

Assessment approaches and practices of university lecturers: a nationwide empirical research

Pratiche valutative nelle università italiane: una ricerca esplorativa a livello nazionale

Beatrice Doria

University of Padova, Dept. of FISPPA, Padova (Italy)

Valentina Grion

University of Padova, Dept. of FISPPA, Padova (Italy)

Omar Paccagnella

University of Padova, Dept. of Statistical Sciences, Padova (Italy)

OPEN ACCESS

Double blind peer review

Citation: Doria, B., et al. (2023). Assessment approaches and practices of university lecturers: a nationwide empirical research. *Italian Journal of Educational Research*, 30, 129-143 <https://doi.org/10.7346/sird-012023-p129>

Corresponding Author: Beatrice Doria
Email: beatrice.doria@phd.unipd.it

Copyright: © 2023 Author(s). This is an open access, peer-reviewed article published by Pensa Multimedia and distributed under the terms of the Creative Commons Attribution 4.0 International, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited. IJEDuR is the official journal of Italian Society of Educational Research (www.sird.it).

Received: February 28, 2023

Accepted: May 25, 2023

Published: June 30, 2023

Pensa MultiMedia / ISSN 2038-9744
<https://doi.org/10.7346/sird-012023-p129>

Abstract

International educational research has for decades considered assessment a strategic process through which to foster more effective teaching/learning practices. Despite growing research evidence, assessment in the university context is still considered almost exclusively as the final moment of teaching activity, relegated to a control and certification function rather than a learning support function. In this context, it would be important to foster greater awareness among lecturers with regard to assessment. Training actions, however, must start from an analysis of the state of the art, which in Italy is only known in a fragmented and local way. In order to bridge this gap, the present research aimed to explore which assessment practices are most widely used at national level by university lecturers. A representative sample of Italian syllabi was analysed within the framework of a mixed sequential research design QUAL → Quant. The results make it possible to construct an overall picture that outlines the prevalence of traditional practices and confirms differences between subject areas. Concluding considerations concern the value of these results in relation to possible future training paths and further perspectives of the work.

Keywords: Syllabi; Assessment; Lecturers; University; Italy.

Riassunto

Da decenni, la ricerca educativa internazionale considera la valutazione un processo strategico attraverso il quale favorire pratiche d'insegnamento/apprendimento maggiormente efficaci. Nonostante le crescenti evidenze di ricerca, la valutazione in contesto universitario è ancora considerata quasi esclusivamente come momento finale dell'attività didattica, relegata ad una funzione di controllo e certificazione, piuttosto che di supporto all'apprendimento. In tale contesto sarebbe importante favorire maggiori consapevolezze dei docenti rispetto alla valutazione, attraverso azioni formative, le quali dovrebbero partire da un'analisi dello stato dell'arte, che in Italia finora non si conosce, se non in modo frammentato e locale. Per colmare tale gap, la presente ricerca ha inteso esplorare quali pratiche valutative vengono maggiormente utilizzate a livello nazionale dai docenti universitari. Nell'ambito di un disegno di ricerca misto sequenziale QUAL → Quant, è stato analizzato un campione rappresentativo di Syllabi italiani. I risultati permettono di costruire un quadro d'insieme che delinea la prevalenza di pratiche tradizionali e conferma differenze fra aree disciplinari. Le considerazioni conclusive riguardano il valore di questi risultati in relazione a possibili futuri percorsi formativi e alle prospettive ulteriori del lavoro.

Parole chiave: Syllabi; Valutazione; Docenti; Università; Italia.

Credit author statement: Il presente contributo rappresenta il prodotto di un lavoro condiviso. L'effettiva stesura dei paragrafi è avvenuta secondo le seguenti attribuzioni: Grion ha scritto il paragrafo "Introduzione" e il paragrafo 2; Doria ha scritto i paragrafi 3 e 4; Paccagnella e Doria hanno scritto il paragrafo 5; Grion, Doria e Paccagnella hanno scritto il paragrafo 6.

1. Introduzione

Già a metà del secolo scorso, Bloom (Bloom & Krathwohl, 1956), nella sua tassonomia delle abilità cognitive collocava la capacità valutativa al livello più elevato fra tali processi, decretandone di conseguenza, un ruolo importante nelle attività di insegnamento e apprendimento. Anderson & Krathwohl (2001), rivisitando la tassonomia, posero il processo valutativo ad un livello più basso rispetto all'apice, costituito ora dal processo creativo, chiarendo però come quest'ultimo implichi la capacità di valutare criticamente e scegliere risorse appropriate in base al contesto per creare nuova conoscenza.

In questa direzione, la ricerca educativa, negli ultimi decenni, ha mostrato un crescente interesse per la valutazione come momento centrale del processo d'insegnamento/apprendimento; così determinante delle esperienze formative degli studenti, da condurre Brown (2014) ad affermare che, per migliorare la qualità della formazione, sarebbe innanzitutto necessario rinnovare le attuali pratiche valutative. Sambell, Brown, e Race (2019) suggeriscono che la valutazione dovrebbe essere caratterizzata da un approccio *fit for purpose*, rilevando, con tale espressione, la necessità che i processi valutativi e di feedback debbano essere realizzati come dinamiche: a) genuinamente integrate nelle attività d'apprendimento, incoraggiando e supportando il successo formativo (funzione formativa della valutazione); b) che coinvolgono pienamente gli studenti; c) che permettono di perseguire gli standard di qualità e i risultati d'apprendimento attesi. Fra queste ultime caratteristiche, l'attenzione alla *funzione formativa* della valutazione e il *coinvolgimento degli studenti* nei processi valutativi a supporto dell'apprendimento rappresentano aspetti di rilevante interesse per numerosi gruppi di ricerca a livello internazionale, alcuni dei quali sono giunti a ridefinire il *Formative assessment* come *Assessment for Learning* (Sambell, MacDowell & Montgomery, 2013), proprio con l'intenzione di enfatizzare il necessario ruolo attivo degli studenti nei processi valutativi.

Il dibattito scientifico su questi temi, negli ultimi decenni ha incentivato la diffusione di pratiche valutative innovative, sia in ambito scolastico che universitario, a livello internazionale. Basti confrontare, a supporto di tale affermazione, la ricca letteratura riferita allo sviluppo delle pratiche di new assessment e di valutazione alternativa (Nicol, 2018, 2020, 2021; Lipnevich et al., 2021; Panadero & Lipnevich, 2022; Sambell, McDowell & Montgomery, 2017; Grion & Serbati, 2019; Foschi & Cecchinato, 2019), così come l'implementazione dell'*Assessment for Learning* (Carless, 2005) della valutazione come apprendimento (Sadler, 1989; Wiliam, 2011; Black & William, 2009; Taras, 2005) o valutazione formante (Trinchero, 2018), della valutazione sostenibile (Boud, 2000; Grion & Serbati, 2019).

D'altra parte, il processo valutativo nell'università deve essere considerato anche in relazione al processo di Bologna, il quale sottolinea l'importanza di sviluppare competenze spendibili durante tutto l'arco della vita. Proprio per questo motivo, le istituzioni universitarie non possono limitarsi a valutare gli esiti dell'apprendimento, ossia il prodotto finale delle attività di insegnamento escludendo il processo attraverso cui ogni studente vi è giunto, perché, così facendo, non sarebbero in grado di verificare le *competenze* apprese dagli studenti. Essendo la competenza un costrutto complesso, necessita di una valutazione che consideri non solo i prodotti dell'apprendimento, ma anche i processi e i progressi dell'apprendimento degli studenti (Marzano et al., 2014; Castoldi, 2016; Varisco, 2004).

Focalizzando l'attenzione sui contesti universitari, va rilevato che in Italia l'interesse verso gli sviluppi della ricerca valutativa risulta relativamente recente (Coggi, 2005; Grion, Serbati, Tino, Nicol, 2017). Soprattutto, ad oggi mancano dati che permettano di delineare un quadro complessivo attraverso cui conoscere quali siano le pratiche valutative messe in atto dai docenti universitari italiani. Le indagini finora condotte si riferiscono a situazioni locali e specifiche (cfr. Coggi, 2005, 2019a, 2019b, 2022; Grion & Tino, 2018; Grion, Serbati, Felisatti & Li, 2019; Pastore, 2012; Serbati, Grion & Fanti, 2019), ma non risultano esserci indagini che permettano di delineare un quadro d'insieme della realtà nazionale.

2. Pratiche valutative alternative

La valutazione assume un ruolo centrale all'interno del contesto universitario, poiché funge da strumento di *lifelong learning*: alcuni studiosi (Boud & Soler, 2016) sottolineano come, per diventare *lifelong learners* efficaci, gli studenti debbano anzitutto essere *lifelong assessors*, ossia essere preparati ad affrontare i compiti valutativi che affrontano attualmente e incontreranno nella loro vita futura. D'altra parte, essere in grado

di valutare in modo pertinente ed equilibrato non è una competenza che nasce spontaneamente, ma va intenzionalmente allenata, considerandola un obiettivo formativo indispensabile di ogni disciplina (Nicol, 2014). I laureati possono agire tale competenza valutativa con autonomia e responsabilità nel momento in cui viene offerta loro l'opportunità di esercitarsi e sperimentarsi in tal senso durante il percorso universitario.

Effettivamente, l'ampia letteratura di settore (Brown, 2014; Nicol, 2021; Grion & Serbati, 2019; Lipnevich et al., 2021; Coggi, 2022) riconosce nella valutazione un momento strategico dell'insegnamento/apprendimento, in grado di migliorare il processo apprenditivo degli studenti (Boud & Falchikov, 2007; Grion & Serbati, 2019; Grion et al., 2022). Tuttavia, la ricerca rileva che per poter mettere in atto una valutazione "efficace" in relazione ai contesti di formazione in cui viene attuata, è necessario utilizzare pratiche alternative (Dochy et al., 1999; Nicol, 2018; 2021; Panadero & Lipnevich, 2022; Grion & Serbati, 2019; Foschi & Cecchinato, 2019), orientate da approcci valutativi partecipativi, finalizzati innanzitutto al supporto dell'apprendimento e sostenibili (Sambell, McDowell & Montgomery, 2017; Carless, 2015; Zhao & Qi, 2022; Boud & Soler, 2016).

A tal proposito, Nicol (2018, 2020, 2021) riporta l'attenzione sul concetto di feedback valutativo come pratica in grado di coinvolgere attivamente gli studenti nell'acquisizione di conoscenze e competenze. Egli introduce in particolare il concetto di "*internal feedback*", ossia la teoria secondo cui il modo più efficace per migliorare l'apprendimento degli studenti, e le loro capacità autoregolative, è quello di sfruttare la loro capacità di generare continuamente feedback valutativi interni grazie al confronto con diverse tipologie di informazioni esterne (libri, produzioni, video, ecc.). In questa prospettiva, il docente ha il compito di agevolare e stimolare i processi valutativi degli studenti attraverso la messa a disposizione e il confronto con una vasta gamma di risorse didattiche.

Un'ulteriore pratica da incentivare è l'autovalutazione (Boud, 2000; Panadero & Alonso-Tapia, 2013). Il processo autovalutativo garantisce infatti non solo la riflessione dello studente circa il proprio compito d'apprendimento, ma soprattutto gli permette di comprendere quali elementi migliorare *durante* il processo formativo (Panadero & Alonso-Tapia, 2013; Boud & Soler 2016; Nicol, 2018; Doria & Grion, 2020).

Allo stesso modo, la valutazione fra pari risulta essere un'ulteriore strategia in grado di coinvolgere attivamente gli studenti nel processo d'apprendimento. La valutazione tra pari attiva l'apprendimento poiché pone gli studenti nella condizione di giocare un ruolo attivo sia nel ruolo di valutatori che di valutati (Li, Liu, & Steckelberg, 2010). Come valutatori, gli studenti esaminano il lavoro dei loro compagni e forniscono feedback; come valutati, ricevono il feedback e agiscono su di esso per migliorare i loro prodotti (Li, Liu & Zhou, 2012). La ricerca sui processi di fornitura e ricezione del feedback conferma che l'impegno attivo degli studenti in entrambi i ruoli ha un effetto diretto o indiretto sul loro apprendimento (Li et al., 2010, 2012; Li & Grion, 2019).

Nonostante le ampie evidenze sull'efficacia di tali pratiche valutative alcuni studi condotti in ambito nazionale ed internazionale (Panadero & Brown, 2017; Lipnevich et al., 2021) hanno evidenziato come i docenti universitari ne facciano un uso residuale, perdendo l'occasione di sfruttare le ampie potenzialità che i processi valutativi possono avere sul miglioramento del processo d'insegnamento/apprendimento (Brown, 2014).

3. Studi internazionali relativi alla valutazione nel contesto universitario

In ambito internazionale sono stati condotti alcuni studi volti ad indagare le pratiche valutative maggiormente utilizzate all'interno dell'istruzione accademica su piano nazionale. Fra gli studi che indagano le pratiche valutative impiegate dai docenti all'interno dell'università, alcuni di essi prendono in considerazione come oggetto di analisi il Syllabus.

Panadero et al. (2019) hanno svolto un'indagine con l'obiettivo di rilevare quali pratiche valutative utilizzano i docenti appartenenti alle università statali spagnole, attraverso l'analisi di 1693 Syllabi appartenenti a tutte le università pubbliche del paese (n=48). I ricercatori, esaminando i Syllabi, hanno riscontrato che i docenti spagnoli benché stiano implementando una varietà di pratiche valutative diverse rispetto agli ultimi due decenni, prediligono ancora approcci valutativi "tradizionali" (Panadero et al., 2019, p. 392), ossia basati su lezioni frontali ed esami finali svolti al termine del corso di studi, con quasi nessun coin-

volgimento degli studenti nel processo di valutazione (Panadero et al., 2019)¹. I ricercatori, inoltre, hanno rilevato che i docenti prediligono la prova scritta finale (70,8%) come strumento valutativo utilizzato per l'assegnazione del voto, piuttosto che pratiche e approcci alternativi come autovalutazione, la valutazione fra pari; questo soprattutto in riferimento ai docenti appartenenti a discipline tecnico-scientifiche, piuttosto che umanistiche e linguistiche. Inoltre, gli stessi autori evidenziano che la maggior parte delle differenze nell'utilizzo di pratiche valutative si riscontra tra facoltà e discipline accademiche, poiché le discipline socio-umanistiche fanno riferimento maggiormente a una valutazione sommativa, di prodotto, a differenza delle discipline sanitarie e scientifico-tecnologiche che fanno maggior riferimento ad approcci ed attività valutativi di tipo formativo e pratico (Panadero et al., 2019).

Panadero et al. (2019) hanno infine sottolineato l'importanza di investire sul miglioramento delle pratiche di valutazione, soprattutto alla luce alle differenze riscontrate con le sue ricerche. Tali differenze potrebbero essere interpretate come “*the strength of certain assessment traditions within areas of knowledge, i.e. faculties*” (p. 392).

Allo stesso modo, Lipnevich e colleghi (2020) hanno esaminato gli approcci e le pratiche di valutazione utilizzate nel contesto universitario statunitense attraverso l'analisi dei Syllabi appartenenti a cinque differenti discipline, quali: inglese, matematica, scienze, psicologia ed educazione. I ricercatori, per analizzare i Syllabi, hanno fatto riferimento al framework teorico proposto da Guskey (2011, 2019) il quale individua tre criteri per classificare gli approcci a cui i docenti universitari fanno riferimento assegnando i voti e mettendo in atto le pratiche valutative, ossia:

- *product criteria*; misurano cosa gli studenti sanno fare in quel preciso momento, prediligendo una valutazione sommativa volta a certificare le conoscenze finali al termine del corso di studi (Brookhart, 2009; O'Connor, 2009). I docenti che favoriscono questa tipologia di valutazione basano i voti sui punteggi ottenuti dagli studenti mediante la somministrazione di test finali o sui prodotti finali;
- *process criteria*; misurano non solo i risultati d'apprendimento degli studenti, ma anche il processo attraverso cui sono arrivati all'acquisizione di competenze e conoscenze. Pertanto, i docenti che considerano non solamente il prodotto finale, ma anche il processo d'apprendimento, utilizzano come elementi valutativi quiz e prove di varia natura in itinere, strumenti di verifica della partecipazione, valutazione formativa;
- *progress criteria*; misurano l'apprendimento in termini di “progresso” da uno stato iniziale ad uno finale, ossia i vantaggi derivanti dalle esperienze di apprendimento produttive del passaggio da una pre-condizione ad una condizione finale. I criteri di progresso fanno quindi riferimento al “guadagno d'apprendimento” e alla crescita educativa, misurando concretamente l'effettivo indice di miglioramento dello studente.

Dall'analisi effettuata mediante questi tre criteri valutativi, i ricercatori hanno rilevato che vi è differenza tra le diverse discipline nell'utilizzo di pratiche valutative incentrate sul prodotto. I risultati dimostrano infatti che le discipline riferite all'ambito educativo e linguistico prediligono una valutazione incentrata sul processo d'apprendimento degli studenti, dando minore peso agli esami finali, piuttosto che discipline psicologiche, matematiche e tecnico-scientifiche, le quali prediligono pratiche valutative incentrate sul prodotto finale.

Infine, un recente studio condotto da Lipnevich e colleghi (2021) ha esaminato il differente utilizzo dei criteri di valutazione proposti da Guskey (2011, 2019) per l'assegnazione del voto, in due contesti universitari: Spagna e USA. Il campione di indagine statunitense includeva 250 Syllabi, redatti da docenti appartenenti a cinque diverse discipline accademiche (educazione, matematica, scienze, psicologia e inglese). Il campione di indagine spagnolo comprendeva 175 Syllabi, selezionati dal database nazionale e facenti riferimento alle medesime discipline accademiche. I risultati ottenuti hanno evidenziato che i docenti universitari statunitensi utilizzano in egual misura criteri di valutazione orientati al prodotto e al processo d'apprendimento (50,94% processo e 49,06% prodotto), a differenza di quelli spagnoli che utilizzano

1 Una descrizione puntale per cosa si intenda con approccio *tradizionale* alla valutazione e/o prova valutativa *tradizionale*, si veda Montgomery (2002) e Nasab (2015).

maggiormente i criteri orientati al prodotto finale: il 70,02% dei docenti dichiara di utilizzare maggiormente una valutazione orientata al prodotto, piuttosto che al processo (29,98%). Inoltre, gli studiosi hanno anche rilevato che sia nel contesto spagnolo che in quello statunitense, le pratiche valutative maggiormente innovative, come *self e peer assessment*, vengono scarsamente utilizzate e che in nessuno Syllabo si dichiara di attivare una valutazione di progresso.

In riferimento alla differenza dell'uso delle pratiche valutative in relazione alle discipline accademiche analizzate dai ricercatori, i risultati indicano che in generale gli accademici spagnoli e statunitensi utilizzano maggiormente approcci valutativi tradizionali - ossia sommativi - rispetto a quelli alternativi (Lipnevich et al., 2021). Da un'analisi più approfondita si evidenzia che i docenti appartenenti all'area educativa utilizzano pratiche valutative differenziate per l'assegnazione del voto finale rispetto ai docenti di matematica. Nel dettaglio, i docenti di area educativa utilizzano sia pratiche valutative sommative che alternative, mentre i docenti di matematica prediligono maggiormente pratiche sommative tradizionali, come ad esempio l'esame scritto finale.

In tutte le precedenti ricerche, gli autori concludono che la valutazione permane una questione ostica e in gran parte ancora poco affrontata (Lipnevich et al., 2021). In effetti, secondo un Rapporto dell'EU (Rauhvargers, Deane & Pouwels, 2009), la maggior parte dei Paesi afferma di utilizzare ancora una valutazione sommativa, disattendendo una raccomandazione del Processo di Bologna che riguarda il superamento dei metodi di valutazione più tradizionali (Commissione Europea/EACEA/Eurydice, 2018; Karran, 2005). Uno dei pilastri di questo processo di cambiamento sembrerebbe essere il percorso verso processi di maggiore consapevolezza del ruolo della valutazione nell'apprendimento permanente. Assumendo questa prospettiva, solo alcuni Paesi si sono orientati a proporre forme valutative più partecipative in cui gli studenti possano sperimentare, e quindi acquisire competenze rispetto agli approcci e alle pratiche valutative.

In tale contesto, il presente studio intende esaminare le pratiche dichiarate dai docenti italiani nei Syllabus con la finalità di definire un quadro di tali pratiche a livello nazionale. L'idea che ha mosso l'indagine è stata quella di iniziare a costruire uno "stato dell'arte" a livello nazionale, come punto di partenza per una successiva elaborazione di indicazioni specifiche per mettere in atto pratiche formative relative alla valutazione, dirette ai docenti italiani.

4. La ricerca

4.1 Obiettivo e domande di ricerca

Il presente studio intende esplorare ed analizzare le pratiche di valutazione utilizzate dai docenti universitari nel contesto nazionale, a partire da quelle dichiarate nei Syllabi degli insegnamenti.

Utilizzando il framework di analisi utilizzato anche da Lipnevich et al. (2021) la ricerca ha inteso individuare:

- gli approcci cui i docenti fanno riferimento per l'assegnazione del "voto d'esame";
- le pratiche valutative che i docenti mettono in atto nell'ambito degli insegnamenti.

Le domande di ricerca cui si è voluto rispondere sono, dunque, le seguenti:

- Q1. Quale approccio viene utilizzato per l'assegnazione del voto finale e quali sono le pratiche di valutazione utilizzate dai docenti universitari nel contesto nazionale?
 - Q1.1 Quale approccio al voto finale utilizzano i docenti universitari?
 - Q1.2 Quali pratiche di valutazione mettono in atto nell'ambito dei loro insegnamenti?
 - Q1.3 Vi sono differenze in relazione all'area disciplinare dei docenti o ad altre loro caratteristiche individuali (genere, affiliazione, ecc.)?

4.2 Campione e procedure di analisi

Per rispondere alle domande di ricerca abbiamo deciso di considerare come oggetto di analisi i Syllabi ma, non essendo reperibile un elenco nazionale di Syllabi, per formare il campione di analisi si è scelto di fare riferimento all'elenco dei docenti italiani così come proposto dal MIUR (a.a.2021/2022) in collaborazione con Cineca². Pertanto, per ciascun docente è stato preso in considerazione un Syllabo in modo casuale.

Il campione di docenti è stato selezionato mediante stratificazione dell'intera popolazione in sottopopolazioni, quali i settori scientifico-disciplinari, a cui appartiene il singolo docente. È stato così individuato un campione rappresentativo di 3008 docenti, appartenenti alle università italiane statali e non statali, corrispondente al 5% della popolazione complessiva ($n= 60158$ docenti, ad esclusione dei docenti a contratto). Ne è risultato un campione così articolato (Tabella 1):

Area scientifico-disciplinare	Numerosità campione	% campione
A1. Scienze matematiche e informatiche	168	5,58%
A2. Scienze fisiche	129	4,30%
A3. Scienze chimiche	155	5,15%
A4. Scienze della terra	56	1,86%
A5. Scienze biologiche	255	8,48%
A6. Scienze mediche	471	15,65%
A7. Scienze agrarie e veterinarie	163	5,43%
A8. Ingegneria civile e Architettura	191	6,36%
A9. Ingegneria industriale e dell'informazione	322	10,69%
A10. Scienze dell'antichità, filologico-letterarie e storico-artistiche	251	8,35%
A11. Scienze storiche, filosofiche, pedagogiche e psicologiche	236	7,83%
A12. Scienze giuridiche	246	8,17%
A13. Scienze economiche e statistiche	272	9,06%
A14. Scienze politiche e sociali	93	3,10%
Totale	3008	100,00%

Tabella 1: Campione docenti per area scientifico-disciplinare

Successivamente, sono stati individuati 3008 Syllabi, selezionandone casualmente uno per ciascun docente facente parte del campione.

I Syllabi, scaricati dai siti di ciascun Ateneo di riferimento, sono stati innanzitutto esaminati per verificare che vi fossero compilate le sezioni che avrebbero potuto includere elementi riguardanti le pratiche valutative utilizzate dai docenti, ossia: a) i risultati di apprendimento attesi; b) le pratiche didattiche; c) le modalità di valutazione.

Dal numero di Syllabi del campione individuato sono stati esclusi 466 Syllabi perché non presenti nel sito ufficiale d'Ateneo e 322 Syllabi perché mancanti di informazioni (uno o più elementi di analisi non erano scritti). I Syllabi analizzabili sono risultati così 2220 per un totale di 508075 parole.

All'interno di un disegno di ricerca misto sequenziale esplorativo QUAL Quant. (Creswell & Plano Clark, 2011; Guest & Fleming, 2014), sui Syllabi è stata compiuta un'analisi dapprima qualitativa, i cui risultati sono stati trattati con analisi statistiche.

L'analisi qualitativa del contenuto, condotta con approccio top-down in riferimento al framework proposto da Lipnevich et al. (2021), è stata svolta mediante il software di analisi testuale AtlasTi 22³. Tale analisi è stata condotta da due giudici indipendenti per il 10% dei documenti. L'analisi si è svolta secondo le seguenti fasi:

2 <https://cercauniversita.cineca.it/>

3 Il software è stato scelto poiché lascia totale libertà al ricercatore circa l'analisi e l'interpretazione dei materiali esplorati e analizzati (Soratto, Pires de Pires & Friese, 2020; Smit, 2021).

1. sono state selezionate le porzioni di testo che descrivevano le pratiche valutative utilizzate dai docenti;
2. le pratiche individuate sono state raggruppate nelle categorie di prodotto, processo e progresso.

Questa prima parte si è conclusa con il raggiungimento di una “sensibilità comune” rispetto al processo di codificazione e il calcolo dell’affidabilità (ICC= 0,99) (Hallgren, 2012; Koo & Li, 2016). L’analisi è quindi proseguita suddividendo il successivo lavoro fra i due giudici che hanno operato indipendentemente sui materiali da analizzare rimanenti, mantenendo lo stesso procedimento di analisi in due fasi sopra descritte. Sono state quindi individuate nei testi i codici relativi alle pratiche valutative e individuate le categorie di “prodotto”, “processo”, “progresso”.

Successivamente, sui dati qualitativi è stata applicata un’analisi statistica a classi latenti (Vermunt & Magidson, 2003). Sulla base di un modello probabilistico che analizza le diverse combinazioni di pratiche valutative dichiarate nei Syllabi, l’analisi a classi latenti ha lo scopo di identificare gruppi (cluster) le cui unità siano tanto più omogenee al loro interno, quanto più eterogenee al loro esterno (Collins & Lanza, 2010).

5. Risultati

Il campione di docenti da cui sono stati selezionati casualmente i Syllabi è composto principalmente da professori associati (44,5%) e di genere maschile (61,35%), che svolgono il proprio ruolo nel nord Italia (52,66%) e in mega università (37,03%), ossia università con più di 40.000 studenti (cfr. Tabella 2)

Variabili	Valori	Fq	%
Ruolo	Ordinario	578	26,04%
	Associato	988	44,50%
	RU	221	9,95%
	RTD	433	19,50%
Genere	F	858	38,65%
	M	1362	61,35%
Area geografica	Nord	1169	52,66%
	Centro	463	20,86%
	Sud	575	25,90%
	Telematica	13	0,59%
Tipologia di Ateneo	Mega	822	37,03%
	Grande	806	36,31%
	Medio	343	15,45%
	Piccolo	76	3,56%
	Politecnico	155	6,98%
	Telematico	15	0,68%

Tabella 2: Campione di docenti

In merito alla domanda di ricerca Q1.1 (*Quale approccio al voto finale utilizzano i docenti universitari?*), i dati evidenziano che tutti i docenti (n=2220) dichiarano di utilizzare una valutazione di prodotto, ossia volta a verificare gli esiti degli apprendimenti degli studenti (100% dei docenti). Solo un terzo dei docenti (n=731; 32,92%) dichiara di mettere in atto una valutazione di processo. In un unico caso si utilizza il criterio del progresso (n=1; 0,04%).

In riferimento alla domanda Q1.2 (*Quali pratiche di valutazione mettono in atto nell’ambito dei loro in-*

segnamenti?), i dati codificati rilevano che le pratiche di valutazione maggiormente utilizzate dai docenti nei loro insegnamenti sono gli esami finali orali (74,40% di occorrenze sul totale) e/o gli esami scritti finali (43,74% di occorrenze sul totale) (Tabella 2). Le altre pratiche valutative dichiarate dai docenti sono utilizzate con minore frequenza, come si evince dalla Tabella 3.

Pratiche di valutazione		Fq delle occorrenze	% delle occorrenze
Valutazione di prodotto	Esame orale finale individuale	1652	74,40%
	Esame scritto finale individuale	971	43,74%
	Esame pratico finale individuale	251	11,31%
	Esame parziale	296	13,33%
Valutazione di processo	Attività individuali in itinere (report, project, relazione orale)	391	17,61%
	Attività di gruppo in itinere (report, project, relazione orale).	150	6,76%
	Altre pratiche di valutazione autentica (prove <i>real life</i> , ecc)	254	11,44%
	Partecipazione e frequenza	115	5,18%
	Pratiche di self-assessment	43	1,94%
	Altre pratiche di valutazione formativa (test senza voto, ecc)	95	4,28%
	Peer assessment	6	0,27%
	Peer Feedback	9	0,41%
Valutazione di progresso	Valutazione diagnostica individuale	1	
Totale occorrenze		4233	100%

Tabella 3: Frequenze e % delle occorrenze relative alle pratiche di valutazione utilizzate dai docenti, classificate in base al criterio di prodotto, processo e progresso

In riferimento alle analisi compiute per rispondere alla terza domanda di ricerca Q1.3 (*Vi sono differenze in relazione all'area disciplinare dei docenti o ad altre caratteristiche individuali?*), sono innanzitutto emersi sei cluster differenti a seconda delle pratiche valutative utilizzate dai docenti (Tabella 4).

Cluster	Dimensione	Descrizione
1	43,6%	Prevalenza netta esame orale (unita raramente a qualche integrazione di esami parziali o altre attività individuali in itinere)
2	36,9%	Prevalenza esame scritto unita frequentemente con integrazione orale (e in misura molto minore esami parziali)
3	7,2%	Esame orale (unita frequentemente con attività individuali in itinere)
4	5,4%	Esame Scritto (unito ad integrazioni soprattutto di valutazione di prodotto e alcune di processo)
5	5,3%	Prove diversificate con una netta prevalenza di prove di processo
6	1,6%	Approcci e prove diversificati

Tabella 4: Cluster individuati con analisi statistica a classi latenti

I primi due cluster fanno riferimento ad approcci valutativi esclusivamente di prodotto. I cluster 3 e 4 mostrano una prevalenza di valutazione di prodotto con alcune pratiche di processo, mentre il cluster 5 è composto prevalentemente da pratiche di processo. Il 6, cluster residuale, riporta approcci e prove diversificati.

L'analisi permette di rilevare che vi sono differenze fra le macro-aree scientifico disciplinari. Anche se tutte utilizzano in netta prevalenza valutazione di prodotto, la macro area Scientifico-Tecnologica presenta i più alti valori nel cluster 5 (8,04%, contro 2,53% dell'area Sanitaria e 2,13% dell'area Umanistico – Sociale), dimostrando in tal modo l'utilizzo di pratiche diversificate e maggiormente orientate alla valutazione

di processo (Tabella 5). Va tenuto presente che i cluster 5 e 6 assieme non coprono neppure il 7% complessivo delle occorrenze.

Macro Area	Cluster						
	1	2	3	4	5	6	Totale
Area Sanitaria	187	158	13	26	10	2	396
	47,22%	39,90%	3,28%	6,57%	2,53%	0,51%	100%
Scientifico - Tecnologica	369	514	91	78	94	23	1169
	31,57%	43,97%	7,78%	6,67%	8,04%	1,97%	100%
Umanistico - Sociale	412	147	57	17	14	10	657
	62,71%	22,37%	8,68%	2,59%	2,13%	1,52%	100%

Tabella 5: Macro aree scientifiche e cluster

Un'analisi più specifica relativa alle differenze tra i settori scientifico-disciplinari permette di constatare che il settore 04 – Scienze della Terra, 08 – Ingegneria civile e architettura e 09 – Ingegneria Industriale e dell'informazione si distinguono dalle altre 11 aree, poiché presentano valori equilibrati nei diversi cluster, derivandone che in questi casi si utilizzano approcci e forme valutative variegata (Tabella 6).

Attraverso una lettura per cluster, possiamo affermare che i settori scientifico-disciplinari che si avvalgono prevalentemente dell'esame orale finale sono il 10 - Scienze dell'antichità, filologico-letterarie e storico-artistiche (cl1, 69,61%) e il 12 – Scienze giuridiche (cl1, 74,48%). I settori che impiegano in maggior misura l'esame scritto finale sono l'1 – Scienze matematiche e informatiche (cl2, 64,46%) e il 13 – Scienze economiche e statistiche (cl2, 62,56%).

Considerando il cluster 5, che è caratterizzato dalla prevalenza di valutazione di processo, si rileva che esso supera il 10% delle frequenze solo per tre settori: l'8 – Ingegneria civile e Architettura (21,57%), il 9 - Ingegneria industriale e dell'informazione (10,34%) e il 4 – Scienze della terra (10,20%).

Settori scientifico-disciplinari	Cluster						
	1	2	3	4	5	6	Totale
A1. Scienze matematiche e informatiche	24 19,83%	78 64,46%	5 4,13%	5 4,13%	8 6,61%	1 0,83%	121 100%
A2. Scienze fisiche	39 42,86%	43 47,25%	3 3,30%	1 1,10%	2 2,20%	3 3,30%	91 100%
A3. Scienze chimiche	50 39,37%	60 47,24%	6 4,72%	4 3,15%	4 3,15%	3 2,36%	127 100%
A4. Scienze della terra	24 48,98%	9 18,37%	5 10,20%	6 12,24%	5 10,20%	0 0,00%	49 100%
A5. Scienze biologiche	88 45,60%	90 46,63%	5 2,59%	4 5,24%	4 1,50%	2 0,75%	193 100%
A6. Scienze mediche	124 46,44%	118 44,19%	5 1,87%	2 1,04%	2 1,04%	6 3,11%	267 100%
A7. Scienze agrarie e veterinarie	63 48,84%	40 31,01%	8 6,20%	12 9,30%	6 4,65%	0 0,00%	129 100%
A8. Ingegneria civile e Architettura	45 29,41%	24 5,69%	38 24,84%	11 7,19%	33 21,57%	2 1,31%	153 100%
A9. Ingegneria industriale e dell'informazione	73 31,47%	83 35,78%	18 7,76%	31 13,36%	24 10,34%	3 1,29%	232 100%

A10. Scienze dell'antichità, filologico-letterarie e storico-artistiche	142 69,61%	38 18,63%	22 5,39%	4 1,96%	6 2,94%	3 1,47%	204 100%
A11. Scienze storiche, filosofiche, pedagogiche e psicologiche	91 48,40%	64 30,04%	23 12,23%	7 3,72%	1 0,53%	2 1,06%	188 100%
A12. Scienze giuridiche	143 74,48%	29 15,10%	10 5,21%	4 2,08%	5 2,60%	1 0,52%	192 100%
A13. Scienze economiche e statistiche	26 12,81%	27 62,56%	11 5,42%	8 8,87%	6 7,88%	5 2,46%	203 100%
A14. Scienze politiche e sociali	36 49,32%	6 21,92%	13 17,81%	2 2,74%	2 2,74%	4 5,48%	73 100%

Tabella 6: Settori scientifico-disciplinari e cluster

Non sembrano esserci differenze di approcci e pratiche né in relazione al genere dei docenti, né riguardo alla posizione accademica. Si evidenziano, invece, in relazione all'affiliazione ad un Ateneo o un altro. Le prime analisi mostrano che i due Politecnici di Torino e Milano si distinguono nettamente dalle altre sei Università (grandi Atenei) finora considerate. Mentre, infatti, questi ultimi si attestano su livelli che ricalcano quelli generali con valori molto alti nei cluster 1 e 2, i due Politecnici non solo mostrano valori che si distribuiscono quasi equamente sui primi cinque cluster, ma anche che il cluster 5 è quello con valori più alti degli altri (Tabella 7 e Tabella 8).

Cluster	Fq.	%
1	4	7,69%
2	11	21,15%
3	5	9,62%
4	12	23,08%
5	17	32,69%
6	3	6%
Totale	52	100%

Tabella 7: Politecnico di Torino

Cluster	Fq.	%
1	9	13,430
2	24	35,82
3	5	7,46
4	8	11,94
5	20	29,85
6	1	1,49
Totale	67	100

Tabella 8: Politecnico di Milano

6. Discussione e conclusione

Il presente studio ha inteso ottenere un quadro d'insieme delle pratiche valutative dei docenti universitari italiani, di cui fino ad oggi non si avevano evidenze. Esso, dunque, si propone come la prima indagine condotta a livello nazionale su questo tema. Le analisi compiute su un campione rappresentativo hanno permesso di ottenere informazioni sull'attuale "stato dell'arte" in relazione a "come si assegna il voto" nelle università italiane.

Si ritiene che, seppure il quadro ottenuto costituisca una fonte d'informazione particolarmente inte-

ressante in relazione alla ricerca italiana riguardante la valutazione in ambito universitario e il *Faculty Development*, tema, quest'ultimo, oggi particolarmente caldo vista l'auspicata costituzione dei *Teaching Learning Center*⁴, esso vada letto e interpretato con alcune precauzioni.

La prima riguarda il fatto che quelle analizzate siano dichiarazioni dei docenti nei Syllabi, dunque, dati secondari e non necessariamente rispondenti alle successive pratiche messe effettivamente in atto. Inoltre, la comunicazione utilizzata dai docenti per descrivere le pratiche valutative risulta, a volte, talmente stringata da non permettere di comprendere con precisione la tipologia di prova proposta. Per ovviare a tale limite, il presente lavoro di ricerca proseguirà con un' esplorazione ulteriore, tramite questionari e interviste sia ai docenti che agli studenti, con successiva triangolazione dei dati. Risulterebbe particolarmente difficile l'esplorazione delle pratiche in modo diretto a livello così ampio.

Una seconda precauzione riguarda il campione. Esso è il risultato di un processo in cui sono state compiute alcune scelte che in qualche modo potrebbero avere escluso l'emergere di risultati diversi. Innanzitutto, esso si basa su un elenco ufficiale di docenti strutturati, piuttosto che su un elenco di tutti i Syllabi degli insegnamenti in Italia, poiché quest'ultimo non esiste. In secondo luogo, il campionamento è stato costruito secondo stratificazioni per aree disciplinari e non considerando altre stratificazioni possibili, quali quella relativa alle posizioni accademiche (PO, PA, RTdB, RTdA, RU...), alla dimensione degli Atenei o al genere. D'altra parte, la scelta fatta è stata dettata da precedenti risultati di ricerca evidenziati a livello internazionale (Lipnevich et al., 2021; Panadero et al., 2019), che rilevavano differenze fra discipline nell'attuazione delle pratiche valutative.

Considerati tali limiti, si possono proporre alcuni elementi di discussione dei risultati.

Prima ancora di considerare specificamente le domande di ricerca, il primo evidente punto di attenzione riguarda il numero di Syllabi che non si sono potuti analizzare (più del 20% del campione) perché non presenti sui siti di Ateneo (aspetto che riguarda addirittura elementi di legalità), o perché non completi. Punto questo già discusso altrove in relazione sia al valore del Syllabus come strumento fondamentale di progettazione didattica e di comunicazione con gli studenti, che specificamente allo scarso peso che molti docenti danno al momento valutativo nel contesto delle attività di insegnamento/apprendimento (Serbati, Picasso, Grion & Grion, 2022).

Quanto alla prima domanda di ricerca Q1, *“Quale approccio viene utilizzato per l'assegnazione del voto finale e quali le pratiche di valutazione utilizzate dai docenti universitari nel contesto nazionale, alla luce della ricerca internazionale sulla valutazione?”*, ciò che emerge dall'analisi qualitativa - evidenza attesa per chi conosce l'università italiana - riguarda la netta prevalenza di coloro che utilizzano solo una valutazione di prodotto. Solamente un terzo dei docenti circa (32,92%) realizza anche una valutazione di processo, e in un solo caso si è vista una valutazione di progresso. Il risultato da noi individuato conferma quanto evidenziato da Panadero et al. (2019) quando riferisce che i docenti universitari prediligono ancora approcci “tradizionali” alla valutazione. È, inoltre sovrapponibile ai risultati riferiti alla situazione spagnola indagata da Lipnevich e colleghi (2021) che presenta poco meno di un terzo di docenti che usano la valutazione di processo; mentre si discosta da quella americana in cui la valutazione di processo è condotta da più della metà dei docenti.

Dati questi risultati, ne emergerebbe l'importanza di una profonda riflessione da parte dei docenti italiani rispetto alla possibilità di “catturare” effettivamente l'apprendimento attraverso le prove valutative che propongono. Se, infatti, per “apprendimento” intendiamo riferirci al costruito di “competenza”, come indicato dal Processo di Bologna (Commissione Europea/EACEA/Eurydice, 2018; Karran, 2005), non possiamo non considerare che una valutazione di prodotto non sia sufficiente. Per valutare la competenza, costruito particolarmente complesso e multi-sfaccettato, non si possono non mettere in atto prove che valorizzino il processo di apprendimento o anche il suo progresso (Marzano et al., 2014; Castoldi, 2016; Varisco, 2004).

Anche per ciò che riguarda la seconda domanda di ricerca (*Quali pratiche di valutazione mettono in atto nell'ambito dei loro insegnamenti?*), i nostri risultati confermerebbero quelli del contesto spagnolo (Panadero et al., 2019). Le pratiche valutative maggiormente presenti si riferiscono agli esami finali scritto e/o orale,

4 Per definizione e approfondire i cenni storici, gli scopi, gli aspetti strutturali, l'organizzazione e le funzioni del *Teaching Learning Center*, si veda Lotti et al., (2022).

ossia a prove valutative tradizionalmente presenti nelle università. In alcuni contesti, si mettono in atto anche prove valutative più coerenti con la recente ricerca sulla valutazione, quali le prove autentiche in funzione sommativa e le prove formative⁵; tuttavia, anche in questo caso, per la maggior parte delle situazioni vi è una netta prevalenza di attività valutative individuali. Pure in relazione a questo, sarebbe necessario che i docenti riflettessero sul fatto che molti compiti e prove esterne all'università, propri ad esempio dei contesti professionali, implicano lavori di gruppo, attività di collaborazione con altri e lo svolgimento di attività con persone diverse e in contesti diversificati. In questi casi, le *conoscenze* e le *abilità* individuali devono essere scelte, mobilitate dall'individuo ed essere "giocate" all'interno di contesti complessi, in cui si attivano (e si possono valutare) le *competenze* delle persone. A questo, l'università dovrebbe preparare, e nell'ambito di simili situazioni dovrebbero essere messe a punto maggiormente le prove valutative (Nicol, 2014; Grion & Serbati, 2019).

In riferimento alla terza domanda di ricerca, i nostri risultati confermano quanto dimostrato da Pandero et al. (2019): poiché le discipline socio-umanistiche fanno riferimento maggiormente ad una valutazione di prodotto, a differenza delle discipline sanitarie e scientifico-tecnologiche che rimandano più frequentemente ad approcci ed attività valutativi di processo (Panadero et al., 2019). Diversi sono i risultati del contesto statunitense: nella ricerca di Lipnevich et al., (2020, 2021), le aree socio-educative utilizzano pratiche valutative maggiormente differenziate, mentre quelle matematico-scientifiche prediligono pratiche tradizionali.

Generalizzando questi risultati, potremmo ipotizzare che la cultura valutativa europea risulta più "tradizionalista" di quella americana, che probabilmente accoglie più favorevolmente e implementa maggiormente i risultati dell'attuale ricerca sulla valutazione.

Per quanto riguarda i risultati particolari rilevati per i due politecnici, che fanno ampio uso di pratiche valutative diversificate e alternative, potremmo ipotizzare che questo sia dovuto o a una "cultura valutativa d'Ateneo" (tutta da dimostrare!), oppure che tali risultati derivino dalla presenza ormai consolidata, nei due politecnici, di *Teaching Learning Center* che hanno contribuito a sviluppare e diffondere una migliore cultura valutativa nei docenti ad essi affiliati. A questa seconda ipotesi riteniamo di dare maggiore peso. In tale senso pensiamo che il presente lavoro possa costituire un contributo importante nel processo di edificazione dei *Teaching Learning Center*, in quanto propone un quadro d'insieme da cui partire per progettare e attivare pratiche formative specifiche per il miglioramento della cultura valutativa dei docenti italiani.

Un'ulteriore riflessione a margine si può proporre focalizzando l'attenzione sugli approcci e le pratiche valutative messe in atto in particolare nelle aree socio-educative, che si occupano della formazione degli insegnanti. Se, infatti, consideriamo il modellamento come la pratica formativa più efficace nella formazione degli insegnanti (Korthagen, Loughran & Russell, 2006), ne deriva che se i docenti universitari che operano in questi settori propongono prevalentemente o solo pratiche valutative tradizionali, allora anche nella scuola esse continueranno ad essere perpetuate, con il conseguente rinforzarsi di un circolo vizioso che blocca il processo di rinnovamento delle pratiche valutative in tutto il sistema formativo italiano.

È perciò anche in relazione al rinnovamento delle pratiche valutative nella scuola, come contesto in cui vanno ad inserirsi persone formate dall'università, che risulta urgente porre profonda attenzione al miglioramento delle consapevolezze valutative dei professori universitari

References

- Anderson, L. W. & Krathwohl, D.R., et al. (2001). *A taxonomy for learning, teaching and assessing: A revision of Bloom's taxonomy of educational objectives*. New York: Longman.
- Black, P. & William, D. (2009). Developing the theory of formative assessment, *Educational Assessment, Evaluation and Accountability*, 21(1), 5-31. <https://doi.org/10.1007/s11092-008-9068-5>.
- Bloom, B. S., & Krathwohl, D. R. (1956). *Taxonomy of educational objectives: The classification of educational goals*. New York, NY: Longmans Green.

5 Per una definizione di prove autentiche e formative si confronti, ad esempio, Marta & Trincherò (2020), Grion, Serbati & Cecchinato (2022) e Corsini (2023).

- Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in continuing education*, 22(2), 151-167. <https://doi.org/10.1080/713695728>.
- Boud, D., & Falchikov, N. (2007). Introduction: Assessment for the longer term. In *Rethinking assessment in higher education* (pp. 13-23). London: Routledge.
- Boud, D., & Soler, R. (2016). Sustainable assessment revisited. *Assessment and Evaluation in Higher Education*, 41(3), 400-413. <https://doi.org/10.1080/02602938.2015.1018133>.
- Brookhart, S. M. (2009). *Grading*. London: Pearson Education.
- Brown, S. (2014). *Learning, Teaching and Assessment in Higher Education. Global Perspectives*. London: Palgrave Macmillan.
- Carless D. (2005). Prospects for the implementation of assessment for learning. *Assessment in Education*, 12, (1), 39-54. <https://doi.org/10.1080/0969594042000333904>.
- Carless, D. (2015). Exploring learning-oriented assessment processes. *Higher Education*, 69(6), 963– 976. <https://doi.org/10.1007/s10734-014-9816-z>.
- Castoldi, M. (2016). *Valutare e certificare le competenze*. Roma: Carocci.
- Coggi C. (ed.), (2022). *Formare i docenti universitari alla didattica e alla valutazione*. Milano: Franco Angeli.
- Coggi, C. (ed.) (2019a). *Innovare la didattica e la valutazione in Università: Il progetto IRIDI per la formazione dei docenti*. Milano: Franco Angeli.
- Coggi, C. (2019b). Migliorare l'affidabilità degli esami, innovare gli strumenti, adottare strategie formative nel valutare e concorrere all'accountability istituzionale. In C. Coggi (ed.), *Innovare la didattica e la valutazione in Università. Il progetto IRIDI per la formazione dei docenti*. (Vol. 1, pp. 155-200). Milano: Franco Angeli.
- Coggi, C. (2005). Valutare gli studenti. Problemi teorici e prassi nella facoltà. In C. Coggi (ed.), *Per migliorare la didattica universitaria* (pp. 205-238). Lecce: Pensa MultiMedia.
- Collins, L., & Lanza, S. (2010). *Latent class and latent transition analysis: With applications in the social, behavioral, and health sciences*. Hoboken, New Jersey: John Wiley & Sons.
- Commissione europea/EACEA/Eurydice, (2018), *La carriera degli insegnanti in Europa: accesso, progressione e sostegno. Rapporto Eurydice*. Lussemburgo: Ufficio delle pubblicazioni dell'Unione europea.
- Corsini, C. (2023). *La valutazione che educa*. Milano: FrancoAngeli.
- Creswell, J. W. & Plano Clark, V. L. (2011). *Designing and conducting mixed methods research*. Thousand Oaks, California: Sage Publications.
- Dochy, F. J. R. C., Segers, M., & Sluijsmans, D. (1999). The use of self-, peer and co-assessment in higher education: A review. *Studies in Higher education*, 24(3), 331-350. <https://doi.org/10.1080/03-0750799-12331379935>.
- Doria, B. & Grion, V. (2020). L'autovalutazione nel contesto universitario: una revisione sistematica della letteratura. *Form@re – Open Journal Per La Formazione in Rete*, 20(1), 78-92. <https://doi.org/10.13128/form-8247>.
- Foschi, L. C., & Cecchinato, G. (2019). Validity and reliability of peer-grading in in-service teacher training. *Italian Journal of Educational Research*, 177-194. <https://doi.org/10.7346/SIRD-1S2019-P177>.
- Grion, V. & Serbati, A. (2019). *Valutazione sostenibile e feedback nei contesti universitari*. Lecce: Pensa MultiMedia.
- Grion, V., Serbati, A. & Cecchinato G. (2022). *Dal voto alla valutazione per l'apprendimento*. Roma: Carocci.
- Grion, V., Serbati, A., Felisatti, E., & Li, L. (2019). Peer feedback and technology-enhanced assessment as critical issues to foster student learning. *Italian Journal of Educational Research*, 13(numero speciale), 9-14.
- Grion, V., Serbati, A., Tino, C., & Nicol, D. (2017). Ripensare la teoria della valutazione e dell'apprendimento all'università: un modello per implementare pratiche di peer review. *Italian Journal of Educational Research*, 209-226. Retrieved from <https://ojs.pensamultimedia.it/index.php/sird/article/view/2554>.
- Grion, V. & Tino, C. (2018). Verso una "valutazione sostenibile" all'università: percezioni di efficacia dei processi di dare e ricevere feedback fra pari. *Lifelong Lifewide Learning*, 31, 38-55. <https://doi.org/10.19241/lll.v14i31.104>.
- Guest, G., & Fleming, P. (2014). Mixed Methods Research. In G. Guest, E.E. Namey (Eds.), *Public Health Research Methods* (pp. 581-610). London: Sage.
- Guskey, T. R. (2011). Five obstacles to grading reform. *Educational Leadership*, 69(3), 16-21.
- Guskey, T. R. (2019). *Get set, go! Implementing successful reforms in grading and reporting*. Bloomington, Indiana, USA: Solution Tree.
- Hallgren, K. A. (2012). Computing Inter-Rater Reliability for Observational Data: An Overview and Tutorial. *Tutorials in quantitative methods for psychology*, 8(1), 23-34. <https://doi.org/10.20982/tqmp.08.1.p023>.
- Karran, T. (2005). Pan-European grading scales: Lessons from national systems and the ECTS. *Higher Education in Europe*, 30(1), 5–22. <https://doi.org/10.1080/03797720500087949>.
- Korthagen, F., Loughran, J., & Russell, T. (2006). Developing fundamental principles for teacher education programs and practices. *Teaching and Teacher Education*, 22(8), 1020-1041. <https://doi.org/10.1016/j.tate.2006.04.022>.
- Koo, T. K., & Li, M. Y. (2016). A guideline of selecting and reporting intraclass correlation coefficients for reliability research. *Journal of chiropractic medicine*, 15(2), 155-163. <https://doi.org/10.1016/j.jcm.2016.02.012>.

- Li, L., & Grion, V. (2019). The Power of Giving and Receiving in Peer Assessment. *AISHE Journal*, 11(11), 1-17.
- Li, L., Liu, X. Y. & Zhou, Y. C. (2012). Give and Take: A Re-Analysis of Assessor and Assessee's Roles in Technology Facilitated Peer Assessment. *British Journal of Educational Technology*, 43(3), 376-384. <https://doi.org/10.1111/j.1467-8535.2011.01180.x>.
- Li, L., Liu, X., & Steckelberg, A.L. (2010). Assessor or assessee: How student learning improves by giving and receiving peer feedback. *British Journal of Educational Technology*, 41, 525-536. <https://doi.org/10.1111/j.1467-8535.2009.00968.x>.
- Lipnevich, A. A., Guskey, T. R., Murano, D. M., & Smith, J. K. (2020). What do grades mean? Variation in grading criteria in American college and university courses. *Assessment in Education: Principles, Policy & Practice*, 27(5), 480-500. <https://doi.org/10.1080/0969594X.2020.1799190>.
- Lipnevich, A. A., Panadero, E., Gjicali, K. & Fraile, J. (2021). What's on the Syllabus? An analysis of assessment criteria in first year courses across US and Spanish universities. *Educational Assessment, Evaluation and Accountability*, 33(4), 675-699.
- Lotti, A., Serbati, A., Grion, Picasso, F., & Felisatti, E. (2022). Teaching and Learning Centre: Analysis of Key Elements. *Formazione & insegnamento*, 20(2), 75-88. https://doi.org/10.7346/-fei-XX-02-22_06.
- Marta, D. A., & Trincherò, R. (2020). Progettare la valutazione per competenze nell'istruzione superiore. *Giornale Italiano della Ricerca Educativa*, 13(24), 37-53.
- Marzano, A., Calenda, M., Galliani, L. & Petolicchio, A. (2014). La dimensione referenziale: prodotto, processo, sistema. In L. Galliani, A. M. Notti (Eds.), *Valutazione educativa* (pp. 69-139). Lecce: Pensa MultiMedia.
- Montgomery, K. (2002). Authentic tasks and rubrics: Going beyond traditional assessments in college teaching. *College teaching*, 50(1), 34-40.
- Nasab, F. G. (2015). Alternative versus traditional assessment. *Journal of Applied Linguistics and Language Research*, 2(6), 165-178.
- Nicol, D. (2014). Guiding principles of peer review: unlocking learners' evaluative skills. In C. Kreber, C. Anderson, N. Entwistle, J. McArthur (Eds.), *Advances and Innovations in University Assessment & Feedback* (pp. 197-224). Edinburgh: Edinburgh University Press.
- Nicol, D. (2018). Unlocking generative feedback via peer reviewing. In V. Grion, & A. Serbati (Eds.), *Assessment of Learning or Assessment for Learning? Towards a culture of sustainable assessment in HE* (pp. 73-85). Lecce: Pensa MultiMedia.
- Nicol, D. (2020). The power of internal feedback: exploiting natural comparator processes. *Assessment & Evaluation in higher education*, 46(5), 756-778. <https://doi.org/10.1080/02602938.2020.1823314>.
- Nicol, D. (2021). *Guiding learning by activating students' inner feedback*. London: Times Higher Education.
- O'Connor, K. (2009). Grades: When, why, what impact, and how?. *Education Canada*, 50(2), 38-41.
- Panadero, E., & Lipnevich, A.A. (2022). A review of feedback models and typologies: Towards an integrative model of feedback elements, *Educational Research Review*, 35, 1-22. <https://doi.org/10.1016/j.edu-rev.2021.100416>.
- Panadero, E., & Alonso-Tapia, J. (2013). Self-assessment: Theoretical and practical connotations. When it happens, how is it acquired and what to do to develop it in our students. *Electronic Journal of Research in Educational Psychology*, 2(30), 551-576. [10.14204/ejrep.30.12200](https://doi.org/10.14204/ejrep.30.12200).
- Panadero, E., & Brown, G. T. L. (2017). Teachers' reasons for using peer assessment: Positive experience predicts use. *European Journal of Psychology of Education*, 32(1), 133-156. <https://www.jstor.org/stable/44951865>.
- Panadero, E., Fraile, J., Fernández Ruiz, J., Castilla-Estévez, D., & Ruiz, M. A. (2019). Spanish university assessment practices: examination tradition with diversity by faculty. *Assessment & Evaluation In Higher Education*, 44(3), 379-397. <https://doi.org/10.1080/02602938.2018.1512553>.
- Pastore, S. (2012). Silent assessment? Cosa pensano della valutazione gli studenti universitari. *Italian Journal of Educational Research*, 62-73. Retrieved from <https://ojs.pensamultimedia.it/index.php/sird/article/view/240>.
- Rauhvargers, A., Deane C., Pouwels, W. *Bologna Process Stocktaking Report 2009*, ISBN: 978-90-403-0295-4 NUR 84, Flemish Ministry of Education and Training, Brussels, 2009, 144 pages.
- Sadler, D.R. (1989). Formative assessment and the design of instructional systems. *Instructional Science*, 18(2), 119-144. <https://doi.org/10.1007/BF00117714>.
- Sambell, K., Brown, S., & Race, P. (2019). Assessment as a locus for engagement: priorities and practicalities. *Italian Journal of Educational Research*, 45-62. Retrieved from <https://ojs.pensamultimedia.it/index.php/sird/article/view/3268>.
- Sambell, K., McDowell, L., & Montgomery, C. (2013). *Assessment for learning in higher education*. London: Routledge.
- Serbati, A., Grion, V., & Fanti, M. (2019). Caratteristiche del peer feedback e giudizio valutativo in un corso universitario blended. *Giornale Italiano della Ricerca Educativa*, 12, 115-137. <https://doi.org/10.7346/SIRD-1S2019-P115>.

- Serbati, A., Picasso, F., Grion, B., & Grion, V. (2022). Learning outcomes and constructive alignment in the Mega-Universities Syllabi: which “promises” to students? *Form@re - Open Journal Per La Formazione in Rete*, 22(2), 61-77. <https://doi.org/10.36253/form-13022>.
- Smit, B. (2021). Introduction to ATLAS. ti for Mixed Analysis. In A.J. Onwuegbuzie, & R.B. Johnson (Eds.), *The Routledge Reviewer's Guide to Mixed Methods Analysis* (pp. 331-342). London: Routledge.
- Soratto, J., Pires de Pires, D. E., & Friese, S. (2020). Thematic content analysis using ATLAS. ti software: Potentialities for researchs in health. *Revista brasileira de enfermagem*, 73(3). <https://doi.org/10.1590/0034-7167-2019-0250>.
- Taras, M. (2005). Assessment - Summative and Formative - some theoretical reflections. *British Journal of Educational Studies*, 466-478. <https://doi.org/10.1590/0034-7167-2019-0250>.
- Trincherò, R. (2018). Valutazione formante per l'attivazione cognitiva. Spunti per un uso efficace delle tecnologie per apprendere in classe. *Italian Journal of Educational Technology*, 26(3), 40-55. <https://doi.org/10.17471/2499-4324/1013>.
- Varisco, B.M. (2004). *Portfolio. Valutare gli apprendimenti e le competenze*. Roma: Carocci.
- Vermunt, J., & Magidson, J. (2003). Latent class models for classification. *Computational Statistics & Data Analysis*, 41(3-4), 531-537. [https://doi.org/10.1016/S0167-9473\(02\)00179-2](https://doi.org/10.1016/S0167-9473(02)00179-2)
- William, D. (2011). Formative Assessment: Getting the Focus Right. In D. William, *Educational Assessment* (pp. 283-289). New Jersey: Lawrence Erlbaum Associates, Inc.
- Zhao, C. G., & Qi, Q. (2022). Implementing Learning Oriented Assessment (LOA) Among Limited Proficiency EFL Students: Challenges, Strategies, and Students' Reactions. *TESOL Quarterly*. <https://doi.org/10.1002/tesq.3167>.