

Peer evaluation at university: a comparison between two teaching approaches

Valutazione tra pari all'università: due approcci didattici a confronto

Alessandra La Marca^a, Valeria Di Martino^b, Elif Gülbay^{c,1}

^a *Università degli Studi di Palermo*, alessandra.lamarca@unipa.it

^b *Università degli Studi di Catania*, valeria.dimartino@unict.it

^c *Università degli Studi di Palermo*, elif.gulbay@unipa.it

Abstract

In the field of higher education, assessment tools must be renewed and diversified, including specific technologies designed for assessment. This paper aims to enhance the reflection on those teaching practices that are most effective for the integration of new technologies for formative assessment and peer evaluation at university. In particular, we want to check whether the evaluation forms based on the use of the *Peergrade* platform are more effective in a face-to-face interaction teaching approach rather than on remote interactions between one lesson and the next one. The research is conducted on a sample of 203 students attending the course of Experimental Pedagogy at the University of Palermo and at the University of Catania and it uses a mixed methods research methodology in order to investigate the qualitative and quantitative differences between the two teaching and evaluation approaches.

Keywords: teaching in higher education; evaluating in higher education; peer assessment; feedback; Peergrade.

Sintesi

Nell'ambito della didattica universitaria occorre rinnovare e diversificare gli strumenti di valutazione includendo sempre più anche tecnologie progettate per la valutazione. Il presente contributo mira a contribuire alla riflessione sulle pratiche didattiche che risultano più efficaci per l'integrazione delle nuove tecnologie per la valutazione formativa e la valutazione tra pari in università. In particolare si vuole verificare se l'utilizzo di tali forme valutative tramite la piattaforma *Peergrade* sia più efficace in un approccio didattico basato sull'interazione face-to-face o su interazioni a distanza tra una lezione e la successiva. La ricerca, condotta su un campione di 203 studenti frequentanti il corso di Pedagogia Sperimentale dell'Università degli Studi di Palermo e dell'Università degli Studi di Catania, utilizza una metodologia di ricerca mixed methods per indagare le differenze qualitative e quantitative tra i due approcci didattici e valutativi.

Parole chiave: didattica universitaria; valutazione in università; valutazione tra pari; feedback; Peergrade.

¹ Questo articolo è il risultato del lavoro congiunto dei tre autori. In particolare La Marca ha scritto i paragrafi 2 e 4; Di Martino i paragrafi 1, 6 e 7; Gülbay i paragrafi 3, 5 e 8.

1. Introduzione

Nel contesto sempre più diversificato della didattica universitaria, le funzioni formative ed euristiche della valutazione sono diventate prominenti (Topping, 2018). La valutazione, infatti, non può e non deve essere ricondotta ad un momento separato di verifica del raggiungimento degli obiettivi di un programma di studio con funzione prevalentemente o esclusivamente certificatoria o selettiva (Grion, Serbati, Tino, & Nicol, 2018), totalmente gestito dal docente e in cui il discente riveste un ruolo passivo (Coggi, 2005). È necessario, piuttosto, un approccio che valorizzi la centralità dello studente, stimolandone la capacità di elaborare un giudizio critico autonomo, svincolato dal giudizio del docente.

Negli ultimi decenni numerose ricerche internazionali (Evans, 2013; Henderson, Ajjawi, Boud, & Molloy, 2019; Irons, 2007) e più recentemente anche nazionali (Balzaretto & Vannini, 2018; Binetti & Cinque, 2016; Coggi, 2019; Coggi & Ricchiardi, 2018; Pandolfi, 2017) hanno attenzionato in modo particolare la valutazione formativa nell'ambito della didattica universitaria. Tale valutazione costituisce una parte fondamentale del processo di apprendimento, in quanto, attraverso una condivisione dei traguardi da raggiungere funge da orientamento, consente un monitoraggio costante della produzione dello studente e, allo stesso tempo, lo aiuta a riflettere sulle conoscenze e competenze apprese e sulle procedure per gestirle autonomamente (Coggi & Ricchiardi, 2018; Vertecchi, 1976).

Una delle principali ragioni per cui è importante focalizzare l'attenzione sulla valutazione formativa in università è legata al forte impatto che essa ha sulla qualità dell'apprendimento (Henderson et al., 2019). Come si evince da varie ricerche, infatti, modalità valutative di natura formativa influenzerebbero l'esperienza universitaria degli studenti (Bartram & Bailey, 2010), il loro approccio allo studio (Struyven, Dochy, & Janssens, 2005), la qualità del loro apprendimento e i risultati ottenuti (Henderson et al., 2019).

La valutazione formativa è legata all'uso sistematico del feedback, recentemente definito come un processo dialogico in cui il valutato e il valutatore interagiscono (Jonsson & Panadero, 2018), e anche come un "process through which learners make sense of information from various sources and use it to enhance their work or learning strategies" (Carless & Boud, 2018, p. 1). Un feedback è efficace "se si riferisce a tutte le competenze attese, se è rapido, preciso nelle informazioni, riferito a criteri chiari e condivisi, se esplicita i traguardi da raggiungere e se motiva intrinsecamente chi studia ad attivarsi verso apprendimenti ulteriori" (Coggi & Ricchiardi, 2018, p. 27).

La gestione del feedback risulta oggi una questione di grande interesse soprattutto in riferimento ad aule con elevati numeri di studenti, situazioni sempre più frequenti nelle università odierne. È anche in tal senso che l'indagine sul ruolo del feedback attivato fra studenti, risulta una questione di fondamentale importanza per il miglioramento della didattica universitaria, dal momento che la capacità del docente di fornire feedback per ogni studente è limitata (Topping, 2018). Queste prospettive forniscono a docenti e studenti rilevanti opportunità per spostarsi da modelli trasmissivi a modelli più partecipativi (Nicol, 2018) non solo nel processo didattico, ma anche in quello valutativo, che si configura come momento fondamentale per l'acquisizione di conoscenze e competenze disciplinari e trasversali.

L'uso delle tecnologie digitali a sostegno della valutazione tra pari rende possibile processi altrimenti difficilmente attuabili (Tonelli, Grion, & Serbati, 2018), aumentando l'efficacia dell'apprendimento (Henderson et al., 2019) e favorendo un atteggiamento positivo degli studenti nei confronti della valutazione tra pari grazie alle condizioni per garantire l'anonimato (Foschi, Cecchinato, & Say, 2019).

Mercader, Ion, e Diaz-Vicario (2020) fanno notare che mentre esistono numerose ricerche che prendono in considerazione le caratteristiche della progettazione didattica e dell'ambiente di apprendimento quando il feedback è fornito dai docenti, poco esplorato è invece l'ambito di ricerca che prende in considerazione tali aspetti nella valutazione tra pari.

Inserendosi in quest'ultima direzione, la ricerca presentata in questo contributo è tesa ad indagare se la valutazione tra pari (tramite l'ausilio della piattaforma *Peergrade*) sia più efficace in setting didattici face-to-face (in presenza) o a distanza (tra una lezione e la successiva). Nei paragrafi seguenti, dopo una breve rassegna sulla valutazione tra pari e sulla tecnologia utilizzata (*Peergrade*), sarà descritto l'approccio metodologico utilizzato e saranno riportati i risultati ottenuti.

2. La valutazione tra pari: benefici e condizioni

La valutazione tra pari è generalmente definita come un processo in cui gli studenti valutano il lavoro reciproco in base a criteri di valutazione concordati (Li, 2019). Gli studenti di solito svolgono entrambi i ruoli di valutatore e valutato. Come valutatori, valutano il lavoro dei colleghi e forniscono feedback critici per aiutare i colleghi a migliorare i propri prodotti. Come valutati ricevono e rispondono al feedback dei loro pari per migliorare il proprio lavoro. Ricevere feedback da prospettive multiple, a volte contrastanti tra loro, e al contempo produrre feedback, aiuta gli studenti ad assumere un ruolo attivo. Gli studenti, rivedendo e valutando il lavoro prodotto da uno o più colleghi, inevitabilmente riflettono sul proprio lavoro, rinforzano la propria conoscenza della disciplina, apprendono modi diversi di eseguire lo stesso compito, maturando una capacità autovalutativa (Nicol & MacFarlane-Dick, 2006). Pertanto, il coinvolgimento attivo in entrambi i processi conduce ad un apprendimento profondo e significativo, anche grazie alla negoziazione di significati che segue la formulazione e condivisione di un feedback (Li, Liu, & Zhou, 2012).

I benefici della valutazione tra pari sui risultati accademici degli studenti e sullo sviluppo cognitivo sono stati ampiamente descritti in letteratura (Li, 2019). Tra gli effetti positivi possiamo evidenziare: il maggiore coinvolgimento degli studenti (Topping, 2018); la promozione del pensiero critico (Novakovich, 2016); l'incremento della motivazione ad apprendere (Topping, 2018); lo sviluppo della competenza metacognitiva (Henderson et al., 2019); il potenziamento della capacità di autoregolare il proprio apprendimento (Nicol & Macfarlane-Dick, 2006); una più profonda comprensione (Nicol, Thomson, & Breslin, 2014); la promozione dell'apprendimento permanente, aiutando gli studenti a valutare realisticamente i loro risultati (Lu & Law, 2012). I feedback ricevuti dai pari, infatti, aiutano gli studenti a rivedere e a migliorare il proprio lavoro, consentendo di identificare i propri punti di forza e di debolezza, sviluppare competenze metacognitive e altre abilità trasferibili (Cho & MacArthur, 2010).

Numerose ricerche hanno anche messo in evidenza che gli studenti apprendono di più attraverso la produzione di feedback sul lavoro dei pari, che non ricevendone da questi ultimi (Cho & MacArthur, 2010; Foschi et al., 2019; Nicol et al., 2014). In questo modo gli studenti maturano una delle competenze trasversali più importanti per la loro vita personale e professionale: la capacità di sviluppare giudizi, di utilizzare criteri valutativi e di crearne di propri (Panadero, Broadbent, Boud, & Lodge, 2019).

Al fine di massimizzare l'impatto del feedback sull'apprendimento degli studenti, devono essere attenzionati alcuni aspetti fondamentali legati ad esempio al ruolo dei soggetti coinvolti nel fornire e ricevere il feedback, il contenuto stesso dei feedback, il numero di volte in cui si richiede di fornire i feedback e ritornare sull'elaborato, la progettazione stessa dei compiti (Mercader et al., 2020). La comprensione di queste condizioni risulta quindi cruciale per incidere positivamente sugli esiti di apprendimento degli studenti (Ion, Sánchez Martí, & Agud, 2019).

I processi di valutazione tra pari sono mediati da fattori relativi a credenze, valori e motivazioni che gli studenti attribuiscono, in generale, alla valutazione, e, più nello specifico, alla formulazione e ricezione di feedback (Henderson et al., 2019). Gli studenti percepiscono un feedback come efficace quando è reso attraverso commenti facilmente utilizzabili e sufficientemente dettagliati (Mercader et al., 2020) e quando sono coinvolti nella scelta dei criteri valutativi (Biggs, 1999). Anche le esperienze pregresse, positive o negative, possono influenzare esperienze successive di valutazioni tra pari (Yu & Hu, 2017). Alcune ricerche mettono poi in luce l'impatto positivo della guida dell'insegnante durante la valutazione tra pari (Zhu & Carless, 2018) e il ruolo di una formazione specifica sul feedback nell'influenzare la produzione di feedback specifici e costruttivi e, indirettamente, incidere anche sull'apprendimento significativo (Filius et al., 2018; Topping, 2018).

L'impatto del feedback sugli esiti degli studenti è legato anche a fattori connessi al curriculum, alla progettazione e alla valutazione delle attività didattiche (Henderson et al., 2019; Mercader et al., 2020). In una meta-analisi, Evans (2013) nota che l'efficacia del feedback dipende molto dalla forza e dalla coerenza del progetto di valutazione complessivo. Inoltre, Boud (2000) raccomanda che il feedback tra pari faccia parte del processo di insegnamento e apprendimento, mentre Filius et al. (2018) sottolineano la necessità di fornire abbastanza tempo per completare il processo di revisione tra pari.

3. Peergrade: una tecnologia per la valutazione tra pari

È interessante ricordare che da quando la tecnologia è stata ampiamente utilizzata all'Università, nel decennio iniziale di questo nuovo millennio, i test oggettivi a scelta multipla sono tornati di moda come formati di valutazione rapidi ed efficaci (Laici & Pentucci, 2019).

Negli ultimi anni sono state avanzate proposte innovative per utilizzare la tecnologia nei processi di valutazione che sono più avanzati (Davies, 2009; Liu & Li, 2014), dettati dalla necessità di progettare strumenti di valutazione che richiedono agli studenti di intraprendere compiti di natura complessa e che, pertanto, presuppongono tecniche e strumenti di valutazione altrettanto complessi (Boud, 2000). L'uso delle tecnologie digitali a sostegno della valutazione tra pari consente di gestire con flessibilità i tempi e i numeri elevati di studenti (Filius et al., 2018).

Nel panorama attuale delle tecnologie disponibili in questo ambito, Peergrade, il cui uso è semplice e intuitivo, costituisce una valida possibilità. Si tratta di una piattaforma online gratuita per consentire le sessioni di feedback tra studenti, progettata per facilitare ai docenti l'impostazione e la gestione di una valutazione tra pari e in cui gli studenti vengono valutati dai pari anche per la loro capacità di fornire dei feedback precisi e pertinenti.

Più in dettaglio, Peergrade integra vari strumenti, ad esempio la realizzazione e condivisione di rubriche di valutazione, il sistema di comunicazione tra pari, lo strumento di assegnazione e consegna di compiti e il sistema di valutazione formativa e sommativa.

Il docente crea una rubrica di valutazione che, a seconda dei casi e degli obiettivi, può essere predisposta con la collaborazione dei discenti.

La rubrica viene condivisa su Peergrade e fornisce agli studenti i termini e i criteri del dibattito valutativo.

L'assegnazione del compito prevede le seguenti possibilità:

- definizione data di inizio e consegna dei lavori;
- indicazione della data iniziale e finale della valutazione tra pari;
- descrizione dell'attività/compito da svolgere anche tramite un file allegato;
- indicazione del numero di lavori che ciascuno studente dovrà valutare.

Dopo l'avvenuta consegna dei lavori, Peergrade provvederà a distribuirli in modo casuale agli studenti, in modo che questi possano valutare il lavoro dei colleghi garantendone l'anonimato. I feedback vengono forniti all'autore del lavoro, dando vita a un dialogo di ordine metacognitivo e valutativo, sollecitandoli a rendere conto del loro operato e delle loro scelte. Inoltre, il processo di valutazione tra pari prevede un'ulteriore attività specificamente dedicata alla revisione delle valutazioni da parte dei valutati. La dimensione dialogica del processo valutativo emerge in particolare in quest'ultima fase, volta a generare un confronto, richiedendo ai valutati di esprimere le proprie considerazioni in merito alle valutazioni e ai feedback ricevuti. Viene infine fornita la possibilità di commentare con un testo libero. In questa fase di revisione delle valutazioni e dei feedback si concretizza maggiormente la dimensione formativa del processo valutativo: essere consapevoli che la valutazione sarà oggetto di confronto è uno stimolo a fornire feedback più accurati e produttivi.

Tale dibattito e i dati statistici forniti da Peergrade, rendono trasparente al docente il processo di apprendimento e di valutazione, fornendogli dati e informazioni che altrimenti sarebbero difficilmente esplicitabili e consentendogli al contempo di verificare attendibilità e validità dei criteri di valutazione su cui è costruita la rubrica. Peergrade infatti calcola un *Feedback Score* che responsabilizza ulteriormente i valutatori: chi partecipa a un processo di valutazione tra pari riceve una valutazione non solo per i propri elaborati, ma anche per la capacità di fornire dei feedback precisi e pertinenti.

L'efficacia della valutazione tra pari potrebbe essere ostacolata da problemi di parzialità (gli amici danno buoni voti agli amici) a causa della pressione dei pari o della paura di ritorsioni. Per questo motivo Peergrade indirizza distribuendo casualmente e in modo anonimo le valutazioni agli studenti per la valutazione e consente loro di contestare il feedback con cui non sono d'accordo, richiedendo la revisione e la rivalutazione da parte del loro docente.

4. Metodologia

4.1. Il campione

L'uso della piattaforma Peergrade è stato sperimentato con 203 studenti tramite l'elaborazione di un compito rivolto agli studenti di Pedagogia sperimentale del terzo anno

del corso di laurea in Scienze dell'Educazione e della Formazione dell'Università di Catania (89) e del secondo anno del corso di laurea in Scienze della Formazione Primaria dell'Università degli Studi di Palermo (114).

4.2. L'impianto didattico e valutativo

I corsi di Pedagogia Sperimentale dei due atenei, pur afferendo a corsi di laurea differenti, erano caratterizzati dallo stesso Syllabus, sia in termini di contenuti e *learning outcomes* che di testi di riferimento adottati. In entrambi i casi è stata utilizzata la piattaforma Peergrade ed è stata fornita la stessa consegna. Differivano tuttavia nell'ambito dell'approccio didattico utilizzato:

- *face-to-face* presso l'Università degli Studi di Palermo (Unipa), le attività sono state svolte durante la lezione, il docente stabiliva manualmente i tempi per la consegna dell'elaborato, per la valutazione dei pari e per il feedback sulla valutazione ricevuta;
- *a distanza* presso l'Università degli Studi di Catania (Unict), le attività sono state svolte nell'arco di una settimana tra una lezione e la successiva, gli studenti avevano quindi la possibilità di gestire con maggiore flessibilità e autonomia i tempi da dedicare a ciascuna fase della consegna.

4.3. La consegna

La consegna prevedeva la stesura di un abstract di massimo 400 parole a partire da un contributo di ricerca.

Gli studenti hanno dovuto revisionare tre elaborati anonimi dei colleghi secondo criteri valutativi condivisi, descritti nel sottoparagrafo che segue e, infine, esprimere un feedback sui giudizi ricevuti dai colleghi.

Gli stessi studenti svolgono dunque due ruoli dal momento che al contempo sono gli autori sia degli elaborati (abstract) che dei feedback forniti agli elaborati di altri colleghi di cui non conoscono il nome. Tutti gli studenti sono stati informati che avrebbero ricevuto dai loro colleghi dei feedback in forma anonima sulla base della rubrica condivisa.

4.4. Lo strumento

Il software consente l'elaborazione di uno strumento di valutazione condiviso nella piattaforma chiamata *rubric* che offre la possibilità di invitare i pari a riflettere su alcuni aspetti dell'elaborato tramite la possibilità di formulare varie tipologie di quesiti: a risposta aperta e a risposta multipla. La rubric, appositamente elaborata per la consegna, è costituita complessivamente da 114 quesiti raggruppati in tre sezioni (Figura 1):

- la sezione 1 riguarda due quesiti a risposta aperta relativi all'individuazione di punti di forza e agli eventuali aspetti che possono essere migliorati, esplicitandone le modalità;
- la sezione 2 è costituita da sei quesiti relativi alle modalità con cui sono state esplicitate le fasi principali di una ricerca empirica in ambito educativo. Le opzioni di risposta sono fornite su scala Likert a 5 punti;
- la sezione 3 è costituita da cinque quesiti che prevedono risposte dicotomiche (si/no) rispetto ad alcuni elementi che non dovrebbero essere presenti in un abstract ben formulato.

Sezione 1 (Aperte)

- Menziona cosa il tuo collega ha fatto bene spiegandone il perché.
- Menziona qualcosa che il tuo collega potrebbe migliorare. Come?

Sezione 2 – (Scala Likert)

- Come sono esplicitati la motivazione e gli obiettivi della ricerca?
- Come è eventualmente esplicitata l'ipotesi della ricerca?
- Eventualmente come è esplicitato il campione di ricerca?
- Com'è esplicitata la metodologia e gli strumenti di ricerca?
- Come sono esplicitati i risultati e la loro interpretazione?
- Complessivamente, come reputi la possibilità di comprendere la ricerca nella sua complessità (fasi cui si fa riferimento nei quesiti precedenti) senza entrare troppo nel dettaglio?

Sezione 3 (Dicotomica)

- Nell'abstract sono presenti riferimenti bibliografici?
- Nell'abstract sono presenti figure e/o tabelle?
- Nell'abstract sono riportate interpretazioni e/o considerazioni personali?
- Nell'abstract sono presenti informazioni aggiuntive rispetto a quanto riportato nella descrizione della ricerca (file)?
- È stato rispettato il limite delle 400 parole?

Figura 1. Articolazione della rubrica valutativa adottata su Peergrade.

4.5. La raccolta dei dati

La piattaforma Peergrade consente ai docenti di disporre di una grande quantità di dati sul processo di apprendimento sia rispetto ad uno specifico compito sia, più in generale, sull'andamento di un corso. I dati sono visualizzabili dai docenti in tempo reale con un'interfaccia grafica interattiva e intuitivamente comprensibile.

Inoltre i dati, sia relativi ad un singolo compito che relativi all'intero corso possono essere facilmente scaricati dalla sezione *Summary* in vario formato (.csv e .xlsx).

5. Analisi dei dati quantitativi

L'elaborazione dei dati è stata condotta attraverso il software IBM SPSS 25. Dall'analisi dei dati emerge che dei 203 studenti che hanno eseguito la consegna, 49 (24.14%) non hanno effettuato nessuna revisione del lavoro dei pari, 4 studenti (1.97%) ne hanno effettuata solo una, 2 studenti (0.99%) ne hanno effettuato 2, mentre 148 studenti (72.91%) hanno eseguito la consegna in tutte le sue parti, eseguendo dunque tre revisioni del contributo dei propri colleghi (Figura 2).

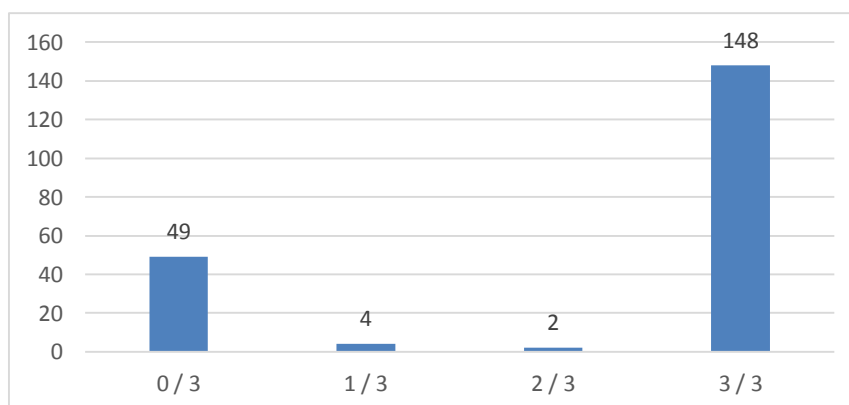


Figura 2. Frequenze degli studenti che hanno completato 0, 1, 2 o 3 revisioni tra pari.

Peergrade restituisce per ciascuno studente tre tipologie di punteggi:

- il *submission score* relativo al punteggio attribuito all'elaborato sulla base della rubric condivisa;
- il *feedback score* che consiste nel punteggio relativo alla competenza valutativa del soggetto nel valutare i pari;
- il *combined score* che è la combinazione dei due punteggi precedenti, in cui di default pesano ciascuno al 50%.

Si riportano in Figura 3 le statistiche descrittive relative ai primi due punteggi. Nella Figura 4 è invece riportata la distribuzione dei punteggi del Combined Score.

	<i>N</i>	<i>Minimo</i>	<i>Massimo</i>	<i>Media</i>	<i>Deviazione std.</i>
<i>Submission score</i>	180	57	97	81.37	7.274
<i>Feedback score</i>	70	25	100	74.36	19.869

Figura 3. Statistiche descrittive relative al Submission Score e Feedback Score del campione.

Come si nota facilmente dalla Figura 3, il punteggio relativo al Submission score risulta maggiore rispetto a quello di Feedback probabilmente per il fatto che raramente, se non addirittura mai, gli studenti, nella loro carriera scolastica ed accademica sono stati valutati per la loro competenza valutativa.

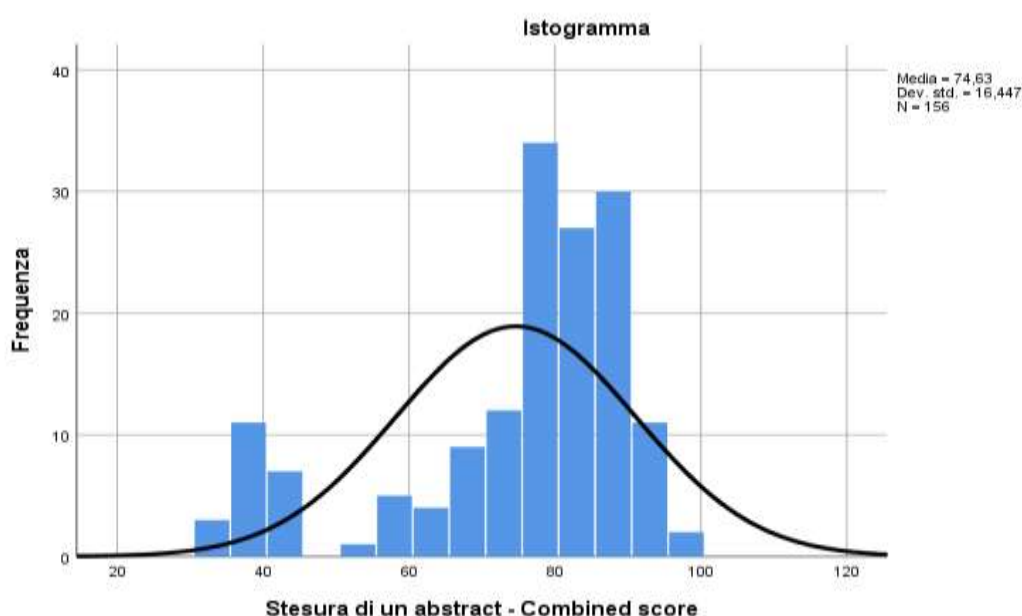


Figura 4. Distribuzione di frequenza del Submission score nel campione.

Avendo utilizzato lo stesso compito, con la stessa piattaforma all'interno dello stesso corso universitario ma con modalità differenti, eravamo interessati a capire se la metodologia adottata potesse avere delle ricadute sui due punteggi. Pur consapevoli del fatto che una tale ipotesi richiederebbe un impianto sperimentale ben più rigoroso, le statistiche descrittive riportate in Figura 5 mettono in luce un punteggio medio nell'ateneo palermitano di ben 10 punti superiore a quello catanese.

Ateneo		<i>N</i>	<i>Media</i>	<i>Deviazione std.</i>	<i>Media errore standard</i>
<i>Combined score</i>	Unict	46	67.33	21.351	3.148
	Unipa	110	77.69	12.825	1.223

Figura 5. Statistiche descrittive del combined score nei due sub-campioni.

Le differenze dei punteggi tra i due sottogruppi del campione risultano statisticamente significative sulla base del test ANOVA condotto sul *combined score*: $F(1.154)=3.736$ $p<0.001$. Ulteriori ricerche che tengano conto anche dei livelli di partenza degli studenti che frequentano i due corsi, sarebbero tuttavia necessarie per ipotizzare una differenza riconducibile alla metodologia applicata e quindi fornire alcune indicazioni più specifiche per la pratica didattica universitaria.

6. Analisi dei dati qualitativi

Le analisi qualitative sono state condotte con feedback forniti dagli studenti rispetto alle due domande aperte della sezione 1 della rubric sulla piattaforma Peergrade (Figura 1). Esse si riferiscono essenzialmente all'individuazione dei punti di forza e di debolezza dell'elaborato e all'individuazione di eventuali suggerimenti di miglioramento.

6.1. Il processo di codifica

Per la codifica dei segmenti sono state adottate le classificazioni esplicitate da Hattie e Timperley (2007) relative a feedback sul compito, feedback sui processi attivati nella risoluzione del compito, feedback per l'autoregolazione, feedback sul sé in quanto persona. La categoria riguardante il feedback sul compito è stata poi ulteriormente codificata secondo la classificazione individuata da Cho, Schunn, e Wilson (2006): direttivo, non direttivo, apprezzamento, critiche, riassunto e off-task. I codici utilizzati e una loro breve descrizione sono riportati in Figura 6.

<i>TIPOLOGIA DI FEEDBACK</i>	<i>DEFINIZIONE</i>
<i>SUL SE</i>	Espressione di un giudizio positivo (o in alcuni casi anche negativo) sul prodotto o sull'autore; non offre alcuna informazione sul compito e quindi raramente genera ulteriore coinvolgimento e azioni da parte degli studenti.
<i>PER L'AUTOREGOLAZIONE</i>	Incoraggia gli studenti a confrontare e utilizzare l'informazione ricevuta, generando una propria riflessione; consente di re-intervenire sulla situazione.
<i>SUI PROCESSI</i>	Esplicita la relazione tra i vari elementi dell'elaborato; riguarda anche le strategie utilizzate per risolvere un compito e la correzione di strategie fallimentari con altre più efficaci per raggiungere l'obiettivo.
<i>SUL COMPITO</i>	Riguarda le modalità in cui è stata svolta l'attività prevista, distinguendo le risposte giuste da quelle incorrette ed eventualmente evidenziando possibili soluzioni che possono essere intraprese.
<i>Direttivo</i>	Suggerisce uno specifico cambiamento all'autore.
<i>Non direttivo</i>	Suggerisce un cambiamento non specifico che potrebbe essere applicato a qualsiasi elaborato; commenta un dettaglio senza suggerire un cambiamento.

<i>Apprezzamento</i>	Descrive lo scritto o parte dello scritto positivamente, includendo note incoraggianti.
<i>Critica</i>	Fornisce critiche o valutazioni negative dello scritto o di parte dello scritto, mette in luce un'area non molto sviluppata, non fornisce suggerimenti per il miglioramento.
<i>Riassunto</i>	Riassume i punti principali dello scritto o di parte dello scritto.
<i>Off-task</i>	Commenti ambigui, non pertinenti o che non rientrano nelle precedenti categorie.

Figura 6. Descrizione della classificazione dei feedback utilizzati per la codifica dei dati.

Complessivamente sono stati analizzati 452 commenti forniti dagli studenti tramite il software statistico MAXQDA 2020. In totale sono stati codificati 1077 feedback, la cui lunghezza può variare da poche parole a diverse frasi. Tutti i segmenti sono stati codificati in maniera indipendente da due ricercatori sulla base delle caratteristiche dei feedback riportate in Figura 6. Per ciascuna tipologia di codice è stata calcolata la misura kappa di concordanza di Cohen che varia da moderata ad alta con valori compresi tra 0.53 e 0.96. È stata poi discussa l'attribuzione dei codici discordanti sino al raggiungimento di un accordo.

Come si evince dalla Figura 7 la maggior parte dei codici sono riferiti ai feedback degli studenti relativi al compito (85.6%). Le percentuali di frequenza relative dei due atenei non si discostano molto dalla percentuale totale.

	<i>Catania</i>		<i>Palermo</i>		<i>Totale</i>	
	N	%	N	%	N	%
SUL COMPITO	269	92.5	653	83.1	922	85.6
SUI PROCESSI	5	1.7	71	9.0	76	7.1
PER L'AUTOREGOLAZIONE	1	0.3	13	1.7	14	1.3
SUL SE	16	5.5	49	6.2	65	6.0
<i>Totale</i>	291	100.0	786	100.0	1077	100.0

Figura 7. Frequenze percentuali e nette relative al campione e ai due sub-campioni.

Nei sottoparagrafi che seguono saranno esaminati più nel dettaglio i dati relativi alle varie tipologie di feedback.

6.2. Feedback sul compito

I feedback relativi al compito costituiscono la categoria maggiormente codificata, dal momento che rappresentano l'85.6% delle codifiche totali.

	<i>Catania</i>		<i>Palermo</i>		<i>Totale</i>	
	N	%	N	%	N	%
Direttivo	42	15.6	197	30.2	239	25.9
Non direttivo	22	8.2	23	3.5	45	4.9
Apprezzamento	168	62.5	302	46.2	470	51.0
Critiche	23	8.5	66	10.1	89	9.7
Riassunto	2	0.7	28	4.3	30	3.2
Off-task	12	4.5	37	5.7	49	5.3
<i>Totale</i>	269	100.0	653	100.0	922	100.0

Figura 8. Frequenze percentuali relative al feedback sul compito nei due sub-campioni di studenti dell'Università di Catania e di Palermo.

Nei due sub-campioni le frequenze percentuali si discostano di quasi dieci punti percentuali (Figura 8) e dal test Kolmogorov-Smirnov a due campioni risulta che le differenze sono statisticamente significative: $D(181)=4.569$ $p<0.001$. Ne consegue dunque che gli studenti dell'Università di Catania forniscono feedback sul compito in maniera significativamente superiore rispetto ai colleghi dell'Università di Palermo.

Dal momento che i feedback di questa tipologia sono molto numerosi, si è reso necessario classificare ulteriormente tale categoria secondo quanto proposto da Cho et al. (2006).

- Feedback direttivi.

Il 25.9% dei feedback relativi al compito sono stati codificati come direttivi. Si tratta cioè di suggerimenti di specifici cambiamenti relativi all'elaborato. Questa tipologia di feedback è stata fornita in maniera significativamente maggiore, $D(181)=3.261$ $p<0.001$, nell'ambito dell'approccio didattico face-to-face (Unipa: 197/653; 30,2%) rispetto alla modalità a distanza (Unict: 42/269; 15,6%). Si tratta di commenti particolarmente validi dal punto di vista didattico, poiché se presi in considerazione dall'autore potrebbero condurre ad un miglioramento dell'elaborato ("Non ha spiegato in maniera chiara gli strumenti di ricerca e potrebbe migliorare questo aspetto, magari soffermandosi ed esplicitandoli" P22). In alcuni casi vengono anche esplicitate le ragioni per cui viene dato un determinato consiglio ("L'aspetto da migliorare è l'ordine di presentazione delle fasi e caratteristiche di una ricerca, ad esempio il campionamento utilizzato viene riportato soltanto alla fine. Inoltre viene trascurato l'elemento relativo all'interpretazione dei dati, aspetto che al contrario è nodale per la comprensione di una ricerca" P96).

Alcuni dei feedback direttivi (52/239; 21,8%) riguardano suggerimenti di natura prettamente linguistica: revisioni di errori grammaticali e/o di battitura, di aspetti lessicali e di organizzazione strutturale del testo ("Nella parte iniziale dell'abstract si potrebbe migliorare l'utilizzo della punteggiatura e la formulazione dei periodi, in quanto talvolta quest'ultimi risultano prolungati" P88).

Tuttavia si riscontra che l'11,3% dei feedback sul compito di natura direttiva riguardano suggerimenti errati, che rischierebbero dunque di far incorrere anche il collega in errore ("Era utile inserire qualche considerazione personale" C27). In questi casi potrebbe risultare particolarmente utile la funzione di *flag* prevista dal software Peergrade, attivabile da chi riceve il commento, che notifica al docente il problema intervenendo per chiarire eventuali controversie. Si tratta di una funzione che tuttavia è stata utilizzata complessivamente solo due volte.

- Feedback non direttivi.

Una piccola percentuale dei feedback sul compito (45/922; 4,9%) ha un carattere non direttivo, vale a dire che riguarda un'osservazione non specifica, soffermandosi su aspetti molto generali ("Avrebbe potuto approfondire uno degli argomenti un po' di più" C9). Nonostante questa tipologia di feedback sia stata riscontrata in percentuali più alte tra gli studenti catanesi (22/269; 8,2%) rispetto ai colleghi palermitani (23/653; 3,5%), le differenze tuttavia non risultano essere significative: $D(181)=0.841$ $p=0.479$.

- Feedback di apprezzamento.

Tra i feedback relativi al compito, nel 51% dei casi si tratta di feedback di apprezzamento, vale a dire di osservazioni incoraggianti che riguardano una parte o l'intero elaborato. Non si riscontrano differenze significative nella distribuzione di frequenze dei due atenei: $D(181)=0.364$ $p=0.921$.

Si tratta per lo più di apprezzamenti generali (298/470; 63.4%), in cui cioè non è possibile individuare elementi specifici riferiti al testo (“L’abstract riassume in modo chiaro e preciso il contesto della ricerca” C64) o in cui viene espressamente dichiarato che nell’elaborato non ci sia nulla che possa essere migliorato (“Secondo me è stato fatto un buon lavoro e non trovo punti di criticità da dover migliorare” P39). Questa tipologia di feedback, poco utile dal punto di vista didattico, è fornito in maniera significativamente maggiore dagli studenti catanesi (153/168; 91.1%) rispetto ai colleghi palermitani (145/302; 48.0%): $D(181)=2.276$ $p<0.001$.

In percentuale nettamente inferiore si riscontrano feedback di apprezzamento specifici (118/470; 25.1%), in cui è possibile rilevare un riferimento preciso al testo o ad una parte di esso (“Ha espresso correttamente e in modo dettagliato soprattutto gli strumenti utilizzati nella ricerca e i vari punteggi del questionario che erano previsti” C56) e in cui in alcuni casi ne vengono esplicitate le ragioni (“Il collega ha inserito un’adeguata parte introduttiva, in cui sono presentati l’argomento, i momenti e gli strumenti della ricerca, in modo tale da fornire le informazioni principali della ricerca ed invogliare il lettore ad approfondire la lettura dei concetti dell’articolo in maniera più specifica” P16). Questa tipologia di feedback di apprezzamento è utilizzata in maniera significativamente maggiore nell’ateneo palermitano (109/302; 36.1%) rispetto a quello catanese (9/168; 5.4%): $D(181)=3.5$ $p<0.001$.

Il 9.1% (43/470) dei feedback di apprezzamento riguarda aspetti linguistici, quali ad esempio la correttezza grammaticale (“Hai scritto in maniera grammaticalmente corretta” C15), l’adeguatezza lessicale (“Ha utilizzato una terminologia esatta e coerente al testo di riferimento” P98), la chiarezza e scorrevolezza espositiva (“È stato chiaro e conciso” SP/Unict). Anche questo tipo di feedback è stato utilizzato in maniera significativamente maggiore, $D(181)=1.448$ $p=0.030$, nell’approccio didattico basato sull’interazione face-to-face (Unipa: 38/302; 12.6%) rispetto a quello a distanza (Unict: 5/168, 3%).

Infine, in una piccola percentuale di feedback (11/470 2.3% di cui 10/11 Unipa) viene apprezzato ciò che effettivamente costituisce un errore nel componimento di un elaborato (“L’elaborato del/la mio/a collega ha come punto di forza la rielaborazione personale del testo della ricerca” P73). Si tratta di uno dei rischi della valutazione tra pari, ma è anche un elemento che può essere colto in un’ottica formativa dal docente titolare del corso per rilevare delle misconcezioni e/o aspetti ancora non chiari e ritornare su di essi.

- Feedback critici.

Nel 9.7% dei casi, il feedback relativo al compito riguarda critiche, esprime cioè una valutazione negativa dell’elaborato o di parte di esso. Nonostante le frequenze percentuali relative a questa tipologia di feedback risultino leggermente superiori tra gli studenti dell’ateneo palermitano (66/653 pari a 10.1% a fronte di 23/269 pari al 8.5% tra gli studenti catanesi), le differenze tuttavia non risultano statisticamente significative: $D(181)=0.625$ $p=0.830$.

Più nel dettaglio, nel 67.4% dei casi vengono fornite chiare indicazioni per risalire in maniera precisa il problema evidenziato (“Manca l’inserimento dell’ipotesi di ricerca” C61) mettendo a volte in evidenza anche le cause che possono averlo generato (“Si è dedicato/a maggiormente a scrivere delle informazioni non necessarie, piuttosto che esplicitare in modo più dettagliato le fasi della ricerca” P29). Le differenze tra le percentuali di frequenza dei due atenei non risultano in quest’ambito statisticamente significative: $D(181)=0.762$ $p=0.607$.

Nel 12.4% dei casi i feedback critici evidenziano invece problemi di carattere generale, in cui dunque non è direttamente individuabile l'ambito dell'elaborato in cui è necessario intervenire al fine di migliorarlo ("Inoltre mancano alcuni concetti fondamentali e a loro discapito ne sono stati inseriti altri che io ritengo superflui" P88).

Una parte dei feedback critici (13.5%) riguarda aspetti linguistici di varia natura: ortografia, lessico, punteggiatura, organizzazione strutturale del testo ("Non viene usata in maniera opportuna la punteggiatura e i periodi sono troppo lunghi" P39).

Infine il 6.7% dei feedback critici evidenzia come problematici aspetti che in realtà sono corretti ("All'interno dell'abstract non è stato menzionato il nome dell'autore della ricerca" P45). Anche in questo caso la funzione di segnalazione fornita dal software potrebbe risultare utile per risolvere problematiche di questo genere (Foschi et al., 2019).

Le percentuali di frequenza di queste tipologie di feedback non si discostano molto tra i due atenei, difatti il test di Kolmogorov-Smirnov a due campioni non evidenzia differenze significative per nessuna delle sub-categorie².

- Feedback riassuntivo.

Una percentuale pari al 3.2% dei feedback complessivi relativi al compito consiste semplicemente nel ricapitolare i punti principali dell'elaborato o di una porzione di esso ("Il collega si è focalizzato bene sulle parti più importanti della ricerca infatti, dopo aver presentato il tipo di ricerca, ha spiegato quali sono gli obiettivi della ricerca e chi è la popolazione di riferimento. Dopo di che, ha descritto la procedura di valutazione della ricerca e i dati emersi dalla valutazione" P37). Per tale tipologia di feedback non si riscontrano differenze significative tra i due sub-campioni, $D(181)=0.950$ $p=0.328$, pur essendo frequente con percentuali differenti tra gli studenti catanesi e palermitani (rispettivamente 2/269 pari a 0.7% e 30/922 pari a 4.3%).

- Off-task.

Infine una percentuale pari al 5.3% dei feedback sul compito sono completamente fuori tema, o perché poco chiari ("Potrebbe essere migliorata l'integrità del testo" P42) o perché non pertinenti ("Semplicemente avrebbe potuto lasciare maggiore spazio" C73). Nel campione catanese, inoltre, si riscontra una percentuale pari al 66.7% (8/12) dei feedback di questa categoria costituiti essenzialmente da segni di punteggiatura, utilizzati perché il software richiedeva l'inserimento di almeno un carattere per poter concludere la consegna. Ciò induce a supporre che gli studenti abbiano dedicato meno tempo e importanza alla formulazione dei feedback. Complessivamente non si riscontrano differenze significative tra le frequenze di questa tipologia di feedback nei due sub-campioni: $D(181)=0.475$ $p=0.978$.

6.3. Feedback sui processi

I feedback sui processi riguardano la relazione tra le varie componenti dell'elaborato, così come le strategie utilizzate per risolvere il compito e la correzione di eventuali strategie fallimentari con altre più efficaci per raggiungere l'obiettivo ("Penso che in questo abstract la caratteristica migliore è la forma schematica. Infatti leggendo ho avuto l'impressione che il riassunto fosse diviso in sezioni, in parti che mettono in risalto le parti salienti della

² Più nello specifico: $D(181)=0.164$ $p=0.998$ nel caso di feedback critici di carattere generale; $D(181)=0.505$ $p=0.961$ nel caso di feedback critici di carattere linguistico; $D(181)=0.206$ $p=0.999$ nel caso di feedback critici che di fatto sono corretti.

ricerca” P6; “Personalmente avrei seguito le linee guida dell’esempio abstract riportato sul libro” C54; “Penso che questo sia dovuto a una lettura troppo veloce del testo o magari a una scelta poco accurata delle informazioni da inserire e ciò ha portato, qualche volta, a trattare gli argomenti a metà” C65; “Potrebbe, quindi, lavorare di più sulla comprensione (e di sintesi) mediante una lettura più attenta del testo, magari inserendo le fasi o gli scopi della ricerca all’interno di mappe concettuali e sviluppando capacità di concentrazione” P13).

L’analisi dei dati evidenzia notevoli differenze nella distribuzione di frequenza relativa all’utilizzo di questa tipologia di feedback nei due sub-campioni, dal momento che è stato rilevato solo in 5/2991 casi nell’ambito della didattica tradizionale (Unict, 1.7%) mentre si riscontra in 71/786 nell’ambito dell’interazione face-to-face (Unipa, 9%). Il test Kolmogorov-Smirnov a due campioni avvalorava la significatività di tale differenza: $D(181)=2.523$ $p<0.001$.

6.4. Feedback per l’autoregolazione

Dall’analisi dei dati emerge una bassa percentuale, pari a 1.6%, di feedback per l’autoregolazione degli apprendimenti. Si tratta di feedback che incoraggiano gli studenti all’utilizzo delle informazioni e che possono contribuire a generare una riflessione che potrebbe indurre l’individuazione di informazioni importanti, il confronto con le proprie e il re-intervento sulla situazione (“Per migliorare questi aspetti dovrebbe rileggere il testo in modo tale da accorgersi degli errori commessi per correggerli e mettersi nei panni del lettore medio che non sa nulla sull’argomento, in modo tale da capire cosa deve specificare o meno per rendere facilmente comprensibile il testo” P93).

Con la sola eccezione di un caso, questa tipologia di feedback è quasi esclusivamente utilizzata dagli studenti dell’ateneo palermitano (16/17). Si potrebbe quindi ipotizzare che ci possa essere un legame con la metodologia didattica utilizzata, face-to-face, anche se ciò meriterebbe ulteriori approfondimenti.

6.5. Feedback sul sé

L’analisi dei dati evidenzia che, rispetto al totale dei feedback forniti, una percentuale pari al 6.0% riguarda i feedback sul sé, vale a dire espressioni di giudizio positivo, o in alcuni casi anche negativo, sull’autore (“Il mio collega presenta ottime capacità riassuntive e buone competenze linguistiche” P62).

La letteratura mette in luce che, dal momento che tali tipologie di feedback non offrono alcuna informazione sul compito, risultano poco utili dal punto di vista didattico, poiché raramente generano ulteriore coinvolgimento e azione da parte degli studenti (Hattie & Timperley, 2007).

Nonostante tale tipologia di feedback sia leggermente più frequente nel campione palermitano (49/786; 6.2%) rispetto a quello catanese (16/291; 5.5%), le differenze non risultano tuttavia essere significative: $D(181)=0.822$ $p=0.509$.

7. Discussione

L'analisi dei dati ha consentito di mettere in luce che l'introduzione di una nuova tecnologia per la valutazione tra pari richiede contestualmente anche un ripensamento della didattica tradizionale.

I dati quantitativi, ricavabili direttamente da Peergrade, evidenziano punteggi, relativi sia al compito che alla capacità di fornire feedback, significativamente maggiori quanto tale tecnologia è utilizzata in presenza durante le lezioni rispetto a quando è utilizzata a distanza.

L'analisi qualitativa ha consentito poi di esaminare più nel dettaglio le tipologie di feedback fornite dagli studenti nei due contesti. Sulla base della classificazione delle tipologie di feedback operata da Hattie e Timperley (2007), i dati mettono in evidenza il prevalere di feedback sul compito (85.6%), rispetto a quelli relativi ai processi (7.1%), all'autoregolazione (1.3%) e al sé (6%). Nel sub-campione palermitano, nel quale il compito è stato eseguito durante la lezione, si riscontrano frequenze significativamente maggiori di feedback relativi ai processi e all'autoregolazione. Nel sub-campione catanese, nel quale la consegna è stata svolta a casa tra una lezione e la successiva, invece si riscontrano frequenze significativamente maggiori di feedback sul compito. Non si evidenziano invece differenze significative nelle frequenze dei feedback relativi al sé.

L'elevata quantità di feedback relativi al compito ha reso necessario uno sguardo più approfondito su tale categoria. L'ulteriore codifica dei feedback sul compito secondo la classificazione adoperata da Cho et al. (2006), ha messo in luce il prevalere in maniera significativamente maggiore dei feedback direttivi e di apprezzamento specifici e linguistici nell'ambito dell'utilizzo del software con modalità face-to-face. Si tratta di feedback particolarmente utili per attivare il coinvolgimento degli studenti e condurli al miglioramento del compito (Coggi & Ricchiardi, 2018; Mercader et al., 2020). Al contrario, nel caso di svolgimento del compito a distanza tra una settimana e la successiva, si evidenziano frequenze significativamente maggiori di feedback di apprezzamento generale e non direttivi.

Pur consapevoli dei limiti legati alla numerosità del campione e alla necessità di un disegno di ricerca sperimentale più rigoroso, l'analisi dei dati quantitativi e qualitativi induce a concludere che l'introduzione di nuove tecnologie per la valutazione tra pari sia più efficace nell'ambito dell'interazione didattica face-to-face. Ciò è in linea con quanto emerge dalla letteratura (Zhu & Carless, 2018) secondo cui il docente può svolgere un importante ruolo di guida durante lo svolgimento del compito, chiarendo eventuali dubbi, ma anche scandendo i tempi delle varie fasi della consegna ed evitare che gli studenti dedichino ad esempio poco tempo alla formulazione dei feedback, come probabilmente è avvenuto nell'ambito dell'interazione a distanza.

8. Conclusioni

Il presente contributo offre nuovi spunti di riflessione sulla necessità di progettare adeguatamente le metodologie didattiche con le quali introdurre nuove tecnologie per la valutazione tra i pari nella didattica universitaria. Infatti, sebbene recentemente vi sia stato un crescente interesse di ricerca rispetto all'utilizzo del feedback e della valutazione tra pari anche nell'ambito della didattica universitaria (Li, 2019; Topping, 2018), non si è riscontrata altrettanta attenzione alle metodologie didattiche attraverso le quali

implementarle per ottenere feedback di qualità e, in prospettiva, anche migliorare gli esiti di apprendimento (Mercader et al., 2020). Quest'ultimo aspetto, in particolare, potrebbe essere approfondito in ulteriori ricerche che potrebbero prendere anche in considerazione, triangolandole, la valutazione tra pari, l'autovalutazione dello studente e la valutazione del docente. Potrebbero anche essere indagati i benefici della produzione di feedback sul piano metacognitivo e sulle percezioni degli studenti rispetto alla valutazione tra pari.

Riferimenti bibliografici

- Balzaretti, N., & Vannini, I. (2018). Promuovere la qualità della didattica universitaria. La Formative Educational Evaluation in uno studio pilota dell'Ateneo bolognese. *Journal of Educational, Cultural and Psychological Studies*, 18, 187–213.
- Bartram, B., & Bailey, C. (2010). Assessment preferences: a comparison of UK/international students at an English university. *Research in Post-Compulsory Education*, 15(2), 177–187.
- Biggs, J. (1999). *Teaching for Quality Learning at University*. Buckingham: SRHE and Open University Press.
- Binetti, P., & Cinque, M. (2016). *Valutare l'Università & Valutare in Università. Per una "cultura della valutazione"*. Milano: FrancoAngeli.
- Boud, D. (2000). Sustainable assessment: rethinking assessment for the learning society. *Studies in continuing education*, 22(2), 151–167.
- Carless, D., & Boud, D. (2018). The Development of Student Feedback Literacy: Enabling Uptake of Feedback. *Assessment & Evaluation in Higher Education*, 43(8), 1315–1325.
- Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and instruction*, 20(4), 328–338.
- Cho, K., Schunn, C. D., & Wilson, R. W. (2006). Validity and Reliability of Scaffolded Peer Assessment of Writing from Instructor and Student Perspectives. *Journal of Educational Psychology*, 98(4), 891–901.
- Coggi, C. (Ed.). (2005). *Domande di qualità: le istanze degli studenti universitari*. Lecce: Pensa MultiMedia.
- Coggi, C. (Ed.). (2019). *Innovare la didattica e la valutazione in Università: Il progetto IRIDI per la formazione dei docenti*. Milano: FrancoAngeli.
- Coggi, C., & Ricchiardi, P. (2018). Developing effective teaching in Higher Education. *Form@re - Open Journal Per La Formazione in Rete*, 18(1), 23–38.
- Davies, P. (2009). Review and reward within the computerised peer- assessment of essays. *Assessment & evaluation in higher education*, 34(3), 321–333.
- Evans, C. (2013). Making sense of assessment feedback in higher education. *Review of educational research*, 83(1), 70–120.
- Filius, R. M., de Kleijn, R. A., Uijl, S. G., Prins, F. J., van Rijen, H. V., & Grobbee, D. E. (2018). Strengthening dialogic peer feedback aiming for deep learning in SPOCs. *Computers & Education*, 125, 86–100.

- Foschi, L., Cecchinato, G., & Say, F. (2019). Quis iudicabit ipsos iudices? Analysis of skill development in a teacher training course through peer-and self-assessment. *Italian Journal of Educational Technology*, 27(1), 49–64.
- Grion, V., & Serbati, A. (2017). *Assessment for Learning in Higher Education. Nuove prospettive e pratiche di valutazione all'università*. Lecce: Pensa MultiMedia.
- Grion, V., Serbati, A., Tino, C., & Nicol, D. (2018). Ripensare la teoria della valutazione e dell'apprendimento all'università: un modello per implementare pratiche di peer review. *Italian Journal of Educational Research*, 19, 209–226.
- Hattie, J., & Timperley, H. (2007). The Power of Feedback. *Review of Educational Research*, 77(1), 81–112.
- Henderson, M., Ajjawi, R., Boud, D., & Molloy, E. (Eds.). (2019). *The Impact of Feedback in Higher Education: Improving Assessment Outcomes for Learners*. Cham: Palgrave Macmillan.
- Ion, G., Sánchez Martí, A., & Agud Morell, I. (2019). Giving or receiving feedback: which is more beneficial to students' learning? *Assessment & Evaluation in Higher Education*, 44(1), 124–138.
- Irons, A. (2007). *Enhancing learning through formative assessment and feedback*. New York, NY: Routledge.
- Laici, C., & Pentucci, M. (2019). Il feedback con le tecnologie all'università: una rassegna sistematica. *Form@re - Open Journal Per La Formazione in Rete*, 19(3), 6–25.
- Li, L. (2019). ENGAGE Students in formative peer assessment to support learning. *Italian Journal of Educational Research*, 12, 63–70.
- Li, L., Liu, X., & Zhou, Y. (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.
- Liu, X., & Li, L. (2014). Assessment training effects on student assessment skills and task performance in a technology-facilitated peer assessment. *Assessment & Evaluation in Higher Education*, 39(3), 275–292.
- Lu, J., & Law, N. (2012). Online peer assessment: Effects of cognitive and affective feedback. *Instructional Science*, 40(2), 257–275.
- Mercader, C., Ion, G., & Díaz-Vicario, A. (2020). Factors influencing students' peer feedback uptake: instructional design matters. *Assessment & Evaluation in Higher Education*. <https://doi.org/10.1080/02602938.2020.1726283> (23.03.2020).
- Nicol, D. (2018). Unlocking generative feedback through peer reviewing. In V. Grion & A. Serbati (Eds.), *Valutare l'apprendimento o valutare per l'apprendimento? Verso una cultura della valutazione sostenibile all'Università* (pp. 47-59). Lecce: Pensa Multimedia.
- Nicol, D. J., & MacFarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, 31, 199–218.
- Nicol, D., Thomson, A., & Breslin, C. (2014) Rethinking feedback practices in higher education: a peer review perspective. *Assessment & Evaluation in Higher Education*, 39(1), 102–122.

- Novakovich, J. (2016). Fostering critical thinking and reflection through blog- mediated peer feedback. *Journal of Computer Assisted Learning*, 32(1), 16–30.
- Panadero, E., Broadbent, J., Boud, D., & Lodge, J. M. (2019). Using formative assessment to influence self-and co-regulated learning: the role of evaluative judgement. *European Journal of Psychology of Education*, 34(3), 535–557.
- Pandolfi, L. (2017). La valutazione della didattica universitaria in prospettiva formativa: linee evolutive e partecipazione degli studenti. *Excellence and innovation in learning and teaching*, 2, 34–50.
- Peergrade. <https://www.peergrade.io/> (ver. 23.03.2020).
- Struyven, K., Dochy, F., & Janssens, S. (2005). Students' perceptions about evaluation and assessment in higher education: a review. *Assessment & Evaluation in Higher Education*, 30(4), 325–341.
- Tonelli, D., Grion, V., & Serbati, A. (2018). L'efficace interazione fra valutazione e tecnologie: evidenze da una rassegna sistematica della letteratura. *Italian Journal of Educational Technology*, 26(3), 6–23.
- Topping, K. (2018). *Using peer assessment to inspire reflection and learning*. New York, NY: Routledge.
- Vertecchi, B. (1976). *Valutazione formativa*. Torino: Loescher.
- Yu, S., & Hu, G. (2017). Understanding university students' peer feedback practices in EFL writing: Insights from a case study. *Assessing Writing*, 33, 25–35.
- Zhu, Q., & Carless, D. (2018). Dialogue within peer feedback processes: Clarification and negotiation of meaning. *Higher Education Research & Development*, 37(4), 883–897.