

Performance ed esami integrati Un'esperienza nell'ambito delle Università di Verona e Padova

Performance and integrated exams An experience in the Universities of Verona and Padova

STEFANO NOVENTA • PIERMATTEO ARDOLINO • GIUSEPPE FAVRETTO

Nel presente lavoro vengono riassunti alcuni risultati di una ricerca relativa alle performance di uno stesso esame in differenti corsi universitari. Vengono investigati fattori quali il genere, l'età, la frequenza alle lezioni, l'essere o meno studente lavoratore, la media dei voti, il diploma. In particolare, vengono discussi possibili effetti negativi sulle performance dovute all'uso degli esami integrati. Tali risultati forniscono interessanti spunti di riflessione riguardo ad alcune tendenze della didattica e della valutazione particolarmente diffuse nell'Università Italiana. Le analisi sono state effettuate utilizzando i modelli misti lineari, noti anche come multilevel o modelli lineari gerarchici.

In the present work, some of the findings of a research relative to the performances of the same exam in different academic curricula are outlined. Several factors like gender, age, class attendance, working students, grade average and school-leaving certificate were investigated. A particular focus is given to the possible negative impact on performances of an integrated assessment structure. Results offer some cause for reflection about some tendencies in teaching and assessment in Italian University. Analysis have been carried by using linear mixed-models, also known as multilevel or linear hierarchical models.

Parole chiave: docimologia, misurazione, valutazione, esami integrati, test a scelta multipla, modelli a effetti misti

Key words: assessment, measurement, evaluation, integrated exams, multiple-choice test, mixed-effects models

Performance ed esami integrati Un'esperienza nell'ambito delle Università di Verona e Padova

Introduzione

L'analisi delle problematiche del sistema universitario europeo ha evidenziato come esso sia rimasto a lungo isolato, tradizionalista e di conseguenza slegato dalla società e da una vera e propria prospettiva internazionale (Commissione delle comunità europee, 2003). Tale critica, riportata al contesto nazionale, ha comportato nell'ultimo decennio una serie di mutamenti che hanno interessato la struttura e l'assetto dell'Università Italiana, configurandosi in un vero e proprio cambio di paradigma, definito come un passaggio *dall'Università delle Conoscenze all'Università delle Competenze* che ha coinvolto il sistema universitario a numerosi livelli, dalla qualità dell'offerta formativa e della didattica, fino alla preparazione dei futuri ricercatori (si veda per esempio, Zaggia, 2008; Galliani, 2011).

Tra i punti sollevati, è stata sottolineata l'importanza dell'interdisciplinarietà in termini di formazione, didattica e ricerca dato che "le attività universitarie, in particolare in materia di insegnamento, tendono a restare organizzate, e spesso compartimentate, in funzione del quadro disciplinare tradizionale" (Commissione delle comunità europee, 2003).

Negli ultimi anni, infatti, l'approccio interdisciplinare ha ricevuto un notevole riscontro nella letteratura internazionale (si vedano per esempio Shoemaker, 1989; Walker, 1995; Fink, 2003) configurandosi come uno dei capisaldi per la formazione e lo sviluppo di nuovi ruoli e competenze la cui trasversalità e il cui ampio spettro appaiono più idonei alla risoluzione dei problemi della società moderna. In particolare, è opinione diffusa che l'aggregazione di corsi e la formazione di curricula e indirizzi didattici basati su tematiche complementari, interrelate e affini favorisca la comprensione delle materie potenziando e rafforzando la rete di connessioni e conoscenze di studenti e futuri ricercatori.

Tale idea è stata quindi un'importante linea guida da quando, con l'introduzione dei CFU, è emersa la necessità di ridistribuire e ridefinire il contenuto dei corsi e degli indirizzi di laurea in termini di carico lavorativo richiesto agli studenti. In numerose ex facoltà prima, e dipartimenti poi, è emersa pertanto nell'ultimo decennio la tendenza ad accorpare corsi che precedentemente erano indipendenti con l'introduzione di una particolare modalità di esame definita "integrato", nella quale in sostanza gli esami diventano moduli e i voti finali ottenuti in due o più corsi, in genere tenuti da differenti docenti, vengono accorpati in un unico punteggio finale (una media semplice o ponderata sui crediti formativi a seconda dei criteri adottati).

Nel caso specifico della presente ricerca, sono stati confrontati i risultati conseguiti nello stesso esame di Organizzazione Aziendale sostenuto dagli studenti di sei diverse coorti provenienti da differenti lauree magistrali e triennali afferenti a due diversi Atenei per verificare la presenza di effetti dovuti a variabili di carattere generale o anagrafico come il genere, l'età, la frequenza alle lezioni, il voto medio, l'essere o meno studenti lavoratori e il diploma posseduto al momento dell'iscrizione, ed eventuali effetti di trattamento dovuti alla modalità di esame (con o senza orale) e alla presenza di un esame integrato o meno.

Anche se i risultati ottenuti nell'analisi delle performance in un singolo esame, per quanto su un discreto spettro di studenti e corsi, non possono certo essere generalizzati all'intero quadro dei corsi integrati italiani in tutti gli indirizzi e le aree, è nostra opinione che essi forniscano interessanti spunti di riflessione critica nei confronti dell'impostazione della didattica e delle prove di misurazione sulla base delle quali vengono oggi costruite le valutazioni di merito. *Assessment drives learning* è un noto adagio, ma è importante anche ricordare come la valutazione sia un "momento di bilancio riguardo la validità del lavoro svolto e della direzione di quello da svolgere" (Notti, 2002).

1. Dati e metodo

Lo studio osservazionale è stato iniziato nel 2010, contestualmente alla somministrazione di un questionario contenente alcuni item relativi al genere, all'età, alla frequenza alle lezioni, all'essere o meno studente lavoratore, alla media dei voti e al diploma. Grazie alla continuità didattica è stato inoltre possibile recuperare i voti degli studenti degli anni precedenti fino al 2006. Nel presente lavoro sono stati quindi presi in considerazione e incrociati due differenti database. Il primo (1683 studenti, M=603, F=1080), contenente i risultati della prova scritta degli studenti dell'esame di Organizzazione Aziendale nei corsi di Scienze della Formazione, Ingegneria Informatica, Scienze dei Servizi Sociali ed Economia, dell'Università degli studi di Verona e Consulenti del lavoro dell'Università degli studi di Padova. I dati, come si può vedere in Tabella 1, riguardano nel complesso il periodo 2006-2012, ma variano a seconda dell'attivazione dei corsi nei vari dipartimenti. La tabella riporta inoltre le numerosità, la modalità di esame e la presenza o meno di un esame integrato. I corsi elencati, con l'eccezione di Consulenti del Lavoro sono stati al momento tutti soppressi o cambiati nel loro assetto didattico.

Coorte	Integrato	Modalità d'esame	Numerosità	Periodo
<i>Scienze della Formazione 1</i>	No	Scritto più orale	310	2006-2009
<i>Scienze della Formazione 2</i>	Sì	Scritto	64	2010-2012
<i>Ingegneria Informatica</i>	No	Scritto	396	2007-2012
<i>Scienze dei Servizi Sociali</i>	No	Scritto	78	2008-2011
<i>Consulenti del lavoro</i>	Sì	Scritto più orale	762	2006-2012
<i>Economia</i>	Sì	Scritto	73	2011-2012

Tab. 1: Coorti e modalità di esame, integrazione, numerosità e periodo

Nel secondo database (416 studenti, età media 22.0 ± 0.1 , M=146, F=270) sono invece contenute le anagrafiche e le risposte al breve questionario fornite dagli studenti iscritti ai suddetti corsi nel biennio 2010-2012. Si noti che il curriculum di Scienze della Formazione è spezzato in due coorti a causa di un cambiamento *in itinere* sia nella modalità di esame, passata da una modalità che contemplava scritto più orale a una modalità solo scritta, che nella presenza di un corso integrato al quale è stato successivamente accorpato.

Nei periodi elencati gli studenti hanno seguito lo stesso esame (Organizzazione aziendale) tenuto dallo stesso docente con lo stesso libro di testo e le stesse slide, ma con principale differenza il fatto che tali corsi fossero o meno integrati da un altro corso o avessero una

modalità d'esame che poteva prevedere o meno una successiva prova orale. La prova d'esame scritta consisteva, per tutti gli studenti indipendentemente dalla presenza o meno di un'integrazione o di un successivo orale, in un test a scelta multipla formato da 30 item estratti a random da una item bank costruita sul testo principale del corso. Gli item sono stati scritti nel formato convenzionale dei multiple-choice (Burton et al., 1991; Haladyna et al., 2002) e il test è stato costruito per valutare il programma coperto in circa 4 CFU.

L'analisi complessiva delle performance e della loro dipendenza da eventuali fattori quali integrazione, modalità di esame, genere, sessione e appello, area di appartenenza è stata quindi effettuata sui voti dell'ultima prova scritta nota di ogni studente, contenuti nel primo database, e i risultati di tale analisi sono stati incrociati con quelli del secondo database contenente invece le anagrafiche di 416 studenti raccolte nel periodo 2010-2012 e riguardanti età, voto medio, diploma posseduto al momento dell'iscrizione, se frequentante e se lavoratore. Le analisi sono state effettuate tramite linear mixed-models (si veda per esempio, Pinheiro & Bates, 2000) con il software statistico R (R Development Core Team, 2013) e i pacchetti *lme4* (Bates & Maechler, 2009) and *languageR* (Baayen, 2008).

Tale tipologia di modelli statistici, noti anche come multilevel o modelli gerarchici, è molto efficace nel trattare dati altamente sbilanciati e descrivere la presenza di effetti fissi come il genere o il trattamento controllando al contempo gli effetti sulla variabile dipendente dovuti a fattori random o annidati quali l'appartenenza degli studenti a diverse prove d'esame in diverse aree. Come si evince dalla Tabella 1, infatti, i dati sono altamente sbilanciati, con differenti numerosità nelle diverse coorti. In questi frangenti i test delle ipotesi effettuati con la distribuzione F, come tutti i metodi lineari che si basano sui minimi quadrati ordinari, sono più sensibili alle violazioni delle assunzioni di normalità e di omoschedasticità; limitazione che invece i modelli misti possono gestire.

2. Risultati e discussione

In Tabella 2 sono riportati i principali risultati ottenuti mediante i modelli misti per entrambi i database. Si noti che il secondo dataset è costituito da studenti che hanno svolto l'esame nel periodo 2010-2012 mentre il primo contiene anche gli studenti nel periodo 2006-2009 per il quale non erano disponibili anagrafiche e risposte al questionario. Le distribuzioni dei voti nei due periodi risultano diverse ($p < .01$ con il test di permutazione), infatti, come si nota in Tabella 1, nel periodo 2010-2012 l'esame non integrato di Formazione 1 non è presente; inoltre, gli esami di Ingegneria Informatica e di Scienze dei Servizi Sociali sono in fase di cessazione, per cui risentono dell'accumularsi degli studenti meno abili o motivati. Ne consegue che i risultati del secondo database si riferiscono a un sottocampione costituito per il 76% da studenti che hanno seguito un esame integrato e per il 24% dalle code statistiche degli esami non integrati. La generalizzazione dei risultati del secondo database va effettuata quindi alla luce di questa considerazione.

È inoltre importante osservare come fattori quali l'area e il diploma non siano stati riportati in Tabella 2 perché la loro varianza stimata risultava nulla o trascurabile e la loro non significatività verificata tramite test di *Likelihood-ratio*. Se trattati come fattori random, così che i loro livelli possano essere considerati un campionamento da una popolazione più vasta, una stima di varianza nulla suggerisce come non vi sia una distribuzione alla base. Per quanto tale assenza di variabilità potrebbe essere dovuta alla difficoltà di ripartire in troppi livelli la varianza osservata, soprattutto in disegni sbilanciati, è interessante notare che, anche se trattati

come fattori fissi e considerando quindi un'analisi focalizzata nel cercare differenze rispetto a quegli specifici livelli (come in un'analisi della varianza) essi risultano non significativi. Questo suggerisce che la variabilità dovuta all'appartenenza degli studenti a diverse aree o alla provenienza da differenti istituti sia effettivamente non incisiva nel presente campione nello spiegare eventuali differenze di voto.

Come si evince dalla Tabella 2, non emergono inoltre effetti significativi in vari fattori quali l'età degli studenti, la modalità di esame oppure la frequenza alle lezioni, e le maggiori fonti di variabilità in entrambi i database sono intrinseche alle sessioni d'esame e agli appelli, mentre il punteggio finale è ben predetto dalla media dei voti, dalla presenza o meno di un esame integrato (livello di riferimento: assenza di integrazione) e dall'essere o meno uno studente lavoratore (livello di riferimento: studente lavoratore). Vi è inoltre una tendenza del genere con le femmine che otterrebbero circa mezzo punto in più dei maschi (livello di riferimento: femmine).

Effetti	Database 1 (N=1683)		Database 2 (N=416)	
	Varianza	Dev.St.	Varianza	Dev.St.
Esame	2.25	1.50	1.73	1.31
Sessione	0.86	0.93	5.11	2.26
Residui	18.69	4.32	16.36	4.04
Fissi	Stima	Err.St.	Stima	Err.St.
Età	/	/	0.04	0.10
Media Voto	/	/	0.59 (***)	0.11
Integrazione	-2.38 (***)	0.50	-0.09	0.61
Mod. esame	-0.65	0.51	0.18	0.53
Genere	-0.47 (.)	0.25	0.19	0.45
Lavoratore	/	/	1.12 (*)	0.48
Frequentante	/	/	-0.07	0.50

Tab. 2: Principali risultati dei mixed linear models nei due database
Indicatori di significatività: (.) al 10%, (*) al 5%, (***) allo 0.1%

Il risultato più interessante, da un punto di vista formativo e didattico è probabilmente il fatto che la presenza di un'integrazione con un altro esame venga stimata ridurre di oltre due punti il punteggio degli studenti. Si noti che l'integrazione è significativa nel primo database, ma non nel secondo. Tale risultato è coerente con il fatto che, come osservato in precedenza, il secondo database è costituito in prevalenza da studenti che hanno effettuato esami integrati e dalle code statistiche degli esami non integrati perciò l'effetto dell'integrazione scompare e viene stimato come non significativo.

Osservando inoltre il secondo database si nota come la media dei voti degli studenti si confermi un buon predittore della loro performance all'esame, mentre lo stesso non si verifica per l'età. Allo stesso modo, mentre la frequenza alle lezioni non sembrerebbe avere un particolare effetto sul voto finale, l'essere uno studente lavoratore comporterebbe una diminuzione del voto finale stimata in circa un punto. Per quanto tali risultati siano ristretti al

campione del secondo database, non è implausibile una loro generalizzazione. Da un lato, infatti, la media dei voti è un predittore della performance già noto in letteratura e il coefficiente di regressione ($\beta=0.59$) indicato in Tabella 2 è stato stimato aggregando gli studenti indipendentemente dalla presenza di integrazione e dalla stratificazione per area che è risultata non significativa. Dall'altro lato, l'essere studente lavoratore si configura nel secondo database come un aggravio, ed è ragionevole supporre che tale risultato sia generalizzabile: la percentuale maggiore di lavoratori ($\chi^2_{(1)} = 24.24, p < .01$) si ha infatti tra gli studenti che non integrano. Su 100 studenti non sottoposti a integrazione circa il 50% è lavoratore, mentre nei 316 che integrano lo è il 27%. Non è quindi irragionevole presumere che, se l'essere studente lavoratore costituisce un aggravio per quelli che non integrano, a maggior ragione può costituire un ulteriore aggravio per coloro che devono integrare. Pur nei limiti della presente generalizzazione è un risultato che merita attenzione.

Discorso a parte riguarda invece la tendenza del fattore genere, verificabile nel primo database ma non nel secondo. L'esistenza di un *gender bias* è stata infatti discussa a lungo e investigata in letteratura, soprattutto economica, e un'ampia porzione di ricerche indicherebbe come le donne ottengano voti inferiori nei test a scelta multipla o più in generale nei test standardizzati (si veda per esempio Siegfried, 1979). A giustificazione di ciò è stato, per esempio, ipotizzato che gli uomini siano abili nell'utilizzare le conoscenze per risolvere nuovi problemi mentre le donne avrebbero migliori doti verbali, abilità di scrittura e memorizzazione (Lumsden & Scott, 1987). Tuttavia, ricerche più recenti hanno osservato che classi con alte percentuali di donne progrediscono più rapidamente (Beekhoven et al., 2003) e che i punteggi delle donne sono in genere migliori durante il percorso scolastico al punto che il *gender gap* sarebbe ormai stato colmato e rovesciato negli ultimi anni (Goldin et al., 2006). Nel caso in cui la tendenza indicata in Tabella 2 fosse indice di un effetto reale, secondo le analisi tramite modelli misti sarebbero infatti le donne ad avere un voto superiore di circa mezzo punto.

Un'altra interessante osservazione riguarda gli effetti random. La variabilità dell'esame non deve essere confusa infatti con la variabilità nella difficoltà dei singoli test. In primo luogo, essa sintetizza l'annidamento di diversi test in diversi appelli perciò non tiene conto del loro andamento temporale, i punteggi ottenuti infatti dagli studenti nel primo appello sono significativamente superiori al secondo appello ($t_{(1237.7)} = 7.9, p < .001$). Un effetto simile si trova a livello delle sessioni: nella sessione estiva i punteggi risultano migliori che nelle altre due sessioni ($p < .001$ con correzione di Bonferroni nei contrasti multipli). In secondo luogo, il punteggio medio del singolo esame non corrisponde alla reale media di ciascun test somministrato in quanto sono stati considerati solo gli ultimi voti degli studenti e non i loro precedenti tentativi. A questo va inoltre aggiunto il fatto che punteggi più estremi e ai margini della distribuzione di cui viene stimata la deviazione standard, corrispondono a sessioni d'esame con numerosità particolarmente basse e aventi ampi errori standard. Nonostante ciò, la variabilità associata al singolo esame è quantificata all'incirca in 1.5 punti, sufficientemente stabile da considerare affidabile e valida la misurazione dei test.

Una nota finale, modellare l'eteroschedasticità raffina i modelli ma non cambia in modo sostanziale le stime e le significatività. Inoltre, le assunzioni fondamentali per applicare i modelli misti sono rispettate, i residui sono leggermente platicurtici ma distribuiti in modo simmetrico attorno allo zero. Gli effetti random sono distribuiti normalmente. Infine, le correlazioni tra gli effetti fissi suggeriscono non vi siano problemi di multicollinearità.

Conclusioni

Nel presente lavoro sono stati incrociati due database. Il primo, contenente i voti ottenuti nel periodo 2006–2012 da 1683 studenti afferenti a diverse aree didattiche e sottoposti alla prova scritta di Organizzazione Aziendale; il secondo, contenente le anagrafiche associate ai 416 studenti che hanno sostenuto l'esame nel periodo 2010–2012 e ai quali è stato inoltre somministrato un breve questionario riguardante fattori quali il genere, l'età, la frequenza alle lezioni, l'essere o meno studente lavoratore, la media dei voti, il diploma.

Poiché in linea generale ci si può aspettare una variabilità nelle performance associate a differenti aree e corsi (Beekhoven et al., 2003), sono stati applicati i modelli lineari misti per cercare di contenere la variabilità legata a fattori quali l'area, l'esame e la sessione in presenza di dati altamente sbilanciati. L'uso di fattori fissi e random permette infatti di pulire i contributi di fattori anagrafici o di trattamento, quali il genere, la presenza di un'integrazione e la modalità d'esame, la frequenza alle lezioni e così via, tenendo conto al contempo del contributo di variabili annidate che descrivono le fluttuazioni delle performance a livello individuale, del singolo esame, dell'area di indirizzo o di eventuali andamenti temporali.

L'aspetto più saliente emerso dall'analisi di entrambi i database riguarda la presenza di pochi fattori realmente predittivi della performance degli studenti: diploma, frequenza alle lezioni, genere, età, modalità di esame, e la stessa area di corso non sembrano avere effetti significativi sulle performance o quanto meno tali effetti sono oscurati da fattori più incisivi. Le differenze più notevoli emergerebbero infatti in relazione all'essere o meno studenti lavoratori, con gli studenti lavoratori che prenderebbero circa un punto in meno di quelli non lavoratori, alla media dei voti dello studente che si conferma come un buon predittore delle performance, agli andamenti temporali associati alle sessioni e agli appelli, con voti superiori nel primo appello e nella sessione estiva.

L'effetto più interessante e che emerge invece nel primo database riguarda la presenza di un'integrazione che viene stimata in un peggioramento della performance di oltre due punti su trenta. Quali siano le cause all'origine di tale differenza, esse non possono essere desunte dalla natura dei dati a disposizione. Le strategie di apprendimento sono infatti suscettibili a numerosi fattori di natura sociale, psicologica e pedagogica, non controllabili con dati anagrafici e di performance. Tuttavia alcune osservazioni potrebbero rivelarsi utili spunti di riflessione. In termini generali si potrebbe infatti parlare di distorsione della performance e quindi di *washback effect*, nella sua accezione negativa (per delle review recenti si veda, Pan, 2009; Ahmad & Rao, 2012), nella fattispecie quindi un effetto di modulazione dell'apprendimento dovuto alla modalità finale di esame. Più difficile è però delinearne le cause effettive. Una differenza così pronunciata potrebbe essere indice di un approccio più superficiale nelle strategie di apprendimento dovuto a un aumento del carico di lavoro, inteso come numero di ore di lavoro e CFU complessivi, oppure dovuto alla percezione dell'aumento del carico di lavoro (si veda per esempio, Chambers, 1992; Kember, 2004). Non necessariamente in antitesi alle ipotesi precedenti, la differenza di metodo di apprendimento potrebbe dipendere da un aumento del carico cognitivo, e quindi avere una spiegazione in termini di processi non necessari, di memoria di lavoro, e di risorse cognitive, attenzione divisa e ridondanza (si veda per esempio van Merriënbor & Sweller, 2005). Al contempo, una spiegazione plausibile potrebbe essere individuata in termini di *work avoidance*, intesa come 'la preoccupazione di concludere un lavoro con il minimo sforzo possibile' (Meece et al., 1988). Gli studenti potrebbero essere indotti ad accettare un voto inferiore pur di concludere un carico lavorativo che ritengono troppo oneroso per concretizzarsi in un solo voto finale. Considerando

infatti lo storico di tutti i voti presi degli studenti all'esame scritto, risulta in effetti che una percentuale statisticamente significativa di studenti in più ripete l'esame nei corsi nei quali non c'è integrazione. L'aumentato carico cognitivo potrebbe quindi disincentivare gli studenti meno abili.

In conclusione, per quanto i risultati del presente studio osservazionale non siano in alcun modo generalizzabili all'intero contesto nazionale, in quanto limitati a sei differenti corsi in due sole università, essi suggeriscono l'importanza di tutelare categorie di studenti quali gli studenti lavoratori, e di approfondire l'impatto che i corsi integrati hanno sul rendimento e l'apprendimento degli studenti negli atenei italiani. Il problema di strutturare efficacemente corsi interdisciplinari e di creare delle valide forme di misurazione e quindi di valutazione di tali esami va probabilmente ben oltre la natura segmentata e compartimentata dei metodi di insegnamento e delle procedure di valutazione spesso utilizzate nei nostri Atenei. La complessità di un approccio integrativo e interdisciplinare è infatti vasta, non soltanto in termini valutativi ma anche educativi e pedagogici. La multidisciplinarietà, intesa come esposizione e rappresentazione parallela di competenze, compartimentate e indipendenti da parte di diversi docenti, non implica infatti la capacità dello studente di integrare quanto appreso (Klein, 2005). Sono richiesti approcci che, mescolando aspetti di apprendimento collaborativo, lavoro di gruppo, comunitario e multiculturale, a seminari ed esperienze tematiche, ridefiniscano i concetti stessi di studio, apprendimento e valutazione (si veda per esempio Newell, 2001), trasformando lo studente in un interlocutore attivo e ridefinendo le abilità e le competenze degli insegnanti per renderli in grado di mantenere i fili di un approccio integrato e di rovesciare da negativo a positivo ogni effetto di *washback*. Un aspetto quest'ultimo che non dovrebbe essere trascurato all'interno di una corretta pianificazione che traduca "le finalità educative in possibilità operative controllabili, misurabili e gestibili didatticamente" (Marzano, 2002).

Riferimenti Bibliografici

- Ahmad, S., & Rao, C. (2012). A review of the pedagogical implications of Examination Washback. *Research on Humanities and Social Sciences*, 7(2), 11-20.
- Baayen, R. H. (2008). *Analyzing Linguistic Data: A Practical Introduction to Statistics Using R*. New York: Cambridge University Press.
- Bates, D., & Maechler, M. (2009). lme4: Linear mixed-effects models using Eigen and Eigenfaces [Computer software manual]. Estratto da <http://CRAN.R-project.org/package=lme4>.
- Beekhoven, S., De Jong, U., & Van Hout, H. (2003). Different courses, different students, same results? An examination of differences in study progress of students in different courses. *Higher Education*, 46, 37-59.
- Burton, S.J., Sudweeks, R.R., Merrill, P.F., & Wood, B. (1991). *How to Prepare Better Multiple-Choice Test Items: Guidelines for University Faculty*. Brigham Young University Testing Services and The Department of Instructional Science.
- Chambers, E. (1992). Work-load and the quality of student learning. *Studies in Higher Education*, 17(2), 141-153.
- Commissione delle comunità europee (2003). Il ruolo delle università nell'Europa della conoscenza. *Comunicazione della commissione*, 58. Bruxelles.
- Fink, L. (2003). *Creating significant learning experiences: An integrated approach*. San Francisco: Jossey-Bass.
- Galliani, L. (2011). La riforma universitaria e le responsabilità dei docenti. In L. Galliani (ed.), *Il Docente Universitario. Una professione tra ricerca, didattica e governance degli atenei* (pp. 3-9). Atti della VIII

- Biennale Internazionale sulla Didattica Universitaria, Padova, 2 e 3 dicembre 2010. Lecce: Pensa MultiMedia.
- Goldin, C., Katz, L. F., & Kuziemko, I. (2006). The Homecoming of American College Women: The Reversal of the College Gender Gap. *Journal of Economic Perspectives*, 4(3), 133-156.
- Haladyna, T.M., Downing, S.M., & Rodriguez, M.C. (2002). A Review of Multiple-Choice Item-Writing Guidelines for Classroom Assessment. *Applied Measurement in Education*, 15(3), 309-334.
- Kember, D. (2004). Interpreting student workload and the factors which shape students' perceptions of their workload. *Studies in Higher Education*, 29(2), 165-184.
- Klein, J.T. (2005). Integrative learning and interdisciplinary Studies. *AAC&U peer Review*, 7(2), 8-10.
- Lumsden, K. G., & Scott, A. (1987). The Economics Student Reexamined: Male-Female Differences in Comprehension. *The Journal of Economic Education*, 18(4), 365-375.
- Marzano, A. (2002). La valutazione degli obiettivi. In C. Coggi, M. Notti, *Docimologia* (pp. 45-86) Lecce: Pensa MultiMedia.
- Meece, J. L., Blumenfeld, P. C., & Hoyle, R. (1988). Students' goal orientations and cognitive engagement in classroom activities. *Journal of Educational Psychology*, 80, 514-523.
- Newell, W.H. (2001). Powerful pedagogies. In reinventing ourselves: Interdisciplinary education, collaborative learning, and experimentation in higher education (B.L. Smith & McCann, J., eds). McCann, 196-211. Bolton, MA: Anker Press.
- Notti, A.M. (2002). Origine e sviluppo della docimologia. In C. Coggi, M. Notti, (2000), *Docimologia* (pp. 11-44) Lecce: Pensa MultiMedia.
- Pan, Y.C. (2009). A review on washback and its pedagogical implications. *VNU Journal of Science, Foreign Languages*, 25, 257-263.
- Pinheiro, J.C., & Bates, D.M. (2000). *Mixed effects models in S and S-plus*. New York: Springer Verlag.
- R Development Core Team (2013). *R: A language and environment for statistical computing*. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, URL <http://www.R-project.org>.
- Shoemaker, B. J. E. (1989). Integrative Education. A Curriculum for the Twenty-First Century. *OSSC Bulletin Series*. Eugene, Oregon: Oregon School Study Council.
- Siegfried, J. J. (1979). Male-female differences in economic education: A survey. *Journal of Economic Education*, 10(1), 1-11.
- van Merriënbor, J. J. G. & Sweller, J. (2005). Cognitive Load Theory and Complex Learning: Recent Developments and Future Directions. *Educational Psychology Review*, 17(2), 147-177.
- Walker, D. (1995). Integrative education. *Research roundup*, 12(1), 1-5.
- Zaggia, C. (2008). *L'Università delle competenze. Progettazione e valutazione dei corsi di laurea nel processo di Bologna*. Milano: FrancoAngeli.