) over means (learning experiences), a utilitarian emphasis on average effects,

---

and socio-cultural understanding of the problem at hand; and (5) democratization and socio-political critique.

and methodological objectivity and distance. This evaluation as intervention would not disturb the status quo.

*Democratic evaluation: Intervention as disturbance or disquiet*

The democratic evaluation team would design an evaluation study and communicative process that are significantly different from the familiar policy-oriented evaluation described just above. Following the democratic ideas of Ernest House (House 2014; House-Howe 1999), these evaluators would seek to equitably include representatives from *all* important stakeholder groups as key audiences for the evaluation – from decision makers and administrators to science education experts and teachers, and also to students participating in the program and their parents, along with relevant community and media representatives. All audiences would be consulted regarding their priorities for evaluation questions, and respectful dialogues among the various stakeholder audiences would be encouraged as vehicles for learning about each other's program standpoints and perspectives. Advisory boards comprising representative stakeholders may be established. The evaluation's purposes would include both substantive learning about the quality and effectiveness of the *Science for All* program, as disaggregated by relevant student sub-groups, and advancement of democratic values of equity, inclusion, and social justice in the contexts at hand. Relatedly, key evaluation questions would ask how well the program serves students from various socio-cultural and demographic groups, and in particular, how well the program serves students who are least well served in the relevant contexts. Key evaluation questions would also engage both the quality of the learning experience and the magnitude and contextual importance of the learning outcomes. And criteria for judging program quality would focus on equity and fairness for all students, in addition to educational soundness in access to program opportunities, experiences, and outcomes. A variety of methodologies are likely to be employed, to gather data relevant to the full range of evaluation questions. And the evaluation team would communicate and consult with stakeholder groups throughout the evaluation process (orally and in writing), keeping stakeholders informed of the evaluation's progress and seeking their input on key evaluation decisions.

In terms of the three evaluation-as-intervention influences on the program and its context discussed above, in this democratic evaluation, (1) participants in evaluative conversations about the program span the full range of program stakeholders; (2) these conversations are focused on the programmatic and educational interests and concerns of various stakeholders; and (3) the relationships established in the evaluation aspire to be respectful and dialogic. Therefore, the values advanced in this evaluation include inclusion, equity, and educational quality in program access, experience, and outcome for all students, especially those under-served. This evaluation as intervention would likely disturb the status quo, in ways necessary for our societies to reach their full democratic ideals.

### 3. Reprise

Not all readers will agree with my support for a democratic approach to evaluation, as is expected and desired in our pluralistic societies. Even so, I still encourage all readers to recognize the interventionist strands of their own work as evaluators; to reconfigure these strands as necessary to enact a defensible evaluation practice (from each reader's point of view); and also to name, claim, and justify the values that are advanced by our respective evaluation practices.

### References

- T.A. ABMA (2006), *The Social Relations of Evaluation*, in I.F. SHAW-J.C. GREENE-M.M. MARK (eds.), *The Sage Handbook of Evaluation*, Sage, London 2006, pp. 184-199.
- P.R. BRANDON-N.L. SMITH-C. TRENHOLM-B. DEVANEY (2010), *Evaluation Exemplar: The Critical Importance of Stakeholder Relations in a National, Experimental Abstinence Education Evaluation*, in "American Journal of Evaluation", 31 (4/2010), pp. 517-531.
- L.-E. DATTA-P.G. GRASSO (eds.) (1998), *Evaluating Tax Expenditures: Tools and Techniques for Assessing Outcomes* ("New Directions for Evaluation", 79 (1998)).
- E.R. HOUSE (2014), *Evaluating: Values, Biases, and Practical Wisdom*, Information Age Publishing, Charlotte NC 2014.
- E.R. HOUSE-K. HOWE (1999), *Values in Evaluation and Social Research*, Sage, Thousand Oaks CA 1999.
- W.R. SHADISH-T.D. COOK-L.C. LEVITON (2001), *Foundations of Program Evaluation: Theories of Practice*, Sage, Thousand Oaks CA 2001.
- C.H. WEISS (1998), *Evaluation*, Prentice Hall Inc., Upper Saddle River NJ 1998<sup>2</sup>.



Giulio Sapelli

BREVE PENSIERO SULLA VALUTAZIONE  
DI UN ECONOMISTA ROUSSOIANO

**Abstract**

*The essay highlights the close connection that is in place between a specific mode of evaluation and a specific economic approach, namely the neo-classic or marginalist approach; both are incapable of understanding the historical and social singularities and are responsible for clamorous interpretative and predictive failures.*

«Nos jugemens sont encores malades, et suyvent la depravation de nos meurs. Je voy la pluspart des esprits de mon temps faire les ingenieux à obscurcir la gloire des belle et genereuses actions anciennes, leur donnant quelque interpretation vile et leur controuvant des occasions et des causes vaines.

Grande subtilité! Qu'on me donne l'action la plus excellente et pure, je m'en vois y fournir vraysemblablement cinquante vitieuses intentions. Dieu sçait, à qui les veut estendre, quelle diversité d'images ne souffre nostre interne volonté! Ils ne font pas tant malicieusement que lourdement et grossierement les ingenieux à tout leur medisance».

(M. DE MONTAIGNE, *Essais*, I, XXXVII: *Du jeune Caton*, in ID., *Oeuvres completes*, Gallimard, Paris 1962, p. 226).

Perché si valuta? Come si valuta? E che cosa vuol dire “valutare”? A queste tre domande non solo si può ma si deve rispondere insieme, perché le questioni che esse pongono sono tutte intimamente legate in un ordito finissimo e fittissimo che brevemente cercherò di dipanare.

Cominciamo col distinguere tra valutazione ontologica e valutazione teleologica. La prima ha di mira la verità, ovvero ciò che i valutatori credono che la verità sia. Una verità che può essere anche senso comune e che va quindi definita in senso antropologico come *mores*, costumi, opinioni. Questo concetto ontologico di verità richiama quindi al

sistema di valori e, parsonianamente, alle latenze culturali che fondano l'ordine. Fin dal tempo di Durkheim e della sua opera *La divisione sociale del lavoro*, emerse chiaramente che il problema vero delle società non è tanto il conflitto quanto l'ordine. L'ordine che consente alle società di stare insieme e di non frantumarsi. E in questo senso, dunque, valutare un comportamento, un testo, uno stile di vita richiama alla conformità o al conformismo inestricabile dalla *Gesellschaft* e non solo dalla *Gemeinschaft*. In quest'ultima la verità valutata è immediata, già data, è immanente all'essere sociale e la solitudine anticonformistica non è ammessa. Nella *Gesellschaft* la conformità che si chiama verità serve per fondare il sistema di ruoli, perché i ruoli sociali altro non sono che il fascio di aspettative che si generano nei confronti di chi di quei ruoli è al comando. In questo senso la valutazione può anche essere uno strumento fondamentale per attribuire, donare, conservare sistemi di *status* e abbassare il grado dei vari conflitti di *status*, tema classico della sociologia generale novecentesca.

Naturalmente esiste anche un altro concetto ontologico di valutazione riferito alla verità. E mi riferisco alla verità che rifiuta la "civiltà delle buone maniere" perché la reputa ipocrita, convenzionale, così come ci insegna la vita di Rousseau, il suo rifiuto dei salotti parigini, il suo passeggiare solitario che fonda con Jean Paul ben più di uno stile di vita. Fonda un modello di sottrazione alla valutazione, forma altissima di verità se pensiamo che sempre la verità si raggiunge per sottrazione, mai per addizione. Questi non son tempi di passeggiate solitarie. Oggi alle gabbie d'acciaio profetizzate da Max Weber si è aggiunta la valutazione. Valutazione che non è più soltanto l'opinione della società delle buone maniere ma è uno strumento di potere e dunque di allocazione delle risorse che fondano l'ordinamento societario.

Questa riflessione richiama al secondo aspetto fondamentale della valutazione, ossia quello teleologico. Certo anch'esso ha dei modelli di riferimento, quindi si sovrappone ed è *embedded* nei sistemi di potere, ma nasconde questo suo essere con una retorica ordinativa rispetto al fine. I mezzi devono essere adeguati ai fini e, una volta scontato che si sia assestato in modo più o meno stabile un orizzonte dei fini, la valutazione tutta s'esercita sui mezzi. Questo comporsi teleologico della valutazione richiama tre questioni fondamentali.

La prima è chi decide quale sia il fine. E quindi ogni decisione richiama al problema e all'assetto del potere, e alla legge dell'oligarchia per cui è sempre una minoranza organizzata che domina una maggioranza disorganizzata. La seconda questione è quella di come si decide la congruità del mezzo rispetto al fine. E qui forse la sottolineatura va fatta sull'invadenza sempre più ossessiva della valutazione di questo *trade off* con strumenti, metodi, financo ideologie, di tipo nomotetico e non idiografico, quantitativo anziché qualitativo, misurabile anziché incommensurabile. E questo richiama naturalmente a un concetto di efficacia e di efficienza che condiziona (un'altra gabbia d'acciaio) tutti i prodotti della vita umana associata e individuale valutata appunto attraverso un concetto di verità che richiama alla razionalità strumentale anziché alla verità sapienziale, che per sua intima natura misurabile non è mai.

E infine c'è la terza questione, il prevalere di una valutazione dei mezzi separata dal fine. Separazione che esalta l'approccio nomotetico e quantitativo, che matematizza la vita senza più trovare nel numero pascalianamente l'esistenza di Dio, ma invece esaltando un nuovo paganesimo surrettizio che riduce comportamenti, valori, sistemi di

vita, addirittura simboli archetipali, a razionalità matematizzate, proceduralizzate ed esaltate grazie al potente mezzo tecnico che consente di ridurre tutto, della vita, a dato statistico.

Quello che ho fin qui detto, di fatto, è una metafora della valutazione *mainstream* degli e negli studi economici. Sono tempi, questi, in cui gli economisti neoclassici girano ubriachi in bicicletta valutando e allocando disgrazie in ogni minuto secondo, convinti che nessuno li valuterà mai. Ed è forse questa la forma più arrogante della valutazione neoclassica (per dirla in gergo economico). A questo punto si potrebbero reclutare schiere di volontari in grado di allocare pertinentemente al discorso sin qui fatto in termini molto generali concetti che nominano cose e realtà come concorsi universitari, dipartimenti universitari, cattedre universitarie, *papers* universitari e di banche d'affari (le monografie son quasi finite perché richiedono qualcosa che non si valuta più, cioè la cultura alta o generale), e tutto ciò che può essere assimilato al sistema educativo e allocativo dominante tanto nelle agenzie di formazione dei ruoli sociali quanto nelle popolazioni organizzative dei quasi-mercati imperfetti.

In definitiva, allorché ci si affaccia dal mezzo al fine e dall'analisi della forma, per dirla con Lukács, a quella dell'anima, la valutazione apre dinanzi a noi il problema del potere che diviene formula per affermare un senso comune che matematizzando il divenire del sapere lo riduce a forma più consona alla reificazione consustanziale all'ultimo capitalismo. In esso si è perduto ogni pensiero dell'essere sociale che sia relativo alla dimensione idiografica e quindi alle "scienze dello spirito", per dirla con lo storicismo tedesco, che possano in tal modo riempire di senso la vita dei soggetti.

La valutazione oggi imperante opera potentemente per la riduzione del soggetto a individuo e quindi a mezzo scambiabile nel processo di reificazione capitalistica della produzione di merci per mezzo di merci.

Si presenta come essenza dell'a-valutazione predittiva e ontologica mentre, invece, ne è l'incarnazione apocalittica in senso de-possessivo e de-privativo della soggettività...

Tale valutazione imperante oggi scarnifica la stessa relazione dell'essere con il passato e la cultura delle società naturali. Essa informa – ancora e sempre – il presente dei costrutti economici dominati, in effetti, dalle relazioni personali anziché dai ruoli. Ecco il mondo con i piedi in terra e a testa in su che non si vuol vedere... Ecco il differire delle imprese – grandi, medie, piccole e piccolissime ch'esse siano – dalle ipostatizzazioni neoclassiche riduzionisticamente economicistiche: avevano insegnato tale distinzione gli scritti di Le Play nell'Ottocento a coloro che sapevano cogliere – nella divisione sociale del lavoro – il volto dell'associarsi umano, e nel primo terribile Novecento l'aveva insegnato, per insegnarlo sino a oggi, il grande Chajanov, senza il magistero del quale nulla potremmo comprendere delle imprese modellate sulle società naturali e da esse governate nella loro vita nei quasi-mercati in cui agiscono.

Se tutto questo è vero – e lo è – ben s'intende, allora, perché un approccio neoclassico in economia, l'economia come disciplina, sia consustanzialmente estraneo al dibattito sulle forme della vita e non possa che condurre a fallimenti interpretativi deprecabili dal punto di vista del dovere deontologico della ricerca della verità oltreché alla clonazione di personalità autoritarie che sono un grande pericolo per la stessa riproduzione sociale.

Dalla merce non nasce mai la società. Tutto ciò deriva non solo dalla dominazione ideologica del *mainstream* marginalistico, ma altresì dall'ignoranza oggi dilagante nelle

università, in primo luogo nelle discipline economiche. Un'ignoranza che si è diffusa in tutto il mondo unitamente a un riduzionismo economicistico che ha effetti esilaranti e devastanti.

Insomma, ciò che Allan Bloom scriveva nel 1987 nella sua straordinaria e profetica opera, *The Closing of the American Mind*, e ch'egli riferiva principalmente agli studenti, osservando – in un tipico esercizio di sociologia comprendente – la decadenza del sistema educativo nordamericano, oggi potrebbe ben riferirsi non più agli studenti, quanto, invece, ai professori universitari in primo luogo e a quelli di economia in primissimo luogo, oltreché agli operatori managerialisti di questo sempre più potente settore della divisione sociale del lavoro.

In questo senso il discorrere sulla valutazione ci conduce inevitabilmente al destino sociale, ossia al ruolo che gli studi e la formazione umanistica avranno nel divenire storico. È difficile pensare a una riproduzione della società civilizzata (e la civilizzazione non è la cultura antropologicamente intesa...), per esempio, senza che i classici continuino a “parlare” agli esseri umani, oppure allorché il destino della società sia inteso come fondato sul predominio dell'economico anziché su quello del sapere, come in effetti, invece, è.

Ma la valutazione oggi imperante, ecco l'effetto diabolico, mette il mondo a testa in giù e pone quindi il valore personale in guisa di strumento allocativo anziché in guisa di percorso formativo, ponendo in pericolo, ripeto, lo stesso riprodursi dell'umanità come forma vivente associata.

Il futuro sarà quello, forse, dei percorsi non incrociantsi dei pensatori solitari.

Un destino senza dubbio migliore del vivere a testa in giù.

# PRATICHE



Patrizia Lemma

## INTERVENTI DI PROMOZIONE DELLA SALUTE: UNA VALUTAZIONE APPROPRIATA

### **Abstract**

*This paper examines some of the issues underlying the notion of “appropriateness” in the context of health promotion evaluation. Appropriateness depends on an alignment of evaluation practices with the tenets of contemporary health promotion, including the notions of participation, community control, and respect for people not as unthinking objects of research but as partners in knowledge development. This means characterizing health promotion programmes as a process of social change rather than a “dose of treatment.”*

Negli Stati Uniti della seconda metà degli anni '60 la valutazione diviene ufficialmente un passaggio portante nella progettazione dei programmi di promozione della salute. È Donald T. Campbell, autore di un fondamentale testo che discute dell'applicazione del metodo sperimentale nella ricerca sociale (cfr. Campbell-Stanley 1963), ad affermare che da quel momento i programmi d'intervento sarebbero stati disegnati per rispondere a ben definiti problemi e poi scientificamente valutati attraverso sperimentazioni controllate, in modo da eliminare quelli che fossero falliti, per replicare solo quelli che avessero dimostrato di avere ottenuto successi (1969). Questo rassicurante scenario si è però sbiadito con il passare del tempo e in un recente documento dell'Organizzazione mondiale della salute (WHO 2009) un gruppo di esperti, individuato tra i partecipanti alla settima Conferenza Internazionale di promozione della salute di Nairobi, definendo i punti che devono essere sviluppati in ciascun piano che si proponga di implementare azioni di promozione della salute, individua così il quinto ed ultimo punto: «misurare cosa capita e perché» e lo accompagna con la domanda «come possiamo documentare i cambiamenti prodotti?».

Quelli che si contrappongono sono i presupposti stessi su cui si basa la valutazione delle azioni avviate: da una parte il programma è considerato come la risposta razionale a un problema adeguatamente investigato e compreso, e la valutazione è lo strumento per ricondurlo alle attese; dall'altra esso rappresenta invece il “suggerimento” prescelto, perché considerato il più adeguato per sviluppare le potenzialità presenti nell'insieme dato dalle persone e dal contesto, e la valutazione diviene allora lo strumento sia per cogliere dove e perché tale suggerimento è stato raccolto, sia per sostenere le scelte che si faranno nel tempo (cfr. Bezzi 2001 e Palumbo 2001).

Questo breve testo si propone di delineare gli elementi centrali di questo dibattito, abbracciando la tesi che la valutazione non possa essere solo considerata come una faticosa, e necessaria, attività tecnica ma vada invece affrontata a partire da una riflessione intorno ai principi che la debbono guidare. In anni recenti, infatti, dopo un lungo periodo in cui i professionisti del campo della Sanità Pubblica hanno cercato di aumentare la loro credibilità abbracciando ciecamente il paradigma positivista, la letteratura è molto cresciuta e, nell'affrontare il tema della valutazione, è andata oltre il dibattito intorno ai meriti di una tecnica o di un metodo sugli altri, esplorando questioni fondamentali quali le caratteristiche che rendono un approccio valutativo appropriato a esaminare, e giudicare, interventi di promozione della salute.

### 1. *Appropriato rispetto a cosa?*

L'appropriatezza di un approccio valutativo è intimamente legata alla natura stessa dell'oggetto che è sottoposto a valutazione (cfr. Sprigett 2001). Eppure l'ambito della clinica e della prevenzione delle malattie come quello della promozione della salute vengono tutti comunemente analizzati utilizzando lo stesso paradigma valutativo: quello positivista.

#### 1.1. *Tra positivismo e costruttivismo*

Lo stato di malattia è, abbastanza concordemente, considerato come una condizione oggettiva dell'individuo di cui sono largamente ormai noti sia i fattori che hanno concorso al suo insorgere sia il suo naturale processo evolutivo. È questa evoluzione, visualizzabile come un processo pressoché lineare, tradizionalmente definito come la storia naturale della malattia, che s'intende contrastare attraverso le azioni terapeutiche o di diagnosi precoce che si sono dimostrate efficaci. Tale dimostrazione di efficacia trova il suo "gold standard" nel trial controllato randomizzato (RCT), disegno di osservazione sperimentale caratterizzato dall'assegnazione casuale dei soggetti ai gruppi a confronto. Il metodo di osservazione definito come sperimentale vuole studiare la regolarità con cui un evento (quale l'assunzione di un farmaco, o l'esposizione a un programma di comunità) si presenta associato a un risultato favorevole atteso che ci si è, quindi, preparati a osservare.

Attraverso una successione di osservazioni che mostreranno l'associazione e la variazione concomitante tra la presenza dell'evento e il risultato in osservazione sarà possibile affermare, con un definito margine di probabilità, che il primo può essere individuato come causa del secondo. Conseguentemente, anche nel caso l'osservazione sia tesa alla valutazione di efficacia di un intervento in comunità, sarà possibile prevedere con quale probabilità, qualora l'intervento sia riproposto in un'analogia popolazione, si riprodurrà il risultato atteso. L'obiettivo ultimo è quindi quello di definire leggi generali: attraverso il ragionamento ipotetico-deduttivo (cfr. Popper 1959), che governa l'osservazione sperimentale, "la verità" è svelata e ciò che si è dimostrato efficace sarà replicabile in altre realtà. L'idea di giungere a un'accurata rappresentazione del mondo, conseguente al progressivo accumularsi delle conoscenze che derivano dalla possibilità di

generalizzare i risultati prodotti dalle osservazioni sperimentali, guida l'approccio positivista e caratterizza il modello della ricerca in medicina (cfr. Cochrane 1978).

La salute, al contrario, è un costrutto sociale la cui misura richiede un preventivo accordo sulle variabili che s'intendono osservare. La sua promozione passa poi attraverso un complesso di azioni che, agendo su individui e contesto, mette in atto un processo sociale i cui risultati sono il frutto della complessa relazione che si viene a creare tra interventi, individui e ambiente. Tale processo sociale ricerca la partecipazione della comunità e costruisce alleanze con i diversi attori sociali presenti nel territorio: questo allo scopo di aumentare le reali possibilità di controllo sui determinanti di salute, sia da parte dei singoli che della comunità nel suo complesso, e permettere la sostenibilità nel tempo delle azioni avviate (cfr. WHO 2009). Un appropriato approccio valutativo dovrà quindi essere così flessibile da accogliere queste peculiarità, e sempre più forti sono i dubbi che il disegno osservativo sperimentale abbia queste caratteristiche e possa quindi essere considerato come lo studio più idoneo a costruire evidenza di efficacia per gli interventi condotti in comunità.

Crisi del modello sperimentale o, più in generale, crisi dell'approccio positivista alla ricerca psico-sociale? A entrare in crisi è, infatti, la convinzione che anche nel campo dell'azione di comunità, come in quello biologico, sia possibile affermare l'esistenza di una verità oggettiva che, non essendo direttamente osservabile, possa essere svelata attraverso la "forza metodologica" dell'osservazione sperimentale. Emerge cioè l'idea che non esista una verità indipendente dalla relazione che si crea tra azioni messe in atto, soggetti coinvolti e contesto, e che la realtà che descriviamo sia sempre "il prodotto di una costruzione sociale" che prende forma durante l'esperienza diretta: dall'attenta raccolta dei dati non si dedurrebbe la realtà ma la si genererebbe attraverso un processo induttivo (cfr. Guba-Lincoln 1989). Sul tema della valutazione dei programmi di promozione della salute l'idea di fondo è che essi si trasformino nel contatto con ogni specifico contesto e che quindi, per definirne l'impatto, sia necessario tenere conto delle interpretazioni che, durante il processo, i soggetti della comunità forniscono rispetto al mutare o meno della realtà.

Della contrapposizione tra positivismo e costruttivismo è intrisa la letteratura che di valutazione di efficacia si è occupata in questi anni: da quella più recente emerge però anche la convinzione che, dinnanzi all'inconciliabilità delle due posizioni, occorra far avanzare un approccio pragmatico che, a partire dagli elementi di crisi del positivismo, e raccogliendo elementi di sfida proposti dal costruttivismo, ponga le basi metodologiche che consentano l'avanzare della conoscenza sull'impatto delle azioni condotte con la comunità attraverso una loro appropriata valutazione.

Per valutare impatti delle azioni bisognerà allora fare luce nella "scatola" dell'incontro tra il programma d'intervento, i diversi gruppi di persone e i differenti contesti, per cercare di identificare i meccanismi che tra questi elementi si vengono a creare. L'obiettivo non sarà quindi più quello di confermare, una volta per tutte, l'efficacia di una determinata azione nel produrre l'atteso risultato, attraverso un definito meccanismo, ma ci si dovrà invece sforzare di capire perché, con un determinato gruppo di persone e in un determinato contesto, in presenza di quell'*input* si sia ottenuto un certo risultato. Cumulare conoscenza richiederà quindi l'individuazione di quelle

tipologie d'insiemi di programma, persone e contesto, che abbiano prodotto determinati risultati, in un paziente e impegnativo processo di astrazione qualitativa che porti alla luce i diversi meccanismi agenti (cfr. Pawson-Tilley 1997).

### 1.2. *Valutare empowerment*

Nel 1995 l'Organizzazione Mondiale della Sanità istituisce un gruppo di lavoro europeo centrato sulla valutazione in promozione della salute. Il testo, che il gruppo di lavoro produce alcuni anni dopo, si apre con un'introduzione che definisce i principi a cui deve rispondere la valutazione per definirsi appropriata alle iniziative di promozione della salute. Oltre a dover essere delineata in modo da accogliere la complessità che caratterizza i suoi interventi, richiamando gli elementi che sono stati fin qui discussi, e disegnata attraverso la collaborazione delle diverse discipline, allo scopo di ampliare i punti di vista e le procedure utilizzate, la valutazione dovrà essere partecipata, coinvolgendo nei modi appropriati i diversi attori sociali interessati, in un processo che costruisca, nei soggetti e nella comunità, le capacità utili a promuovere salute (cfr. Rootman 2001). La centralità, nella letteratura che si occupa di promozione della salute, del processo attraverso il quale i soggetti accrescono le capacità necessarie al controllo sulle azioni e decisioni che riguardano la propria vita, e che nella letteratura anglosassone è detto di "*empowerment*", è continuamente sottolineata nel citato documento dell'Organizzazione mondiale della salute (cfr. WHO 2009).

Spesso di *empowerment* si discute come di un concetto astratto dimenticando che, seppure il cuore del processo sia rappresentato dalla percezione di poter influenzare il decorso degli eventi, questa è la conseguenza delle aumentate capacità di controllo della propria vita. Capacità che devono poter essere individuabili e trasferibili, attivando un processo che contribuisca a modificare positivamente la storia degli individui e della comunità alla quale essi appartengono. Progettare e valutare *empowerment* richiede allora l'identificazione di tali aree e, tra gli autori che hanno maggiormente lavorato in questa direzione, troviamo Glenn Laverack (2001), che individua una lista di quelli che definisce "domini" di influenza attraverso i quali l'*empowerment* si esprime, e che rappresentano una sorta di sua traduzione operativa (vedi tabella).

<b>Domini</b>	<b>Descrizione</b>
Partecipazione comunitaria	Partecipazione in attività di piccoli gruppi o in più larghe organizzazioni
Sviluppo di <i>leadership</i> locali	Connessa alla precedente mette in evidenza responsabilità locali nell'avvio e nella conduzione
Utilizzo di strutture organizzative	Attivo utilizzo delle strutture presenti e l'eventuale attivazione di nuove
Valutazione dei problemi	Capacità di individuare problemi, possibili soluzioni e azioni necessarie
Mobilizzazione delle risorse	Avvenuta mobilizzazione di risorse interne e negoziazione di risorse esterne

Chiedersi il perché	Capacità di valutare criticamente le cause dei problemi e il modo in cui la comunità stessa o agenti esterni ne favoriscano gli effetti negativi
Legami con altri	Capacità di ricercare e gestire <i>partnership</i> , interne ed esterne alla comunità, finalizzate allo sviluppo
Consulenti esterni	Capacità di interagire in modo efficace con agenti esterni per poi rendersi da questi autonomi
Gestione dei programmi	Capacità di condurre le azioni necessarie al raggiungimento degli obiettivi.

Descrizione sintetica dei nove domini dell'*empowerment* di comunità (Laverack 2001).

Intorno a tali capacità fondamentali, che rappresenterebbero gli elementi costituenti l'*empowerment*, andrebbero progettate le azioni di sviluppo di comunità, e andrebbe poi valutata, attraverso tecniche osservative e narrative, l'acquisizione di queste capacità stesse. All'interno di ognuna di queste aree è infatti possibile descrivere un *continuum* che, partendo dall'osservazione dell'assenza della capacità, individua quegli elementi che, presentandosi, mostrano un progredire all'interno del processo del suo sviluppo.

Lo stesso autore parla poi di "*double tracking*" sostenendo che, sia nella progettazione sia nella valutazione dei progetti di promozione della salute, è possibile delineare due strade parallele: il processo per raggiungere gli obiettivi di salute e quello per incrementare l'*empowerment* (cfr. Laverack 2008). Due strade che variamente s'intrecciano, anche in ragione dell'approccio seguito nella progettazione, e che devono essere entrambe tenute sotto controllo: contemporaneamente ma separatamente. Solo così sarà possibile cumulare conoscenza sul ruolo svolto dallo sviluppo dell'*empowerment* nel raggiungimento degli obiettivi di salute.

## 2. Valutazione e processi decisionali

In altro testo si è già discusso di come al variare del modo di interpretare il coinvolgimento della comunità (il ruolo che gli si assegna e le finalità per cui è ricercato) si modifichi la maniera in cui le diverse fasi della progettazione sono interpretate e condotte: al variare dell'approccio alla progettazione dell'incremento dei livelli di salute di una comunità, a cambiare sono anche le finalità per cui è condotta l'azione valutativa (cfr. Lemma 2007).

Il tradizionale approccio agli interventi in comunità, caratterizzato da un basso livello di partecipazione poiché centrato sulle scelte condotte da esperti, disegna il processo di progettazione come una sequenza lineare: alla definizione del quadro conoscitivo della realtà in cui si condurrà l'intervento, ritenuta adeguata da coloro che hanno la responsabilità della conduzione delle azioni, segue l'assunzione delle scelte e la loro successiva attuazione e valutazione. Questo modello direttivo, almeno nella sua più semplice interpretazione, appare governato da una razionalità orientata a un predefinito scopo, e il processo decisionale che mette in atto è teso a individuare e adottare i mezzi

che permettano, nel modo ritenuto il migliore possibile, cioè più efficace e meno costoso, di raggiungere il fine dato.

Quello che si richiede alla valutazione è quindi di prevedere gli esiti in base alle premesse: a partire dalla conoscenza del teatro d'intervento, e delle possibili soluzioni al problema identificato, si chiede di individuare gli obiettivi che è possibile perseguire. Il suo ruolo è di giudicare a priori la bontà di un programma, insieme alla sua realizzabilità, valutandone quindi l'efficacia teorica e pratica e prevedendo la modificabilità della situazione in seguito alla "somministrazione" dell'intervento. Dagli obiettivi discenderanno poi gli indicatori, rispetto ai quali saranno attuate le misurazioni che permetteranno di verificare il corretto procedere verso le mete, e il loro successivo raggiungimento. Alla valutazione, in questo secondo momento, si chiede quindi di individuare quelle variabili significative che permetteranno di monitorare il corretto svolgersi del programma e il raggiungimento degli obiettivi, completando così il ruolo predittivo che le era stato assegnato (cfr. Bezzi 2001).

Solo dopo aver impostato il disegno valutativo che conduce la valutazione, spesso un soggetto esterno la cui neutralità è vista come garanzia della necessaria obiettività, raccoglierà i dati che gli indicheranno i risultati che nel tempo saranno raggiunti. La letteratura anglosassone parla di "*summative evaluation*", cioè di valutazione riepilogativa, che quindi si propone di riassumere i risultati ottenuti e di decidere se il progetto meriterà in futuro di essere replicato in situazioni analoghe. A questa si affianca la "*formative evaluation*", che accompagna lo sviluppo del progetto ed aiuta i decisori a governarlo dandogli "la forma" migliore per poter raggiungere gli obiettivi definiti: dato infatti il valore normativo che, in questo approccio, assume il programma predefinito, il mancare dei risultati previsti non potrà che essere ricondotto alla mancata adesione alla prassi stabilita.

Questo modello si fonda però su alcuni presupposti: la reale possibilità da parte del decisore di individuare con chiarezza un problema e di poter identificare tutti gli interventi, alternativi o meno, che possano condurre al suo superamento; ma anche quella di poter prevedere tutte le conseguenze che deriveranno dalla selezione di ogni alternativa, e di poter quindi definire tutti gli obiettivi il cui conseguimento dovrebbe condurre alla soluzione del problema; non dimenticando poi l'assunto che l'attuazione degli interventi scelti rappresenti una mera esecuzione di quanto deciso (cfr. Jullien 1998). Solo all'interno di questa impostazione, sempre più ritenuta come semplicistica, possono avere senso i ruoli predittivo e di monitoraggio che gli approcci direttivi assegnano alla valutazione.

Questa presunzione, oltre a dare per assodato che siano realmente ricostruibili tutti i mezzi che possono essere messi in gioco, si sostiene inoltre su di una concezione "a-relazionale" degli attori sociali. Emerge però sempre con maggiore forza la posizione di chi sottolinea come, tutte le volte che entri in gioco l'intenzionalità dell'agire umano, si debba considerare infranto il sogno di poter mettere in luce leggi causali, e ci si debba invece accontentare di porre in evidenza, nel mutare dei fenomeni, soli andamenti tendenziali. Bisogna accettare che non solo potrà accadere che attori sociali diversi interpretino gli stessi dati di contesto in modo differente, agendo quindi diversamente,

ma anche che lo stesso attore sociale modifichi la sua azione, dinnanzi agli stessi dati di contesto, in risposta al bilancio tratto dalla sua passata esperienza (cfr. Marradi 1996).

Nel dibattito intorno agli elementi che caratterizzano la “post-modernità” (cfr. Giddens 1994) il tema della “crisi del progetto” emerge costantemente. Le sue radici sarebbero anche da ricercare nel venir meno di un altro tra i presupposti su cui si fonda la conduzione della progettazione in termini di razionalità assoluta: la condivisione, tra i soggetti coinvolti nell’azione di sviluppo di comunità, del quadro di certezze e valori da perseguire nel campo della salute. Saltando questo presupposto l’ampia partecipazione dei diversi attori sociali al processo decisionale risponderà allora alla necessità di ricostruire il senso stesso che fa da cornice all’azione: e questo non potrà non avere conseguenze sul ruolo assegnato alla valutazione all’interno dei processi decisionali.

I diversi attori sociali, per diversità di valori e d’interessi, possono essere portatori non solo di visioni differenti del problema ma anche, seppure in quota diversa, dei mezzi per risolverli. Assegnare a queste diversità in campo il ruolo di risorsa richiederà allora che il processo decisionale sia, prima di tutto, in grado di scegliere tra le definizioni alternative del problema proposte dai diversi attori sociali: non pensando di prenderle in considerazione tutte per selezionare “la migliore”, ma solo di individuare, tra quelle vagliate, la prima che soddisfi a sufficienza i criteri su cui la scelta si sta basando. In questo quadro la razionalità non scompare ma da sostanziale, capace cioè di dettare le soluzioni per ogni problema, diviene procedurale: razionali non saranno cioè più le decisioni ma il processo di loro assunzione. Tale modello non può che essere guidato da una concezione costruttivista della valutazione che sposta l’accento dai temi della conoscenza (che in un’accezione positivista è sempre vista come preliminare, obiettiva e neutrale) e della decisione (assunta solo da chi ha la competenza tecnica e istituzionale) a quelli della comunicazione, della negoziazione e della partecipazione (cfr. Bobbio 1996 e Stame 2001). In questo procedere verso una progettazione interattiva a razionalità limitata accade allora che la valutazione non sia più solo il frutto di certezze scientificamente fondate, quanto piuttosto il prodotto di una negoziazione: essa diviene lo strumento sia per comprendere la situazione e definire i problemi, attraverso l’interpretazione che ne danno i diversi attori sociali, sia per chiarire ciò che dal programma si potrà ottenere.

La valutazione passa allora da una concezione previsionale a una di prefigurazione, operando attraverso la definizione di scenari che, sintetizzando le informazioni di cui si è in possesso, mettano in luce i vincoli e le opportunità derivanti dal contesto in cui si opera: solo infatti partendo da una base comune si potrà sviluppare un processo decisionale che possa realmente individuare condivisi obiettivi e risultati attesi. Questo richiede però che tra gli attori si crei una reale dimensione collaborativa e si sviluppino efficaci processi comunicativi: quanto più verrà garantita la produzione di processi di formazione delle opinioni, tanto più le prefigurazioni saranno affidabili (cfr. Stame 2001).

Con il ridursi delle certezze la valutazione diventa quindi un’attività di supporto strategico, che coinvolge i diversi attori, e che si propone di rilevare le conseguenze, previste e non previste, delle azioni messe in atto. Parlare di valutazione “partecipata” vuol dire allora ricercare la collaborazione della comunità non solo per raccogliere le

informazioni necessarie a valutare ma, soprattutto, per interpretarle e dare a queste un senso (cfr. Fetterman e Wandersman 2005). Una valutazione “costruttiva” che assuma quindi il ruolo di indagare il processo di attuazione del progetto, proponendosi di spiegare se e perché, in quella situazione, un determinato risultato può essere considerato un successo e quindi decidere, insieme ai diversi attori sociali, cosa stia funzionando proponendo modifiche che andranno nuovamente rilette in questo processo circolare.

### Bibliografia

- C. BEZZI (2001), *Il disegno della ricerca valutativa*, Franco Angeli, Milano 2001.
- L. BOBBIO (1996), *La democrazia non abita a Gordio. Studio sui processi decisionali politico-amministrativi*, Franco Angeli, Milano 1996.
- D.T. CAMPBELL-J. STANLEY (1963), *Experimental and Quasi-Experimental Designs for Research*, Rand McNally, Chicago 1963.
- D.T. CAMPBELL (1969), *Reforms as Experiments*, in “American Psychologist”, 24 (1969), pp. 409-429.
- A.L. COCHRANE (1978), *L’inflazione medica*, trad. it. E. Coffano, Feltrinelli, Milano 1978.
- D. FETTERMAN-A. WANDERSMAN (2005), *Empowerment Evaluation Principles in Practice*, The Guilford Press, New York 2005.
- A. GIDDENS (1994), *Le conseguenze della modernità*, trad. it. M. Guani, Il Mulino, Bologna 1994.
- E.G. GUBA-Y.S. LINCOLN (1989), *Fourth Generation Evaluation*, Sage, Newbury Park 1989.
- F. JULLIEN (1998), *Trattato dell’efficacia*, trad. it. M. Porro, Einaudi, Torino 1998.
- G. LAVERACK (2001), *An Identification and Interpretation of the Organizational Aspects of Community Empowerment*, in “Community Development Journal”, 36 (2001), pp. 40-52.
- G. LAVERACK (2008), *Health Promotion in Action – From Local to Global Empowerment*, Palgrave Macmillan, Basingstoke 2008.
- P. LEMMA (2007), *Promuovere salute nell’era della globalizzazione*, Unicopli, Milano 2007.
- A. MARRADI (1996), *Due famiglie un insieme*, in C. CIPOLLA-A. DE LILLO (a cura di), *Il sociologo e le sirene. La sfida dei metodi qualitativi*, Franco Angeli, Milano 1996, pp. 167-178.
- M. PALUMBO (2001), *Il processo di valutazione*, Franco Angeli, Milano 2001.
- R. PAWSON-N. TILLEY (1997), *Realistic Evaluation*, Sage, London 1997.
- K.R. POPPER (1959), *The Logic of Scientific Discovery*, Hutchinson, London 1959; trad. it. M. Trinchero, *Logica della scoperta scientifica*, Einaudi, Torino 2010.
- I. ROOTMAN (2001), *Introduction to the Book*, in I. ROOTMAN-M. GOODSTADT-B. HYNDMAN-D. MCQUEEN-L. POTVIN- J. SPRINGETT-E. ZIGLIO (a cura di) *Evaluation in Health Promotion – Principles and Perspectives*, WHO, Copenhagen 2001, pp. 3-6.
- N. STAME (2001), *Tre approcci principali alla valutazione: distinguere e combinare*, in M. PALUMBO, *Il processo di valutazione*, Franco Angeli, Milano, pp. 25-45.

- J. SPRIGETT (2001), *Appropriate Approaches to the Evaluation of Health Promotion*, in “Critical Public Health”, 11 (2/2001), pp. 139-151.
- WHO (WORLD HEALTH ORGANISATION) (2009), *Nairobi Call to Action. A Primer for Mainstreaming Health Promotion*, WHO, Nairobi 2009.



Paola Ricchiardi  
Emanuela M. Torre

## VALUTAZIONE DELLA SCUOLA E DEL SISTEMA SCOLASTICO: QUALITÀ FORMALE E QUALITÀ EFFETTIVA<sup>1</sup>

### Abstract

*Assessment practices have currently intensified in several countries, according to different approaches. Since the late 1960s in particular, the European school systems have been affected by periodic international surveys, with comparative analyses of learning outcomes. These were complemented by increasingly complex system assessments, which have taken into account, since the 1990s, even contextual and procedural variables. The evaluation questions have therefore gained an increasing importance within the international debate on the quality of education. Assessment takes the leading role in guiding the system towards quality. In some cases, however, the risk is that of proceeding in a merely formal way, without promoting “real quality.” In this essay, we will identify which strategies of school evaluation may be usefully activated to promote educational success. To this end, the essay suggests the development of more holistic evaluations, their extension to more qualitative aspects of effective learning environments, and the encouragement of a better use of the collected data.*

### 1. Introduzione

Da alcuni decenni il tema della valutazione in ambito scolastico è oggetto di riflessione da parte degli studiosi e dei legislatori nei suoi diversi aspetti: dall'accertamento degli esiti degli studenti, alla rilevazione delle performance dei docenti e delle strategie di management dei dirigenti scolastici, fino alla valutazione interna ed esterna delle organizzazioni scolastiche, del sistema scolastico nel suo complesso e delle politiche educative. Dalla fine degli anni '60, in particolare, i sistemi scolastici europei sono stati interessati da periodiche rilevazioni internazionali (IEA, IAEP), con analisi comparative degli esiti di apprendimento. A queste si sono affiancate valutazioni sempre più complesse del sistema, che hanno preso in considerazione, a partire dagli anni '90, anche variabili contestuali e processuali (OCSE-PISA, IEA-TIMMS, IEA-PIRLS). Le questioni valutative hanno dunque acquisito importanza crescente nel dibattito internazionale sulla qualità dell'educazione e dell'istruzione, sia rispetto ai metodi, sia soprattutto rispetto all'uso degli esiti rilevati, con l'intento di comprendere come la valutazione possa

---

<sup>1</sup> Il presente contributo è frutto del lavoro di approfondimento e ricerca comune delle due autrici. In particolare a P. Ricchiardi vanno attribuiti i paragrafi 1, 2, 3.2 e a E.M. Torre vanno attribuiti i paragrafi 3, 3.1., 3.3, 3.4, 4.

diventare un'adeguata guida per orientare il sistema verso la qualità<sup>2</sup>. Il rischio, in alcuni casi, è attualmente quello che le procedure di gestione e rilevazione della qualità restino, invece, ancorate a logiche formali: a diversi livelli (dalla scuola all'università)<sup>3</sup> e in diversi campi, tali procedure sono infatti state sottoposte ad un processo di critica, che ne mette in luce i limiti. In questo contributo ci proponiamo quindi di focalizzare l'attenzione sui modelli di valutazione che vengono attualmente considerati dalla ricerca internazionale come i più utili per promuovere la "qualità reale", vale a dire per favorire il successo formativo di tutti gli studenti.

## 2. Le criticità nella valutazione del sistema scolastico: quali scopi, quali mezzi, quale equità?

Le pratiche di analisi e valutazione si sono attualmente intensificate in diversi Paesi, secondo approcci differenti. Da un lato, con l'incremento delle autonomie scolastiche, la valutazione è divenuta infatti una strategia necessaria di monitoraggio e controllo dell'efficacia dei modelli didattici e delle pratiche adottate a livello locale<sup>4</sup>. Dall'altro, con la diffusione dell'approccio *evidence based*<sup>5</sup> anche nella ricerca educativa, è incrementata la richiesta di fondare decisioni importanti di tipo organizzativo ed educativo su dati fattuali e non solo su speculazioni di carattere teorico.

Nelle valutazioni di efficacia dei sistemi scolastici ha assunto un ruolo fondamentale la rilevazione dei risultati degli studenti, rispetto alla quale sono emerse nel tempo però una serie di questioni importanti, specie se gli esiti degli alunni diventano fonte unilaterale per apprezzare la qualità del sistema nel suo complesso. Alcune preoccupazioni sono più di carattere tecnico e riguardano la scelta di strumenti, metodi, competenze e processi cognitivi da privilegiare nelle rilevazioni, con implicazioni non irrilevanti sugli esiti (es.: focus sui processi cognitivi di base, come la memorizzazione e la comprensione; o attivazione prevalente dei processi cognitivi superiori come il ragionamento, la capacità critica o la creatività). Altre questioni sono connesse agli aspetti operativi: si tratta, per esempio, dell'utilizzo effettivo dei dati raccolti. Altre ancora sono più di carattere etico e sono legate agli scopi di tali procedure, anche dal punto di vista del rispetto dell'equità. Di particolare interesse, a tal proposito, è il conflitto valoriale, generato dalla possibile discrasia tra le istanze istituzionali e le condizioni individuali degli alunni, tra gli standard da raggiungere, stabiliti a livello nazionale, e le condizioni di partenza di alcuni gruppi di apprendenti<sup>6</sup>. La richiesta formale di raggiungere obiettivi di apprendimento distanti

---

<sup>2</sup> Cfr. J. SCHEERENS, *Effective Schooling: Research, Theory and Practice*, Cassell, London 1992.

<sup>3</sup> Un'interessante analisi di tali problematiche con riferimento specifico alla didattica universitaria è presentata in A.F. WALL-D. HURSH-J.W. RODGERS, *Assessment for Whom: Repositioning Higher Education Assessment as an Ethical and Value-Focused Social Practice*, in "Research & Practice in Assessment", 9 (2014), pp. 5-17.

<sup>4</sup> Cfr. K.F. NAYIR-G. MCNAMARA, *The Increasingly Central Role of School Self-Evaluation in Inspection System across Europe: The Case of Ireland*, in "Turkish Journal of Education", 3 (1/2013), pp. 48-59.

<sup>5</sup> Cfr. J. HATTIE, *Visible Learning: A Synthesis of over 800 Meta-analyses Relating to Achievement*, Routledge, London 2013.

<sup>6</sup> Cfr. N. POPE-S.K. GREEN-R.L. JOHNSON-M. MITCHELL, *Examining Teacher Ethical Dilemmas in Classroom Assessment*, in "Teaching and Teacher Education", 25 (5/2009), pp. 778-782.

dalle reali competenze degli allievi in determinati contesti, ha visto l'incremento, per esempio in Italia, del fenomeno del *cheating*, ovvero della tendenza a falsare gli esiti per attendere formalmente alle richieste<sup>7</sup>. L'esigenza di giungere a standard formalmente stabiliti rischia inoltre di mettere in crisi l'istanza di personalizzazione. La normativa di diversi Paesi consente di differenziare le programmazioni e di conseguenza le richieste rivolte a studenti svantaggiati o con particolari difficoltà, utilizzando parametri valutativi differenti a seconda delle condizioni di partenza dei soggetti e delle caratteristiche personali (cognitive, emotive, socio-culturali...), nel rispetto del principio di equità, ovvero dell'istanza di mettere tutti nelle medesime condizioni di riuscita. Tale personalizzazione va però in contrasto con il conseguimento di buoni esiti nelle rilevazioni standardizzate, alle quali, in alcuni casi, possono essere connessi meccanismi di premialità dell'istituto scolastico o del singolo docente. C'è inoltre il rischio che i docenti focalizzino l'attenzione sulle competenze oggetto della valutazione standardizzata e attribuiscono «una minor importanza a necessità evolutive e pedagogiche di più ampio respiro degli studenti»<sup>8</sup>.

Un altro nodo chiave è lo scarso utilizzo degli esiti di indagini nazionali ed internazionali, che costituisce invece una sfida rilevante, secondo il Rapporto OECD 2013: «Le tecniche di analisi e valutazione dovrebbero essere messe al servizio degli obiettivi pedagogici e di apprendimento degli studenti e dovrebbero contribuire a [...] migliorare la pratica didattica e l'apprendimento degli studenti»<sup>9</sup>.

### 3. Alcune proposte per una valutazione efficace

Gli sviluppi della ricerca valutativa da un lato e le esperienze internazionali di più lunga tradizione dall'altro individuano, a partire dalle criticità rilevate, alcune strategie utili ad attivare procedure di valutazione efficaci nel promuovere il miglioramento degli interventi educativi e formativi attuati.

#### 3.1. Il rispetto dell'equità sociale nella valutazione

Fondamento metodologico di ogni processo scientifico di valutazione è la ricerca valutativa<sup>10</sup>. Quest'ultima non si propone solo di generare teorie e produrre conoscenze<sup>11</sup>, ma è orientata anche a informare l'azione, a facilitare i processi decisionali,

<sup>7</sup> Cfr. D. RAVITCH, *The Dead and Life of the Great American School System: How Testing and Choice Are Undermining Education*, Basic Books, New York 2010 e P. LUCISANO, *INVALSI: meglio cercare di comprendere che valutare. Che cosa ha prodotto l'attuale Sistema di valutazione nazionale?*, in A.M. NOTTI, *A scuola di valutazione*, PensaMultimedia, Lecce 2014, pp. 79-103.

<sup>8</sup> OECD [ORGANISATION FOR ECONOMIC CO-OPERATION AND DEVELOPMENT], *Synergies for Better Learning: An International Perspective on Evaluation and Assessment. Summary*, OECD, Paris 2013, pp. 2-3.

<sup>9</sup> *Ibidem*, p. 2.

<sup>10</sup> Cfr. E. SCHUMAN, *Evaluative Research. Principles and Practice in Public Service and Social Actions Programs*, Russell Sage Foundation, New York 1967; S. MATHISON, *What Is the Difference between Evaluation and Research and Why Do We Care?*, in N.L. SMITH-P.R. BRANDON (a cura di), *Fundamental Issues in Evaluation*, The Guilford Press, New York-London 2008, pp. 183-196.

<sup>11</sup> Cfr. M. PATTON, *Qualitative Evaluation and Research Methods*, Sage, London, 2002<sup>3</sup>.

anche a livello politico, e a favorire il miglioramento e il cambiamento. A questo scopo si fonda sulle conoscenze e sulla riflessione relativa alle pratiche attuate<sup>12</sup>. Gli studi più recenti in tale ambito evidenziano alcune questioni, su cui è interessante soffermarsi, nella logica di individuare prassi atte a coniugare qualità formale e qualità reale. Tali riflessioni riguardano in particolare il ruolo che la valutazione può avere nell'apprezzare e promuovere trasformazioni che garantiscano l'equità sociale. In questa direzione si è sviluppato un recente approccio, definito "trasformativo". Esso si focalizza sulla rilevazione dell'entità dei cambiamenti prodotti da una determinata azione educativa (in termini di comportamenti, atteggiamenti, processi sociali, acquisizione di competenze), piuttosto che sugli esiti finali rapportati a standard di riferimento. L'approccio trasformativo presta particolare attenzione ai gruppi più svantaggiati e al rispetto dell'equità sociale attraverso l'analisi delle diseguaglianze strutturali, il coinvolgimento dei diversi *stakeholders* e il rispetto delle dimensioni culturali e contestuali specifiche, considerate determinanti ai fini della valutazione<sup>13</sup>.

Tali riflessioni richiamano istanze più generali, quali il diritto all'istruzione per tutti, l'assicurazione del successo formativo, la personalizzazione dei percorsi educativi, il rispetto delle differenze individuali. Si tratta di aspetti che, soprattutto nel caso di contesti particolarmente svantaggiati, rischiano di essere lasciati in ombra da modelli esclusivamente centrati sulla valutazione ex-post di esiti misurati con riferimento a standard nazionali e internazionali.

L'approccio trasformativo propone inoltre l'integrazione sul campo, in tempi adeguati, delle indicazioni scaturite dalle rilevazioni condotte<sup>14</sup>.

### 3.2. *Un miglior utilizzo dei dati: integrazione di School Effectiveness e School Improvement*

Un processo di valutazione della scuola può uscire dalla formalità, se i dati rilevati sono funzionali ad un miglioramento effettivo. I primi modelli su cui si fondava l'analisi dell'efficacia scolastica erano di derivazione economica o socio-organizzativa. Lo schema teorico di base era focalizzato essenzialmente sulla funzione di produzione (input-output) e non riusciva ad apprezzare adeguatamente le diverse dimensioni coinvolte nel processo formativo<sup>15</sup>. I modelli successivi hanno assunto uno schema più complesso che tiene in considerazione più elementi (contesto-input-processo-output), variamente articolati, e quindi potenzialmente più capaci di guidare il sistema verso la qualità. Ne costituisce un valido esempio il *CIPP Model* di D.L. Stufflebeam sviluppato nei primi

---

<sup>12</sup> Cfr. L. ARTHUR-E. COX, *From Evaluation to Research*, in "International Journal of Research & Method in Education", 36 (2/2013), pp. 1-14.

<sup>13</sup> Cfr. D.M. MERTENS, *Transformative Research & Evaluation*, Guilford, New York 2009; D.M. MERTENS-A.T. WILSON, *Program Evaluation Theory and Practice: A Comprehensive Approach*, Guilford, New York 2012.

<sup>14</sup> Cfr. M. BAMBERGER-M. SEGONE, *How to Design and Manage Equity Focused Evaluations*, Unicef, New York 2011 (<http://www.mymande.org/?q=content/how-design-and-manage-equity-focused-evaluations>, ultima visita dicembre 2014).

<sup>15</sup> Si veda, ad es., M.B. MILES-M. EKHOLM, *What Is School Improvement?*, in W.G. VAN VELZEN *et al.*, *Making School Improvement*, ACCO, Leuven 1985, pp. 33-67.

anni '70 e ancora utilizzato in diverse proposte valutative, anche in ambito italiano<sup>16</sup>. Tali modelli rimangono comunque sterili se i sistemi di indicatori che da essi derivano non producono azioni trasformative nel sistema. È opportuno quindi non perdere di vista la connessione stretta che deve sussistere tra valutazione e cambiamento, condizione perché le procedure di assicurazione della qualità attualmente richieste, su più livelli, anche al sistema scolastico non rimangano circoscritte al rispetto di imposizioni burocratiche, ma sempre possano essere utilizzate per perseguire obiettivi di miglioramento del sistema stesso.

Tale connessione si è realizzata nel tempo in particolare attraverso l'integrazione progressiva tra i due principali filoni di studio che si sono occupati del tema, denominati rispettivamente *School Effectiveness*<sup>17</sup> e *School Improvement*<sup>18</sup>.

Il filone della *School Effectiveness* si propone di rilevare l'effettivo raggiungimento da parte di un istituto scolastico di un determinato insieme di obiettivi, attraverso adeguate strategie di ricerca empirica. La valutazione si fonda sulla comparazione degli esiti raggiunti dagli studenti al termine di un certo percorso, adeguatamente riproporzionati sulla base delle condizioni specifiche dei diversi contesti. Tiene poi in considerazione aspetti legati, ad esempio, alla soddisfazione dei docenti e alla capacità della scuola di rispondere ai bisogni della comunità. L'assunto di base è che le modalità di funzionamento dell'istituto scolastico, le specifiche condizioni organizzative, educative e di contesto che caratterizzano la singola scuola possano determinare la qualità della sua offerta formativa<sup>19</sup>.

Il secondo filone (*School Improvement*) si fonda, invece, su un processo dinamico di cambiamento, per favorire il miglioramento delle prestazioni della scuola nel tempo<sup>20</sup>.

Il crescente interesse per entrambi i punti di vista ha portato negli ultimi quarant'anni, soprattutto nelle valutazioni longitudinali, a percepire l'utilità di una loro integrazione<sup>21</sup>,

<sup>16</sup> Cfr. D.L. STUFFLEBEAM, *The Relevance of the CIPP Evaluation Model for Educational Accountability*, in "Journal of Research and Development in Education", 5 (1/1971), pp. 19-25 e D.L. STUFFLEBEAM-A.J. SHINKFIELD, *Evaluation, Theory, Model & Applications*, Jossey Bass, San Francisco 2007.

<sup>17</sup> Cfr. D. REYNOLDS-C. TEDDLIE, *The International Handbook of School Effectiveness Research*, Falmer Press, London 2002.

<sup>18</sup> Cfr. D. HOPKINS, *School Improvement for Real*, Routledge Falmer, London 2001.

<sup>19</sup> Alcuni autori ampliano l'analisi, affermando che per ottenere un quadro completo dell'*Educational Effectiveness* occorre considerare tre aspetti: la *School Effectiveness*; la *Teaching Effectiveness*, che riguarda l'attività del docente, e la *System Effectiveness* ossia le condizioni organizzative e di management che a loro volta dipendono anche dalle politiche nazionali o sovranazionali (cfr. J. SCHEERENS, *What Is Effective Schooling? A Review of Current Thought and Practice*, International Baccalaureate Organization, Genève 2013, <http://doc.utwente.nl/87298/>, ultima visita dicembre 2014).

<sup>20</sup> Cfr. J. SCHEERENS, *What Is Effective Schooling? A Review of Current Thought and Practice*, ed. cit.

<sup>21</sup> I due filoni di ricerca trovano infatti i propri fondamenti in due paradigmi di riferimento differenti. Come rileva Hopkins (nel suo *School Improvement for Real*, ed. cit.), quello della *School Effectiveness* utilizza una metodologia di tipo quantitativo, si propone di fotografare le caratteristiche formali di una scuola in un dato momento e tende a focalizzare le proprie analisi sulle situazioni di eccellenza. Il paradigma che fa capo al secondo filone (*School Improvement*) privilegia invece metodologie qualitative, approfondisce le dinamiche dei processi organizzativi che contraddistinguono una scuola, si preoccupa di utilizzare gli esiti scolastici come elementi di riflessione per il miglioramento della qualità formativa. È evidente che le differenze metodologiche rendono questi modelli complementari, con la conseguenza che la loro integrazione può utilmente migliorare la valutazione dei sistemi scolastici; cfr. T. WRIGLEY, *Rethinking*

consentendo così di valorizzare i dati da rilevare nell'ottica di un miglioramento della scuola. Tale connessione ha trovato un riconoscimento formale all'inizio degli anni '90<sup>22</sup>, e ha visto, secondo l'ampia rassegna di ricerca di D. Hopkins *et al.*<sup>23</sup>, cinque fasi principali.

I primi contributi si sono focalizzati sull'analisi estensiva dei risultati degli allievi e delle caratteristiche organizzative delle scuole, al fine di individuare relazioni significative tra i due ordini di fattori. Gli esiti emersi da ampi studi, come quelli sintetizzati nel famoso Rapporto Coleman (1966)<sup>24</sup>, sono poi stati approfonditi con ricerche su singole istituzioni scolastiche (seconda fase). Si tratta perlopiù di istituti considerati di eccellenza rispetto agli apprendimenti degli alunni, nonostante la situazione socio-economica svantaggiata dell'utenza. Lo scopo di tali indagini era di individuare le condizioni che favoriscono il successo a scuola, nonostante i fattori di rischio. Proseguendo in questa linea, gli studi afferenti alla terza fase si sono focalizzati sull'identificazione di scuole di qualità, delle quali si sono analizzate e confrontate le condizioni complessive di funzionamento in modo da riconoscere i tratti ricorrenti e comuni. Ciò ha consentito la strutturazione e sperimentazione di modelli di scuole di successo. Famosi sono gli studi realizzati all'interno dei progetti *Success for All*<sup>25</sup> e *High Schools That Work*<sup>26</sup>, volti a prevenire l'insuccesso scolastico a partire dalle evidenze di ricerca. Le indagini valutative si sono poi focalizzate (quarta fase) sulla rilevazione dell'efficacia dei cambiamenti introdotti in scuole connesse in rete, che condividono linee guida ed obiettivi di miglioramento e mettono in comunicazione i professionisti provenienti da più istituzioni scolastiche. La constatazione infine che le ricerche sulla *School Effectiveness* hanno portato alla realizzazione e sperimentazione di modelli di miglioramento della scuola (*School Improvement*)<sup>27</sup> ha messo in luce l'opportunità di effettuare il medesimo passaggio a livello di sistema (quinta fase). Tale linea di ricerca è sostenuta dalle rilevazioni internazionali di TIMSS e PISA, che forniscono dati comparativi interessanti per il miglioramento di sistema.

Risulta ancora da incrementare l'utilizzo efficace dei dati emersi dalle valutazioni. Per far fronte a tale sfida occorre che la selezione degli elementi oggetto di rilevazione sia già

---

*School Effectiveness and Improvement: A Question of Paradigms*, in "Discourse: Studies in the Cultural Politics of Education", 34 (1/2013), pp. 31-47.

<sup>22</sup> L'integrazione tra i due filoni è sancita formalmente dall'articolo di D. REYNOLDS-D. HOPKINS-L. STOLL, *Linking School Effectiveness Knowledge and School Improvement Practice: Towards a Synergy*, in "School Effectiveness and School Improvement", 4 (1/1993), pp. 37-58, nel quale gli autori analizzano i due paradigmi e le loro possibili sinergie, principalmente in termini metodologici (cfr. T. WRIGLEY, *Rethinking School Effectiveness and Improvement: A Question of Paradigms*, ed. cit.).

<sup>23</sup> Cfr. D. HOPKINS, *Every School a Great School*, Open University Press, Maidenhead 2007.

<sup>24</sup> J.S. COLEMAN *et al.*, *Equality of Educational Opportunity*, National Center for Educational Statistics, Washington 1966 (<http://files.eric.ed.gov/fulltext/ED012275.pdf>, ultima visita dicembre 2014).

<sup>25</sup> R.E. SLAVIN-N.A. MADDEN-L.J. DOLAN-B.A. WASIK-S. ROSS-L. SMITH-M. DIANDA, *Success for All: A Summary of Research*, in "Journal of Education for Students Placed at Risk", 1 (1/1996), pp. 41-76.

<sup>26</sup> [http://www.sreb.org/page/1078/high\\_schools\\_that\\_work.html](http://www.sreb.org/page/1078/high_schools_that_work.html) (ultima visita dicembre 2014).

<sup>27</sup> Citiamo alcuni studi a titolo esemplificativo: R.F. ELMORE, *School Reform from the Inside Out*, Harvard Education Press, Cambridge 2004; S.C. STRINGFIELD-M.E. YAKIMOWSKI-SREBNICK, *The Promise, Progress, Problems and Paradoxes of Three Phases of Accountability: A Longitudinal Case Study of the Baltimore City Public Schools*, in "American Educational Research Journal", 42 (2005), pp. 43-75.

orientata a favorire un miglioramento. È necessario inoltre che la comunicazione dei risultati raggiunti non si limiti a comparazioni, ma possa diventare stimolo per il cambiamento effettivo, sia mettendo a disposizione del singolo istituto scolastico gli esiti più analitici, a partire dai quali rivedere traguardi da raggiungere e strategie da utilizzare, sia curando il coinvolgimento degli insegnanti, non più “sotto esame” ma attori del processo di miglioramento.

### 3.3. L'autovalutazione per il miglioramento

Nella valutazione della qualità del sistema scolastico si è vista una transizione progressiva da una focalizzazione sulla qualità formale, legata alla logica della soddisfazione del cliente, ad una logica autovalutativa, che coinvolge principalmente docenti e dirigenti. Un'ampia letteratura di ricerca in merito mette in luce il legame tra l'incremento degli sforzi di autovalutazione e il miglioramento dell'offerta formativa della scuola, delle procedure e anche degli esiti di apprendimento. Diverse ricerche hanno evidenziato come l'attuazione di pratiche autovalutative di istituto possa favorire il miglioramento della didattica e dell'apprendimento degli studenti<sup>28</sup>. Rassegne di ricerca rilevanti a tale proposito sono state prodotte all'interno dell'*International School Inspection Project*, che coinvolge sei Paesi europei<sup>29</sup>, con lo scopo di individuare quali approcci alla valutazione della scuola possano favorire maggiormente il miglioramento della stessa<sup>30</sup>. Da tali rassegne emerge che la variabile intermedia, che spiegherebbe l'incremento degli esiti degli studenti a fronte dell'applicazione di pratiche autovalutative, è il cosiddetto *process use*. Quest'ultimo può essere definito come il cambiamento in termini cognitivi, di comportamento, organizzativi o di linee progettuali, derivante, in maniera diretta o indiretta, dal coinvolgimento degli *stakeholders* (nel nostro caso in particolare degli insegnanti e dei dirigenti scolastici) nel processo di valutazione e autovalutazione. Esso si verifica quando le persone coinvolte nella valutazione imparano dal processo stesso e attuano cambiamenti nell'intervento o nelle loro pratiche sulla base di tale processo e non solo a partire dai risultati ottenuti al termine della valutazione<sup>31</sup>.

Gli effetti del *process use* si riscontrano a diversi livelli. Si osservano, ad esempio: lo sviluppo di conoscenze sul sistema valutato e di competenze nella valutazione, che possono essere anche insegnate ad altri; l'integrazione delle pratiche valutative nella progettazione di nuovi interventi; un atteggiamento di maggior disponibilità a trasferire le indicazioni che derivano dal processo valutativo alla pratica professionale. Si rileva in particolare l'acquisizione della capacità di “pensare in maniera valutativa”.

Tali effetti possono essere più o meno ampi a seconda dell'intensità del coinvolgimento dei diversi attori; della motivazione dei partecipanti a impegnarsi nel processo valutativo; delle loro esperienze pregresse nella valutazione; del grado di fiducia rispetto alla possibilità di utilizzare i risultati ottenuti; del supporto organizzativo offerto

<sup>28</sup> Cfr. K. LEITHWOOD-R. STEINBACH-D. JANTZI, *School Leadership and Teachers' Motivation to Implement Accountability Policies*, in “Educational Administration Quarterly”, 38 (2002), pp. 94-119.

<sup>29</sup> Austria, Gran Bretagna, Irlanda, Olanda, Repubblica Ceca, Svezia.

<sup>30</sup> Cfr. M.C.M. EHREN-G. ALTRICHTER-G. MCNAMARA-J. O'HARA, *Impact of Schools Inspections on Improvement of Schools. Describing Assumptions on Casual Mechanisms in six European Countries*, in “Educational Assessment, Evaluation and Accountability”, 25 (1/2013), pp. 3-43.

<sup>31</sup> Cfr. M.Q. PATTON, *Utilization-Focused Evaluation. The New Century Text*, Sage, Thousand Oaks 1997<sup>4</sup>.

al lavoro valutativo e degli eventuali vincoli imposti dall'interno dell'istituzione o dall'esterno<sup>32</sup>.

Il coinvolgimento diretto degli insegnanti nel processo valutativo della scuola può dunque contribuire a renderli più disponibili ad essere posti "sotto osservazione" e a fare un uso proficuo degli esiti che ne derivano.

Emerge però, dagli studi, una difficoltà a realizzare autovalutazioni adeguate<sup>33</sup>. L'efficacia delle strategie di valutazione del proprio operato dipende infatti dagli aspetti considerati, dagli strumenti utilizzati e dalle competenze di autoanalisi dei soggetti. Rispetto alle autovalutazioni dei docenti, per esempio, la ricerca ha messo in luce la difficoltà degli insegnanti meno strumentati di esaminarsi in maniera rispondente al reale. Tali difficoltà possono essere superate costituendo équipe in cui siano presenti anche insegnanti con maggiore esperienza o comunque docenti che presentano migliori capacità di sviluppare la loro professione sulla base dei feed-back ricevuti. La pianificazione comune di azioni migliorative sulla base delle stimolazioni ricevute e l'interazione con un supervisore può consentire a tutti di progredire<sup>34</sup>.

Occorre però anche che il sistema restituisca gli esiti delle autovalutazioni in modo adeguato e significativo per far evolvere la professionalità docente e indirizzare in maniera mirata le occasioni di formazione continua<sup>35</sup>.

#### 3.4. *Valutazione di un complesso insieme di elementi*

Nella valutazione della scuola e del sistema scolastico si sono progressivamente considerati e integrati anche aspetti ulteriori rispetto alle performance degli studenti. Si tratta delle qualità psicosociali degli ambienti di apprendimento che consentono di interpretare meglio gli esiti in ciascun contesto e di pianificare correttivi più adeguati. Tali aspetti, a cui è stata attribuita minor importanza nelle rilevazioni estensive, risultano però importanti per consentire un incremento della "qualità effettiva" e favorire così il successo formativo degli alunni. Anche le indagini PISA hanno messo in evidenza come il clima di classe, le strategie di gestione della disciplina, alti livelli valoriali e di impegno degli studenti siano connessi con elevati risultati di apprendimento<sup>36</sup>. Può essere dunque

---

<sup>32</sup> Cfr. J.B. COUSIN, *Process Use in Theory, Research and Practice*, Jossey Bass, San Francisco 2007 e S. JACOB-L. OUVARD-J.-F. BÉLANGER, *Participatory Evaluation and Process Use within a Social Aid Organization for At-risk Families and Youth*, in "Evaluation and Program Planning", 34 (2011), pp. 113-123.

<sup>33</sup> Cfr. K.F. NAYIR-G. MCNAMARA, *The Increasingly Central Role of School Self-Evaluation in Inspection System across Europe: The Case of Ireland*, ed. cit.

<sup>34</sup> Cfr. P. ANTONIOU-L. KYRIAKIDES, *The Impact of a Dynamic Approach to Professional Development on Teacher Instruction and Student Learning: Results from an Experimental Study*, in "School Effectiveness and School Improvement", 22 (2011), pp. 291-311.

<sup>35</sup> Cfr. L.E. FRASE-W. STRESHLY, *Lack of Accuracy, Feedback, and Commitment in Teacher Evaluation*, in "Journal of Personnel Evaluation in Education", 8 (1994), pp. 47-57 e L.A. NOAKES, *Adapting the Utilization-Focused Approach for Teacher Evaluation*, in "Journal of MultiDisciplinary Evaluation", 6 (11/2009), pp. 83-88.

<sup>36</sup> L'analisi condotta fin dalle prime rilevazioni (PISA, *Learning for tomorrow's world: First results from PISA 2003*, OECD, Paris 2004) è stata sviluppata nelle successive: sono stati esplorati i diversi fattori che incidono sul successo formativo e si è approfondita l'evoluzione dell'equità nell'istruzione nei vari paesi nel corso degli anni (OECD, *PISA 2012 Results: Excellence Through Equity: Giving Every Student the Chance* 120

utile un impiego più frequente di strumenti appositamente predisposti per rilevare tali aspetti, come il sussidio svedese GAVIS (*Goals, Attitudes and Values in School*). Si tratta di una scala strutturata sulla base del modello di apprendimento di Schwartz<sup>37</sup>, che si propone di rilevare quanto l'ambiente di apprendimento sia in grado di stimolare: la creatività; il piacere e la soddisfazione di ciascuno; lo sviluppo personale e il successo di tutti; la percezione di competenza, di sicurezza (assenza di pericoli, violenze e molestie) e di controllo (presenza di regole, organizzazione e autoregolazione); la cooperazione e il supporto tra gli studenti; la partecipazione di ciascuno nel gruppo; l'assunzione di responsabilità; la possibilità di esprimere le proprie idee sui contenuti trattati e di essere ascoltati<sup>38</sup>.

L'introduzione sistematica di strumenti di questo tipo consentirebbe, nel processo di "riesame" delle procedure e di proposta dei correttivi, di effettuare variazioni che toccano la "qualità sostanziale".

#### 4. Conclusione

Nel corso del presente contributo si è cercato di individuare quali strategie valutative della scuola e del sistema scolastico possono essere utilmente attivate, e come le stesse possano favorire il successo formativo degli alunni anche nei contesti più svantaggiati. Nell'articolo si è sottolineata l'importanza a tal fine di sviluppare una valutazione più olistica, estesa ad aspetti anche più qualitativi dell'efficacia degli ambienti di apprendimento, e di favorire un miglior utilizzo dei dati rilevati, attraverso la diffusione di strategie di autovalutazione, accompagnate da confronto tra pari e supervisione.

---

to Succeed, vol. 2, OECD, Paris 2013, <http://www.oecd.org/pisa/keyfindings/pisa-2012-results-volume-ii.htm>).

<sup>37</sup> Cfr. S.H. SCHWARTZ, *Universals in the Content and Structure of Values: Theoretical Advances and Empirical Tests in 20 Countries*, in M.P. ZANNA (a cura di), *Advances in Experimental Psychology*, Academic Press, London 1992, pp. 1-65 e S.H. SCHWARTZ-K. BOEHNKE, *Evaluating the Structure of Human Values with Confirmatory Factor Analysis*, in "Journal of Research in Personality", 38 (3/2004), pp. 230-255.

<sup>38</sup> Cfr. M. WESTLING ALLODI, *Assessing the Quality of Learning Environments in Swedish Schools: Development and Analysis of a Theory-based Instrument*, in "Learning Environment Research", 10 (3/2007), pp. 157-175 e M. WESTLING ALLODI, *Goals and Values in School: A Model Developed for Describing, Evaluating and Changing the Social Climate of Learning Environments*, in "Social Psychology of Education", 13 (2/2010), pp. 207-235.



Maria Cassella  
Donatella Mutti

## SU ALCUNI ASPETTI DELLA VALUTAZIONE DELLA RICERCA NELLE SCIENZE UMANE. BIBLIOMETRIA E DINTORNI

### Abstract

*Society and policy makers are asking universities and research institutions to be accountable for public investment. Developing performance measures for assessing research quality and impact has therefore become a hot topic, on which a great growing debate has been raging for the past few years. For basic research in STM (Science Technology Medicine) disciplines, there are fairly well established indicators of research outputs, based on publications and citations, but they seem not to be applicable to the humanities and the social sciences. This essay examines the main aspects of research assessment in the humanities, its history, and several issues that have emerged through experiences in Italy and the European countries.*

«Se la bibliometria è una sorta di male necessario nell'epoca della big science e della iperspecializzazione dei profili professionali, conviene allora promuoverne una conoscenza e una pratica il più possibile approfondite, in linea con gli standard e le migliori esperienze internazionali, ma al tempo stesso critiche, in linea con lo status di scienza sociale (e non di scienza esatta o matematica applicata) che le compete».

(NICOLA DE BELLIS, 2014)

### 1. Introduzione

La valutazione della ricerca è un processo decisionale regolato da norme e criteri codificati fondamentale per l'avanzamento della scienza. Processo complesso e multidimensionale<sup>1</sup>, in Italia il tema della valutazione della ricerca ha subito, a partire dal

---

<sup>1</sup> Il riferimento alla valutazione della ricerca come processo multidimensionale è tratto da H. MOED-A. PLUME, *The Multi-dimensional Research Assessment Matrix*, in "Research Trends", 23 (2011), <http://www.researchtrends.com/issue23-may-2011/the-multi-dimensional-research-assessment-matrix/>.

secondo esercizio di valutazione nazionale (Valutazione Qualità della Ricerca), una rapida accelerazione e maturazione. Il mondo accademico è stato coinvolto, giocoforza, in un grande dibattito formale e informale sui temi della valutazione. Sulla necessità/opportunità di una valutazione sistematica sembrano ormai concordare tutti gli attori coinvolti nel tema: docenti e ricercatori, società scientifiche, società professionali, editori, cultori della materia, *policy makers*, enti finanziatori della ricerca ecc. Sui metodi, i tempi e le tecniche di valutazione della ricerca il dibattito è fervente e tocca in modo molto diretto e particolare le scienze sociali e umane, per le quali, per differenti motivazioni storiche ed epistemologiche, le metodologie di valutazione appaiono poco consolidate. Soprattutto le scienze umane e sociali non sono state toccate fino ad oggi dall'approccio bibliometrico.

In questo articolo approfondiremo alcuni temi centrali nella valutazione della ricerca nelle scienze umane: dalla bibliometria come opportunità da non demonizzare, al tema dell'internazionalizzazione e a quello della valutazione delle monografie di ricerca. Prima di entrare nella riflessione sulla bibliometria applicata alle scienze umane ci preme, tuttavia, inquadrare il dibattito sulla valutazione quantitativa della ricerca in un contesto storico e culturale più ampio, ripercorrendo le tappe fondamentali della nascita della "scientometria" e della "bibliometria" quale branca specialistica della scientometria, evidenziando anche il contributo della sociologia della scienza alla riflessione sulla scienza moderna e sui suoi meccanismi di competizione e valutazione.

## 2. *All'origine della bibliometria e degli indici bibliometrici citazionali*

La nascita della bibliometria si colloca concettualmente, e forse non casualmente, nell'ambito della riflessione scientifica e biblioteconomica. Nel 1926 il chimico e matematico Alfred J. Lotka, presidente dell'American Statistical Society, pubblica un articolo sul "Journal of Washington Academy of Sciences" nel quale affronta in maniera pionieristica il problema di valutare la produttività scientifica dei chimici. Lotka analizza la distribuzione degli articoli indicizzati nel periodo 1907-1916 dalla rivista "Chemical Abstracts" e arriva a formulare la seguente legge empirica: «il numero di autori che hanno scritto almeno N articoli è (statisticamente) proporzionale a  $1/N$  alla seconda del numero totale di autori censiti»<sup>2</sup>. L'anno successivo P.L.K. Gross e E.M. Gross, due coniugi, chimici del Pomona college, lanciano sulla rivista "Science"<sup>3</sup> per primi l'idea di utilizzare le citazioni per individuare le riviste scientifiche più rilevanti nel settore della chimica in modo che ogni biblioteca di università possa selezionarle ed includerle nelle proprie collezioni. I Gross individuano il seguente metodo: considerare tutte le riviste contenute nelle citazioni degli articoli pubblicati sul "Journal of the American Chemical Society" e calcolare per ciascuna il numero di citazioni ricevute nei cinque anni successivi. Con lo studio dei Gross nasce la *citational analysis* (o *citation studies*) che

---

<sup>2</sup> L. MODICA, *Passato e futuro della ricerca universitaria: valutare cosa? Valutare come? Valutare perché?*, in P. MICCOLI-A. FABRIS (a cura di), *Valutare la ricerca? Capire, applicare, difendersi*, ETS, Pisa 2012, p. 15.

<sup>3</sup> P.L.K. GROSS-E.M. GROSS, *College Libraries and Chemical Education*, in "Science", 66 (1713/1927), pp. 385-389.

diventerà più tardi una parte fondamentale degli studi bibliometrici. La legge bibliometrica che anticipa l'opera di Garfield è la legge di Samuel C. Bradford, bibliotecario presso il Museo della Scienza di Londra. Bradford descrive la sua legge in un articolo pubblicato nel 1934 sulla rivista "Engineering" e, successivamente, nel volume *Documentation* dimostrando che la maggior parte degli articoli significativi si concentrano in un numero limitato di riviste che di quel settore costituiscono il "nucleo" (*core*). «È comprensibile che tra le leggi bibliometriche la legge di Bradford sia stata il riferimento principale di Garfield, specificatamente per la scelta dei *core journals* da includere nel database dell'ISI»<sup>4</sup>.

A metà degli anni Cinquanta matura negli Stati Uniti il dibattito culturale sulla crescita esponenziale delle pubblicazioni scientifiche<sup>5</sup>. L'incremento nel numero di riviste accademiche, il boom economico e la necessità di sviluppare sistemi di *intelligence* sempre più sofisticati spingevano a cercare una soluzione al problema del recupero dell'informazione che fosse quanto più possibile standardizzata e controllata meccanicamente. Eugene Garfield comincia, quindi, a concepire il Science Citation Index, un registro contenente le riviste "core" in ambito scientifico, completo di indice delle citazioni. L'idea di Garfield era, all'origine, quella di creare uno strumento che servisse ai ricercatori per selezionare le riviste più rilevanti nel proprio settore di ricerca (*core journals*) e mettesse in evidenza le reti di relazioni attive nella comunicazione scientifica. «Gli indici di citazione nascono in quanto nuova metodologia di recupero ed organizzazione dell'informazione; in base a tale metodologia è possibile scoprire i legami esistenti tra i documenti scientifici nel tempo e cogliere i rapporti intrinseci tra ambiti disciplinari e tematiche anche molto distanti»<sup>6</sup>. Solo in un secondo momento Garfield intravvide la strada dell'applicazione degli indici citazionali per la valutazione della ricerca, proponendola in un articolo pubblicato su "Science": «*In effect the system would provide a complete listing, for the publications covered, of all the original articles that had referred to the article in question. This would be clearly particularly useful in historical research, when one is trying to evaluate the significance of a particular work and its impact on the literature and thinking of the period*»<sup>7</sup>. Così progressivamente l'idea dei *core journals* veniva ad identificarsi sempre meno con il concetto di "centralità" e sempre più con i concetti di "rilevanza" e di "qualità". Cinque anni più tardi Garfield fonda l'Institute of Scientific Information (ISI) che, a partire dal 1964, pubblica il "Science Citation Index", successivamente affiancato dal

<sup>4</sup> R. DI CESARE, *Alcune riflessioni su bibliometria e analisi delle citazioni*, in A. VALENTE (a cura di), *Trasmisione d'élite o accesso alle conoscenze? Percorsi e contesti della documentazione e comunicazione scientifica*, F. Angeli, Milano 2002, p. 134.

<sup>5</sup> Ricordiamo i contributi di: H.E. BLISS, *The Organization of Knowledge and the Subject-approach to Books*, Wilson, New York 1933; J.D. BERNAL, *The Social Function of Science*, Routledge, London 1939; ID., *Information Service as an Essential in the Progress of Science*, in *Report of the Proceedings of the 20<sup>th</sup> Conference of ASLIB*, ASLIB, London 1945; ID., *Preliminary Analysis of Pilot Questionnaire on the Use of Scientific Literature*, Royal Society, London 1948; H.G. WELLS, *World Brain*, Doubleday, Doran & Co., New York 1938.

<sup>6</sup> A. VALENTE, *Gli indici di citazioni nel circuito di organizzazione, selezione e comunicazione di conoscenza scientifica*, in ID. (a cura di), *Trasmisione d'élite o accesso alle conoscenze? Percorsi e contesti della documentazione e comunicazione scientifica*, ed. cit., p. 76.

<sup>7</sup> E. GARFIELD, *Citation Indexes for Science*, in "Science", 122 (3159/1955), pp. 108-111.

“Social Sciences Citation Index” (1973), quindi dall’“Arts & Humanities Citation Index”. Più recente è la pubblicazione del “Conference Proceedings Citation Index” (2009) e del “Book Citation Index” (2013)<sup>8</sup>.

Negli stessi anni nei quali Garfield ideava e cominciava a sviluppare commercialmente gli indici citazionali, un fisico statunitense, Derek John de Solla Price, studiava i metodi quantitativi applicati all’analisi della scienza e fondava la “scientometria”, disciplina che studia «la scienza come prodotto sociale e culturale attraverso l’analisi della produzione, diffusione e circolazione dell’informazione scientifica»<sup>9</sup>.

Del 1969 è il primo utilizzo del termine “bibliometria” (*bibliometrics*) che viene messo in evidenza in un articolo pubblicato da Alan Pritchard. Nel “Journal of Documentation” Pritchard dà la seguente definizione di bibliometria: «*the application of mathematics and statistical methods to books and other media of communication*»<sup>10</sup>. Mentre si consolidavano i confini concettuali di “scientometria” e di “bibliometria” e si perfezionavano le tecniche bibliometriche (ci riferiamo, ad esempio, alle teorie relative alla normalizzazione degli indici bibliometrici), sempre negli Stati Uniti il sociologo statunitense Robert K. Merton avvia la riflessione sulla scienza moderna e individua i quattro imperativi categorici che fondano l’etica della scienza e regolano il lavoro dello scienziato: universalismo, comunismo, disinteresse, scetticismo organizzato. Il disinteresse è il terzo imperativo etico della scienza. Il ricercatore deve essere distaccato emotivamente dal suo lavoro e perseguire i propri obiettivi in modo lecito. Nelle situazioni di competizione è raro che gli scienziati adottino mezzi illeciti. La competizione tra studiosi accademici è una condizione della scienza che si autoregola grazie all’applicazione dei principi etici mertoniani. Infatti: «la distribuzione dei riconoscimenti [...] avviene secondo regole competitive di merito, decise collettivamente, dalle comunità scientifiche che si autoregolano»<sup>11</sup>.

La *peer review* e la citazione sono i due processi fondamentali attraverso i quali le comunità di ricerca riconoscono il merito reciproco: le distorsioni nell’applicazione del processo di revisione tra pari e nel comportamento citazionale, anche se presenti, tendono, secondo il principio mertoniano, ad annullarsi. Scrive A. Bonaccorsi, membro dell’Anvur:

«nel sistema scientifico ci sono semplicemente troppi attori e tutti troppo bene informati perché qualcuno possa manipolare la situazione a proprio vantaggio. [...] Anche ammettendo la possibilità di manipolazione di status o clanistica l’idea di fondo è che le comunità sono talmente ampie da spazzare via nel tempo questi fenomeni»<sup>12</sup>.

---

<sup>8</sup> Il “Book Citation Index” è un indice citazionale di circa 30.000 titoli di monografie integrato nel Web of Science. È stato concepito proprio in funzione della valutazione della ricerca nelle scienze umane e sociali che hanno come principale prodotto della ricerca scientifica la monografia.

<sup>9</sup> R. DI CESARE, *Alcune riflessioni su bibliometria e analisi delle citazioni*, in A. VALENTE (a cura di), *Trasmissione d’élite o accesso alle conoscenze? Percorsi e contesti della documentazione e comunicazione scientifica*, ed. cit., p. 133.

<sup>10</sup> A. PRITCHARD, *Statistical bibliography or bibliometrics?*, in “Journal of Documentation”, 25 (4/1969), p. 349.

<sup>11</sup> A. BONACCORSI, *Ancora sulla valutazione nelle aree umanistiche e sociali*, in P. MICCOLI-A. FABRIS (a cura di), *Valutare la ricerca? Capire, applicare, difendersi*, ed. cit., p. 89.

<sup>12</sup> *Ibidem*, ivi.

La competizione porta ineludibilmente con sé il meccanismo della valutazione: dei singoli, dei gruppi di ricerca, delle istituzioni accademiche.<sup>13</sup> Particolarità epistemologiche, varietà di prodotti della ricerca e frammentazione di “posizioni umanistiche” hanno fatto sì che, fino ad oggi, la bibliometria non sia mai stata applicata alla valutazione della ricerca prodotta in area umanistica. Su questo tema rifletteremo nel prossimo capitolo.

### 3. Citazioni, indici citazionali e bibliometria nelle scienze umane

Non è possibile, e non è tra gli scopi di questo articolo, fare riferimento all’immensa letteratura che si è occupata del significato delle citazioni nelle diverse discipline di ricerca, degli indici citazionali (*Impact Factor*, *H-index* e sue varianti, *Eigenfactor* ecc.) e degli usi ed abusi degli indici citazionali nella loro applicazione concreta alla valutazione della ricerca<sup>14</sup>.

In relazione alle discussioni emerse e/o emergenti nella letteratura professionale e nel dibattito accademico sui temi sopra elencati, preme qui approfondire alcune riflessioni. La prima è se la citazione sia una metrica che, a prescindere dagli obiettivi di un esercizio di valutazione della ricerca, possa avere una funzione anche per le *humanities*. Da un lato appare condivisibile la critica – che emerge nella letteratura sul tema – secondo cui la citazione di un articolo o di qualsiasi altro tipo di pubblicazione non indica assiomaticamente un giudizio di qualità, dal momento che gli articoli possono essere citati anche in negativo, le teorie criticate o rigettate (*negational reference*)<sup>15</sup>. Dall’altro ci sembra possa essere logicamente sostenibile che la citazione equivale a una manifestazione di interesse da parte di chi cita e rappresenta una serie di relazioni tra autori appartenenti a una stessa area disciplinare o ad aree disciplinari diverse. «La citazione è una sorta di dialogo, personale e concreto, seppure indiretto e a distanza, con l’autore che si sta citando, al quale si esprime attraverso la citazione, in genere, riconoscimento, accordo»<sup>16</sup>.

Certamente appare diverso il significato delle citazioni tra il segmento delle scienze dure e quello delle scienze umane. Nel primo caso, infatti, la citazione viene utilizzata per

<sup>13</sup> Tre sono i livelli di applicazione nella valutazione della ricerca: micro, meso e macro. Le istituzioni accademiche appartengono al livello macro. La competizione tra istituzioni accademiche viene incoraggiata, oltre che dagli esercizi di valutazione nazionali, anche da alcune note classifiche come quella del Times Higher Education, la classificazione dell’università di Leiden, la classificazione di Shangai e la più recente [Best Global Universities](#) di *U.S. News and World Reports*.

<sup>14</sup> Per un’ottima ed esaustiva sintesi della letteratura sugli indici bibliometrici si rimanda ad A. BACCINI, *Valutare la ricerca scientifica. Uso ed abuso degli indicatori bibliometrici*, Il Mulino, Bologna 2010.

<sup>15</sup> Cfr. M.H. MACROBERTS-B.R. MACROBERTS, *The Negational Reference: on the Art of Dissembling*, in “Social Studies of Science”, 14 (1984), pp. 91-94.

<sup>16</sup> Diversamente dalla citazione, la referenza ha un carattere più impersonale e rimanda alle fonti di una pubblicazione. Cfr. R. DI CESARE, *Alcune riflessioni su bibliometria e analisi delle citazioni*, in A. VALENTE (a cura di), *Trasmissione d’élite o accesso alle conoscenze? Percorsi e contesti della documentazione e comunicazione scientifica*, ed. cit., p. 140.

sostenere o confutare una teoria; uno studioso delle discipline scientifiche deve inevitabilmente fare riferimento a tutti i contributi che rendono verificabile e ripetibile la sua teoria. Le citazioni hanno vita breve nel tempo. Nelle scienze umane, invece, il valore semantico delle citazioni è di «ricapitolazione, di nuova sintesi di contributi anche lontani nel tempo»<sup>17</sup>. Nelle aree umanistiche la funzione della citazione può essere influenzata in modo negativo anche dal pluralismo paradigmatico che caratterizza le scienze umane.

Tale pluralismo rappresenta, infatti, per gli umanisti una ricchezza, un valore da tutelare ma può sollecitare una certa segmentazione nelle citazioni e comportamenti faziosi a favore di un filone di pensiero piuttosto che di un altro. È frequente, infatti, che i membri di una scuola si citino tra loro, ma non citino chi appartiene a scuole di pensiero contrapposte. Il meccanismo si rivela particolarmente perverso per gruppi di ricerca di piccole dimensioni, che si scoprono svantaggiati sia dalla logica quantitativa delle citazioni che dall'impossibilità di adottare una *peer review* realmente anonima e imparziale. Questi comportamenti faziosi, che potremmo definire “scarsamente mertoniani”, tendono al momento, per fortuna, progressivamente ad attenuarsi grazie ai processi di internazionalizzazione della ricerca che inducono le comunità scientifiche ad aprirsi e a collaborare sempre più dentro e fuori dall'accademia. Quanto sopra esposto rivela unicamente che esiste una differenza di significato tra la citazione nelle scienze dure e nelle scienze umane, ma non dimostra che la citazione non ha alcun significato per gli umanisti. Quanto al ruolo della citazione nessuna conclusione sembra, invero, generalizzabile: «chi ha provato ad approfondire il ruolo delle citazioni nella comunicazione scientifica non ha trovato risultati inequivocabili, né ha raggiunto conclusioni generalizzabili: talvolta sembrano prevalere ragioni normative di stampo mertoniano, talaltra prende il sopravvento la funzione puramente cosmetica delle bibliografie. In nessun caso, purtroppo, i filoni d'indagine sul significato delle citazioni hanno incrociato la strada della valutazione della ricerca, dove prevale un modello di conteggio indifferente a contenuti e contesti di provenienza delle *cited references*»<sup>18</sup>.

Esaminiamo ora più da vicino il tema della bibliometria applicata alle scienze umane. Il tema è ampiamente dibattuto da alcuni anni sia sul piano internazionale<sup>19</sup> che sul piano nazionale. In Italia la discussione sulla bibliometria applicata alle aree non bibliometriche (le aree CUN 10-14) si è aperta con la VQR nel momento in cui è nata l'esigenza di affiancare alla revisione tra pari altre metodologie di valutazione per le aree per le quali appunto la bibliometria non era, e non è, del tutto applicabile tanto da suggerire l'adozione di soluzioni pseudo-bibliometriche come l'*informed peer review* ovvero

---

<sup>17</sup> A. BONACCORSI, *Ancora sulla valutazione nelle aree umanistiche e sociali*, ed. cit., p. 92.

<sup>18</sup> N. DE BELLIS, *Introduzione alla bibliometria: dalla teoria alla pratica*, AIB, Roma 2014. La citazione è tratta dall'edizione epub dell'opera.

<sup>19</sup> In tale direzione si era mossa, ad esempio, nel 2000 la European Science Foundation. Cfr. anche il report finale dello European Scoping Project: B. MARTIN *et al.*, *Towards a Bibliometric Database for the Social Sciences and Humanities: a European Scoping Project*, 2010 ([http://www.dfg.de/download/pdf/foerderung/grundlagen\\_dfg\\_foerderung/informationen\\_fachwissenschaften/geisteswissenschaften/esf\\_report\\_final\\_100309.pdf](http://www.dfg.de/download/pdf/foerderung/grundlagen_dfg_foerderung/informationen_fachwissenschaften/geisteswissenschaften/esf_report_final_100309.pdf)).

l'adozione di liste “classificate” di titoli di riviste<sup>20</sup>. Si è discusso di bibliometria in seno all'Anvur che si è espressa in modo favorevole<sup>21</sup>, ne hanno discusso i rapporti finali dei GEV. Per le aree 10-14 i rapporti forniscono un quadro non omogeneo, tra aree (ad esempio l'area 13) che hanno utilizzato sia *peer review* che valutazione bibliometrica e aree (ad esempio l'area 10) che hanno utilizzato come metodo unico di valutazione quello della revisione tra pari. Un picco di interesse rispetto al tema della bibliometria applicata alle scienze umane si è avuto a gennaio 2014 quando l'Anvur ha presentato in un seminario pubblico la proposta di realizzare una banca dati bibliografica/bibliometrica per le scienze umane e sociali. L'idea dell'Anvur va nella direzione di valorizzare la produzione scientifica nelle discipline umanistiche e si innesta in un *bouquet* di iniziative in corso in diversi paesi volte a sviluppare sistemi nazionali di documentazione scientifica e a creare basi dati bibliografiche nazionali. Nel documento *Specifiche preliminari per una base dati bibliometrica italiana nelle aree umanistiche e sociali* l'Anvur ha fornito le specifiche del database che si intenderebbe realizzare: il database includerebbe in via sperimentale le riviste di sola fascia A; gli editori dovrebbero fornire il *full-text* e i metadati degli ultimi dieci anni delle riviste pubblicate; il database sarebbe finalizzato alla ricerca e alla valutazione della ricerca grazie all'estrazione delle citazioni dal *corpus* degli articoli considerati.

Le comunità scientifiche italiane hanno reagito in modo critico di fronte alla proposta di creare un database bibliografico/bibliometrico. Alcune critiche appaiono motivate: 1. non convince la decisione di includere nel perimetro di azione della banca dati le sole riviste di fascia A escludendo, ancora una volta, le monografie e i capitoli di monografie; 2. non viene definito il costo complessivo dell'operazione che sembrerebbe ricadere prevalentemente sugli editori, laddove andrebbero calcolati anche una serie di costi indiretti che si esternalizza sugli autori e sulle biblioteche. Inoltre, la metodologia proposta per la realizzazione del database «presuppone una serie di importanti operazioni preliminari di metodo (tecnica della citazione), di merito (natura e classificazione delle riviste di cui si raccolgono le citazioni) e tecniche (software) con effetti standardizzanti di cui non si valuta né il risvolto culturale né il peso che è addossato a soggetti esterni»<sup>22</sup>.

Sotto il profilo più puramente concettuale il rapporto tra scienze umane e bibliometria appare sì complesso, certamente ancora poco maturo, ma non così conflittuale come potrebbe sembrare dalla discussione in atto. In primo luogo è stato ripetutamente sottolineato – e lo ha ripetutamente sottolineato anche l'Anvur – che la bibliometria è solo una delle possibili metriche per valutare la ricerca e, quando viene utilizzata, lo è in combinazione con altre metodologie di valutazione. Limiti metodologici esistono nella

---

<sup>20</sup> Per *informed peer review* si intende una valutazione qualitativa assistita da liste di riviste, liste di editori o da griglie di valutazione

<sup>21</sup> Cfr. A. BONACCORSI, *Potenzialità e limiti della analisi bibliometrica nelle aree umanistiche e sociali. Verso un programma di lavoro*, 7 marzo 2012 ([http://www.anvur.org/attachments/article/44/valutazione\\_aree\\_umanistiche\\_e\\_sociali.pdf](http://www.anvur.org/attachments/article/44/valutazione_aree_umanistiche_e_sociali.pdf)).

<sup>22</sup> M. CAMMELLI, *Anvur, data base bibliometrica italiana aree umanistica e sociali: note a margine*, 18 gennaio 2014 (<http://www.roars.it/online/anvur-data-base-bibliometrica-italiana-aree-umanistiche-e-sociali-note-a-margine/>).

*peer review*<sup>23</sup>, nell'*informed peer review* e, perfino, nelle più recenti e fortunate metriche alternative del web (*altmetrics*).

Una riflessione critica che viene portata sovente contro l'utilizzo della bibliometria è che l'adozione di indici bibliometrici crea comportamenti adattivi da parte dei ricercatori e favorisce la diffusione della scienza *mainstream*. Data la complessità dei meccanismi che regolano la ricerca questa visione appare riduttiva. Infatti, a nostro avviso, non è tanto, né unicamente la bibliometria che favorisce la scienza *mainstream*, quanto le scelte degli enti finanziatori che tendono a privilegiare temi di tendenza e di impatto sulla società. Le scelte di chi finanzia non sono sempre razionali e i temi di nicchia con scarso impatto sul pubblico o i temi eccessivamente innovativi tendono, nel primo caso, a richiamare scarsa attenzione e, nel secondo, ad essere scarsamente compresi.

Un terzo argomento che viene portato contro l'utilizzo della bibliometria quale metodologia di valutazione idonea a misurare la ricerca di area umanistica è che la monografia che resta ad oggi, nonostante alcuni segnali di cambiamento, il principale prodotto della ricerca in area umanistica è scarsamente presente nella copertura tipologica delle due principali banche dati citazionali, entrambe di natura commerciale: Scopus e Web of Science. Quanto a quest'ultima riflessione, ci sembra che essa possa offrire una duplice chiave di lettura: da un lato, misurare bibliometricamente le scienze umane con gli strumenti attualmente a nostra disposizione è una forzatura e conduce a risultati inadeguati laddove «i sistemi di comunicazione scientifica dovrebbero potersi avvalere di strumenti documentari e indici che includano le realtà nazionali e locali, le differenti scuole e i differenti approcci alla conoscenza»<sup>24</sup>. Dall'altro, le stesse comunità di umanisti sollecitano ormai una più ampia copertura geografica, disciplinare e tipologica sia in Scopus che in WoS. Concretamente, un numero crescente di riviste di area umanistica aspira ad essere indicizzata dalle due principali banche dati citazionali, così come, contestualmente, sta crescendo anche il numero di monografie indicizzate sia in Scopus che in WoS. Segno di un riconoscimento implicito che le basi dati citazionali hanno maturato anche nell'ambito delle scienze umane e del fatto che la bibliometria tende a maturare un significato anche per le discipline umanistiche. In Europa un caso interessante di database bibliometrici per le scienze umane e sociali sono gli archivi spagnoli IN~RECS (Índice de Impacto Revistas Españolas de Ciencias Sociales), IN~RECH (Índice de Impacto Revistas Españolas de Ciencias Humanas) e IN~RECJ (Índice de Impacto Revistas Españolas de Ciencias Jurídicas), a cura del gruppo di ricerca EC3 delle Università di Granada e Navarra.

Come uscire dalla *querelle* della bibliometria applicata alle scienze umane e sociali? Crediamo che un approccio maturo e consapevole rispetto al tema vada perseguito uscendo dai confini della discussione "amatoriale" sui temi della valutazione ed entrando nella sfera della ricerca e della ricerca-azione. In Italia l'Anvur aveva annunciato nel suo programma di lavoro 2013-2015 la nascita di un Centro studi sulla valutazione «cui affidare il compito di raccogliere documentazione e di coordinare, svolgere e pubblicare

---

<sup>23</sup> Del resto l'idea di una presunta oggettività degli indici citazionali scaturisce proprio da un utilizzo poco corretto della *peer review*.

<sup>24</sup> A. VALENTE, *Gli indici di citazioni nel circuito di organizzazione, selezione e comunicazione di conoscenza scientifica*, ed. cit, p. 92.

in modo regolare studi e ricerche sulla valutazione della qualità del sistema universitario, della didattica e della ricerca pubblica nel suo insieme»<sup>25</sup>. L'idea di un Centro studi, diramazione dell'Anvur, è stata criticata dalle comunità scientifiche italiane, in quanto non offrirebbe sufficienti garanzie di autonomia nelle scelte valutative. In Europa sono diffusi i centri specializzati nella valutazione della ricerca con una forte propensione verso la bibliometria; fanno capo a università (ad esempio: il CWTS dell'università di Leiden e la SPRU dell'università del Sussex), a biblioteche (l'ISSRU presso la biblioteca dell'Accademia Ungherese delle Scienze di Budapest) o a consorzi interuniversitari (l'ECOOM in Belgio). È innegabile che in Italia sia pressante l'esigenza di innalzare il livello della discussione sulla valutazione della ricerca attraverso studi e analisi di tipo sociologico e bibliometrico; si potrebbe anche partire da ciò che già esiste ovvero da quei gruppi di ricerca che già lavorano su questi temi per dare vita a un centro interuniversitario di analisi e ricerche che sia indipendente rispetto all'Anvur. In questo contesto problematico, talvolta confuso, ci sembra possa condurre a sviluppi positivi la decisione dell'Anvur di stimolare studi e ricerche sulla valutazione in area umanistica attraverso il finanziamento di sei progetti di ricerca<sup>26</sup> e la realizzazione di quattro gruppi di lavoro tematici: uno per le aree 10-11, uno per l'area 12, uno per l'area 13 e l'ultimo per l'area 14.

Volgendo ora nuovamente la nostra attenzione alla banca dati bibliografica italiana per le scienze umane e sociali crediamo che, nonostante la complessità di realizzazione, essa rappresenti un passaggio obbligato per le discipline umanistiche; infatti, un database di questo tipo avrebbe tra gli altri vantaggi quelli di: 1. accrescere, a livello nazionale e internazionale, la visibilità della ricerca umanistica prodotta in Italia, ponendo fine ad anni di anonimato e di scarsa valorizzazione. Il problema dell'indicizzazione è particolarmente gravoso per i capitoli di monografie e per le opere miscelanee; 2. favorire l'incontro e lo scambio di relazioni tra le comunità di umanisti che coltivano argomenti di ricerca simili o affini.

La realizzazione di una banca dati bibliografica italiana per l'area umanistica e per le scienze sociali non suggerisce in via assiomatica che la bibliometria debba essere applicata alla valutazione della ricerca di area umanistica. Consentirebbe, tuttavia, agli umanisti di utilizzare una serie di servizi avanzati attraverso identificativi persistenti come DOI o ORCID<sup>27</sup> dei quali, al momento, i ricercatori di area umanistica possono usufruire solo in minima parte e solo se le loro pubblicazioni risultano indicizzate in database citazionali come Scopus o WoS o vengono rese accessibili in rete<sup>28</sup>. Le scienze

<sup>25</sup> Agenzia Nazionale di Valutazione del sistema Universitario e della Ricerca, *Programma delle attività dell'Anvur: 2013-2015* (<http://www.roars.it/online/wp-content/uploads/2013/02/Programma...pdf>).

<sup>26</sup> I progetti di ricerca approvati a novembre 2014 dall'Anvur esplorano fondamentalmente tre temi della valutazione della ricerca in area umanistica: la valutazione delle monografie di ricerca, la Library Catalog Analysis e l'utilizzo di Google Scholar come possibile base dati per valutare la ricerca di area umanistica.

<sup>27</sup> DOI è il Digital Object Identifier, un identificativo persistente per le risorse digitali; ORCID è un identificativo persistente per i nomi degli autori.

<sup>28</sup> A questo proposito la creazione dell'ANPrePS (Anagrafe Nazionale dei Professori e dei Ricercatori e delle Pubblicazioni Scientifiche) potrebbe rappresentare il primo passo verso la realizzazione di un database bibliografico per le scienze umane e sociali.

sociali sono invece già orientate a un approccio valutativo multivariato che combina bibliometria e *peer review* sia in quanto le scienze sociali stesse sono in larga parte sia in Scopus che in WoS, sia in quanto esse sono ormai epistemologicamente sempre più vicine alle scienze dure.

C'è da chiedersi se anche le scienze umane siano destinate alla stessa deriva (o evoluzione?) epistemologica (di forma e di contenuti) che ha profondamente mutato le scienze sociali, e se nel giro di pochi anni, per fronteggiare la crisi di identità e il ripetersi degli esercizi di valutazione, anche le scienze umane finiranno per adeguarsi alla logica prevalente del *publish or perish* riducendo i tempi di pubblicazione e adottando l'articolo come forma di pubblicazione prevalente. Quello che sembra certo è che con la visibilità in rete e con l'approccio bibliometrico le scienze umane saranno obbligate vie più a confrontarsi<sup>29</sup>. Infatti, come ha scritto di recente Elio Franzini: «andare contro ciò che la tecnica e la rete mettono a disposizione è miope. Se anche non si vuole aderire al modello [bibliometrico n.d.a], bisogna “criticarlo” ovvero affrontare criticamente la questione»<sup>30</sup>.

#### 4. Oltre la bibliometria: alcuni temi aperti nella valutazione della ricerca in area umanistica

Abbandoniamo ora il dibattito sulla bibliometria nelle scienze umane e allarghiamo la discussione a due temi che appaiono centrali nella valutazione della ricerca per le scienze umane alla vigilia del lancio del terzo esercizio di valutazione nazionale della ricerca (la VQR 2011-2014) e mentre si sta svolgendo il processo di valutazione dei dipartimenti a mezzo della famigerata SUA-RD (Scheda Unica Annuale della Ricerca Dipartimentale).

Il primo tema è quello dell'internazionalizzazione. Orientate a un perimetro nazionale, legate alla valorizzazione dell'ambito locale e nazionale, le scienze umane sono chiamate ad aprirsi all'internazionalizzazione. La valutazione non è l'unico fattore che spinge le scienze umane verso il contesto internazionale, ma certamente sta diventando uno dei principali elementi di spinta. Qui varrebbe forse la pena riflettere un attimo sul rapporto tra causa e effetto, tra fattori interni alla comunicazione scientifica e all'evoluzione della ricerca e fattori che intervengono dall'esterno e a posteriori, come appunto gli effetti di un esercizio nazionale di valutazione.

In primo luogo va detto che il grado di apertura verso l'estero è molto diverso da disciplina a disciplina; ancora una volta le scienze umane denunciano, loro malgrado, una frammentazione e diversificazione di approcci che in alcuni casi rappresenta anche la loro ricchezza. I filosofi, gli psicologi e gli antropologi appaiono condurre le scienze umane verso un approccio internazionale, ma cosa dire, invece, dei giuristi, degli studiosi della lingua e della storia italiana?

---

<sup>29</sup> Si noti che nella VQR 2004-2010 nell'area 11 per la Psicologia e i settori M-EDF sono state realizzate 2.146 valutazioni bibliometriche. Del resto anche le famigerate mediane utilizzate durante l'Abilitazione Scientifica Nazionale sono, a rigore di logica, un criterio bibliometrico, non citazionale, ma comunque quantitativo.

<sup>30</sup> E. FRANZINI, *Ho visto cose... valutazioni di un umanista*, in A. BANFI-E. FRANZINI-P. GALIMBERTI, *Non sparate sull'umanista: la sfida della valutazione*, Guerini e associati, Milano 2014, p. 25.

Altro argomento dibattuto: come misurare il livello di internazionalizzazione dei diversi gruppi di umanisti? Nelle scienze dure sembra abbastanza assiomatico che il grado di internazionalizzazione si possa misurare dal numero di collaborazioni con ricercatori stranieri, dalla partecipazione a conferenze internazionali e/o a comitati scientifici di riviste in lingua inglese. Vale lo stesso anche per gli umanisti?

Il terzo argomento – irrisolto – è relativo al grado di internazionalizzazione di una rivista. In che modo possiamo misurarlo? Non dalla lingua di pubblicazione: almeno non unicamente. Per le scienze dure pubblicare su una rivista in lingua inglese è indice indiscusso di un profilo di ricerca internazionale. Nelle scienze umane l'inglese è solo una delle molteplici lingue lette, parlate e scritte e i rapporti internazionali sono più variati e linguisticamente molteplici. Inoltre, non tutte le riviste e gli editori stranieri possono realmente fregiarsi di un buon livello di internazionalizzazione: «già si intravede così l'importante questione della differenza tra le coppie nazionale-straniero e nazionale-internazionale, per cui vi può benissimo essere una rivista italiana “internazionale”, così come vi sono numerosissime riviste straniere “nazionali»<sup>31</sup>.

Quarto argomento: in che modo valutiamo l'internazionalizzazione del singolo ricercatore, di un dipartimento o di un ateneo? Ha senso valutare il grado di internazionalizzazione di un dipartimento tirando semplicemente le somme di quanti coautori stranieri pubblicano con i docenti afferenti ad un dipartimento o di quanti *visiting professors* vengono ad esso assegnati – due misure queste ultime richieste dall'Anvur nella SUA-RD? Eppure con l'internazionalizzazione, così come con la bibliometria, le scienze umane sono chiamate a confrontarsi: negare valore al contesto internazionale per difendere inutili localismi o vetusti privilegi rischia di marginalizzare e di ridimensionare più di quanto non sia già avvenuto la ricerca umanistica. Certamente il tema è controverso, come si legge anche nel rapporto finale del GEV dell'area 12: «Certo, vi sono differenze tutt'altro che irrilevanti e proprio in occasione della VQR si sono manifestate appieno: non sono pochi, infatti, gli esperti esterni (al GEV) che hanno segnalato difficoltà e dubbi in relazione al parametro dell'internazionalizzazione. Anche a prescindere dal fatto che questo parametro non sia preso in considerazione in altre esperienze di valutazione, segnatamente in quella inglese, gli sforzi effettuati dal GEV per fornirne un'interpretazione, per quanto possibile, “adeguatrice” hanno dato luogo a esiti solo parzialmente soddisfacenti»<sup>32</sup>. Quello che risulta poco chiaro non è unicamente come misurare il parametro dell'internazionalizzazione, ma, soprattutto, come adeguarlo alle particolarità delle diverse discipline di ricerca e alle finalità delle diverse procedure e dei molteplici livelli di valutazione.

Il secondo tema sul quale vorremmo concentrare brevemente la nostra attenzione è quello della valutazione delle monografie di ricerca. Da un lato appare evidente come

<sup>31</sup> A. GRAZIOSI, *La valutazione delle discipline umanistiche in Italia, 1999-2011, nell'esperienza della Società Italiana per lo Studio della Storia Contemporanea*, in P. MICCOLI, A. FABRIS (a cura di), *Valutare la ricerca? Capire, applicare, difendersi*, ed. cit., p. 65.

<sup>32</sup> *Valutazione della Qualità della Ricerca 2004-2010 (VQR 2004-2010). Rapporto finale di area. Gruppo di Esperti della Valutazione dell'Area Giuridica (GEV 12)* ([http://www.anvur.org/rapporto/files/Area12/VQR2004-2010\\_Area12\\_RapportoFinale.pdf](http://www.anvur.org/rapporto/files/Area12/VQR2004-2010_Area12_RapportoFinale.pdf)), p. 19.

esista nella comunicazione scientifica una crisi della monografia accademica di ricerca<sup>33</sup>, crisi economica, innanzitutto, ma anche formale. La riforma dei corsi di laurea ha innescato un processo, forse irreversibile, di revisione nella struttura della monografia accademica di orientamento didattico che diventa più snella per adeguarsi alle mutate esigenze della didattica; gli esercizi di valutazione nazionali e l'abilitazione scientifica nazionale impongono scadenze imperative che riducono i tempi di scrittura, i tempi di pubblicazione diventano più veloci, il digitale consente alle riviste di proliferare e prosperare nelle scienze umane grazie all'utilizzo di piattaforme di pubblicazione come OJS; cosa resta, dunque, della monografia accademica di ricerca? Secondo Robert Darnton il libro scientifico è stato dichiarato morto così tante volte che gode di ottima salute<sup>34</sup>. Così anche Andrea Capaccioni, che in un articolo appena pubblicato su "AIB Studi" scrive: «non è corretto parlare di declino del libro scientifico, bisognerebbe far invece riferimento a una lunga fase di transizione in corso durante la quale gli editori sperimentano nuove soluzioni anche grazie alle opportunità offerte dalla tecnologia digitale e dalla rete»<sup>35</sup>.

Se il destino della monografia di ricerca appare incerto, quanto meno in evoluzione, ancora più incerte sono le metriche per la sua valutazione. Fino all'altro ieri il prodotto monografico era scarsamente presente sia in Scopus che WoS. Di recente entrambe le banche dati hanno dimostrato un'attenzione crescente verso la monografia accademica di ricerca: Thomson Reuters ha lanciato nel 2013 il "Book Citation Index", un indice citazionale di circa 30.000 titoli di monografie integrandolo nel Web of Science. Di contro Elsevier ha lanciato il programma denominato "Scopus Books Enhancement Program" e nel 2014 dichiara di includere in Scopus 420 serie complete di monografie di ricerca per un totale di quasi 50.000 titoli con l'obiettivo di arrivare ai 75.000. Le monografie di ricerca sono indicizzate anche in Google Scholar. Il prestigio di quest'ultimo database non commerciale è, tuttavia, di gran lunga inferiore a quello di Scopus e WoS: i dati sono troppo sporchi per essere ritenuti affidabili e per potere essere utilizzati correttamente<sup>36</sup>.

Sintomi che le basi dati bibliometriche esistenti potranno essere utilizzate anche per valutare la monografia accademica? Può darsi: in attesa che maturi una tale prospettiva si cercano strade diversificate per la valutazione della monografia accademica di ricerca e si elaborano teorie sofisticate che possono sembrare, e talvolta lo sono, meri esercizi di

---

<sup>33</sup> Negli Stati Uniti il calo delle vendite della monografia di ricerca è costante dal 1986. In pochi decenni si è passati da una tiratura di più di duemila copie a poche centinaia.

<sup>34</sup> R. DARNTON, *A Program for Reviving the Monograph*, in "Perspectives on history", 37 (3/1999), <http://www.historians.org/publications-and-directories/perspectives-on-history/march-1999/a-program-for-reviving-the-monograph>. Nell'articolo l'autore esalta il canale digitale come mezzo per rivitalizzare la monografia di ricerca.

<sup>35</sup> A. CAPACCIONI, *La monografia scientifica e le sfide dell'accesso aperto*, in "AIB Studi", 54 (2-3/2014), pp. 201-211.

<sup>36</sup> Tra i numerosi contributi si legga: E. DELGADO LÓPEZ-CÓZAR-N. ROBINSON-GARCÍA-D. TORRES-SALINAS, *The Google Scholar Experiment: How to Index False Papers and Manipulate Bibliometric Indicators*, in "Journal of the Association for Information Science and Technology", 65 (2014), pp. 446-454. Particolarmente evidente è in Google Scholar il problema relativo alla mancanza di un controllo di autorità per i nomi degli autori.

riflessione sulla valutazione in area umanistica. Tra le metodologie “alternative” di valutazione della monografia accademica vi sono:

- la cosiddetta Library Catalog Analysis (LCA), proposta nel 2009 in un articolo di Torres-Salinas e Henk Moed, che misura l’impatto di una monografia verificando la presenza dell’opera nei cataloghi di un insieme selezionato di prestigiose biblioteche<sup>37</sup>;
- il numero di recensioni in riviste prestigiose;
- il numero di download (solo per i volumi in formato digitale) ovvero il numero di volte in cui un libro è stato prestato o scaricato dalla rete (legalmente o illegalmente);
- il *rating* degli editori e/o delle collane nelle quali la monografia viene pubblicata.

Quest’ultima sembra essere a tutt’oggi la via più praticabile e praticata<sup>38</sup>, anche se discutibile, così come possono esserlo le liste di riviste: un *informed peer review* di secondo grado, senza neanche il parametro della *peer review* che possa fungere da criterio di selezione tra un editore o l’altro<sup>39</sup>. Scrive a tal proposito Nicola De Bellis nel suo volume dedicato alla bibliometria: «l’equivoco [...] consiste nel ritenere che le soluzioni bibliometriche e quelle pseudo-bibliometriche siano equivalenti e sostanzialmente intercambiabili o che le seconde siano addirittura superiori alle prime in virtù del giudizio qualitativo degli esperti nell’assegnazione delle classi di merito. [...] Non può esistere una scorciatoia pseudo-bibliometrica alle valutazioni»<sup>40</sup>.

In questo articolo abbiamo provato ad affrontare alcuni temi caldi della valutazione della ricerca in Italia, senza la pretesa di trovare soluzioni ai dubbi amletici che stanno attraversando il dibattito italiano e, in parte, anche quello internazionale. In conclusione ci sembra utile proporre oltre alle criticità sopra esposte qualche suggerimento.

In primo luogo, a nostro avviso, andrebbe incoraggiato l’approccio valutativo basato sulla peculiarità disciplinare. Decisamente convincente ci sembra, ad esempio, la metodologia proposta di recente da un giurista dell’università di Leuven, Alain Laurent Verbeke, per la valutazione dei lavori di ricerca dell’area giuridica. Verbeke elabora un

<sup>37</sup> Cfr. D. TORRES-SALINAS-F.H. MOED, *Library Catalog Analysis as a Tool in Studies of Social Sciences and Humanities: an Exploratory Study of Published Book Titles in Economics*, in “Journal of Informetrics”, 3 (2009), pp. 9-26, <http://www.sciencedirect.com/science/article/pii/S1751157708000527>. La misura dei volumi presenti a catalogo andrebbe fatta separando le copie realmente acquistate dalle biblioteche dagli eventuali doni ricevuti.

<sup>38</sup> Liste (*rating*) di editori sono contenute in: VABB-SHW (Vlaams Academisch Bibliografisch Bestand voor de Sociale en Humane Wetenschappen), delle università delle Fiandre (<https://www.ecoom.be/en/vabb>), a cura dell’ECOOM belga e CRISin, il sistema informativo norvegese. In CRISin ciascuna pubblicazione acquisisce, nel momento stesso in cui viene inserita, un’indicazione qualitativa standard (livello 1 o 2) derivata da classifiche predefinite di riviste ed editori. In Spagna il CSIC ha prodotto la SPI (Scholarly Publishers Indicators), una lista di editori spagnoli e internazionali diversificata per discipline di area umanistica e per le scienze sociali.

<sup>39</sup> In realtà nel VABB-SHW la valutazione delle monografie viene integrata consentendo agli editori di apporre ai titoli un’etichetta *gprc* (*guaranteed peer review content*) validata da una commissione indipendente di esperti.

<sup>40</sup> N. DE BELLIS, *Introduzione alla bibliometria: dalla teoria alla pratica*, ed. cit. La citazione è tratta dall’edizione epub dell’opera.

metodo di classificazione qualitativa dei lavori di ricerca, distinguendo tra pubblicazioni di ricerca, pubblicazioni di ricerca applicata e pubblicazioni divulgative, e discute degli strumenti per realizzarla. La proposta esalta il processo di autovalutazione partendo dal principio che sia il singolo ricercatore a dover valutare in modo coerente e responsabile la propria attività di ricerca e considera ai fini valutativi una serie di indicatori di qualità e di prestigio come, ad esempio: il numero di partecipazioni a conferenze e seminari, i progetti di ricerca approvati, la partecipazione ai comitati scientifici e ai *panel* di area<sup>41</sup>.

In secondo luogo si dovrebbe slegare il finanziamento ordinario delle università dai risultati degli esercizi di valutazione, almeno per quella parte di quota FFO che non riguarda strettamente la ricerca. È quanto ha chiesto di recente il Collegio dei Direttori di Dipartimento de “La Sapienza” in una lettera aperta indirizzata al proprio Rettore<sup>42</sup>. Questa scelta allenterebbe, almeno in parte, la tensione dialettica esistente sulla valutazione della ricerca e consentirebbe di affrontare alcuni nodi cruciali della valutazione nelle scienze umane (e sociali) in modo più maturo e distaccato. Inoltre potrebbe essere utile concentrare gli sforzi dell’autorità politica e dei valutatori non unicamente sui risultati prodotti dalla ricerca, ma anche sull’erogazione dei finanziamenti, ovvero mettere in atto buone pratiche per valutare meglio e in modo più coordinato i fondi da allocare nei progetti di ricerca. Sarebbe opportuno fondare in Italia, come già in Francia<sup>43</sup>, un’agenzia nazionale che abbia tra i suoi compiti quello di guidare e coordinare i finanziamenti ai programmi di ricerca<sup>44</sup>, secondo quel sano principio che vuole che ci sia sempre una valutazione *ex-ante* e una *ex-post*. Si potrebbe argomentare – a ragione – che i fondi erogati per la ricerca in Italia sono sempre più scarsi, ma apriremmo un *cabier de doléances* che necessiterebbe di essere meglio documentato e argomentato.

Infine, per affrontare la crisi di identità delle discipline umanistiche, nel dibattito sulle scienze umane va enfatizzato l’impatto della ricerca in termini sociali (*societal quality*). La ricerca universitaria, infatti, dovrebbe essere inquadrata in una cornice valutativa più ampia di quella offerta dalla sola valutazione ai fini accademici, qualitativa o quantitativa che sia. Sarebbe interessante poter concepire e adottare degli “indicatori sociali” per mettere in evidenza il contributo delle scienze umane e sociali alla “Terza missione”<sup>45</sup> dell’università. Non è un caso che l’Anvur abbia provato ad inserire nella scheda SUA

---

<sup>41</sup> A.L.P.G. VERBEKE, *Beyond Quantity. Classifying and Evaluating Legal Research in a Trusting Environment*, 26 marzo 2014 ([http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=2416240](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=2416240)).

<sup>42</sup> La lettera, scritta a novembre 2014, è stata pubblicata sul blog “ROARS” a dicembre 2014 (<http://www.roars.it/online/sua-rd-la-posizione-del-collegio-dei-direttori-di-dipartimento-della-sapienza/>).

<sup>43</sup> Ci riferiamo all’Agence Nationale de la Recherche (ANR) che in Francia assicura la selezione e il finanziamento dei progetti di ricerca.

<sup>44</sup> Questa riflessione ha una sua deriva: infatti, non sempre la pianificazione della ricerca ha avuto esiti positivi, molte ricerche nascono da intuizioni e idee che è complesso, talvolta controproducente, sottoporre ad una metodica programmazione.

<sup>45</sup> L’idea di “Terza missione” dell’università va ricondotta al pensiero del filosofo Ortega y Gasset. Cfr. J. ORTEGA Y GASSET, *La missione dell’università*, trad. it. A. Savignano, Guida, Napoli 1972. Le funzioni dell’insegnamento universitario sono, infatti, molteplici: l’insegnamento delle professioni, la produzione di nuova ricerca e la trasmissione della cultura. Il sociologo francese Pierre Bourdieu parla, invece, di “capitale culturale” che viene costruito con il contributo delle scienze umane e sociali; cfr. P. BOURDIEU, *Homo Academicus*, Les Éditions de Minuit, Paris 1984, p. 317.

anche una parte dedicata alla “Terza missione”, con esiti per ora incerti. Quello che appare certo è che attraverso la “Terza missione”, attraverso la disseminazione della scienza, la condivisione e la partecipazione ai progetti di ricerca e alla riflessione scientifica, attraverso l’applicazione delle teorie e dei principi espressi dalle scienze umane l’università può sostenere un nuovo umanesimo: le scienze umane sono chiamate a dare il loro contributo.



Angelo Miglietta

LA VALUTAZIONE DELLA QUALITÀ DELLA RICERCA  
NELLE DISCIPLINE MANAGERIALI: BUONI PROPOSITI, CATTIVE  
PRATICHE E URGENTI CAMBIAMENTI<sup>1</sup>

**Abstract**

*The evaluation of university professors' activities has recently become a crucial issue in the academic debate, with a particular consideration for research funding. Beyond the mere financial aspects of such a domain, there is a reputational framework involved in this field that reshapes the power architecture of the Italian universities – and of society as a whole through this. This essay focuses on the evaluation of university professors' activities within the field of managerial sciences. Despite significant differences between the Italian and the Western research evaluation approaches, in the last thirty years the field has seen a deep change in research methodology: the approach developed by mainstream economists at the international level has at last been adopted. This approach highlights mere research activities as the main – if not the only – criterion to evaluate the quality of a good professor. Teaching activities and knowledge transfer have almost lost relevance. Hence, the evaluation of university professors' activities is only based on peer reviews and citation indexes, de facto delivering the evaluation only to self-referential academics. After highlighting the questionable aspects of this narrow approach, the essay shows how research is not currently evaluated on the basis of a serious scientific methodology. Finally, the paper criticizes the evaluation of the quality of research activities based on bad tools provided by what has now become a pseudo-science and attempts to add, to the current evaluation approach, the assessment of non-academic factors such as knowledge transfer skill, awards from non-public or non-academic institutions, and the evaluation of the general culture that good professors and scholars should hold.*

1. *Introduzione e inquadramento della problematica*

Le discipline manageriali si sono lungamente distinte da quelle economiche. Pur mutuando numerosi concetti sviluppati dalla teoria economica, esse hanno sempre svolto un ruolo di rilievo nella definizione di modelli utili per coloro che sono chiamati a gestire le organizzazioni. Anche sul piano epistemologico, tali differenze di finalità euristica hanno determinato per le discipline economiche importanti distinzioni che con un'accettabile approssimazione potrebbero caratterizzarsi come un *orientamento alla modellizzazione quantitativa tipico della teoria economica ricorrendo all'uso dell'econometria* (anche detta economia politica), mentre nelle scienze manageriali ha prevalso l'attenzione alle

---

<sup>1</sup> Ringrazio Enrico Guglielminetti per i suggerimenti e le indicazioni ricevute, e Emanuele Parisi per l'assistenza editoriale.

fattispecie, e anche la modellizzazione – peraltro ben presente – si fonda su categorie logico-qualitative piuttosto che quantitative. Per dirla con Vicari, «agli albori le discipline manageriali erano di sicuro legate alle pratiche manageriali. Il motivo per cui sono nate le business schools [...] e le università di Economia [...] va cercato nella necessità di diffondere le migliori pratiche manageriali, nate nelle imprese di maggiore dimensione»<sup>2</sup>.

Questa distinzione non sembra oggi – e personalmente me ne dolgo – ancora proponibile. Si è assistito, a livello internazionale, a una rapidissima omologazione degli studi manageriali alla metodologia in voga, a partire dalla fine della seconda guerra mondiale, fra gli economisti. Sempre Vicari osserva che «proprio la nascita delle discipline manageriali, dettata da necessità operative, ha prodotto nei professori (di management, osservazione mia) un senso di inferiorità rispetto agli scienziati alle prese con discipline dotate di uno statuto di rigore metodologico costruito in decenni e talvolta secoli [...]. Dagli anni settanta e ottanta del secolo scorso [...] il desiderio di una posizione di più elevato prestigio all'interno degli ambienti accademici ha condotto i ricercatori, soprattutto statunitensi, a richiedere un maggiore rigore nella ricerca sul management. E come spesso accade quando è necessario recuperare da una posizione di svantaggio, la spasmodica ricerca di legittimazione ha condotto gli studi di management a diventare sempre meno rilevanti per il mondo delle imprese»<sup>3</sup>.

Il nostro Paese, con poche eccezioni, è rimasto sostanzialmente al margine di questo fenomeno fino alla fine degli anni '80, principalmente a causa dell'isolamento nazionale. Un isolamento dorato, grazie ai ricchi proventi professionali allora conseguibili e al successo degli studi di economia aziendale presso studenti e imprese. Tutto ciò è stato aggravato dalla modestia del background culturale di alcuni professori di management del nostro Paese, tipico in quegli anni. Essi provenivano dalla laurea in Economia e Commercio, una delle poche – almeno fino alla riforma degli esami di maturità del 1968 – alle quali era possibile accedere anche dopo avere conseguito solo un diploma di scuola media superiore di tipo tecnico, senza dunque possedere basi culturali solide. Tali basi culturali sono conseguibili solo attraverso lo studio delle cosiddette *Humanities*, in particolare le letterature classiche e la filosofia. Questi accademici avevano dunque una formazione che era veramente l'opposto di quanto aveva immaginato la riforma Gentile di inizio secolo scorso per formare le classi dirigenti del nostro Paese.

Gli economisti italiani, forse perché non considerati dalle organizzazioni economiche italiane del tempo, invece, avviarono ben presto un percorso di apertura al contesto internazionale che li ha portati in larga misura a essere omologati già dall'inizio degli anni '80 al *mainstream* dominante, e perciò molto più forti nel sempre esistito confronto con gli studiosi di management. In altre parole, l'isolamento degli aziendalisti italiani dal contesto internazionale, i loro gravi limiti culturali e una loro certa distrazione dall'attività scientifica e accademica dovuta a una troppo abbondante attività professionale li hanno resi colpevolmente indifesi davanti al proprio competitore

---

<sup>2</sup> S. VICARI, *All of us have a problem. Rilevanza e rigore nelle discipline manageriali*, in “Economia e Management”, 3 (2013), p. 4. Ringrazio molto Salvio Vicari per avermi fornito in più occasioni spunti di riflessione sui temi oggetto di questo scritto, ben documentati dal suo articolo qui assai diffusamente citato.

<sup>3</sup> *Ibidem*, pp. 4-5.

naturale nella ripartizione delle risorse finanziarie (ovvero fondi di ricerca e posti di ruolo nelle università) e nel contempo spiazzati rispetto ai criteri di valutazione della ricerca in ambito economico-manageriale fissati dall'Anvur, prontamente e astutamente occupata proprio dai portatori di questo pensiero ritenuto dominante. Che è stato nobilitato, in particolare, dal fatto di essere internazionale, e perciò a priori superiore rispetto a qualsiasi altro pensiero. Consci di questa rivoluzione, alcuni studiosi di management, che per primi si erano aperti alla cultura internazionale, hanno *cavalcato l'onda* e adottato acriticamente le nuove metodologie di valutazione della ricerca, per cercare di acquisire posizioni più forti nel quadro delle proprie discipline specifiche, proprio in quella prospettiva della sudditanza psicologica ricordata prima da Vicari. Il processo è stato persino feroce nelle proprie manifestazioni, perché rapido, non costruito e non condiviso, ma semplicemente imposto sulla base della presunta e indiscutibile superiorità dell'approccio internazionale e della contiguità con gli economisti. Essi possono finalmente ristabilire la loro superiorità grazie a questo *autodafé* dei competitori aziendalisti, che odiosamente hanno sempre i corsi di laurea più frequentati dagli studenti.

In tutto ciò non pochi danni ha subito anche il tema della valutazione della didattica, relegata ai margini della valutazione della qualità di un docente, come pure assai poco viene considerata l'attività di *terza missione*, che invece costituisce uno dei punti di forza della valutazione dell'operato delle università nei Paesi più competitivi sul fronte della ricerca, come Gran Bretagna, Stati Uniti, Israele e Germania. Va peraltro subito sgombrato il campo dall'idea che occorra ritornare al passato degli studiosi di management, alla chiusura nazionalista. La formazione di un bravo professore universitario di discipline manageriali, che per essere tale deve essere anche uno studioso e possibilmente un ricercatore, non può prescindere da un robusto percorso internazionale. Quei professori italiani di management che negli anni '80 guardavano con indifferenza o sufficienza ai dottorati di ricerca, che non mandavano i propri allievi all'estero per completare la loro formazione e che fondavano la loro scuola nel senso di appartenenza e di fedeltà (e magari con un po' di sfruttamento per svolgere la propria attività professionale), alimentando un tanto volgare quanto diffuso nepotismo, non hanno lasciato niente dietro di sé se non il loro narcisismo senescente, e certo non possono essere ricordati oggi dai loro sventurati allievi con stima e rispetto. Ma questi stessi sciagurati, oggi divenuti professori, sentano la responsabilità di fare il loro dovere – omesso dai loro pessimi maestri – e quindi di aiutare i loro allievi a formarsi, con capacità critica, internazionalmente. Ciò serva soprattutto ad evitare che questa sorta di *furia iconoclasta* dell'Anvur porti alla selezione e formazione di una classe di studiosi di management che, seppur perfetti per gli standard internazionali di ricerca, sono incapaci di essere interlocutori credibili della realtà, auspicabilmente evitando manifestazioni di provincialismo ottuso come avere un docente italiano che fa lezione in inglese a un'aula che parla italiano. Al contempo va assicurato che nessuno possa essere incardinato se non è in grado di interloquire con facilità in inglese, nonché senza prima avere integrato il proprio processo di formazione con un robusto percorso internazionale, possibilmente in più di una sede.

Per trattare il tema del lavoro si procede esaminando questi aspetti: in primo luogo la definizione degli obiettivi della valutazione della ricerca, poi le interrelazioni di tale processo con la metodologia di ricerca, da cui si potranno trarre prime conclusioni sulla correttezza del modello oggi prevalente, evidenziandone le contraddizioni e le criticità. Sarà così possibile, infine, proporre un migliore modello di valutazione. Tutto il percorso si svilupperà avendo a riferimento la *fotografia* dell'esistente che è stata appena tratteggiata. Le riflessioni e analisi prodotte dovranno essere sempre lette in questa specifica prospettiva, che è quella degli studiosi di discipline manageriali, e che non può essere in nessun modo estesa in via automatica ad altri campi. Soprattutto vorrei ricordare che in nessun modo il fine di queste riflessioni è quello di ripristinare la situazione pre-esistente ma, al contrario, sottrarsi al danno che i cambiamenti introdotti possano persino peggiorare lo stato, non proprio ottimale, della ricerca nelle discipline manageriali nel nostro Paese.

La soluzione che pare ideale, e sarà oggetto della parte finale del ragionamento, è quella indicata da Deidre McCloskey, che in un suo ormai noto lavoro – per quanto stigmatizzato – ha illustrato *i vizi degli economisti* e il modo per superarli attraverso la celebrazione e il ripristino delle *virtù della borghesia*<sup>4</sup>.

## 2. *Gli obiettivi della valutazione della ricerca nell'ambito delle discipline manageriali: una visione generale e sovranazionale*

La ricerca, come a tutti ben noto e da tutti condiviso, è la causa del progresso dell'uomo. Lo è evidentemente sul piano materiale, attraverso l'acquisizione di nuove conoscenze che si fanno competenze, per consentire il raggiungimento di un benessere superiore a parità di costo o di ridurre il costo per ottenere lo stesso livello di soddisfazione. Spesso, addirittura, i due aspetti si intrecciano, rendendo i benefici della ricerca scientifica ancora più rilevanti e apprezzati. Non sempre peraltro il tutto si risolve in questi benefici universalmente riconoscibili: può accadere che i risultati della ricerca non generino benefici di tipo *paretiano* come quelli ipotizzati. Ciò avviene per esempio se per la società nel suo complesso i risultati di un'attività di ricerca comportano un miglioramento, mentre però per alcuni soggetti ciò si traduce in un peggioramento del benessere percepito. Anche in questo caso, bisogna riconoscere che è la ricerca scientifica il motore del progresso, pur con le contraddizioni che esso inevitabilmente porta con sé. La ricerca scientifica infatti, con un effetto per così dire *collaterale*, favorisce anche il progresso della dimensione non materiale della vita umana nelle società, perché l'incremento della conoscenza si accompagna con l'esaltazione della componente non istintiva della persona umana (non rettiliana, per dirla con la psicologia), favorendo l'intelligenza (nel senso latino del termine). Si può osservare che costante è la relazione fra l'aumento

---

<sup>4</sup> D.N. MCCLOSKEY, *I vizi degli economisti, le virtù della borghesia*, trad. it. R. Merlini, Edizioni IBL Libri, Torino 2014. Il testo originale pubblicato in Olanda risale al 1996. Ringrazio sinceramente Deidre per gli stimoli e le riflessioni e per avermi fatto parte del suo articolato e ben consolidato pensiero sulla qualità della ricerca in materia economica in occasione di un ciclo di seminari tenuti a Milano nell'ottobre 2015, in particolare presso la Scuola di dottorato dell'Università IULM di Milano.

dell'intelligenza e la comprensione delle visioni differenti, il che si traduce in un incremento della tolleranza e si fa fattore di creazione di coesione sociale. La società ha visto l'affermarsi di un'ampia libertà delle persone e di un contesto che favorisce la creatività, il fertile terreno su cui cresce la ricerca<sup>5</sup>.

Questo processo, osservabile nella storia proprio a partire dalla stagione delle scoperte scientifiche del XVII secolo e grazie alle applicazioni tecniche che esse hanno favorito, ha dato impulso alla nascita della società moderna come noi la conosciamo e dell'incredibile benessere che tutto ciò ci ha consegnato, pur nelle contraddizioni, violenze e ingiustizie che caratterizzano la società globalizzata contemporanea. Queste riflessioni non si applicano tuttavia in modo automatico alla ricerca scientifica nei campi riferibili alle scienze non dure. Qui infatti la mancanza di un riscontro concreto al progresso della conoscenza può generare produzione di pensiero e di ideologie, che non necessariamente si sono accompagnate con il progresso della società e l'affermazione anche delle libertà, ma anzi sono sfociate in una involuzione dai tratti purtroppo drammatici: e questo può esser il caso delle discipline economiche e manageriali, soprattutto nella misura in cui si pongono in una prospettiva astrattamente teorica o di empirismo finto, perché basato solo su una metodologia statistico-econometrica<sup>6</sup>.

La conclusione che si trae è dunque che sostenere la ricerca è il modo migliore per favorire il progresso della società, e di conseguenza anche la valutazione della ricerca ha un ruolo fondamentale nelle politiche di un Paese, perché solo un corretto processo valutativo permette una corrispondentemente adeguata promozione della ricerca stessa, attraverso gli incentivi che possono essere messi a disposizione. Sostenere la ricerca in modo corretto, e dunque grazie a una sua corretta valutazione, è allora ancora ancora più importante per un Paese che preoccuparsi di risolvere la disoccupazione giovanile o ridurre il fenomeno della fuga dei cervelli, perché questi due fenomeni trovano una naturale soluzione proprio aprendo ai nuovi talenti le vie della ricerca che crea occupazione non sussidiata. Non è un caso che i Paesi più competitivi, Stati Uniti, Giappone, Germania, Corea del Sud e Israele, si caratterizzino per quote di PIL più elevate di investimenti in R&D. Ma soprattutto Israele e Stati Uniti vantano una ormai consolidata politica di sostegno non generico alla ricerca, perseguita sia con strumenti sia pubblici, sia privati, ma attraverso il sostegno alla nascita di start up e di una sorta di "ecosistema" che risulti favorevole alla loro nascita e sviluppo. La valutazione della ricerca, in questi Paesi soprattutto, e con specifico riferimento alle scienze "dure", tiene conto della capacità di innovazione e valuta la ricerca e le università attraverso parametri come il numero di brevetti e le start up avviate e portate a successo con gli spin off, mentre i criteri che fanno riferimento alle pubblicazioni e alla loro valutazione trovano sempre meno spazio. Diverso è il caso degli studi economici e manageriali, dove invece,

---

<sup>5</sup> Cfr. E. BERTACCHINI-W. SANTAGATA, *Atmosfera creativa*, Il Mulino, Bologna 2012. La prematura scomparsa di Walter ha lasciato un vuoto scientifico incolmabile nel campo dell'economia della cultura e degli studi sullo sviluppo economico attraverso i settori culturali e creativi; lo voglio anche qui ricordare con commozione e affetto, avendo a mente le sue straordinarie doti umane, che sempre accompagnano i grandi studiosi.

<sup>6</sup> Nel successivo paragrafo sono presentati in modo organizzato gli elementi a supporto di questa affermazione.

anche a livello internazionale, la valutazione fa pressoché esclusivo riferimento alla qualità delle pubblicazioni.

### *3. Gli obiettivi della valutazione della ricerca nell'ambito delle discipline manageriali: le specificità italiane*

Nel nostro Paese la materia, come è noto, è trattata dalla Legge 240 del 2010. Una buona approssimazione per comprendere quale sia giudicata una buona ricerca viene offerta dalla lettura del “profilo scientifico del professore ordinario” come definito dal Consiglio Direttivo dell’Anvur: «Il candidato ha una posizione riconosciuta nel panorama internazionale della ricerca [...] ottenuta pubblicando contributi significativi riconosciuti a livello internazionale, ha partecipato a congressi internazionali in qualità di oratore invitato o di membro del Comitato scientifico [...] ha ottenuto riconoscimenti per la sua attività scientifica [...]. Nei settori per i quali sia appropriato, ha mostrato la capacità di trasferire conoscenza al contesto socio-economico esterno all’università»<sup>7</sup>. L’enfasi, come appare chiaro, è tutta posta sul tema della pubblicazione, solo marginalmente e sembrerebbe in via residuale si richiama il trasferimento di conoscenza. Prima ancora di entrare nel merito della bontà di questa scelta, analizzando in particolare il modo come viene valutata la qualità di una pubblicazione, si può già osservare che nel nostro Paese la buona qualità della ricerca è prevalentemente legata alla qualità delle pubblicazioni e alla “internazionalizzazione” dello studioso. Certamente non potrà essere una ricerca naturalmente produttiva di innovazione per il rafforzamento della capacità competitiva, visto il peso marginale del trasferimento della conoscenza. Colpiscono poi: a) l’assenza del richiamo alla verifica della sussistenza di una robusta preparazione culturale (forse presunta dalla qualità delle pubblicazioni); b) il fatto che l’enfasi sull’appartenenza alla comunità scientifica internazionale di riferimento non venga mai collegata alla capacità di relazioni interdisciplinari; c) il fatto che non vi sia alcun riferimento alle capacità didattiche, come se un professore non dovesse anche insegnare.

Merita ora un approfondimento il tema proprio dell’autonomia della ricerca dalla realtà, molto cara all’Anvur ed espressione degli esperti che l’hanno definita e imposta al sistema, beninteso con riferimento alle discipline manageriali. Credo che un ottimo modo per trattare la questione sia proprio fare riferimento al pensiero di Vicari, già ricordato. In particolare Vicari ritiene che occorra distinguere fra rilevanza e utilità dell’attività di ricerca scientifica in campo manageriale: «La prima attiene a un’applicabilità immediata dei risultati, la seconda riguarda invece il fatto che l’aumento di conoscenza ha in ogni caso un’utilità, non sempre immediatamente comprensibile, potendo manifestarsi dopo molto tempo» Sempre secondo lo stesso autore, pertanto, «quando si sostiene che la ricerca non è rilevante, non bisogna dimenticare che essa può

---

<sup>7</sup> ANVUR, *Criteri e parametri di valutazione dei candidati e dei commissari dell’abilitazione scientifica nazionale*, approvato dal consiglio direttivo dell’Anvur del 22 giugno 2011 ([http://www.anvur.org/attachments/article/44/documento01\\_11.pdf](http://www.anvur.org/attachments/article/44/documento01_11.pdf)), pag. 10.

essere si irrilevante, ma in ogni caso è sempre utile»<sup>8</sup>. E quindi «perché si produca conoscenza nel contesto accademico è necessario che l'unico obiettivo di uno scienziato sia appunto la conoscenza e non altro [...] la conoscenza si sviluppa dalla libertà intellettuale del singolo, che vuole oltrepassare i limiti di quanto è già conosciuto. Questo “andare oltre” non può che essere guidato dalla libertà intellettuale e dalla curiosità [...] E il punto di arrivo della conoscenza non sta nell'uso che di essa si vuole fare [...] ma nella capacità [...] di astrarre».

L'Autore si spinge oltre, e sostiene che «la questione della rilevanza non è stata risolta [...] perché nella scienza vi è la necessità di una “chiusura” alle istanze esterne e quindi anche al mondo manageriale [...]. La chiusura a necessità esterne è, infatti, un connotato essenziale dell'attività scientifica, in qualunque campo». Proseguendo si manifesta il cuore del pensiero dell'Autore nella fattispecie, quando afferma che «l'idea di fondo [...] è che la modalità di formazione della conoscenza nel mondo scientifico e in quello manageriale siano del tutto differenti tra loro e che nel campo delle discipline manageriali il campo scientifico e quello professionale non siano due sistemi in connessione tra loro, ma siano due insieme del tutto diversi e anche con poche relazioni»<sup>9</sup>.

Queste posizioni, così ben argomentate, sono veramente fondamentali per il ragionamento che qui si vuole sviluppare. In primo luogo va detto che queste posizioni sono certamente pienamente condivise dalla comunità scientifica internazionale in materia di studi manageriali (e ovviamente fra gli economisti, ispiratori dell'approccio) e sono un caposaldo, una premessa logico concettuale, del lavoro dell'Anvur in materia di valutazione della ricerca (ma anche della qualità di un professore universitario). Sono rappresentate come il “vento nuovo” che spazza via i nepotismi, le arretratezze e una

---

<sup>8</sup> A sostegno di questa tesi, che è perciò condivisibile, cita il caso della scoperta del calcolo binario inventato nel Seicento da Juan Caramuel, e rimase per secoli solo «un gioco teorico senza alcuna rilevanza pratica, e tuttavia la sua utilità è stata manifesta quando altre conoscenze hanno consentito lo sviluppo del computer e l'utilizzo del codice binario per la programmazione» (S. VICARI, *All of us have a problem. Rilevanza e rigore nelle discipline manageriali*, ed. cit., p. 3).

<sup>9</sup> *Ibidem*, pp. 2-4. Questi temi sono stati oggetto di un seminario organizzato dall'OCSE e dall'UE nell'ambito del loro progetto comune HEI (High Education Institutions) a Roma lo scorso mese di ottobre 2014. La *partnership* fra le due istituzioni ha lo scopo di favorire il miglioramento della capacità competitiva del nostro continente, anche nella prospettiva di Horizon 2020, attraverso il rafforzamento del trasferimento tecnologico alle imprese e l'innovazione. All'incontro hanno partecipato figure di vertice del mondo accademico di diverse università italiane, evidentemente (e fortunatamente, stante la visione di metodo scientifico da sviluppata da noi aziendalisti) non tutti esponenti del mondo degli studiosi di management. È stata ribadita la posizione di chi auspica invece una forte rilevanza dell'attività di ricerca applicata, in particolare esprimendo la preoccupazione che l'Europa non riesca a mantenere le proprie posizioni competitive. Quasi tutti i colleghi presenti che si occupano di scienze dure (anche se non ingegneri) hanno ricordato che nelle loro comunità scientifiche la rilevanza non solo non contrasta con il metodo scientifico, ma anzi ne è una determinante e una garanzia di scientificità. Ancora più orientata in questo senso la posizione del mondo delle ingegnerie, come ha ben rappresentato nella sua relazione il Rettore del Politecnico di Torino Marco Gilli, *keynote speaker*, presentando il caso di successo dell'incubatore I3P. Una conclusione importante della giornata è stata riassunta nel motto “*less paper more impact*”, che verrà presentato dagli organizzatori alle autorità politiche europee per indirizzare i fondi di ricerca. Mi pare quindi (e per fortuna per la nostra Europa) in modo difforme dalle finalità perseguite attualmente dall'Anvur, per cui si potrebbe dire, parafrasando il noto adagio di Deng Xiao Ping, che “non importa che prenda il topo, ma solo il colore del gatto”.

certa corruzione morale che aveva infettato la comunità scientifica italiana degli studiosi di management, e non solo. In secondo luogo, e questo è un tema di sostanza ancora più importante, viene giustificato l'attuale metodo di valutazione della qualità della ricerca, che è basato sull'analisi delle pubblicazioni e della partecipazione all'attività della comunità scientifica internazionale dei professori di management, secondo il citato documento Anvur.

Secondo dunque il nuovo approccio alla valutazione della qualità della ricerca, l'obiettivo della valutazione deve essere quello di verificare che uno studioso di management sia un bravo ricercatore, perché parte del proprio sistema scientifico, che ha «come elementi fondamentali le teorie e il metodo: soltanto la teoria che è stata processata attraverso questi due elementi è riconosciuta come appropriata. I ricercatori sono prevalentemente interessati al rigore (in senso metodologico) della dimostrazione scientifica, rispetto alla rilevanza per la realtà aziendale e imprenditoriale». Proseguendo coerentemente Vicari afferma quindi che «il meccanismo che consente di selezionare la classe accademica è quello del metodo scientifico, che è approvato dalla comunità stessa, ed è questo che costituisce l'unico elemento discriminante, valutato attraverso procedure quali la peer review [...]. Se fossero utilizzati altri metodi di validazione della ricerca, come per esempio la praticabilità dei risultati, il giudizio su questo aspetto non potrebbe più essere affidato agli scienziati, ma ai [...] manager, che introdurrebbero immediatamente criteri non accettabili per la comunità scientifica».<sup>10</sup>

È allora chiaro quale è l'obiettivo della valutazione della qualità della ricerca in campo manageriale attualmente vigente nel nostro Paese attraverso le scelte dell'Anvur: verificare la piena appartenenza di uno studioso o di un professore di management ai dettami della comunità scientifica internazionale, misurati in modo preciso, come nel prossimo paragrafo viene illustrato e commentato. Proprio l'opposto di quanto si prefiguravano le *business school* di oltreoceano e le facoltà di economia europee quando nacquero. Forse forzando un po' si potrebbe dire che in questo quadro fra un gruppo di ricerca che lavora con successo intorno a un incubatore universitario (perché produce imprese, occupazione e innovazione) e uno che scrive un *paper* sul suo funzionamento attraverso l'uso del corretto metodo scientifico e lo pubblica su una rivista a elevato *impact factor* e quindi assai prestigiosa, interesserà ovviamente solo il secondo, perché opera secondo le regole del "metodo scientifico". Solo questo sarà il lavoro premiabile, nella prospettiva del modello Anvur.

Mentre ora si può passare, come anticipato, a esaminare i modi con cui si valutano i contributi di ricerca, si rinvia a un paragrafo successivo l'esame della correttezza del metodo scientifico attualmente vigente nelle discipline manageriali, appiattite su quelle sviluppate dagli economisti. Certo che, se si dimostrasse che il metodo scientifico non è tale, o presenta fortissime limitazioni sul piano epistemologico, saremmo in presenza di una metodologia di valutazione e incentivazione della ricerca in materia economico-manageriale che non solo ha obiettivi forse non condivisibili o largamente insufficienti, come sopra si è visto, ma che orienta anche verso un metodo che potrebbe rendere l'economia e le discipline manageriali, che al suo metodo si sono supinamente adeguate, solo pseudo-scienze, al pari dell'astrologia e dell'alchimia.

---

<sup>10</sup> S. VICARI, *All of us have a problem. Rilevanza e rigore nelle discipline manageriali*, ed. cit., pp. 2-4.

#### 4. I modi per la valutazione della qualità della ricerca nelle discipline manageriali

In questo paragrafo si vogliono analizzare i modi di valutazione adottati dall'Anvur per valutare la qualità della ricerca, ferma restando la definizione degli obiettivi di valutazione che sono stati appena presentati, e dunque senza entrare nel merito della correttezza e completezza di questi obiettivi, che verrà analizzata e criticata nel prosieguo del lavoro. Già la semplice analisi dei limiti e inadeguatezze dei modi, che nel seguito si vuole dimostrare, concorre a evidenziare le gravi criticità di tutto l'impianto della valutazione della qualità della ricerca, e purtroppo si dimostrerà nel prosieguo la sua dannosità per il sistema Paese Italia, anche se ben altre sono le soluzioni rispetto al ritorno al passato.

I modi per la valutazione della qualità della ricerca nelle materie economiche (area 13, a cui fanno riferimento anche gli studi di management), sono così efficacemente descritti da Giovanni Federico: «esiste un solo metodo: chiedere ad un altro esperto del tema, detto referee, cosa ne pensa (peer review). Questo è il metodo adottato da tempo per la scelta degli articoli da pubblicare sulle riviste internazionali [...] un esito che si verifica quasi sempre solo dopo una almeno profonda revisione del testo originale [...] una versione più democratica dello stesso principio affida il giudizio sulla qualità del lavoro alla comunità scientifica piuttosto che a due referee. La qualità dell'articolo si misura dal numero di citazioni che ha ricevuto: tanto più è alto, tanto più l'articolo è importante. I GEV hanno usato questi due criteri per tutte le aree così dette bibliometriche (la bibliometria è la disciplina che studia come misurare la qualità dei lavori scientifici) [...] Un articolo è stato definito eccellente se pubblicato in una rivista nel primo quintile (le 20% migliori per impact factor e se ha avuto un numero di citazioni che lo pone nel primo quintile degli articoli)»<sup>11</sup>.

Questi metodi di valutazione sono particolarmente infondati, irrazionali e pretestuosi, per i seguenti motivi:

1. la scelta dei peer reviewers è per sua definizione fondata sul pre-giudizio che qualcuno sia in grado di giudicare i propri pari, ma perciò automaticamente non più pari. E infatti spicca che fra gli esperti di valutazione dell'Anvur, nelle discipline manageriali, siano stati (casualmente?) scelti autorevoli colleghi noti per essere custodi del purismo della ricerca scientifica e della sua autoreferenzialità, secondo le visioni della scienza descritte nel precedente par. 3. Ciò non configge, a ben vedere, ed è anzi coerente con l'idea di obiettivo della valutazione della qualità della ricerca che è (sciaguratamente) oggi in voga. Ma certamente si palesano due criticità: a) il rischio evidente di un pregiudizio anche di tipo ideologico, oltre che in ragione delle appartenenze “di scuola”; per esempio, dopo avere scritto questo articolo, che dispiacerà sicuramente ai portatori della corrente visione di che cosa sia la scienza manageriale, sarà molto più elevato il rischio di una bocciatura dei miei lavori e di quelli dei miei allievi e magari dei colleghi da cui sono stimato e che hanno l'ardire di incautamente citarmi, anche se i giudici miei peer

---

<sup>11</sup> G. FEDERICO, *La Valutazione della Qualità della ricerca italiana: istruzioni per l'uso*, (<http://noisefromamerika.org/articolo/valutazione-qualita-ricerca-italiana-istruzioni-uso>, 24 luglio 2013).

potranno sempre sostenere, assai indignati, di essere ovviamente moralmente al di sopra di queste miserie umane; b) la scelta dei giudici fra propri pari porta automaticamente al venire meno della parità, perché è chiaro che non può essere tale il rapporto fra valutato e valutatore.

2. il ricorso alla misurazione delle citazioni come strumento democratico di valutazione da parte della propria comunità solo in apparenza soddisfa questa esigenza. Forse nel Paese delle Meraviglie (quante volte paiono così i mondi descritti dagli economisti!) le citazioni avvengono per i nobili motivi ricordati da Federico. Nella pratica, come sappiamo molto bene tutti noi ora impegnati a pubblicare secondo le regole dell'Anvur, la cosa si svolge, nel migliore dei casi, secondo un sistema di *marketing della citazione* e di *attenzione agli editors* o ai componenti dei comitati scientifici delle riviste giudicate più interessanti. La prima azione consiste nel creare un vero e proprio *Ponzi Scheme* della citazione, che nei casi più soft si traduce in una disciplina autoimposta a citare colleghi amici, che ricambieranno la cortesia nell'ambito di una *general fairness* o estetica della citazione reciproca, in taluni casi si diviene però oggetto persino di pressioni, neppure troppo garbate, a citare, da parte per esempio dei revisori dei propri articoli proposti alle "prestigiose riviste dei quintili migliori".

Molta influenza sulle citazioni la esercitano anche, ovviamente, gli editor e i membri dei comitati scientifici delle riviste su cui si desidera pubblicare, i cui loro contributi, ovviamente, vanno citati abbondantemente. Questi fenomeni non riguardano, va detto, solo gli autori nazionali, ma anzi sono ancora più rilevanti su scala internazionale. Conoscersi e, purtroppo, scambiarsi *soft grants* è una regola fondamentale per acquistare riconoscibilità internazionale. Qui entra in campo la seconda azione, che prevede di partecipare ai convegni internazionali in cui prevale l'attività di PR, che viene ben corroborata attraverso l'invito ai docenti che influenzano la qualità della pubblicazione a fare poi il *visiting professor* o a tenere seminari (ben pagati) presso la propria università.

Nei casi migliori ciò si persegue anche partecipando a progetti di ricerca in network per ottenere finanziamenti di istituzioni (tipicamente pubbliche, purtroppo). Purtroppo i progetti sono valutati, come sempre, da esperti scelti fra i propri *peers*, alieni da qualsiasi contaminazione con il mondo reale, per garantire la purezza del metodo scientifico. È evidente che tutto ciò porta solo a un avvilitamento fra la valutazione e la qualità della produzione scientifica, che così è tanto migliore quanto più è chiusa. Con buona pace delle analisi epistemologiche di Popper, in particolare sull'importanza della società aperta e sulla necessità che le affermazioni che vogliono dirsi scientifiche e siano quindi in tali termini valutabili godano del requisito della falsificabilità.

3. Occorre considerare le citazioni *ex adverso* che cioè implicano un giudizio negativo. Esse non solo non dovrebbero sommarsi, ma addirittura portare a una riduzione del punteggio complessivo.

In conclusione anche i modi utilizzati per valutare la qualità della ricerca, dietro a un'immagine rigorosa e perciò anche moralmente preferibile, hanno generato un sistema eticamente molto discutibile, che, di fatto, sembra ripetere il metodo della cooptazione baronale preesistente, replicandolo solo su scala più ampia perché internazionale. Purtroppo ciò avviene senza che le figure di riferimento, i giustamente vituperati baroni della *prima repubblica dell'Università*, possano svolgere, come talvolta in passato accadeva,

un ruolo di mitigazione forte di un'autorevolezza basata su una cultura solida e su capacità di mediazione secondo criteri di equità.

### 5. Della (non) scientificità delle discipline economiche, ovvero dei vizi degli economisti

Si può ora fare qualche riflessione su una delle ultime questioni aperte all'inizio del lavoro, a proposito della scientificità del metodo usato nelle discipline economiche e in quelle manageriali che su quelle economiche, come detto, si sono appiattite. L'obiettivo non è chiaramente, fare un trattato sulla corretta metodologia nella ricerca economica, per tanti motivi, fra l'altro il fatto che questo non è il tema dello scritto. Tuttavia parlare del metodo scientifico è particolarmente rilevante quando si ragiona di criteri di valutazione della ricerca, perché, come si è visto, tale processo attualmente è tutto concentrato a verificare che essa sia giudicata in modo positivo dalla comunità scientifica di riferimento, che si proclama chiusa a interferenze esterne, che non sono tenute in conto se non marginalmente. Se si dimostra che la scientificità del metodo non sussiste o presenta gravi limiti e lacune, verrebbe a cadere l'unico e ultimo pilastro rimasto a tenere in piedi l'attuale sistema di valutazione della qualità della ricerca, l'idea che essa sia buona se è riconosciuta tale dalla comunità degli studiosi della materia.

Si può partire da un'affermazione piuttosto chiara della McCloskey, che afferma «oggi l'economia come materia di studio ha un grosso problema: i suoi metodi sono sbagliati e, pertanto, producono risultati sbagliati. È triste che gli economisti [...] credano ciecamente nella validità dei propri metodi, meccanismi e dei loro esiti. Offrono consigli ai governi e criticano l'uno il lavoro dell'altro, come se l'economia fosse una scienza esatta [...]. Sono persone serie e benintenzionate. Non meritano di rappresentare una scienza priva di scoperte scientifiche». E ancora «gran parte della teoria economica sviluppata dal secondo dopoguerra andrebbe ripensata ex novo. Quasi tutte le presunte scoperte “scientifiche” dell'economia andrebbero completamente riesaminate con un altro metodo per ottenere un minimo di credibilità [...] l'economia moderna è diventata molto simile ai giochi che fanno i bambini sulla sabbia [...]»<sup>12</sup>.

Queste pesanti affermazioni sono puntualmente documentate e dimostrate dalla studiosa. I problemi dell'economia, secondo la McCloskey, «nascono da tre plausibili progetti, sviluppati negli anni Quaranta da altrettante menti di prim'ordine [...]. Questi tre progetti [...] significatività statistica, dimostrazioni teoriche e ingegneria sociale non hanno giovato [...] l'essenza comune [dei tre progetti] risiedeva nel tentativo di meccanizzare l'economia». I vizi degli economisti, che inficiano il profilo scientifico della disciplina, «sono le tre cattive abitudini intellettuali portate nell'economia moderna dai tre più grandi pensatori degli anni Quaranta: Lawrence Klein, Paul Samuelson e Jan Tinbergen [...] vale a dire: 1. La convinzione kleiniana che la “significatività statistica”, nel senso tecnico dell'espressione, coincida con la significatività scientifica, 2. La convinzione samuelsoniana che le “prove di esistenza” sviluppate teoricamente alla lavagna siano scientifiche e 3. La convinzione tinbergeniana che le due citate componenti della pseudo scienza si possano applicare alla costruzione dell'economia

<sup>12</sup> D.N. MCCLOSKEY, *I vizi degli economisti, le virtù della borghesia*, ed. cit., pp. 9ss.

politica, per ottenere una sorta d'ingegneria sociale»<sup>13</sup>. La McCloskey, nel suo citato libro, dimostra in modo convincente le sue critiche e formula anche proposte per rifondare la scienza economica.

Ai fini presenti pare particolarmente importante il primo vizio. Esso è il supporto concettuale all'utilizzo delle metodologie econometriche, che sono diventate largamente prevalenti nelle riviste più prestigiose per le pubblicazioni in materie manageriali. Ma, oltre alla critica della McCloskey, va ancora ricordata l'eccezione popperiana, con il suo richiamo alla falsificabilità come criterio per riconoscere la scientificità di un metodo. L'econometria e la statistica, proprio a causa dell'approssimazione che è loro intrinseca, documentata dall'utilizzo proprio della significatività, non sono falsificabili. E oltretutto le correlazioni evidenziate nelle analisi multivariate mutano nel tempo e sono soggette all'alea che altre variabili siano le vere cause delle correlazioni individuate<sup>14</sup>. In particolare, è proprio l'approccio metodologico tipico dell'econometria e della statistica che va messo in discussione. Infatti la costruzione di un modello richiede in primo luogo una pre-valutazione a cura del ricercatore, per sua natura arbitraria, dei legami e collegamenti fra alcune variabili e le altre, ritenute dipendenti. In secondo luogo, presenta numerose criticità l'individuazione dei campioni e delle fonti da prendere come riferimento. È infatti evidente che la scelta del campione da osservare tende a influenzare il tipo di correlazione. In terzo luogo, e questo appare l'aspetto critico più rilevante, in presenza di grandezze che possono essere influenzate da una pluralità di variabili, è arbitrario attribuire un nesso di causalità come invece tende ad affermare un approccio che cerca correlazioni "*statisticamente rilevanti?*". Così per esempio si potrebbe trovare un'elevata correlazione fra l'aumento delle performance di un'impresa e la presenza di un'adequata componente femminile nei consigli di amministrazione, il tutto potrebbe essere correttamente motivato, sulla base delle metodologie statistico-econometriche<sup>15</sup>, tuttavia le buone performance d'impresa potrebbero essere state determinate da fatti completamente diversi, come per esempio l'appartenenza a un settore in crescita o le capacità dell'Amministratore Delegato (istintivamente, fra l'altro, vien da propendere più per questo fatto sostanziale invece che per uno che potrebbe essere soprattutto ideologicamente fondato). Si noti che questo tipo di affermazioni e conclusioni mal si presta al giudizio del popperiano tribunale della falsificazione. Infatti ci si potrebbe trovare in presenza di una correlazione con le buone performance d'impresa giudicata statisticamente rilevante per entrambi le variabili, la presenza femminile nei Board e il tasso di crescita del settore di appartenenza. Dunque una conclusione non può negare l'altra, e quindi falsificarla: da ciò il giudizio di non scientificità.

Questo terzo aspetto è poi ulteriormente aggravato dalla variabilità, nel tempo, dei collegamenti esistenti fra le diverse variabili. Anche se la statistica e l'econometria hanno messo a punto e affinato tecniche sempre più sofisticate per ridurre le criticità ricordate, permane un quadro d'incertezza, se non di confusione, nell'attribuzione di correlazioni

---

<sup>13</sup> *Ibidem*, pp. 7-8 e 13-14.

<sup>14</sup> Diverso è il caso delle discipline biomediche, dove la significatività statistica può offrire un conforto rilevante che si traduce in utili protocolli di terapia.

<sup>15</sup> Intendo presentare gli elementi a supporto di questa affermazione in modo organizzato nel successivo paragrafo.

fra variabili e grandezze dipendenti. Tutto ciò è documentato dalla debolezza delle previsioni, che tuttavia sono poi presentate e utilizzate come se fossero invece fondate e dimostrate. Le conseguenze sono drammatiche nell'applicazione di politica economica, come ad esempio tutta la ricerca economica a supporto della tesi che la riduzione dell'indebitamento pubblico sia una condizione per il rilancio dell'economia.

In conclusione viene meno l'unico motivo che giustifica l'attuale approccio di valutazione della qualità della ricerca in Italia: concepite per verificare che gli studiosi rispettino i dettami della comunità scientifica, si deve amaramente osservare che le pratiche e i metodi di ricerca seguiti a livello internazionali rendono l'economia e le discipline manageriali che a essa si sono conformate solo pseudo scienze, vanificando l'esercizio della valutazione e smontando le "classifiche di merito" dei professori e delle università<sup>16</sup>.

#### 6. Conclusioni: per una buona valutazione della ricerca scientifica nelle discipline manageriali

È possibile adesso produrre una sintesi e formulare le proposte per correggere il sistema di valutazione, beninteso nelle discipline manageriali. Sarà proprio il ricorso alle virtù borghesi della McCloskey, qui di seguito rappresentate in alcune fattispecie operative, a consentire una buona valutazione dell'attività di ricerca, che superi le attuali contraddizioni e si faccia strumento di concorso al progresso della società e in particolare del nostro Paese e dell'Unione Europea.

1. L'introduzione della valutazione è stata particolarmente utile, perché ha dato un impulso impensabile all'internazionalizzazione degli studiosi di management e spinto a svolgere più attività di ricerca; sarebbe un errore gravissimo abbandonare il processo e tornare al passato. Per questo, soprattutto in fase di selezione all'ingresso, dovrà essere documentata una fase rilevante di formazione e ricerca, a livello di dottorato e post doc, presso istituzioni internazionali dai candidati ai concorsi. Ma anche la verifica della qualità della ricerca nel prosieguo della carriera dovrà incentivare la partecipazione al contesto internazionale, particolarmente attraverso periodi di *visiting professor* e mediante la collaborazione con team di ricerca internazionali; ciò richiede evidentemente che parte dei fondi di incentivazione siano destinati a coprire i costi di questo processo di internazionalizzazione permanente.

2. Gli obiettivi della valutazione devono essere allargati al trasferimento delle conoscenze alla società, accettando quindi l'idea che la ricerca debba, ma soprattutto possa, essere rilevante. Ciò consente non solo di trasformare il costo del finanziamento della ricerca in un investimento per rafforzare la capacità competitiva del Paese, ma anche di offrire importanti spunti alla ricerca stessa, sgombrando il campo dall'idea che la realtà non sia rilevante per il metodo scientifico. Opportuni parametri possono essere adottati, fra i quali spiccano la capacità di attrarre finanziamenti per progetti di ricerca e

---

<sup>16</sup> Questo articolo temo mi renderà ancora più odioso a qualche collega, e confesso che la cosa mi rende triste e veramente dispiaciuto, oltre che intimorito per le ripercussioni. Solo per anticipare una critica, comunico di essere comunque uscito non male dalle valutazioni Anvur, posizionandomi in una fascia intermedia.

di assistenza tecnica dal settore privato (non dal settore pubblico e dalle fondazioni bancarie, accomunate da funzioni obiettivo e criteri decisionali estranei a logiche di mercato e perciò di discutibile *accountability*).

Occorrerà pertanto sostenere i privati che destinino risorse per la ricerca, per esempio con l'attribuzione di un extra beneficio fiscale in relazione ai fondi da loro così destinati, in luogo dell'attribuzione di fondi alle università sulla base di classifiche fondate su valutazioni prodotte da organismi burocratici, anche se astutamente concepite e presentate come trasparenti ed efficienti grazie all'internazionalizzazione dell'approccio. Le imprese ne beneficeranno, grazie all'impulso all'innovazione, guidata però dalla disciplina imposta dalla concorrenza cui sono soggette. Pure la capacità di creare spin off e occupazione andrà apprezzata e valorizzata, e anche l'uso del *crowdfunding* è molto promettente: anche qui i fondi per la ricerca dovrebbero almeno in parte tradursi in crediti di imposta per incentivare, senza procedure complesse, questi finanziamenti. L'esempio della medicina e dell'ingegneria è di particolare conforto, come proprio Vicari indica, e ciò dimostra che la contaminazione con la realtà è possibile e anzi necessaria: «quante volte l'osservazione di un bisogno è alla base della scoperta?»<sup>17</sup>.

3. Occorre verificare e misurare la sussistenza di una robusta preparazione culturale che non può mai essere presunta solo dalla qualità scientifica delle pubblicazioni. Ciò vale innanzitutto con riferimento alla fase di ingresso, circoscrivendo le tematiche a un'area specifica scelta dagli studiosi all'interno di ambiti con contenuti e riferimenti alla letteratura predefiniti, coerenti con i temi di ricerca sviluppati. In particolare, per la verifica di un'adeguata preparazione culturale, si deve intendere al momento dell'ingresso nel percorso di carriera l'accertamento della conoscenza della letteratura, anche classica, sicuramente nell'ambito delle discipline economico-manageriali, ma anche in discipline i cui contributi consentano una conoscenza allargata e approfondita – perché trasversale e complementare – delle tematiche tipiche degli studi manageriali (psicologia, antropologia, sociologia, storia economica e dell'impresa, economia politica, analisi economica del diritto, epistemologia, gnoseologia, analisi matematica e statistica). È evidente che tali dotazioni culturali non dovranno più essere oggetto di accertamento nell'ambito delle verifiche sull'attività di ricerca svolta successivamente, durante la vita “adulta” del docente ricercatore. Durante questa fase “adulta” le verifiche dovranno vertere soprattutto sul controllo del mantenimento di una capacità di conoscenza e comprensione della realtà economica in cui operano le organizzazioni oggetto di studi in discipline manageriali. In particolare, da una parte si rende necessaria la verifica del continuo aggiornamento del docente ricercatore sulla letteratura scientifica, secondo un perimetro predefinito. Ma sarà anche necessario, dall'altra, che gli studiosi di management dimostrino una buona conoscenza della realtà economica internazionale, in particolare delle imprese e dei mercati, da conseguire attraverso la sistematica lettura di un quotidiano economico internazionale e uno nazionale. In questo senso cultura è

---

<sup>17</sup> S. VICARI, , *All of us have a problem. Rilevanza e rigore nelle discipline manageriali*, ed. cit., p. 4. L'autore, a conclusione del suo lavoro, afferma poi che «La tesi che intendo sostenere è che la collaborazione è possibile solo a condizione che gli scienziati siano “costretti” a produrre ricerca rilevante, incorporando anche la rilevanza tra i criteri per giudicare una buona ricerca» (*ibidem*, p. 8). Queste affermazioni paiono ben coerenti con le conclusioni di questo scritto.

conoscenza della realtà economica di riferimento e delle opinioni che si formano, da parte di osservatori che possono anche non appartenere alla comunità scientifica. Ma, per giudicare adeguato l'aggiornamento culturale, sarà anche necessario individuare modi per verificare l'aggiornamento delle competenze e conoscenze delle discipline prima ricordate, le cui conoscenze sono imprescindibili per una solida e robusta attività di ricerca nelle scienze manageriali.

4. Nell'attesa che le discipline economiche provvedano al drastico rinnovamento auspicato, cosa che non avverrà in tempi brevi per le difese corporative degli interessi esistenti<sup>18</sup>, bisognerà valutare l'opportunità di separare le discipline manageriali in due parti distinte: il *theoretical management*, che sposa il corrente approccio degli studiosi di management che si sono internazionalizzati e il *practical management*, con l'intento di rispondere ai bisogni per cui nacquero le *business school*, con l'uso di metodologie di ricerca adeguate, e che devono essere fortemente internazionalizzate. Questo approccio richiede oggi non solo l'adozione di una metodologia adeguata e ben diversa da quella corrente, ma anche competenze interdisciplinari di rilievo, prodotte principalmente da filosofia, storia, psicologia, sociologia e antropologia, che devono essere documentare e incentivate. Naturalmente sarà necessaria una buona *accountability*, che dovrà essere comunicata sia agli studenti e alle loro famiglie, sia al mondo delle organizzazioni private, sia *for* che *non-profit*, perché sia trasparente il tipo di preparazione che è fornita e il tipo di ricerca che è svolta. L'attuale sistema di valutazione della qualità della ricerca andrà mantenuto e migliorato per superarne le criticità, ma il suo peso ridotto perché rapportato agli altri indicatori di valutazione, qui proposti. È facile prevedere che sarà risolto il crescente problema lamentato dagli studiosi di management, per i quali «comincia oggi a diffondersi nel mondo imprenditoriale la convinzione dell'inutilità di finanziare università e scuole che producono una ricerca non rilevante. E quella di non volere più assumere, e pagare molto, giovani laureati in campo economico i quali hanno una formazione, dal punto di vista della rilevanza, non dissimile da quella [...] dei laureati in scienze politiche o in filosofia, e che dunque vanno poi formati all'interno delle aziende».

---

<sup>18</sup> Gli interessi in gioco sono immensi. Esiste oggi una chiara tendenza degli studiosi, sia di materie economiche, sia manageriali, a usare il prestigio conseguito attraverso i processi valutativi della qualità della loro ricerca per ottenere incarichi di prestigio nei consigli di amministrazione di importanti società, ovvero a svolgere attività di consulenza a istituzioni governative, nonostante la loro completa impreparazione rispetto alle dinamiche della realtà economica, che il rigore "scientifico" delle pseudo scienze economiche impone per garantire la qualità dello studioso. Giova molto a questi "scienziati" la capacità di influenzare il pubblico attraverso i media e rivendicare una capacità superiore di comprensione della realtà. Oltre a proporsi come consiglieri aspirano spesso ad assumere cariche politiche, possibilmente al di fuori del processo di selezione democratica e sciaguratamente cagionando gravi danni alle popolazioni colpite dalla loro azione. Sulla pochezza di questo approccio e soprattutto sulla sua pericolosità per lo stesso sistema democratico nei Paesi capitalistici mi permetto di rinviare a un interessante articolo, *The power of self-belief*, pubblicato su "The Economist" del 6 dicembre 2014 (<http://www.economist.com/news/finance-and-economics/21635524-new-paper-looks-how-economists-became-so-influential-power-self-belief>), al paper *The Superiority of Economists*, di M. FOURCADE, E. OLLION e Y. ALGAN ([http://pubman.mpdl.mpg.de/pubman/item/escidoc:2071743:2/component/escidoc:2071741/mpifg\\_mpd14\\_3.pdf](http://pubman.mpdl.mpg.de/pubman/item/escidoc:2071743:2/component/escidoc:2071741/mpifg_mpd14_3.pdf), 2014), nonché al mio *Economia, politica e società nella stagione della grande crisi infinita*, in questa stessa rivista (3 (1/2013), pp. 81-89, <http://www.spaziofilosofico.error404.it/wp-content/uploads/2013/01/Miglietta.pdf>).

Imprese che così agendo dimostrano di possedere le virtù borghesi, quelle che consentono di affrontare le sfide di mercato con l'intelligenza della propria azione e delle proprie scelte, così creando il "miracolo" della creazione di valore e benessere per la collettività. Imprenditori e manager che non si fanno ammaliare dalle rappresentazioni degli studiosi "esperti" di materie economiche, gli "*idola fori*" di questo nostro tempo che così ci sconcerta e preoccupa. Forse però, a ben vedere, senza motivo, dato il benessere senza precedenti che è disponibile a un numero di persone mai stato così grande e la quantità delle conoscenze e sviluppo delle tecnologie di cui oggi dispone l'umanità, grazie al libero mercato e alla buona ricerca scientifica. Il patrimonio che la società occidentale deve sapere difendere, correggendo le sue contraddizioni, nell'interesse dell'intera umanità.

Elio Franzini

## INTERROGARE LA VALUTAZIONE NELLE DISCIPLINE UMANISTICHE

**Abstract**

*In line with the democratic ideals, there is a growing call for a greater public involvement in the establishment of science and technology policies. There exists a variety of public participation procedures that aim at consulting and engaging the public, ranging from public hearings to consensus conferences. Unfortunately, a general lack of empirical consideration of the quality of these methods arises from confusion regarding the appropriate benchmarks to evaluate the humanities. Given the difficulty in determining the quality of the output of any participatory exercise, the author suggests the need to consider which aspects of the process are desirable and then measure the presence or quality of these aspects. To this end, a number of theoretical evaluation criteria that are essential for effective public participation are specified. First of all is the need to explore the potential resources of a theory of dialogism and its core concepts for the development of a philosophy and methodology of the humanities. Future research needs to develop instruments to measure these criteria more precisely and identify the contextual and environmental factors that will mediate the effectiveness of the different participatory methods.*

Interrogare la valutazione evitando atteggiamenti ideologici, che cancellino i problemi in precostituiti “vestiti di idee”, è forse esercizio ancora più difficile del valutare stesso. Esempio di tale situazione può essere il veloce resoconto di un recente dibattito sulla valutazione nelle scienze umane, in occasione della presentazione di un libro cui ho collaborato<sup>1</sup>. Il volume, comunque lo si volesse giudicare, conteneva tesi anche provocatorie, in dialogo tra loro e in differente rapporto critico con le istituzioni che alla valutazione sono preposte. L’ambizione del libro non era di poco conto: ci si chiedeva se gli umanisti potessero, o meno, portare un contributo significativo nel sistema epistemologico della valutazione della ricerca, riflettendo al tempo stesso sulla funzione dell’università. Uno dei presentatori, tuttavia, membro del noto sito ROARS, invece di affrontare questi nodi, dichiarò di voler parlare delle “premesse” del libro. Senza dubbio la formula ermeneutica che l’autore può non sapere tutto sul testo che ha scritto ha la sua verità metodologica: ma subito dubitai che il presentatore, ingegnere, conoscesse Gadamer e la tradizione dell’ermeneutica, ipotizzando premesse di cui non ero stato consapevole. Premesse che si rivelarono essere la legge Gelmini e l’istituzione dell’odiata Anvur.

---

<sup>1</sup> A. BANFI-E. FRANZINI-P. GALIMBERTI, *Non sparate sull’umanista. La sfida della valutazione*, Guerini, Milano 2014.

Mettiamo tra parentesi l'esempio, constatando che le condizioni di possibilità per un discorso sulla valutazione nelle discipline umanistiche erano state ritenute, e da un "addetto ai lavori", la legge Gelmini, votata a conclusione del 2010, e l'istituzione dell'Anvur, che risale, pur inattuata per anni, al 2006. Ebbene, se ci si ferma a discutere su tali piani, ed è questa la mia premessa, il discorso può anche chiudersi: ripetendo per l'ennesima volta l'esempio dell'inserimento di "Suinicoltura" tra le riviste di fascia A, si finisce semplicemente per non affrontare con serietà un discorso complesso: è soltanto un cedimento all'ideologia, quella che identifica il nemico e trova molto più facile cercare di abbatterlo invece che dialogare, proponendo elementi di comune discussione. Che tutto ciò accada a opera di professori universitari è particolarmente grave, dal momento che esercitano una professione che prende avvio da un processo valutativo (per legge si entra nei ruoli solo attraverso un concorso, come è noto) e che quotidianamente mette in atto azioni valutative: preparazione delle tesi e loro discussione, esami, concorsi e via dicendo sono operazioni di valutazione che esistevano ben prima della legge Gelmini e dell'Anvur. Anzi, sono operazioni sulle quali non sarebbe sbagliato riflettere, se non altro perché il sigillo del tempo le ha, in un certo senso, "legittimate": va così osservato che nessuno di tali atti è "anonimo", che le persone si vedono e si parlano, mettono in comune conoscenze, cercando soluzioni praticabili. Quando la distanza, o la complessità del sistema, rendono necessario ampliare e dilatare i processi, subentra, come nel giudizio su articoli da pubblicare, un sistema di anonima valutazione, calmierato tuttavia dal fatto che chi assegna un lavoro in lettura conosce, spesso bene, entrambi i soggetti in gioco, e può quindi operare una scelta ponderata.

I docenti universitari, per "mestiere", valutano e dunque non possono prendersela con leggi, normative e persone quando da valutatori professionisti, magari all'apice della carriera, si trasformano in "valutati". Possono, debbono, discutere i metodi, cercare di capire, non prendersela con nemici immaginari attraverso vignette da cabaret. Il problema non è di poco conto: il passaggio da una valutazione "locale" a una nazionale, o comunque più generalizzata, crea molti più problemi che soluzioni. Là dove non è possibile la valutazione diretta, "mettendoci la faccia", e dove l'anonimato non è "calmierato" dalla conoscenza delle situazioni, quando i "prodotti" da assegnare sono centinaia e l'arbitro conosce poco, male o per nulla valutatore e valutato, non potendo certo mettersi a leggere tutti i lavori che distribuisce a fini valutativi, la questione diventa davvero complessa. E non la si risolve in modo astratto. Senza volermi ripetere, ritengo che la *peer review* generalizzata non risolva tutti i problemi, come ha dimostrato la VQR. I punti critici sono stati numerosi, e si riproporranno, anche se si sono raggiunte delle precise consapevolezza, ovvero che i revisori devono essere più consci dei criteri da utilizzare e il giudizio non può essere limitato a un voto con un commento facoltativo. Allo stesso modo, la lista dei revisori deve essere resa nota, anche se non associata al prodotto giudicato, così come bisogna tener conto dei dibattiti seri che sino ad oggi si sono sviluppati, non ritenendo tuttavia la valutazione soltanto una questione tecnica, ai soli tecnici affidata. Alcuni documenti, come quello dell'Area 10, vanno tenuti in seria considerazione, sin dalle premesse, là ove si ricorda che «i settori umanistici per la multiformità dei loro linguaggi, delle espressioni e dei metodi sono difficilmente riducibili a misurazioni univoche ma non sono impossibili da valutare. È auspicabile, al

contrario, l'elaborazione di un modello che tenga conto della singolarità della ricerca umanistica, della varietà dei prodotti che la trasmettono e delle vie attraverso le quali gli studiosi ne riconoscono la qualità scientifica. Una valutazione che intenda premiare le ricerche migliori o che voglia indicare soluzioni al superamento delle criticità deve avvalersi di criteri confrontabili, ma non deve scadere nella fissità di parametri predefiniti che spingono inevitabilmente all'omologazione, alla mortificazione degli slanci innovativi e, molto spesso, all'adattamento opportunistico»<sup>2</sup>.

Nessun documento avrà il valore di un farmaco. Vi sarà sempre qualcuno che lo riterrà velenoso, e risponderà con un altro documento, considerando che gli umanisti non sono più, ormai, una famiglia unitaria e che anche nel loro alveo le culture si intersecano, senza che sia possibile trovare sintesi percorribili. Cercare quindi gli ancora vivi punti comuni, le dinamiche che qui si instaurano nel rapporto tra diacronia e sincronia, può essere un importante lavoro preliminare, se non altro per analizzare i differenti impatti “sociali” che i prodotti umanistici possono avere, anche nella progressiva costruzione di realtà culturali complesse, dove il singolo contributo può essere compreso solo nell'insieme e nel tempo, all'interno cioè di dinamiche che un processo “veloce” come la valutazione tende a non considerare, vanificando di conseguenza lo specifico significato storico delle scienze umane. Di fronte a questa potenziale dicotomia – valutazione da svolgersi sul piano storico e diacronico e necessità di avere risultati contingenti, che servano a una valutazione con scopi immediati - le responsabilità degli errori, le criticità emerse, i litigi infiniti non sono soltanto conseguenze di un sistema mal congegnato o di leggi e delle loro ideologie. Derivano piuttosto da una più generale crisi di identità in cui si dibattono le discipline umanistiche, crisi intorno alla quale bisogna riflettere, considerando che l'università è cambiata perché è mutata la società, come sono stati rivoluzionati i sistemi comunicativi. La “rete”, le riviste *on line*, e in *open access*, hanno reso “globale” la possibilità della comunicazione, e ciò ha comportato, come nel passato è accaduto con altre lingue, la necessità di una lingua *koiné*, che rendesse più agevole la comunicazione stessa. L'inglese ha qui trionfato anche grazie a una sorta di imperialismo sociale ed economico: ma tale imperialismo riguarda il globo, le sue reti sia finanziarie sia informative, non la valutazione e le sue agenzie, che si limitano a prendere atto di una situazione.

Non sempre, e non certo per responsabilità “esterne”, i saperi umanistici hanno acquisito la consapevolezza del cambiamento che in questi ultimi anni si è verificato nell'articolazione dei propri studi e nel loro ruolo nell'ambito generale del sapere e dell'organizzazione della didattica e della ricerca universitaria, dei differenti rapporti che si sono instaurati con altre metodologie scientifiche e degli interni dissidi: dialogare significa avere coscienza delle differenze, ma significa anche non radicalizzarle in fratture assolute. Le differenze possono anche derivare da una base condivisa: non sono “metafisiche”, bensì “storiche” e possono essere ritenute una ricchezza, non un limite invalidante. Il problema, in sintesi, è in primo luogo quello di mettere in discussione un'identità che non esiste più, per cercare di costruirne una nuova, probabilmente più

---

<sup>2</sup> Il documento dell'Area 10 può essere letto nella sua interezza in vari siti delle società scientifiche di area 10. Lo si può per esempio leggere nel sito della società di linguistica applicata: <http://www.aitla.it/eventi/altri-eventi/seminario-nazionale-delle-consulte-scientifiche-dellarea-10-documento-finale/>.

conflittuale, attraversata, come avrebbe detto Wittgenstein, da giochi linguistici molto diversi tra loro, indice di situazioni storiche e culturali forse inaspettate.

Per acquisire un'identità, bisogna tuttavia guardare le situazioni, descriverne le specificità, cercando di non generalizzare. La valutazione non è un processo imposto da nemici esterni, quali sono disegnati Miur e Anvur (spesso confusamente assimilati, come se fosse possibile ignorare che anche soggetti istituzionali possono essere in conflitto tra loro), ma un meccanismo generale che, come già accennato, riguarda la quotidianità. Per esempio, si ha spesso a che fare con nuclei di valutazione interni a volte dilettanteschi, spinti da una logica di contrapposizione rispetto alle valutazioni nazionali imbarazzante e improponibile. Tutto ciò non ha favorito un dibattito approfondito sulla cultura della valutazione e sulle sue criticità interne ed esterne. Criticità che è doveroso, in particolare per gli umanisti, sottolineare, comprendendo i pericoli dei "pieni" normativi, dei loro linguaggi tecnicizzati e scarsamente comprensibili. Criticità tuttavia, che, se analizzate e discusse, possono condurre a soluzioni percorribili, a sperimentazioni. Per esempio, ritengo si possa cercare, per non perdersi nelle norme imposte, di stabilire alcune regole chiare e semplici per rendere il più possibile uniforme, e non soltanto sul piano formale, il lavoro dei Nuclei di valutazione, giudicando i dati raccolti "dopo" che i processi sono stati attuati, quando il giudizio nasce storicamente da un piano di comparazione.

Ipotizzare una valutazione ex-post, curare che le singole sedi mettano in atto processi chiari, trasparenti e testati, istituiscano osservatori della ricerca, siano soggetti di proposta e non referenti passivi e polemici, permette di porre le basi per una valutazione che rispecchi più fedelmente il lavoro dei singoli all'interno delle strutture di cui sono parte, con giudizi non astratti, bensì connessi agli obiettivi che sono stati posti, sulla ricerca come sulla didattica.

Un'analisi ponderata dei "prodotti", del loro impatto culturale e sociale, delle differenti tipologie, indagini in loco, valutazioni tra pari: sono processi che implicano tempi non brevi e la capacità di mettere a confronto modelli diversi, che non si limitino al dato numerico dell'*impact factor* o alla misurazione delle citazioni. Le discipline umanistiche insegnano non la lentezza in sé, ma il significato di una progettualità storica, che nasce attraverso il dialogo e la comprensione di come le tradizioni si formano. Provare senza punire, provare seguendo modelli diversi, pur con obiettivi comuni, testando i risultati e le circostanze, significherebbe usare senza violenza la valutazione, costruire modelli complessi per realtà complesse e variegate. Uscire dagli schemi ideologici e proporre, mettendoli in atto, modelli valutativi che tengono conto della specificità di ciò che va valutato è forse una strada, un modo per interrogare i processi, senza pretendere risposte immediate. E, nel frattempo, lavorare in sintonia, in polemica, anche aspra, ma civile e non ideologica, con coloro che, per legge, hanno necessità di riscontri immediati. Diacronia e sincronia hanno un diverso rapporto, ovviamente, con il tempo: ma il tempo è la base comune, il medesimo piano su cui dialogare.

Questo è, probabilmente, il punto fondamentale, già accennato, ma che merita di essere ripreso: la valutazione, essendo legata anche all'erogazione di fondi "premiali", non può accettare tempi "lunghi" e di conseguenza pretende che i risultati della ricerca siano giudicabili nell'immediato o almeno in tempi brevi. I vari, e ormai molteplici, "prodotti" della ricerca delle discipline umanistiche possono invece concretizzarsi solo in

tempi lunghi, posti all'interno del corso della storia e delle sue interpretazioni, originando "effetti" che non sono misurabili istantaneamente, che generano sedimenti e stratificazioni, i quali, per essenza, rigettano ogni "misurazione". Il contrasto con la sincronicità valutativa è dunque, per le discipline umanistiche, o per la loro gran parte, non accidentale, ma concettualmente inevitabile. La coscienza di un conflitto non può tuttavia condurre al rifiuto di dimensioni dialogiche: al contrario, può essere l'occasione per costruire strumenti in grado di evidenziare le differenze tra i vari modi di considerare il "fattore tempo", con il fine di "relativizzare" i processi valutativi cercando le opportune mediazioni.

Non si tratta più di essere "a favore" o "contro", ma di accettare la differenza e le diverse strade che "legittimano" la valutazione: non si vuole far passare un'ottica "buonista", ma solo ipotizzare che la fase infantile e conservativa dell'opposizione sia giunta al termine. È ora di uscire dall'adolescenza, per evitare che un ribellismo astratto finisca per "delegittimare" chi, come gli umanisti, ha bisogno, per la valutazione, di meccanismi complessi, non astratti e numerici, dove i "modelli" sono, per loro natura, imprecisi e fallibili. Dove il concetto "chiave" – quello di qualità – è di per sé incerto nella sua definizione e, di conseguenza, nella sua valutazione. I filosofi possono senza dubbio dare di tale concetto diverse accezioni, ma non possono disgiungere il nome dalle "cose" e dai "processi" in cui viene inserito, considerando che è necessario, per comprenderne le ricadute fattuali, cercando di far dialogare diacronia e sincronia, operare una sempre più raffinata "tipologizzazione" dei cosiddetti prodotti, delle sedi di pubblicazione, delle loro finalità sociali, didattiche, scientifiche, degli strumenti che le varie sedi utilizzano per valutare le proprie strutture di didattica e di ricerca, creando modelli "integrati", valutati a posteriori da un organo centrale, che entri nel merito dei processi messi in atto<sup>3</sup>.

Non esistono certo, per fortuna, e non purtroppo, soluzioni miracolistiche: c'è invece un dibattito internazionale che va seguito, considerando che tale dibattito mette in campo soluzioni plurime, mai impositive. Il concetto di qualità non è un parametro formale: la sua trasformazione in "valore" è un processo che richiede tempo, che non vive di sola contingenza. Tuttavia, non sempre un percorso valutativo deve terminare in un'assiologia, che è certo il fine, senza che tuttavia "valore" e "valutazione" debbano di necessità sempre coincidere durante il processo. Con la valutazione, dunque, non siamo all'interno di un percorso assoluto e metafisico, in cui sono in questione i valori ultimi e supremi, ma di una serie di azioni sperimentali che valutano in prima istanza le circostanze sulla base delle quali si può instaurare un valore, ponendo i giudizi all'interno di un quadro storico-sociale loro proprio. La valutazione – si pensi per esempio al mercato dell'arte – origina un concetto di "valore" che è spesso soltanto "economico", e che non coincide di necessità con il valore storico di un'opera, che potrà essere "misurato" soltanto in ben più ampie dimensioni temporali, a volte lunghe secoli. È quel che accade nelle scienze umane: la valutazione, in questi ambiti scientifici, giudica un

---

<sup>3</sup> Non si può dimenticare che, prima di essere costretti dalle normative esterne, gli umanisti non brillavano certo per autocoscienza valutativa: il dibattito, in Italia, è stato generato "per reazione". Non esistendo un passato di cui avere nostalgia, e non essendovi neppure un'autentica "storia" su cui riflettere, forse il lavoro deve ancora prendere avvio.

impatto “sociale” contingente, non il senso storico del valore di una ricerca o di un suo prodotto. Ma ciò significa che quella che i filosofi chiamano “assologia” non conosce una sola scala, bensì vive di percorsi stratificati: e la valutazione della ricerca si riferisce soltanto a un gradino di tale scala, cioè a un valore che abbia una ricaduta immediata in uno specifico contesto socio-culturale. Si tratta allora di valutare “bene” questo gradino, ma avendo precisa coscienza di ciò che esso è.

Le domande, dunque, e in conclusione, che pone “Spazio filosofico”, sono le domande “giuste”, quelle da cui ripartire. Giuste, in primo luogo, perché sono “domande”, perché l’atteggiamento interrogante è quello corretto. Le cose hanno un significato inseparabile dal nostro sguardo: la filosofia cerca il significato oggettivo dei vari modi con cui esse si esprimono, costruendo percorsi dove i concetti sono i momenti costitutivi nella verità di un campo di ricerca. Gli oggetti non sono i loro significati, anche se loro tramite sono conosciuti, appresi, tematizzati: il significato “si riferisce” a un oggetto, ma certo non si identifica con esso. Un oggetto ha una pluralità di significati possibili: indagarli, interrogarli è lo stile della ricerca, il suo modo di chiarificare, cercandone l’essenza, i concetti intorno ai quali, in questa interrogazione, si è storicamente formato il pensiero delle cosiddette scienze umane.

Per porre in atto tale genesi è necessario un atteggiamento critico, consapevole cioè della crisi, anche economica e sociale, che si sta attraversando e degli strumenti razionali che possono fronteggiarla, almeno sul piano dell’analisi concettuale. In uno dei suoi ultimi scritti, dedicato a Seneca, Diderot si pone la domanda su quale sia l’oggetto della filosofia, offrendo una risposta che ne delinea stile e significato: «Legare gli uomini attraverso un commercio di idee e l’esercizio di una mutua beneficenza», tenendo presente che essa «non ci ordina di tormentarci». Il filosofo, continua, si sosterrà soltanto «attraverso la grandezza delle cose»<sup>4</sup>. Diderot non aggiunge altro e non rivela come valutare tali “grandezze”: ma, seguendo il suo pensiero, sappiamo che esse non si rivelano mai in astratto, bensì interpretando i molteplici orizzonti di senso delle cose stesse. È qui, nella loro effettualità e necessità, nella loro concreta “vita”, che potrà emergere la complessità qualitativa dei loro significati, che si rivelerà all’interno di un’articolata rete di relazioni intersoggettive, storiche, sociali, rete che contribuirà a determinarne il senso.

Seguendo allora un modello illuministico, bisogna forse trovare il coraggio di affermare che la valutazione, ove vi sia commercio di idee, condiziona sempre, ha sempre condizionato e sempre condiziona, la produzione di pensiero: i processi sociali – si pensi proprio alla genesi della *Enciclopedia* diderotiana – hanno sempre inciso sui processi culturali. Una valutazione che non cerchi di incidere in essi – che è ciò che fanno quotidianamente i professori – rischia di essere a priori fallimentare. Le motivazioni che la valutazione segue non sono separabili dai contenuti spirituali degli oggetti valutati. Tradurre l’idea astratta della valutazione in processi (per lo più verbali, e verbalmente schematici) deve cercare di comprenderne tutti gli spessori possibili e i

---

<sup>4</sup> La citazione è tratta da una delle ultime opere di Diderot, pubblicata nel 1778 (ma con la data dell’anno successivo) con il titolo *Essai sur la vie de Sénèque et sur le régnes de Claude et Neron*. In seguito il titolo venne semplificato in *Essai sur les régnes de Claude et Neron*. Qui si cita dall’edizione pubblicata con questo titolo da Hermann, Paris 1986, p. 121.

mutui legami, nella consapevolezza, tuttavia, che tali processi sono un orizzonte aperto e in divenire, che si realizzano in vari livelli di narrazioni e argomentazioni.

La valutazione, in conclusione, non è un'ontologia, bensì un percorso che le scienze umane devono analizzare nella sua storicità, nella consapevolezza che la sintesi ha senso solo in una direzione "bachtiniana", come risultato di un dialogo, in una logica della comunicazione spirituale che è alla base dei saperi umanistici: una logica che si fonda su ciò che Bachtin chiamava "comprensività", che significa affermazione della necessità di una coscienza storica che, forte di un senso di continuità tra passato e avvenire, sia anche autentica e consapevole percezione del presente.

Affermare che la valutazione non ha un significato ontologico significa dedurre che non può avere lo scopo di costituire una domanda fondamentale sull'essere della ricerca scientifica. Questo è l'equivoco che paradossalmente spesso unisce valutatori e valutati, forse non sempre consapevoli che non siamo di fronte a un sistema chiuso con significati profondi, emergenti o nascosti, bensì a una "genesì di senso" (da fenomenologo quasi direi: "a una fenomenologia dell'esperienza") di cui bisogna descrivere le specifiche processualità, che non definiscono l'essere "eterno" delle cose, bensì soltanto un "modo" per guardarle, con i suoi precisi limiti e confini. La sua funzione è "critica", è cioè quella di interrogarsi su come si dispiegano nelle "cose stesse" i suoi significati complessi, stratificati, mediati, articolati. Per tale motivo non può venire "entificata" e "tecnicizzata", come qualcosa di separato dai processi attraverso i quali si esplicita. Interrogare tali processi significa manifestare metodi storici, dunque dialogici e non entificabili, che hanno la funzione primaria di rispondere a esigenze sociali (per non dire economiche: il problema infatti viene enfatizzato nel momento in cui diminuiscono le risorse da destinare alla cosiddetta ricerca "di base"). Il percorso della valutazione non può mai essere autoreferenziale, posto in un mondo di enti chiusi che lo allontana dall'esperienza soggettiva e intersoggettiva della ricerca. Deve invece essere sempre in comunicazione con il tempo della storia, con quel piano in cui le cose sono per noi.

Per evitare una tecnicizzazione dei processi valutativi, un predominio cioè della valutazione sulla realtà da valutare, va dunque tenuta viva la storicità dell'interrogazione, senza che venga occultata, attraverso processi valutativi applicati come vestiti, la specificità della ricerca umanistica. La valutazione non può mai cadere vittima di "sedimentazioni" che, nell'oggi, e nell'orgia dell'attualità e della attualizzazione, fanno dimenticare l'attenzione ai processi del senso, a ciò che ha condotto fino al presente. Se questa è la funzione epistemologica della valutazione, se si accetta che essa debba incidere sulla ricerca stessa, che nasca non in un'astratta libertà, ma in un dialogo critico vincolato dalle situazioni e dalle circostanze sociali, la domanda decisiva, e finale, può forse cambiare: queste operazioni sono, *oggi*, messe in atto correttamente? Le ultime decisioni di Anvur non sembrano, detto con assoluta semplicità, e sempre in spirito di dialogo, andare in questa direzione. Inserire le recensioni a volumi, ma solo quelle pubblicate su riviste di fascia A, tra i fattori di valutazione, oltre al rischio di perpetuare gli equivoci sulla validità di tali liste, non è processo che "funzioni" con uguale significato nei vari settori umanistici. Rischia, di conseguenza, di essere un ulteriore

elemento di confusione, di divisione e di arbitrio, senza portare alcun contributo per determinare il “valore” delle monografie.

Ancora più discutibile la questione, molto tecnica, di quella valutazione della ricerca dipartimentale chiamata SUA-RD, non banale dal momento che i suoi esiti concorreranno a definire una parte della quota premiale del Fondo di Funzionamento Ordinario (FFO). Non se ne possono qui discutere i singoli passaggi, ma stupisce che si sia chiesto di fare in poche settimane quel che sarebbe difficile concludere in mesi. Come scrive con saggezza Paola Galimberti, «le procedure di valutazione sono esercizi complessi, che devono essere costruiti con attenzione, attraverso un percorso condiviso e soprattutto cercando di non strangolare gli atenei nelle attività burocratiche. Gli indicatori vanno testati e validati (certamente non in 15 giorni e nemmeno in tre mesi), perché fungendo poi da linea guida andranno ad influire pesantemente sulla attività scientifica di ricercatori, Dipartimenti ed Atenei»<sup>5</sup>.

La necessità attuale è dunque quella, poco battagliera, ma forse utile, di trovare mediazioni condivise, avendo però anche il tempo, prima di attivarle, di una reale sperimentazione. Se si ritiene invece la valutazione solo un veloce processo tecnico, non attento a quel che viene valutato, secondo un approccio “scientifico”, non potrà svilupparsi un’autentica discussione sui suoi modi e sulle sue finalità. Modi e finalità che potranno realizzarsi solo cogliendone gli aspetti “intersoggettivi” e “comunitari”, considerando la valutazione come un’esperienza “comune” dove viene fatto agire un “discorso” – un *logos* non poi così distante da quello che vediamo all’opera nelle ultime pagine del *Sofista* platonico – che nel descrivere le qualità della cose riveli al tempo stesso la “comunanza” sociale che sta a base di tutti i processi intersoggettivi. Valutare, nelle scienze umane, non è una pratica tecnica, tantomeno quantitativa, bensì un’opera per “far proprio” un prodotto di ricerca. Per usare il linguaggio della tradizione fenomenologica siamo all’interno di una “relazione motivazionale” fra persone e cose: gli oggetti sono qui “stimoli”, instaurano una comunicazione che crea una rete di relazioni temporali complesse.

Si può osare, su queste basi, una perentoria conclusione generale. Le assolutizzazioni ontologiche, che ossificano e ipostatizzano i diversi modi temporali di valutazione, conducono alla falsificazione dei significati complessi, alla loro “entificazione”. Per tale motivo, la valutazione non può cadere né nell’ovvietà naturalistica di un’evidenza immediata estranea a una genesi costitutiva, né nella mascheratura di una perdita di senso delle concrete comunità di ricerca di fronte alla volontà o al sistema delle cose, al falso sapere che si presenta come rinuncia al sapere nell’elogio dell’elenco e del fatto. Ma evitare questi pericoli implica atteggiamenti “critici” – un esercizio razionale del “giudizio” – nei confronti delle certezze irriflesse che ancora abitano sia nei valutatori sia nei valutati.

---

<sup>5</sup> P. GALIMBERTI, *SUA RD - Le responsabilità dei Direttori di Dipartimento*, 3 gennaio 2015 (<http://www.roars.it/online/sua-rd-le-responsabilita-dei-direttori-di-dipartimento/>).

# STUDI



Steve Jacob  
Wilfried Seyive Affodegon

## CONDUCTING QUALITY EVALUATIONS: FOUR GENERATIONS OF META-EVALUATION

### **Abstract**

*At the end of the 1960s, Michael Scriven coined the concept of meta-evaluation. In the strictest sense of the term, a meta-evaluation is the evaluation of an evaluation. From a theoretical point of view, it can be used to assess the value of one or more evaluations. In a practical sense, it helps the planning of evaluations. Thanks to its increasingly widespread use, meta-evaluation has emerged as a method of quality assurance for evaluations. To meet the requirement for quality, more and more summative and formative meta-evaluations have been conducted and published over the last twenty years by both evaluators and scholars. This article includes a theoretical overview of meta-evaluation and reviews its foundations, methods, practices, and objects.*

### *Introduction*

Evaluation, as an applied science, entails a perpetual quest for improvement, as evaluators seek the codes and instruments that will allow them to ensure the quality and validity of their conclusions and recommendations. Given the range of evaluation objectives, a number of quality issues are present. First, the evaluation must centre on a specific need for information. Second, it must lead to a judgment on public actions that is based on explicit criteria. Third, it must generate useful, evidence-based recommendations. And last, it must provide information as an input for the decision-making process. As a result, the literature on evaluation quality is rich and constantly evolving. It outlines four evaluation quality assurance approaches. The first is the structural approach developed by Schwartz and Mayne (2005), which involves the elaboration of standards and other directing principles to orient evaluative practice. The second is the systemic approach, which consists of ensuring the reliability of information collection mechanisms during the evaluative process (Bornmann *et al.* 2006). Third, unlike the systemic approach which focuses on the information collection system, the formative approach allows evaluators to ensure the quality of the information at the time of its collection and production. Fourth, the summative approach is similar to the formative approach in the sense that it also concentrates on information quality, but only once it has been produced (Daigneault 2008). These different approaches use precise instruments for their operationalization. One of the methods which allows evaluators to link together several quality assurance approaches is without doubt meta-evaluation.

Meta-evaluation, also referred to as “second-level evaluation”, is an application of the formative and summative quality assurance approaches (Daigneault 2008) because it allows evaluators to ensure, both before and after an evaluation, the quality of their work. Also, meta-evaluation mobilizes instruments such as standards and directing principles for its implementation. In this way, it may appear to be a structural approach.

The origin of the term “meta-evaluation” is attributed to Michael Scriven and dates back to the 1960s (Cook 1978; Reineke and Welch 1986; Stufflebeam and Shinkfield 2007; Stufflebeam 2001a; 2011). Meta-evaluation is defined as being the evaluation of the evaluation and indirectly of the evaluator (Scriven 1991). In an editorial entitled *Meta-Evaluation Revisited*, Scriven explains, “I published my first article about ‘meta-evaluation’ (Scriven 1969), a term I had invented somewhat earlier in a report to the Urban Institute, who had asked me for help in dealing with the non-comparability of the evaluations they had commissioned for several housing projects” (Scriven 2009, p. iii). The main rationale for the existence of meta-evaluation is to respond to criticisms and concerns about the value of evaluations. Reineke and Welch (1986) find expectations on the subject of meta-evaluation in Stufflebeam’s writings. On the one hand, evaluators are increasingly required to demonstrate the quality of their work; on the other hand, it is fitting that they should evaluate their own work. Meta-evaluation allows evaluators to meet this double expectation.

Despite the simplicity of the concept’s definition, the operationalization of meta-evaluation has taken place in several stages, and has included the development of several tools for its implementation and several theoretical approaches for its analysis. Nevertheless, it seems essential to develop characteristics common to the profession to better determine the issues of meta-evaluation (Cooksy and Caracelli 2008) which is now presented by several authors as a professional obligation (Hanssen *et al.* 2008; Jacob and Boisvert 2010; Stufflebeam and Shinkfield 2007; Stufflebeam 2011). When we speak of meta-evaluation, three recurrent questions emerge:

- (i) exactly what it is,
- (ii) how it can be justified, and
- (iii) when and how it should be used (Scriven 2009 p. iii).

Based on a review of the literature, our research will attempt to answer these questions. The objective of this exploratory study is to describe meta-evaluative practice.

A systematic review was used to create an inventory of relevant publications. For the purposes of our research, we explored the Ariane 2.0 Article Search database (the research interface at the Université Laval library), the Web of Science (1990-2013), the Worldwide Political Science Abstracts (1975-2013), the International Bibliography of the Social Sciences (1951-2013) and PAIS International (1972-2013). We identified publications which evaluated the quality of evaluations using meta-evaluation as their method. In all, 28 articles were selected for the research project. The second author of this study, under the supervision of the first, read the publications based on the keywords and the abstracts of the various documents identified in order to select those which were truly adapted to the research objective.

With respect to the data and its analysis, the selected articles were read by the second author to exhaustively identify and extract the elements listed in the reading grid. Two readings of each article helped the authors to fill the grid. Generally, the second author

systematically copied the passages of articles which allowed him to complete the various sections of the grid. When the grids were completed, they were validated by the first author to ensure that the results were checked for reliability.

In the following sections, we will present the epistemological foundations for meta-evaluation, and will then describe the aims and theoretical approaches of meta-evaluation, along with the types of meta-evaluation associated with the evaluation management cycle. Last, we will discuss the four generations of meta-evaluation.

### 1. *Epistemological foundation of meta-evaluation*

More than just the evaluation of evaluation (in the sense of a match between the objectives and results of the evaluation), meta-evaluation is perceived as an evaluation of the quality of an evaluation (concerning the objectives, the evaluation process, and the results and use of the evaluation, etc.) (Chapman 2012; Jacob and Desautels 2014; Patel 2002; Reneike and Welch 1986; Stufflebeam 2001a). In fact, evaluation quality is a professional requirement, in terms of the technical production aspects of the evaluation (methodological rigour, coherence between the different phases, justification of the conclusions, etc.). This requirement acquires extra meaning when the key concern of evaluation use is considered. The quality of the evaluation is also an ethical requirement for the evaluator, personally, and for all the participants in the evaluation (Desautels and Jacob 2012). Evaluators must ensure that they act in accordance with their responsibilities, and with the highest respect for all the involved stakeholders. It is finally a corporatist requirement because the evaluator accomplishes his mission in view of the rules and codes which regulate the profession. Extending the idea of evaluation quality, Stufflebeam (2001a) refines the concept of meta-evaluation and defines it “as a procedure for describing an evaluation activity and judging it against a set of ideas concerning what constitutes good evaluation” (p. 134). This definition refers to the characteristics of a good evaluation which we find in the Standards for Educational Evaluation of the Joint Committee (1981; 1988; 1994; 2003; 2011) and the directing principles of several professional associations (American Evaluation Association [AEA] 1995; 2003; Société suisse d'évaluation [SEVAL] 2000; Société canadienne d'Évaluation [SCE] 2012; etc.). Stufflebeam and Shinkfield (2007) provide the most explicit definition of the concept by reaffirming the necessity of a good evaluation as an ethical principle and a professional obligation. According to these authors, “meta-evaluation is defined [...] as the process of delineating, obtaining, and applying descriptive information and judgmental information – about pertinent criteria – including the evaluand's utility, feasibility, propriety, and accuracy and its systematic nature, competent execution, integrity, respectfulness, and social responsibility – in order to guide the evaluation and report its strengths and weaknesses” (p. 651). From the perspective of this definition, it is possible to identify the characteristics of the concept of “meta-evaluation”. These characteristics revolve around its aims, its theoretical approaches, its criteria of analysis, its practices and its object, making it possible to answer the three principal questions relative to meta-evaluation outlined above.

## *2. Aims of meta-evaluation*

Meta-evaluation is found in the register of evaluation practices for reasons of accountability (normative aim) and decision-making (instrumental aim) (Coosky and Caracelli 2005; Reneike and Welch 1986; Stufflebeam 2001a; 2011). Accountability concerns the evaluator. Meta-evaluation is a method of self-evaluation for the evaluator. It helps to improve evaluators' practices because it allows them to look retrospectively on their own actions so as to improve their practices (Cook 1978; Reneike and Welch 1986; Stufflebeam 2011). In this way, "as professionals, evaluators need meta-evaluations to assure the quality of their evaluations, provide direction for improving individual studies as well as their developing evaluation approaches, and earn and maintain credibility for their services among both clients and other evaluators" (Stufflebeam 2001a, p. 184). In this retrospective process, evaluators learn to base the results of their work on reliable and valid norms so that their results are not contested. In a normative perspective, meta-evaluation can orient the norms for current evaluations and contribute to the elaboration of norms for future evaluations. The decisional perspective concerns the clients of evaluation. Meta-evaluation makes it possible to improve the robustness of data and consequently the use of evaluation results (Coosky and Caracelli 2005; Reneike and Welch 1986; Stufflebeam 2001a; 2011). In an instrumental perspective, it appears as a decision-making tool for evaluation users.

## *3. Theoretical approaches of meta-evaluation*

In relation to these two aims, two theoretical approaches orient the meta-evaluators in how they conduct their examination of the evaluation and provide instructions on how to conduct it: these two theoretical approaches are the "Evaluator-Centred Meta-Evaluation" (Stufflebeam 2001a; 2011) and the "Client-Centred Meta-Evaluation" (Reneike and Welch 1986). In the evaluator-centered approach, methodological rigour is the principal characteristic considered. Rigour is the essence of the meta-evaluation put forward by Scriven and Stufflebeam in their work. The main question is "to what extent does the evaluation meet the standards of evaluation and the directing principles of the profession?". This approach finds its inspiration in evaluative approaches such as the Questions and Methods-Oriented approach, mainly in the form of Objectives-Based Studies, Outcome Evaluation as Value-Added Assessment, Experimental Studies, Cost-Benefit Analysis Approach, Case Study Evaluations, etc. (see Stufflebeam 2001b). On the other hand, the client-centered approach is derived from theoretical approaches to evaluation such as the Client-Centered Studies/Responsive Evaluation of Stake (1967; 1975; 1999); and the Utilization-Focused Evaluation of Patton (1980; 1982; 1994; 1997) (see Stufflebeam 2001b). This approach has been popularized by Reneike and Welch (1986). In this approach, meta-evaluation is conceived as a negotiation game between the client, the evaluator and the meta-evaluator. Client-centered meta-evaluation is more focused on communication, notably the credibility of the evaluator, the characteristics of the client, the report and its presentation, etc. than methodology (i.e. evaluation design,

choice of measurement instruments, observations, sampling, analysis, etc.) (Reneike and Welch 1986).

#### 4. *Type of meta-evaluation*

We have identified three types of meta-evaluation in the literature. First, the type of meta-evaluation often conducted prior to an evaluation in a formative perspective (formative meta-evaluation). Secondly, the type often conducted at the end of an evaluation in a retroactive perspective (summative meta-evaluation). Finally, the type conducted simultaneously to an evaluation (concurrent meta-evaluation) for either a formative or summative purpose. Formative meta-evaluation helps evaluators plan their evaluation. It precedes the evaluation and helps to improve its quality (Hanssen *et al.* 2008; Stufflebeam 2001). It helps in the selection of the evaluation's objectives, from design to results and the desired impacts of the evaluation (Stufflebeam 2011). Summative meta-evaluation helps users to determine the quality of the evaluation by revealing its strengths and weaknesses (Coosky and Caracelli 2005; Stufflebeam 2001; 2011). Concurrent meta-evaluation was developed by Cook (1978) and then later by Hanssen *et al.* (2008), who affirm:

“the concurrent meta-evaluation differs from both formative and summative meta-evaluations because concurrent meta-evaluation (a) is conducted simultaneously with the development and implementation of a new evaluation method; (b) has both formative and summative components; (c) is comprehensive in nature; and (d) includes multiple, original data collection methods” (p. 575).

The literature on meta-evaluation remains dominated by the writings of Stufflebeam. He certainly is the author who has spent the most time reflecting on the operational implementation of meta-evaluation. Based on his writings, the objectives of meta-evaluation relate (1) to the objectives of the evaluation, (2) to design, (3) to process and (4) to the results of the evaluation. While formative meta-evaluation demonstrates separate interest in these various objectives, summative meta-evaluation focuses on all of the objectives together (Stufflebeam 2011). Figure 1 presents a look at meta-evaluation as well as the principal authors involved in the development of this method of quality assurance.

Evaluation must meet a quality requirement (Baker 1983; Green and Attkisson 1984; Mark and Pines 1995). At the heart of the profession, meta-evaluation is the method able to outline quality issues in evaluation. If meta-evaluation ensures the quality of the evaluation before and after its implementation, evaluators aim to use the lessons it provides to answer the question: What are the principal challenges and problems to overcome in order to conduct a quality evaluation?

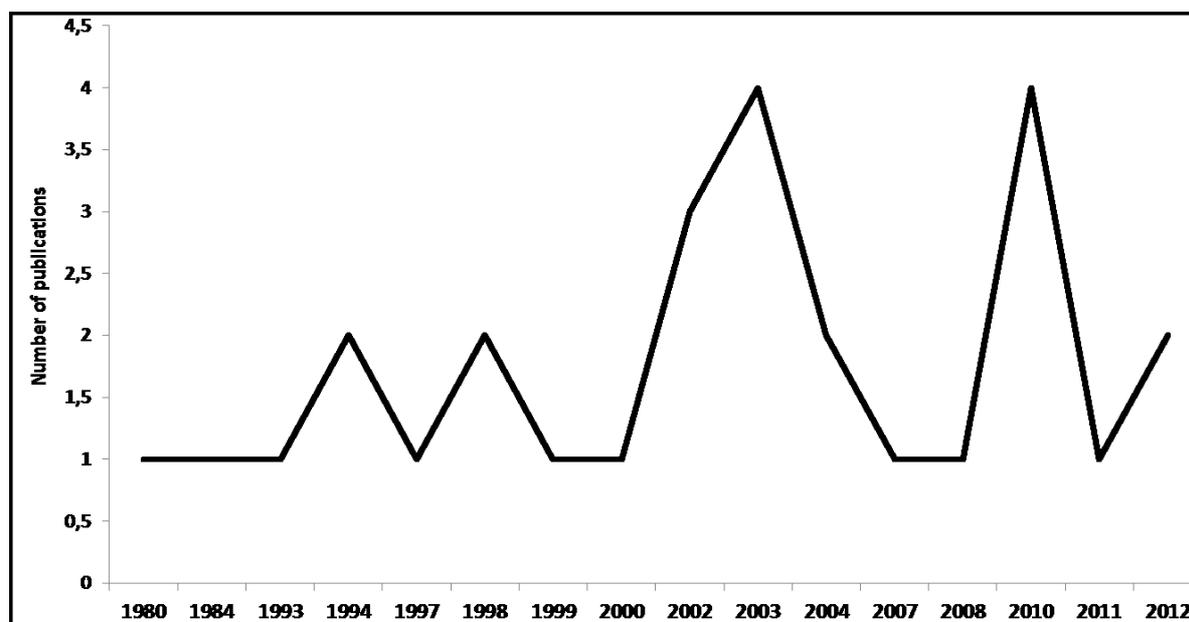
Figure 1: A Look at Meta-Evaluation

<u>Definition</u>	Objects	Instruments	Type of Meta-evaluation	Aims	Theoretical Approaches
<p>“Meta-evaluation is defined [...] as the process of delineating, obtaining, and applying descriptive information and judgmental information – about pertinent criteria – including the evaluand’s utility, feasibility, propriety, and accuracy and its systematic nature, competent execution, integrity, respectfulness, and social responsibility – in order to guide the evaluation and report its strengths and weaknesses” (Stufflebeam and Shinkfield 2007, p. 651).</p>	<p>(i) Evaluation objectives (ii) Evaluation design (iii) Evaluation processes (iv) Evaluation results (Stufflebeam 2011).</p>	<p><u>Standards</u> Standards for Educational Evaluation Joint Committee: utility, feasibility, propriety, and accuracy (1981; 1988; 1994; 2003; 2011) <u>Directing Principles</u> Directing principles of professional associations: systematic nature, competent execution, integrity, respectfulness, and social responsibility (AEA 1995; 2003) and other directing principles from other professional associations (SEVAL 2000; SCE 2012; etc.).</p>	<p><u>Formative Meta-evaluation</u> This helps evaluators plan their evaluation. It precedes the evaluation and helps to improve its quality. (Hanssen <i>et al.</i> 2008, Stufflebeam 2001).</p>	<p><u>Normative Aim (Accountability)</u> Meta-evaluation allows evaluators to retrospectively examine their own actions so as to improve their practices. This is an instrument of self-evaluation for the evaluator (Reneike and Welch 1986; Stufflebam 2011).</p>	<p><u>Evaluator-Centred Approach</u> Methodological rigour is the principal characteristic. To what extent does the evaluation address the evaluation standards and the directing principles of the profession? (Scriven 1966; 2009; Stufflebeam 2001; 2002; 2011; Stufflebeam and Shinkfield 2007)</p>
			<p><u>Summative Meta-evaluation</u> This allows users judge the quality of the evaluation and highlights its strengths and weaknesses (Coosky and Caracelli 2005; Stufflebeam 2001 2007).</p>	<p><u>Instrumental Aim (Decisional)</u> Meta-evaluation helps to increase the defense of the robustness of the data and consequently the use of results in evaluation (Coosky and Caracelli 2005; Reneike and Welch 1986; Stufflebeam 2001; 2011)</p>	<p><u>Client-Centred Approach</u> In addition to methodology, this approach focuses on communication, specifically the credibility of the evaluator, the characteristics of the client, the report and its presentation, etc. (Reneike and Welch 1986)</p>
			<p><u>Concurrent Meta-evaluation</u> This occurs simultaneously to the evaluation with the goal of evaluating a new evaluative approach (Cook 1978; Hanssen <i>et al.</i> 2008)</p>	<p><u>Normative Aim and Instrumental Aim</u></p>	<p><u>Evaluator-Centred Approach and Client-Centred Approach</u></p>

### 5. Four generations of meta-evaluation

Figure 2 shows trends in the number of publications relative to meta-evaluation. A review of these texts highlights the emergence, beyond the first generation, of three more generations of meta-evaluation.

Figure 2: Curve of Evolution of the Number of Articles by Year 1980-2012 (N=28)



*Development Period.* Since its appearance, meta-evaluation has evolved over four decades. The 1970s saw the first generation of meta-evaluation. During this period, evaluation researchers attempted to clarify the concept. In this way, research laid the theoretical foundations for meta-evaluation (Scriven 1969; Stufflebeam 1974). The title *Can Meta-Evaluation Give a Direction for Research on Evaluation?* by Gowin and Millman (1978) clearly evokes the nature of the discussion during this period. The primary concern of researchers was to answer the following questions: (i) What is meta-evaluation? (ii) Why meta-evaluation? (iii) When should we conduct meta-evaluation? Few reports of meta-evaluation were conducted (Baker *et al.* 1980). During this time, related research was published in the form of communications at scholarly or professional association conferences to validate the concept so as to operationalize it. One example is Gray (1978), who made a presentation on meta-evaluation at the annual meeting of the American Educational Research Association in Toronto (Canada). This is also the case for Gracia and Kapes (1982) who conducted a meta-evaluation to synthesize the studies conducted from 1968 to 1979 on the effect of participating in vocational education and of which the results were presented in 1982 at the annual meeting of the Southwest Educational Research Association.

*Propagation Period.* The second generation of meta-evaluation corresponds to the first period of the curve which begins in the early 1980s and ends in the early 1990s. Only a few meta-evaluations were conducted during this period (3 in the sample of this study). Although the theoretical reflections continued (Martin 1982; Reineke and Welch 1986),

the practice of meta-evaluation had acquired a firmer framework. The main question raised in this era was “How do we conduct meta-evaluation?” The answer to this question is found in the Standards for Educational Evaluation of the Joint Committee (1981; 1988) which became available for use by both evaluators and meta-evaluators. However, the popularity of these instruments remained embryonic, leaving it to researchers to use their own knowledge of evaluation to enact the criteria serving to conduct meta-evaluations (Green *et al.* 1980; White *et al.* 1984; Boyd and Windsor 1993).

*Enrolment Period.* The third generation of meta-evaluation goes from the mid-1990s and ends in the 2000s. During this period the production of knowledge on meta-evaluation doubled (7 meta-evaluations in the sample of our study) compared to the Propagation Period. The practice of meta-evaluation also took off, with many different instruments delimiting the boundaries of the field, including the AEA guidelines (1995) which made meta-evaluation a professional obligation. A more refined and operational version of the Standards for Educational Evaluation of the Joint Committee (1994) appeared, strengthening practice. Specifically, the program evaluation models meta-evaluation checklist (based on the Program Evaluation Standards) of Daniel L. Stufflebeam (1999) offered a framework for responding to the recurrent concern of “How do we conduct a meta-evaluation?”.

*Maturity Period.* Once meta-evaluation had been codified, the fourth generation of meta-evaluation was reflected in the two peaks of the curve which showed a net progression of meta-evaluative practice (18 meta-evaluations in the sample of our study). Instruments became more operational (Joint Committee 2003; 2011; AEA 2003). Many other professional evaluation associations adopted or adapted the model of the AEA and the Joint Committee concerning standards or directing principles (SEVAL 2000; SCE 2012; Mbaïrewaye and Jacob 2012) to strengthen the practice of meta-evaluation. In 2010, Scriven published *Evaluating evaluations: A meta-evaluation checklist*, confirming the maturity of this period.

### *Conclusions*

Our objective was to review the literature on meta-evaluation to create an overview of its epistemological foundations, its aims, the different types of meta-evaluation, and the main theoretical approaches. We also reviewed the evolution of meta-evaluation over the last four decades. This review shows that, from a conceptual point of view, meta-evaluation developed from being simply an evaluation of the results of an evaluation (in other words, of the match between the objectives and results) to an evaluation of the quality of the evaluation based on its design, implementation and impact, among other factors. In this way, meta-evaluation has developed through four generations that have shaped its form and defined its boundaries (or characteristics). The first decade of this development, the 1970s, corresponds to the development of the concept and its epistemological foundations. The 1980s saw the enrolment of researchers and theoretical discussion of ways to implement the concept. During the 1990s, meta-evaluation took on a professional dimension as the practice became popularized within several professional networks. Last, the 2000s cemented the support of researchers, professional

associations and government bodies, who increasingly turned to meta-evaluation to assess the quality of an evaluator's work.

### *Bibliographical References*

- AMERICAN EVALUATION ASSOCIATION (1995), *Guiding Principles for Evaluators*, in “New Directions for Program Evaluation”, 66 (1995), pp. 19-26.
- AMERICAN EVALUATION ASSOCIATION (2003), *Guiding Principles for Evaluators* (Ratified by the AEA Membership, July 2004) <http://www.eval.org/p/cm/ld/fid=51>.
- J.R. BAKER *et al.* (1980), *Meta Evaluation of the Saginaw Township Middle School Enrichment Center Project*, Saginaw Township Community Schools, Saginaw MI 1979-80.
- F. BAKER (1983), *Quality Assurance and Program Evaluation*, in “Evaluation & the Health Professions”, 6 (2/1983), pp. 149-160.
- L. BORNMANN-S. MITTAG- H.D. DANIE (2006), *Quality Assurance in Higher Education. Meta-Evaluation of Multi-Stage Evaluation Procedures in Germany*, in “Higher Education: The International Journal of Higher Education and Educational Planning”, 52 (4/2006), pp. 687-709.
- N.R. BOYD JR.-R.A. WINDSOR (1993), *A Meta-Evaluation of Nutrition Education Intervention Research among Pregnant Women*, in “Health Education Quarterly”, 20 (3/1993), pp. 327-345.
- L.S. CHAPMAN (2012), *Meta-Evaluation of Worksite Health Promotion Economic Return Studies: 2012 Update*, in “American journal of health promotion”, 26 (4/2012), pp. TAHP 1-TAHP 12.
- T.D. COOK-C.L. GRUDER (1978), *Metaevaluation*, in “Evaluation Quarterly”, 2 (1/1978), pp. 5-51.
- L.J. COOKSY-V.J. CARACELLI (2005), *Quality, Context, and Use. Issues in achieving the goals of metaevaluation*, in “American Journal of Evaluation”, 26 (1/2005), pp. 31-42.
- L.J. COOKSY-V.J. CARACELLI (2008), *Metaevaluation in Practice: Selection and Application of Criteria*, in “Journal of MultiDisciplinary Evaluation”, 6 (11/2008), pp. 1-15 ([http://journals.sfu.ca/jmde/index.php/jmde\\_1/article/view/211/223](http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/211/223)).
- P.M. DAIGNEAULT (2008), *L'examen de la qualité des évaluations fédérales: une méta-évaluation réussie?*, “The Canadian Journal of Program Evaluation”, 23 (2/2008), pp. 191–224.
- G. DESAUTELS-S. JACOB (2012), *The Ethical Sensitivity of Evaluators: A Qualitative Study Using a Vignette Design*, in “Evaluation. The International Journal of Theory, Research and Practice”, 18 (4/2012), pp. 438-451.
- G. GARCIA JR.-J.T. KAPES (1982), *A Review and Analysis of the Meta-Evaluation Study: “The Effects of Participating in Vocational Education”*, Paper presented at the Annual Meeting of the Southwest Educational Research Association.
- P.J. GRAY (1978), *A Technology for Program Documentation and Evaluation: A Pilot Meta-Evaluation*, Paper presented at the Annual Meeting of the American Educational Research Association.
- R.S. GREEN-C.C. ATTKISSON (1984), *Quality Assurance and Program Evaluation: Similarities*

- and Differences*, in “American Behavioral Scientist”, 27 (5/1984), pp. 552-582.
- C.E. HANSEN-F. LAWRENZ-D.O. DUNET (2008), *Concurrent Meta-Evaluation. A Critique*, in “American Journal of Evaluation”, 29 (4/2008), pp. 572-582.
  - S. JACOB-Y. BOISVERT (2010), *To Be or Not to Be a Profession: Pros, Cons and Challenges for Evaluation*, in “Evaluation”, 16 (4/2010), pp. 349-369.
  - S. JACOB-G. DESAUTELS (2014), *Assessing the Quality of Aboriginal Program Evaluations*, in “Canadian Journal of Program Evaluation”, 29 (1/2014), pp. 62-86.
  - JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION (1981), *Standards for evaluations of educational programmes, projects and materials*, McGraw-Hill, New York NY 1981.
  - JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION (1988), *The personnel evaluation standards*, Sage, Newbury Park CA 1988.
  - JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION (1994), *The programme evaluation standards*, Sage, Beverly Hills CA 1994<sup>2</sup>.
  - JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION (2003), *The student evaluation standards*, Corwin Press, Thousand Oaks CA 2003.
  - JOINT COMMITTEE ON STANDARDS FOR EDUCATIONAL EVALUATION (2011), *The Program Evaluation Standards*, Sage, Newbury Park CA 2011.
  - M.M. MARK-E. PINES (1995), *Implications of Continuous Quality Improvement for Program Evaluation and Evaluators*, in “American Journal of Evaluation”, 16 (2/1995), pp. 131-139.
  - P.H. MARTIN (1982), *Meta-Analysis, Meta-Evaluation and Secondary Analysis*, Washington DC, Eric 1982.
  - M.H. MBAÏREWAYE-S. JACOB (2012), *Cultural Dimensions of the African Evaluation Guidelines*, in J.-C. BARBIER-P. HAWKINS (eds.), *Evaluation Cultures Sense-making in Complex Times*, Transaction Publishers, New Brunswick NJ, pp. 205-237.
  - M. PATEL (2002), *A Meta-Evaluation, or Quality Assessment, of the Evaluations in this Issue, Based on the African Evaluation Guidelines*, in “Evaluation and Program Planning”, 25 (4/2002), pp. 329-332.
  - R.A. REINEKE-W.W. WELCH (1986), *Client-Centered Meta-Evaluation*, in “American Journal of Evaluation”, 7 (3/1986), pp. 16-24.
  - R. SCHWARTZ-J. MAYNE (2005), *Assuring the Quality of Evaluative Information: Theory and Practice*, in “Evaluation and Program Planning”, 28 (1/2005), pp. 1-14.
  - M. SCRIVEN (1969), *An Introduction to Meta-Evaluation*, in “Educational Products Report”, 2 (5/1969), pp. 36-38.
  - M. SCRIVEN (1991), *Evaluation thesaurus*, Sage, New York 1991<sup>4</sup>.
  - M. SCRIVEN (2009), *Meta-Evaluation Revisited*, in “Journal of MultiDisciplinary Evaluation”, 6 (11/2009), p. iii-viii ([http://journals.sfu.ca/jmde/index.php/jmde\\_1/article/view/220/215](http://journals.sfu.ca/jmde/index.php/jmde_1/article/view/220/215)).
  - SOCIÉTÉ CANADIENNE D'ÉVALUATION, *The Program Evaluation Standards: A Guide for Evaluators and Evaluation Users*, Sage, Thousand Oaks CA 2011<sup>3</sup>.
  - SOCIÉTÉ SUISSE D'ÉVALUATION (2000), *Standards d'évaluation*, [http://www.seval.ch/fr/documents/SEVAL\\_Standards\\_2001\\_fr.pdf](http://www.seval.ch/fr/documents/SEVAL_Standards_2001_fr.pdf).
  - D.L. STUFFLEBEAM (2001a), *The Metaevaluation Imperative*, in “American Journal of Evaluation”, 22 (2/2001a), pp. 183-209.
  - D.L. STUFFLEBEAM (2001b), *Evaluation Models*, in “New Directions for Evaluation”, 89 (2001), pp. 7-98.

- D.L. STUFFLEBEAM-A. SHINKFIELD (2007), *Evaluation Theory, Models, & Applications*, John Wiley, San Francisco CA 2007.
- D.L. STUFFLEBEAM (2011), *Meta-Evaluation*, in “Journal of MultiDisciplinary Evaluation”, 7 (15/2011), pp. 99-158.
- L.T. WHITE-D. ARCHER-E. ARONSON-L. CONDELLI-B. CURBOW-B. MCLEOD *et al.* (1984), *Energy Conservation Research of California’s Utilities. A Meta-Evaluation*, in “Evaluation Review”, 8 (2/1984), pp. 167-186.



SUGLI AUTORI/  
*ABOUT THE AUTHORS*

Wilfried Seyive Affodegon è Dottorando in Scienze Politiche presso l'Università Laval di Québec.

<https://www.pol.ulaval.ca/?pid=176>

Maria Cassella è Coordinatore della Biblioteca “Norberto Bobbio” dell'Università di Torino.

[http://www.dcps.unito.it/unitoWAR/page/dipartimenti8/D072/D072\\_personale\\_batch\\_BasicBook\\_Tecnici\\_IT2?id=179861](http://www.dcps.unito.it/unitoWAR/page/dipartimenti8/D072/D072_personale_batch_BasicBook_Tecnici_IT2?id=179861)

Peter Dahler-Larsen è Professore di Valutazione presso l'Università di Copenaghen.

[http://polsci.ku.dk/english/staff/academic\\_staff/?pure=en/persons/184767](http://polsci.ku.dk/english/staff/academic_staff/?pure=en/persons/184767)

Elio Franzini è Professore Ordinario di Estetica presso l'Università di Milano.

<http://dipartimento.filosofia.unimi.it/index.php/elio-franzini>

Jennifer C. Greene è Professore di Metodologie della Ricerca Quantitativa e Valutativa presso l'Università dell'Illinois.

<http://education.illinois.edu/people/jcgreene>

Enrico Guglielminetti è Professore Ordinario di Filosofia Teoretica presso l'Università di Torino.

<http://www.unito.it/persone/eguglie>

Steve Jacob è Professore di Scienze Politiche presso l'Università Laval di Québec.

<http://www.pol.ulaval.ca/?pid=222>

Patrizia Lemma è Professore Ordinario di Igiene Generale e Applicata presso l'Università di Torino.

[http://eduprof.campusnet.unito.it/do/docenti.pl/Show?\\_id=plemma](http://eduprof.campusnet.unito.it/do/docenti.pl/Show?_id=plemma)

Riccardo Manzotti è Ricercatore di Psicologia presso la IULM di Milano.

[www.consciousness.it](http://www.consciousness.it)

Donna M. Mertens è Professore di Metodi di Ricerca e di Valutazione presso la Gallaudet University di Washington DC.

<https://www.gallaudet.edu/Documents/EDU/DonnaMertensVitaAug10.pdf>

Angelo Miglietta è Professore Ordinario di Economia delle Aziende e dei Mercati Internazionali e di Economia dell'Unione Europea presso la IULM di Milano.

<http://www.iulm.it/wps/wcm/connect/iulmit/iulm-it/personale-docenti/dipartimento-di-marketing-comportamenti-comunicazione-e-consumi/professori-ordinari/miglietta-angelo>

Paolo Moderato è Professore Ordinario di Psicologia Generale presso la IULM di Milano.

<http://www.iulm.it/wps/wcm/connect/iulmit/iulm-it/personale-docenti/dipartimento-di-marketing-comportamenti-comunicazione-e-consumi/professori-ordinari/moderato-paolo>

Donatella Mutti è Responsabile dei Servizi di Supporto alla Ricerca presso il Dipartimento di Filosofia e Scienze dell'Educatione dell'Università di Torino.

[http://www.dfe.unito.it/unitoWAR/page/dipartimenti7/D070/D070\\_personale\\_batch\\_BasicBook\\_Tecnici\\_IT2?id=182113](http://www.dfe.unito.it/unitoWAR/page/dipartimenti7/D070/D070_personale_batch_BasicBook_Tecnici_IT2?id=182113)

Paola Ricchiardi è Ricercatore di Pedagogia Sperimentale presso l'Università di Torino.

<http://www.unito.it/persone/paola.ricchiardi>

Giulio Sapelli è Professore Ordinario di Storia Economica presso l'Università di Milano.

<http://www.giuliosapelli.it/>

Emanuela M. Torre è Ricercatore di Pedagogia Sperimentale presso l'Università di Torino.

<http://www.unito.it/persone/etorre>