Università Commerciale Luigi Bocconi
Econpubblica
Centre for Research on the Public Sector

SHORT NOTES SERIES

La valutazione dell'attività didattica nelle Università

Paola Profeta

Short note n. 4

November 2008

www.econpubblica.unibocconi.it

La valutazione dell'attività didattica nelle Università

Paola Profeta[1], Università Bocconi

Novembre 2008

La valutazione della scuola e delle università è in questi giorni al centro di un acceso ed intenso dibattito. E' indubbio che l'abilità del corpo docente a tutti i livelli (maestre, professori, docenti universitari) nello svolgere la propria attività è essenziale per migliorare il sistema di istruzione italiano e i risultati di apprendimento dei nostri studenti, e che i margini di miglioramento siano molto ampi. Una corretta valutazione della performance del corpo docente può rappresentare uno strumento importante per migliorare la qualità della didattica, agendo come incentivo per i docenti sottoposti a valutazione e come riferimento per un sistema di pagamento "meritocratico", in base alla capacità e all'impegno di ogni singolo docente. Ma quali sono gli strumenti più appropriati per valutare le performance dei docenti? Quali sono i criteri e i parametri che dovrebbero essere alla base di questi strumenti? Come si misura la "bravura" di un insegnante?

Il tema è molto ampio e molto complesso. Molti studiosi si sono appassionati alla ricerca di una risposta a queste domande, senza raggiungere accordi unanimi. In particolare, mentre è da tutti riconosciuto che la valutazione sia essenziale, la discussione sulle modalità per la sua implementazione è aperta. In questa nota mi soffermo sulle considerazioni emerse nella letteratura, focalizzando l'attenzione al caso della valutazione della didattica dei docenti universitari. Una pluralità di posizioni caratterizza la discussione, che, nonostante riceva un'attenzione minore, è in realtà più controversa di quella sui criteri di valutazione dell'attività di ricerca.

La valutazione dell'attività didattica dei docenti universitari è un elemento cruciale della misurazione delle abilità di insegnamento e della performance generale dei docenti, importante soprattutto per le analisi sull'efficienza del sistema educativo, le decisioni sulle progressioni di carriera e sulle retribuzioni dei docenti, laddove si pensi ad un legame tra performance e retribuzione.

Uno degli strumenti principali adottato ormai da quasi tutte le Università per misurare la qualità della didattica è la valutazione dei docenti da parte degli studenti, basata sulla somministrazione di questionari, ai quali gli studenti rispondono tipicamente in forma anonima.

L'uso di questo strumento nasce nel cosiddetto approccio "*marketing*" adottato dagli amministratori di *colleges* e di università (Bush, Ferrell & Thomas, Jr, 1998, Wright 2000),

---

[1] Dipartimento di Analisi Istituzionale e Management Pubblico e Econpubblica, Università Bocconi, Via Röntgen 1, 20136 Milano. Tel:0258365309, fax:0258365318,paola.profeta@unibocconi.it. Ringrazio Giampaolo Arachi, Roberto Artoni, Alessandra Casarico e Alessandro Sommacal per i loro commenti.

secondo cui gli studenti sono i consumatori di un servizio offerto dall'Università e, come tali, sono chiamati ad esprimere la loro opinione sulla soddisfazione del servizio ricevuto, allo scopo di migliorare la qualità della didattica (Marsh, 1991). Grazie alle informazioni ricevute dagli studenti, i docenti possono capire come migliorare la qualità delle loro lezioni universitarie. E' solo attraverso i feedback dagli studenti che si possono acquisire informazioni essenziali per migliorare la qualità della docenza.

E' chiaro però che l'Università non è un luogo in cui si "vendono" una serie di insegnamenti, ma è soprattutto sede di istruzione superiore da cui gli studenti dovrebbero uscire avendo studiato, imparato e accumulato capitale umano. Spesso questi due obiettivi, soddisfare il "cliente-studente" e accrescere in modo efficace e consapevole la sua conoscenza, possono essere in conflitto. Da qui i dubbi e gli aspetti critici che emergono sulle valutazioni della qualità didattica fornite degli studenti.

Sarebbe impossibile e molto al di là delle mie competenze fornire qui una rassegna esaustiva della enorme letteratura che si è occupata dell'efficacia della valutazione da parte degli studenti nel misurare la qualità della docenza universitaria. Mi soffermerò quindi solo su alcuni aspetti. In particolare, evidenzierò tre gruppi di criticità associate a questo metodo di valutazione: 1) difficoltà nel cogliere in modo appropriato la qualità della didattica; 2) rischi che ne conseguono; 3) altri problemi di carattere più specifico.

1. La valutazione da parte degli studenti riesce a cogliere la qualità della didattica?

Nel primo gruppo di criticità rientrano i fattori che possono influenzare le valutazioni degli studenti, ma che non sono indicatori delle performance didattiche del docente. Tra questi emergono i seguenti:

a) Non è detto che lo score della valutazione da parte degli studenti sia indicatore di quanto gli studenti hanno imparato (O'Connell & Dickinson, 1993). L'evidenza empirica è contrastante: alcuni studiosi evidenziano una correlazione tra le valutazioni della didattica da parte degli studenti e la performance degli studenti (Cohen, 1981, Feldman, 1989), altri invece trovano una relazione addirittura negativa. Yunker e Yunker (2003) per esempio ottengono che gli studenti di classi i cui insegnanti hanno avuto valutazioni più alte hanno risultati peggiori di quelli di classi i cui insegnanti sono stati giudicati meno capaci (si veda anche Rodin & Rodin, 1972). Secondo Koon e Murray (1995) la correlazione tra il voto degli studenti all'esame finale e il voto dato all'insegnante è solo di 0.3. Inoltre, una volta finita l'università, nel mondo del lavoro gli studenti sono spesso giudicati meno preparati proprio nelle materie in cui avevano espresso la più alta soddisfazione, segnale che il legame tra valutazioni degli studenti e qualità dell'insegnamento può essere debole.
b) Esiste una forte asimmetria informativa, perché gli studenti non hanno la conoscenza necessaria per giudicare in maniera appropriata i loro insegnanti (Olshavsky & Spreng, 1995). Per esempio, non sanno se il professore ha fornito loro il materiale rilevante e se lo ha fatto in modo corretto. In questo contesto di asimmetria informativa, caratteristiche diverse dalla qualità dell'insegnamento possono influenzare il giudizio che gli studenti esprimono nei confronti dei loro professori, quali il carico di studio e, come sottolineerò tra breve, i voti (Marsh, 1987). Allo stesso modo, gli studenti tipicamente premiano le abilità comunicative del docente, indipendentemente dal loro

apprendimento, dal materiale e dal formato della lezione (Williams e Ceci, 1997), oppure premiano il sistema di votazione o addirittura l'aspetto estetico (Wright, 2000).

c) Alcune evidenze sottolineano l'esistenza di un *gender bias* nelle valutazioni: i risultati di valutazioni di docenti donne e uomini sono spesso sistematicamente diversi. Sulla base di un esperimento, Sinclair e Kunda (2000) mostrano che, quando gli standard sono elevati e il corso è difficile, l'avversione verso i docenti donne è molto più forte di quella verso i docenti uomini. L'interpretazione dei due studiosi è che di fronte ad una situazione per loro critica, le persone usano inconsciamente stereotipi negativi per ridurre la validità della difficoltà e aumentare la stima di sé stessi (si veda anche Kaschak, 1978). Nella stessa direzione vanno le evidenze di distorsioni razziali nelle valutazioni.

d) E' difficile misurare la qualità dell'insegnamento sulla base di un'immediata approvazione da parte degli studenti, perché si tratta di valore aggiunto che i ragazzi apprezzeranno solo nel tempo, in particolare sul lavoro o quando, per affrontare situazioni reali, avranno bisogno delle capacità analitiche, anche formative del loro carattere, sviluppate da alcuni insegnamenti (Alauddin e Tisdell, 2000). Può capitare addirittura che gli studenti preferiscano uno stile di insegnamento che poi risulta negativo per il loro apprendimento, come nel caso di corsi molto facili che non stimolano le capacità critiche (Strom, Hocevar, e Zimmer, 1990).

e) La misura della qualità non può essere uguale per tutti gli studenti, perché la stessa metodologia di insegnamento può essere efficace con uno studente, ma non con un altro. In particolare, corsi *undergraduate* e *graduate* andrebbero valutati in modo specifico e differenziato (Whitworth, Price e Randall, 2002). Spesso i corsi *undergraduate* hanno risultati sistematicamente migliori dei corsi *graduate*.

f) La misura della qualità non può essere uguale per tutti gli insegnamenti, anche all'interno della stessa Facoltà, perché ogni insegnamento richiede specifiche metodologie e gli studenti potrebbero avere predisposizioni diverse nei loro confronti. Alcuni elementi in particolare che caratterizzano la materia insegnata risultano sistematicamente correlati a diversi giudizi degli studenti: (i) la familiarità degli studenti con la materia (Jacobs e Kozlowski, 1985); (ii) la tipologia della materia, qualitativa o quantitativa, teorica o applicata: le materie quantitative e teoriche sono tipicamente associate a valutazioni peggiori da parte degli studenti, a causa anche, ma non solo, di una valutazione di "difficoltà" del corso (Yurkiewics, Hoefer e Byrne, 2006); (iii) la metodologia didattica: le lezioni frontali hanno tipicamente meno successo di quelle basate su *case studies* e *working groups* (Yurkiewics, Hoefer e Byrne, 2006). Whitworth, Price e Randall (2002) mostrano che, anche analizzando solo i corsi economici, la materia gioca un ruolo fondamentale nel determinare i risultati delle valutazioni da parte degli studenti, suggerendo quindi che potrebbe essere problematico utilizzare queste valutazioni per classificare l'intera Facoltà. Nella stessa direzione, l'analisi di Green, Calderon e Reider (1998) suggerisce che un portfolio più ampio di strumenti e non solo i giudizi degli studenti dovrebbe essere utilizzato per una valutazione complessiva della Facoltà.

2. Quali rischi possono derivare dall'uso delle valutazioni degli studenti come strumento per stabilire la performance dei docenti universitari?

Il secondo elemento da approfondire è quello dei rischi che possono derivare dall'uso della valutazione della didattica da parte degli studenti, quando questa non è un buon indicatore

della performance degli insegnanti. Questi rischi dipendono in particolare dalla possibilità che gli studenti effettuino le loro valutazioni considerando le aspettative di voto, e rende particolarmente critico l'uso di questo strumento nelle valutazioni di performance dei docenti per le decisioni di "*tenure*".

a) Numerosi studi empirici condotti sulle valutazioni effettuate dagli studenti in varie università statunitensi, soprattutto nelle facoltà di economia e business, evidenziano che gli studenti tendono a dare voti più alti ai docenti di corsi in cui si aspettano di ottenere voti più alti (Krautmann e Sander, 1997, Isely e Singh, 2005, Stapleton e Murkinson, 2001, Whitworth, Price e Randall, 2002). Si potrebbe pensare che voti più alti agli esami siano una prova della capacità del docente di insegnare meglio. In realtà l'evidenza empirica mostra che lo stesso docente, che insegna lo stesso corso in classi diverse ottiene valutazioni migliori nella classe la cui distribuzione dei voti è più alta. Inoltre, esiste una correlazione maggiore tra voto all'esame e voto dato al docente se consideriamo il voto relativo dello studente (voto in quell'esame rispetto agli altri esami) piuttosto che il voto assoluto dello studente (Rice, 1988; Greenwald e Gillmore, 1997). Queste evidenze suggeriscono che è plausibile pensare che si tratti di un comportamento distorto, espressione di quella che viene definita la "*leniency bias hypothesis*", cioè l'attitudine ad apprezzare e valutare positivamente coloro che ci approvano e a non amare chi ci critica. Gilbaugh (1982) riporta che 360 su 518 studenti intervistati alla San Jose State University (circa il 70%) ha ammesso che la loro valutazione degli insegnanti è stata influenzata dal voto che pensavano di ricevere, e Marsh (1987) riporta che una proporzione simile di professori riconosce che il voto e la difficoltà del corso possano essere un *bias* nelle valutazioni degli studenti. Le conseguenze di questa *leniency of grading* sono enormi. Il rischio maggiore è la diffusione di fenomeni di "*dumb down*" degli insegnamenti. Ryan et al. (1980) riportano i risultati di un'indagine in cui il 38% dei professori ha ammesso di fare lezioni più facili per migliorare le valutazioni degli studenti (si veda anche il caso descritto da Sacks, 1986). Questo fenomeno è particolarmente serio nel momento in cui sono valutati allo stesso modo docenti "*tenured"* e *"untenured"*: il rischio è che i docenti sotto esame di *tenure* siano incentivati a "*dumb down"* il loro corso, dare voti alti per aumentare la soddisfazione degli studenti e avere valutazioni migliori che li aiuteranno ad ottenere la *tenure.*
b) Un altro rischio, soprattutto per professori "*untenured"* è che la valutazione didattica da parte degli studenti possa limitare la libertà accademica dei docenti, che cercheranno di evitare di entrare in questioni controverse (di tipo politico, religioso ecc.) e daranno attenzione più alla forma che alla sostanza di quello che dicono.

3. Quali modalità della valutazione da parte degli studenti sono più appropriate?

Infine, ci sono problemi più pratici o specifici. Oltre alle intuitive difficoltà di come somministrare il questionario in classe per evitare risposte inaccurate, frettolose, o fasulle, e come aumentare l'interesse e la responsabilità degli studenti nel completare i questionari, la letteratura su questi temi ha evidenziato alcuni problemi di più difficile soluzione, quali i seguenti:

a) Controversia tra l'approccio multidimensionale, in cui contano i risultati separati delle diverse domande poste agli studenti nel questionario alle quali corrispondono

specifiche dimensioni di valutazione[2] e l'approccio unidimensionale (Abrami 1989, Abrami e d'Apollonia, 1997), in cui la valutazione del docente è sintetizzata da un unico indice globale di performance o da una somma pesata di diversi indici.

b) Pro e contro dell'anonimato dei questionari: i questionari sono tipicamente anonimi, per evitare il rischio che gli studenti non esprimano le loro vere opinioni per paura di essere identificati. In questo modo è però impossibile verificare se il rispondente è una persona che ha frequentato regolarmente, oppure in maniera sporadica, e se la sua valutazione sia legata alla frequenza, così come identificare il caso di studenti che si lamentano in modo "cronico". Alcuni studiosi (Wright, 2006) suggeriscono di mantenere l'anonimato solo nei confronti del docente, ma non nei confronti degli amministratori. Questa proposta però potrebbe comunque limitare la sincerità di espressione dello studente.

c) Inoltre la dimensione della classe può influenzare la valutazione: in classi con pochi studenti il docente riesce a stabilire un maggior legame con gli studenti che è tipicamente valutato in modo positivo (Neath, 1996).

Alcune riflessioni

Nonostante queste criticità, la valutazione da parte degli studenti continua a essere utilizzata come strumento per misurare l'abilità dei docenti universitari. Questo utilizzo è giustificato da alcuni potenziali vantaggi delle valutazioni degli studenti:

a) danno tipicamente risposte stabili e coerenti (gli studenti sono d'accordo nelle valutazioni);
b) sono facili da usare e da analizzare;
c) sembrano uno strumento più oggettivo di altri (per esempio lettere di giudizio da parte di osservatori esterni), anche perché si risolvono in un numero (o una serie di numeri);
d) è difficile pensare ad un'alternativa che sia esente da problemi. Per esempio, affidarsi al parere di esperti nel campo, come suggerito da Cahn (1986), può risultare problematico, perché gli esperti tendono a non essere d'accordo tra di loro (Marsh e Roche,1997).

Le criticità che sono emerse, tuttavia, stanno portando le università a orientarsi verso un utilizzo più limitato delle valutazioni da parte degli studenti. E' ormai condiviso che la complessità della valutazione della didattica universitaria non può essere affrontata basandosi esclusivamente sulle valutazioni da parte degli studenti (si vedano le efficaci riflessioni di Beltratti, 2007). Due principi sono condivisi dagli esperti (Huemer, 2007): (i) utilizzare le valutazioni degli studenti correggendo per alcune ben note distorsioni; (ii) integrare le valutazioni degli studenti con altri strumenti di valutazione della didattica.

Circa il primo principio, le correzioni che tipicamente ricevono maggiore attenzione sono:

- il *leniency of grading*;
- la differenziazione delle valutazioni per docenti *tenured e non-tenured*

[2] Seguendo l'idea che l'insegnamento sia multidimensionale, Marsh (1982) propone una "*Student's evaluations of educational quality*" (SEEQ) basata su 9 dimensioni: "*learning, enthusiasm, organization, group interaction, individual rapport, breadth, examinations, assignments, and overall*".

- la differenziazione delle valutazioni per materie diverse, in particolare per la materie quantitative e qualitative, teoriche e applicate
- anche se è ancora aperta la controversia tra il riferimento a un singolo *score* complessivo oppure agli *scores* nelle diverse dimensioni di valutazione, appare condivisa la necessità di pesare in modo diverso queste dimensioni. In particolare, fattori che misurano "il rapporto individuale" tra docente e studenti o la "difficoltà" del corso dovrebbero avere peso pressocché nullo.
- includere nella valutazione finale quanto emerge dai giudizi liberi dati in forma scritta (e non numerica) e scontare il risultato quando un voto negativo è stato dato senza una spiegazione, oppure con una spiegazione basata sulla difficoltà del corso, sulla personalità del docente o sui voti.

Circa il secondo principio, gli ulteriori strumenti che dovrebbero integrare le valutazioni da parte degli studenti nel giudicare la performance della docenza universitaria includono: il parere di esperti che assistono alle lezioni (dando feedback ai docenti) e che analizzano il materiale del corso; la somministrazione di questionari o interviste a studenti prossimi alla laurea o a ex-studenti, che possono valutare l'importanza che l'insegnamento ha avuto per loro dopo l'università (sul lavoro, o su altri elementi di lungo periodo); discussioni tra studenti in presenza degli amministrativi competenti; i feedback degli assistenti alla didattica; il confronto delle performance di studenti che svolgono lo stesso esame ma hanno seguito il corso con docenti diversi in classi parallele.

Per concludere, valutare la didattica universitaria è molto complesso. I giudizi degli studenti dovrebbero essere considerati con le dovute cautele, e il processo di valutazione nel suo complesso dovrebbe fondarsi anche su un'attenta ulteriore analisi di almeno tre elementi: il materiale del corso, la materia oggetto del corso e la tipologia di lezione.

Bibliografia

Abrami, P. (1985) Dimensions of effective college instruction, *Review of Higher Education* 8:211-228.

Abrami, P. (1989) "How should we use student ratings to evaluate teaching?", *Research in Higher Education*, 30, 2, 221-227.

d'Apollonia, S. e Abrami; P. C. (1997) "Navigating Student Ratings of Instruction," *American Psychologist* 52 (1997): 1198-1208.

Alauddin, M. e Tisdell, C. (2000) "Changing academic environment and teaching of economics at the University-level: Some critical issues analysed with the held of microeconomics", *Economic Papers*, 19, 1, 1-17.

Beltratti, A. (2007) "Ricerca e Didattica: Valutazione e Incentivi", *Rivista Italiana degli Economisti*, 2, 205-214.

Bush, V., Ferrell, O.C., & Thomas, Jr., J. L. (1998) “Marketing the business school: An exploratory investigation”, *Journal of Marketing Education* 20, 16-23.

Cahn, S. (1986), “Saints and Scamps: Ethics in Academia”, *Totowa, NJ: Rowman & Littlefield.*

Cohen, P. (1981) “Student Ratings of Instruction and Student Achievement: A Metaanalysis of Multisection Validity Studies.” *Review of Educational Research* 51, no. 3, 281-309.

Feldman, K. (1989), “The Association Between Student Ratings of Specific Instructional Dimensions and Student Achievement: Refining and Extending the Synthesis of Data from Multisection Validity Studies.” *Research in Higher Education* 30, no. 6, 583-645.

Gilbaugh, J. W. (1982) "Renner Substantiated," *Phi Delta Kappan* 63, 428.
Green Calderone e Reider (1998)

Green, B., Calderon, T., Reider, B. (1998), "A content analysis of teaching evaluation instruments used in accounting departments", *Issues in Accounting Education*, Vol. 13 ,1, 15-30.

Greenwald, A. e Gillmore, M. (1997) "Grading Leniency Is a Removable Contaminant of Student Ratings," *American Psychologist* 11: 1209-17.

Huemer, M. (2007) “Student Evaluations: A Critical Review”, mimeo University of Colorado

Isely, P. e Singh, H. (2005), “Do Higher Grades Lead to Higher Student Evaluations?” *Journal of Economic Education* 36, no. 1, 29-32.

Jacobs, R. e Kozlowski, S. (1985) “A Closer Look at Halo Error in Performance Ratings.” *Academy of Management Journal* 25, no. 1, 201-212.

Kaschak, E. (1978) “Sex Bias in Student Evaluations of College Professors”, *Psychology of Women Quarterly*, 2(3), 1978, 235-242.

Koon and Murray (1995) “Using multiple outcomes to validate student ratings of overall teaching effectiveness”, *Journal of Higher Education*, 66, 61-81.

Krautmann, A. e Sander, W. (1999). “Grades and Student Evaluations of Teachers.” *Economics of Education Review* 18 (1999): 59-63.

Marsh, H. (1991) “Multidimensional students' evaluations of teaching effectiveness: A test of alternative higher-order structures” *Journal of Educational Psychology*, 83 (2), 285-296.

Marsh, H. (1987). “Students' evaluations of university teaching: Research findings, methodological issues, and directions for future research”, *International Journal of Educational Research,* 11, #3, 253-388.

Marsh, H. e Roche, L. (1997) “Making students’ evaluations of teaching effectiveness effective” *American Psycologist* 52: 1187-97.

Neath, I. (1996) “How to Improve Your Teaching Evaluations Without Improving Your Teaching”, *Psychological Reports,* 78, 1363-1372.

O'Connell, D. Q. & Dickinson, D. J., (1993). “Student ratings of instruction as a function of testing conditions and perceptions of amount learned”, *Journal of Research and Development in Education* 27, (1), 18-23.

Olshavsky, R. W. & Spreng, R.A. (1995). “Consumer satisfaction and students: Some pitfalls of being customer driven” *Journal of Consumer Satisfaction Dissatisfaction and Complaining Behaviors*, 8, 69-77.

Rice, Lee. "Student Evaluation of Teaching: Problems and Prospects," *Teaching Philosophy* 11 (1988): 329-44.

Rodin, M. & Rodin, B., (1972). “Student evaluations of teachers”, *Science*. 177, 1164-1166.

Ryan, J., Anderson,J. e Birchler, A. (1980) "Student Evaluations: The Faculty Responds," *Research in Higher Education* 12: 317-33.

Sacks, P. (1986). *Generation X Goes to College,* Chicago IL

Sinclair, L. e Kunda, Z. (2000) “Motivated Stereotyping of Women: She's Fine if She Praised Me but Incompetent if She Criticized Me”, *Personality and Social Psychology Bulletin*, 25(11) pp. 1329-1342.

Stapleton, R. e Murkison, G. (2001) “Optimizing the Fairness of Student Evaluations: A Study of the Correlations Between Instructor Excellence, Study Production, Learning Production, and Expected Grades.” *Journal of ManagementEducation* 25, no. 3, 269-291.

Strom, B., Hocevar, D. & Zimmer, J., (1990). “Satisfaction and achievement: Antagonists in ATI research on student-oriented instruction”. *Educational Research Quarterly* 14, (4), 15-21.

Whitworth, J., Price, B. e Randall, C. (2002) “Factors That Affect College of Business Student Opinion of Teaching and Learning.” *Journal of Education for Business,* 282-289.

Williams, W.M. & Ceci, S.J. (1997). "How'm I doing?" Problems with student ratings of instructors and courses. *Change,* 29, 13-23.

Wright, R.E. (2000). “Student evaluations and consumer orientation of universities. Journal of Nonprofit and Public Sector Marketing” Vol 8, 33-40.

Wright, R. (2006) “Student evaluations of faculty: concerns raised in the literature and possible solutions”, *College Student Journal*

Yunker, P. & Yunker, J. (2003). “Are student evaluations of teaching valid? Evidence from an analytical business core course”, *Journal of Education for Business*. Vol. 78, 313-317.

Yurkiewics, Hoefer e Byrne (2006) “Implications based upon preliminary analyses of a student evaluation of teaching database”, w.p. Lubin School of Business, Pace University.