Italian Journal of Educational Technology

27
VOLUME

Formerly published as:
TD Tecnologie Didattiche

1
ISSUE

| ISSN (print) 2532-4632
| ISSN (online) 2532-7720

2019

Consiglio Nazionale delle Ricerche

Istituto per le Tecnologie Didattiche

Italian Journal of Educational Technology / Volume 27 / Issue 1 / 2019

Quis iudicabit ipsos iudices[1]? Analisi dello sviluppo di competenze in un percorso di formazione per insegnanti tramite la valutazione tra pari e l'autovalutazione

Quis iudicabit ipsos iudices?[2] analysis of skill development in a teacher training course through peer- and self-assessment

Laura Carlotta Foschi*, Graziano Cecchinato and Filippo Say

Department of Philosophy, Sociology, Education and Applied Psychology (FISPPA), University of Padova (UNIPD), lauracarlotta.foschi@phd.unipd.it*, graziano.cecchinato@unipd.it, filippo.say@gmail.com

*corresponding author

HOW TO CITE Foschi, L.C., Cecchinato, G., & Say, F. (2019). Quis iudicabit ipsos iudices? Analisi dello sviluppo di competenze in un percorso di formazione per insegnanti tramite la valutazione tra pari e l'autovalutazione. *Italian Journal of Educational Technology*, *27*(1), 49-64. doi: 10.17471/2499-4324/1019

SOMMARIO Lo studio descritto nel presente contributo ha preso in esame il grado di concordanza fra le valutazioni tra pari, le autovalutazioni e le valutazioni dei docenti del corso in merito a unità di apprendimento realizzate dagli insegnanti in servizio coinvolti in un percorso di formazione. Questo confronto ha consentito di indagare, da un lato, lo sviluppo di competenze valutative nell'ambito dell'approccio metodologico proposto con l'intervento di formazione e, dall'altro, la validità di queste procedure di valutazione in un ambito nel quale sono poco esplorate, come quello della formazione degli insegnanti in servizio. Vengono illustrati: l'oggetto, le metodologie didattiche e le finalità dell'intervento formativo; le modalità di conduzione dei processi di valutazione tra pari, autovalutazione e valutazione dei docenti e le potenzialità degli strumenti digitali adottati per tali processi (Peergrade); le analisi statistiche condotte e i risultati ottenuti. Da questi ultimi emergono buoni e ottimi gradi di concordanza fra le valutazioni degli insegnanti e quelle dei docenti del corso, le cui implicazioni vengono discusse.

PAROLE CHIAVE Valutazione tra Pari; Autovalutazione; Formazione degli Insegnanti in Servizio; Peergrade.

ABSTRACT This study examines the level of agreement between peer-assessments, self-assessments

[1] "Chi giudicherà i giudici?"

[2] "Who will judge the judges?"

and assessments from course trainers regarding lesson plans produced by in-service teachers involved in a training course. This comparison allows us to investigate the development of assessment skills in relation to the methodological approach proposed in the training course. In addition, it permits us to validate these assessment procedures in a field where they are little explored, namely in-service teacher training. In this paper, the following aspects are illustrated: the topic, teaching methods and aims of the training course; the peer- and self-assessment processes and potential of the digital tools adopted (Peergrade); the statistical analysis performed; and the results obtained. The results show substantial and almost perfect levels of agreement between the teachers' peer- and self-assessments, and those from their trainers, and the implications of these outcomes are discussed.

KEYWORDS Peer-assessment; Self-assessment; In-Service Teacher Training; Peergrade.

1. INTRODUZIONE

Gli Autori di questo contributo promuovono interventi di formazione nelle scuole volti a condividere con gli insegnanti coinvolti l'approccio didattico della Flipped Classroom (Mazur, 1997; Baker, 2000; Lage, Platt, & Treglia, 2000). Questa metodologia integra l'uso di risorse e ambienti educativi digitali con l'adozione di pratiche di apprendimento attivo in classe (Bishop & Verleger, 2013; Keengwe, Onchwari, & Oigara, 2014; Cecchinato, 2014; Talbert, 2017) che, nella nostra impostazione, sono ispirate al *Challenge Based Learning* (Schwartz, Lin, Brophy, & Bransford, 1999; O'Mahony et al., 2012).

In sostanza, conformemente ai cambiamenti nei processi di apprendimento indotti dallo sviluppo dei nuovi media – l'interattività del web, l'immersività degli ambienti di realtà aumentata e virtuale, la narrazione emergente dei videogiochi e dei serious games (Gee, 2003; Jenkins, Purushotma, Weigel, Clinton, & Robison, 2009; Anolli & Mantovani, 2011) – si propone agli insegnanti coinvolti nell'attività di formazione di orientare la propria pratica didattica in modo da favorire un maggiore protagonismo dei discenti nel costruire attivamente i propri percorsi di conoscenza. Nello specifico, gli insegnanti in formazione sono sollecitati a progettare "sfide cognitive" con le quali impegnare i propri studenti nell'affrontare i contenuti disciplinari.

Attraverso l'attività di formazione si propone di sostituire il classico ciclo di insegnamento *Lezione – Studio – Verifica*, con uno basato anch'esso su tre fasi indicate con *Lancio – Conduzione – Chiusura* di una "Sfida" (Cecchinato & Papa, 2016). Gli insegnanti sono coinvolti in un percorso che alterna incontri in presenza e attività online per un periodo di 5 - 7 mesi durante i quali progettano e realizzano attività e risorse didattiche e che, come sviluppo finale, prevede la progettazione e la conduzione con la classe di *Unità di Apprendimento* (UdA) basate su questo approccio.

Si tratta di un processo complesso perché necessita, per ognuna delle tre fasi, di un significativo cambiamento concettuale. Nella fase del *Lancio della sfida* è richiesto il passaggio da un'impostazione conoscitiva di carattere deduttivo a una di carattere induttivo. Gli insegnanti sono chiamati ad abbandonare "l'enunciazione" dei contenuti disciplinari per adottare un approccio problematizzante, che presenta agli studenti casi concreti da affrontare e risolvere, derivando da questo processo le conoscenze teoriche. Devono perciò ideare e proporre alla classe, conformemente con il proprio ambito disciplinare, un problema concreto, una domanda su un tema controverso, un caso di studio che susciti interesse, curiosità intellettuale e quindi motivazione a intraprendere un percorso di apprendimento. Si tratta di un approccio che, pur avvalendosi di diverse strategie, fa leva su due costrutti fondanti (Cecchinato & Papa, 2018):

- la *dissonanza cognitiva* (Festinger, 1962) o il *conflitto cognitivo* (Piaget, 1974), cioè la produzione negli studenti di un'inconsistenza fra ciò che sanno e ciò che la sfida propone, inconsistenza

che dovrebbe far sorgere in loro l'esigenza di ricomporla;

- l'attivazione delle *conoscenze pregresse* (Ausubel, Novak, & Hanessian, 1968), che consiste nell'avviare il processo di apprendimento con strategie che stimolino gli studenti a mettere in campo quello che già conoscono, condizione essenziale per un *apprendimento significativo* (Novak, 2002; Mayer, 2002).

Nella fase della *Conduzione della sfida* è richiesto il passaggio da un'impostazione trasmissiva a una costruttivista. Dopo aver attivato, con il *Lancio della sfida*, la curiosità intellettuale (Berlyne, 1960) e la motivazione ad apprendere (Ryan & Deci, 2000), si avranno studenti impegnati ad affrontare la sfida proposta. Diviene a questo punto conseguente predisporre in classe un *setting* didattico orientato alle pratiche di apprendimento attivo. Gli studenti sono quindi chiamati a mettere in atto, sia pur con forme e modalità adeguate alla loro età e al contesto didattico, le strategie cognitive e le procedure di indagine proprie della disciplina che si sta affrontando. Si tratta di sostenere il valore del dubbio, la capacità di porre in discussione quanto si sta analizzando incoraggiando gli studenti a porsi domande appropriate, a formulare ipotesi attendibili, a escogitare metodi per verificare le proprie supposizioni, ad argomentare efficacemente le proprie asserzioni. Le metodologie alle quali far ricorso possono essere varie, attingendo alla quasi secolare tradizione dell'*active learning* (Dewey, 2007), ma anche alle pratiche originali che si producono grazie ai nuovi media (Jonassen, 2008).

La fase della *Chiusura della sfida* richiede l'adozione di un diverso paradigma valutativo, richiede cioè di spostarsi da pratiche di valutazione di carattere sommativo verso altre con valenza prevalentemente formativa (Carless, 2007; Sadler, 1989), passaggio in grado di produrre considerevoli benefici sull'apprendimento (Black & Wiliam, 1998). Lo svolgimento di pratiche di apprendimento attivo richiede una fase conclusiva di riflessione su e valutazione di quanto prodotto. Si tratta di attività che nella nostra proposta solitamente avvengono chiedendo agli studenti, singolarmente o in gruppo, di presentare alla classe i propri percorsi di apprendimento. Questi processi, che hanno lo scopo di condividere con la classe le conoscenze raggiunte, consentono allo stesso tempo di attuare pratiche valutative più significative rispetto alle tradizionali prove di verifica. Gli studenti non si relazionano con l'insegnante che deve certificare la corretta comprensione dei contenuti di studio, ma svolgono un'esposizione in favore dei propri compagni, mettendosi in gioco con i pari di riferimento, processo che dovrebbe risultare maggiormente significativo e coinvolgente, in particolare quando si attiva la valutazione tra pari e l'autovalutazione (Topping, 1998).

Progettare e condurre attività didattiche che si ispirano a questo approccio è significativamente sfidante per gli insegnanti coinvolti nel percorso di formazione, soprattutto per chi non ha in precedenza adottato pratiche ispirate al costruttivismo. Gli insegnanti pertanto vengono coinvolti, prima di procedere alla progettazione delle UdA, in processi finalizzati a ricostruire e definire, attraverso metodologie attive e con il nostro sostegno, le caratteristiche di una "buona" UdA, grazie ad attività che richiedono l'individuazione, fra molte, di quelle corrispondenti all'approccio proposto e allo studio di esempi concreti di "buone" UdA. Le UdA che gli insegnanti progettano al termine del percorso di formazione sono inserite in un processo che prevede la valutazione tra pari, l'autovalutazione e la valutazione da parte dei docenti del corso.

Lo studio presentato in questo articolo prende in esame questo processo di valutazione nell'ambito di un intervento di formazione rivolto a insegnanti in servizio in un Istituto Comprensivo. L'obiettivo è indagare se il percorso di formazione abbia contribuito a far maturare negli insegnanti coinvolti buone competenze valutative nell'ambito dell'approccio metodologico proposto, analizzando il grado di concordanza fra le valutazioni da loro stessi fornite (autovalutazione e valutazione tra pari) e le corrispondenti valutazioni dei docenti del corso.

Le sezioni seguenti descrivono l'approccio adottato nei processi di valutazione tra pari e autovalutazione, inquadrando le scelte metodologiche e tecnologiche nello stato dell'arte del settore. Segue la descrizione della ricerca, in termini di obiettivi, contesto, metodo, risultati, discussione e conclusioni.

2. VALUTAZIONE TRA PARI E AUTOVALUTAZIONE

L'autovalutazione fa riferimento a quell'insieme di modalità attraverso cui un individuo è chiamato a valutare la propria esperienza di apprendimento, il proprio apprendimento, le proprie prestazioni e a riconoscerne i significati personali attribuiti (Panadero, Brown, & Strijbos, 2016). L'autovalutazione può focalizzarsi su diversi aspetti; nel nostro studio abbiamo considerato il prodotto realizzato, ovvero le UdA progettate dagli insegnanti coinvolti nella formazione.

La valutazione tra pari fa riferimento invece a quell'insieme di modalità attraverso cui un individuo è chiamato a valutare e fornire *feedback* relativamente al lavoro dei propri pari, nonché a riceverne a sua volta da questi ultimi in merito al proprio lavoro (Nicol, Thomson, & Breslin, 2014).

In letteratura si sottolinea l'importanza dei *feedback* testuali di revisione (Nicol et al., 2014; Davies, 2006) e del coinvolgimento dei valutati nello sviluppo dei criteri di valutazione (Li et al., 2016). Nella ricerca in esame abbiamo però deciso di utilizzare procedure di valutazione quantitative basate su criteri da noi predisposti. Considerando la natura dell'attività in esame, il nostro target di riferimento, l'organizzazione del corso, nonché le indicazioni presenti in letteratura (e.g., Bloxham & West, 2004; Miller, 2003; Somervell, 1993), abbiamo ritenuto opportuno procedere in questo modo, pur lasciando la possibilità di esprimere anche *feedback* testuali in maniera facoltativa. Abbiamo infatti considerato che le problematiche che tipicamente si possono riscontrare in merito alla valutazione tra pari (Liu & Carless, 2006; Kaufman & Schunn, 2011; Sadler & Good, 2006; Magin, 2001; Falchikov, 1995) non si sarebbero verificate nel nostro contesto. Si tratta, infatti, di adulti e nello specifico di insegnanti abituati a valutare e che quindi difficilmente avrebbero sperimentato una riluttanza nel valutare gli altri; l'anonimato (sia del valutatore che del valutato) assicurato dall'ambiente online utilizzato avrebbe escluso il possibile "friendship bias"; l'assegnazione randomizzata dei valutatori e dei valutati, nonché il fatto che la valutazione non ha valore formale avrebbero mantenuto sotto controllo la pressione e la competizione talvolta indotta dalla valutazione tra pari. Inoltre, se da un lato vi sono indicazioni per l'informalità delle procedure di valutazione (e.g., Yan & Brown, 2017), dall'altro vi sono anche indicazioni per l'utilizzo di criteri di valutazione espliciti (e.g., Bloxham & West, 2004; Miller, 2003; Harris, 2011) ed è in questi secondi termini che abbiamo deciso di procedere. Per le finalità che ci siamo proposti abbiamo quindi ritenuto opportuno strutturare il processo di valutazione definendone con chiarezza i criteri, onde evitare ambiguità su quello che gli insegnanti sono chiamati a valutare e possibile autoreferenzialità, intesa come chiusura entro parametri di giudizio personali e poco condivisibili. La scelta di fornire espliciti criteri è stata effettuata anche in considerazione del fatto che non si trattava di valutare un'attività didattica consueta, bensì una del tutto nuova rispetto alle esperienze pregresse degli insegnanti. In linea con quanto appena messo in luce si rivelano utili anche gli *exemplars* (Sadler, 1987), ovvero gli esempi concreti di "buone" UdA, messi a disposizione degli insegnanti. Fornire modelli di riferimento risulta infatti particolarmente efficace nel chiarire gli obiettivi e i livelli di qualità richiesti, oltre ad offrire un valido standard col quale confrontare il proprio lavoro (Orsmond, Merry, & Reiling, 2002).

Richiedere agli insegnanti di procedere alla valutazione del lavoro altrui e del proprio risponde a specifiche motivazioni, prima fra tutte quella che questo tipo di attività è in grado di promuovere un apprendimento significativo e lo sviluppo di specifiche competenze nell'ambito dell'approccio metodologico proposto (Lynch, McNamara, & Seery, 2012; Sluijsmans, Brand-Gruwel, & van Merriënboer, 2002, Poon, McNaught, Lam, & Kwan, 2009). Abbiamo inoltre ritenuto importante far sperimentare agli insegnanti queste strategie di valutazione perché potessero valutarne le potenzialità per un uso in classe con i propri studenti. Si è più volte riscontrato, infatti, come esse aumentino il coinvolgimento (e.g., Bloxham & West, 2004; Brown & Harris, 2013), promuovano il pensiero critico (e.g., Lynch et al., 2012), incrementino la motivazione ad apprendere (e.g., Topping, 2005), sviluppino la metacognizione (e.g., Vickerman, 2009;

Wen & Tsai, 2006) e l'autoregolazione (e.g., Nicol & Macfarlane-Dick, 2006; Panadero, Tapia, & Huertas, 2012). Sperimentare in prima persona queste pratiche valutative, soprattutto nei percorsi di formazione per insegnanti, consente a questi ultimi di acquisire le competenze necessarie a utilizzarli in modo produttivo con i propri studenti (Yilmaz, 2017; Cheng, M. M. H, Cheng, A. Y. N., & Tang, 2010).

La prospettiva teorica dalla quale muoviamo, infine, sostiene che sia più importante valutare il lavoro dei pari, che non ricevere da questi valutazioni. Infatti, se in passato la ricerca ha messo soprattutto in luce come i *feedback* ricevuti dai pari possano aiutare gli studenti a rivedere e migliorare il proprio lavoro (e.g., Cho & MacArthur, 2010; Topping, 1998), la letteratura recente sottolinea come gli studenti apprendano di più attraverso la produzione di *feedback* sul lavoro dei pari, che non ricevendone da questi ultimi (e.g., Nicol et al., 2014; Cho & MacArthur, 2011). L'atto di revisionare, infatti, attiva un processo riflessivo con il quale gli studenti confrontano il proprio lavoro con quello dei pari e realizzano come possono migliorare il proprio. Revisionare, pertanto, non solamente migliora la *performance*, ma anche la capacità di autoregolare il proprio apprendimento, nonché favorisce una comprensione più profonda dell'oggetto di conoscenza (Nicol et al., 2014).

3.LA SCELTA DI PEERGRADE PER LA VALUTAZIONE TRA PARI E L'AUTOVALUTAZIONE

Per lo svolgimento delle attività online del corso è stato utilizzato Moodle ma, nonostante in questo ambiente siano presenti strumenti che consentono sia la valutazione tra pari che l'autovalutazione, come il *Workshop*, per questi processi ci si è avvalsi di Peergrade[3] in virtù delle sue specifiche caratteristiche. In questo ambiente, infatti, in linea con la letteratura (e.g., Nicol, 2010), sono implementate funzionalità avanzate che lo caratterizzano per concepire la valutazione non come un processo a senso unico, ma come un dialogo fra valutati e valutatori orientato al miglioramento dei processi di apprendimento. In particolare, in Peergrade, i processi di valutazione tra pari e di autovalutazione non si esauriscono con la condivisione delle valutazioni, come di norma avviene, ma prevedono 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, che è volta a generare confronto su tutte le valutazioni, richiedendo ai valutati di esprimere le proprie considerazioni in merito alle valutazioni e ai *feedback* ricevuti.

Una delle modalità con le quali si può produrre il confronto è attivando la funzione "Segnalazioni" (*Flags*), cioè la possibilità di contrassegnare specifiche valutazioni o *feedback* ricevuti (sia da un proprio pari che da un docente del corso) per richiedere chiarimenti in merito. Quando chi ha ricevuto una valutazione si avvale di questa funzionalità, viene inviata una notifica al docente del corso affinché intervenga in modo opportuno. Può per esempio modificare la valutazione (propria o del pari del valutato), oppure confermarla, motivando, in ogni caso, la propria scelta. Le "Segnalazioni" sono visibili al valutatore, così come le risposte del docente ai feedback contrassegnati sono visibili a valutatori e valutati. I valutati e i valutatori possono inoltre confrontarsi tra loro sulle segnalazioni (ricevute ed effettuate), nonché esprimere al docente, in risposta alla sua risoluzione di queste ultime, le proprie considerazioni in merito.

Al di là delle eventuali "Segnalazioni", la fase di revisione delle valutazioni prevede che i valutati esprimano, sulla base di sollecitazioni proposte da Peergrade stesso, le proprie considerazioni complessive sulle valutazioni e sui *feedback* ricevuti, considerazioni che saranno visibili ai rispettivi valutatori. In particolare i valutati sono chiamati a indicare quanto hanno trovato utili i *feedback* ricevuti, scegliendo tra 5 diversi descrittori. A questo proposito viene esplicitato che un *feedback* utile dovrebbe avere le seguenti caratteristiche: costruttività, specificità,

[3] https://www.peergrade.io/. Si tratta di una piattaforma online dedicata a condurre e coordinare le diverse fasi di valutazione tra pari e autovalutazione.

giustificazione e gentilezza. Queste caratteristiche sono in linea con quelle indicate in letteratura per valutare la "bontà" di un *feedback* (e.g., Hattie, 2016). Viene infine fornita la possibilità di commentare con un testo libero. È quindi in questa fase di revisione delle valutazioni e dei *feedback* che si concretizza maggiormente la dimensione formativa del processo valutativo: essere consapevoli che la valutazione sarà oggetto di confronto agisce come stimolo a fornire *feedback* più accurati e produttivi; l'analisi, la valutazione e la possibilità di segnalare *feedback* e valutazioni richiede riflessione ed elaborazione da parte dei valutati; il confronto che segue produce una negoziazione di significati che favorisce l'apprendimento (Nicol, 2010; Price, Handley, & Millar, 2011; Carless et al., 2011).

Inoltre, chi partecipa a un processo di valutazione tra pari riceve una valutazione non solo per i propri elaborati, ma anche per le valutazioni fornite. Peergrade calcola infatti un "*Feedback Score*" che responsabilizza ulteriormente i valutatori.

Un altro aspetto da considerare è che, nella fase di settaggio delle impostazioni di Peergrade, viene espressamente consigliato di stabilire che ogni partecipante debba valutare almeno 3 elaborati dei pari. Questo aspetto è rilevante perché la letteratura ha evidenziato (e.g., Cho & MacArthur, 2010) che ricevere *feedback* da diversi pari costituisce un fattore in grado di migliorare la qualità dei propri elaborati in misura maggiore rispetto a quando si ricevono feedback da un solo pari. Infine, Peergrade prevede che l'autovalutazione del proprio elaborato avvenga solo dopo aver valutato quelli dei propri pari. È importante sottolineare questo aspetto poiché la letteratura ha messo in luce come gli studenti apprendano di più attraverso la produzione di *feedback* sul lavoro dei pari, che non ricevendone da questi ultimi (e.g., Nicol et al., 2014; Cho & MacArthur, 2011); ne consegue che anteponendo la valutazione dei pari all'autovalutazione, quest'ultima dovrebbe rivelarsi più consapevole e competente.

4. OBIETTIVI DELLA RICERCA

La ricerca presentata in questo contributo è tesa a indagare se il percorso di formazione proposto dagli Autori abbia contribuito a far maturare negli insegnanti coinvolti competenze valutative nell'ambito dell'approccio metodologico proposto. A questo fine, abbiamo preso in esame il grado di concordanza fra le valutazioni degli insegnanti (valutazione tra pari e autovalutazione) e quelle dei docenti del corso sulle UdA prodotte. Questo confronto, da un lato, ci consente di indicare se il corso ha consentito lo sviluppo di competenze valutative e, dall'altro, di indagare la validità di queste procedure di valutazione in un ambito in cui sono poco esplorate, come quello della formazione degli insegnanti in servizio. Relativamente al primo aspetto, in considerazione del fatto che la competenza valutativa è ritenuta una competenza di alto livello (Anderson et al., 2001; Krathwohl, 2002) e che in letteratura vi sono evidenze a supporto del grande potenziale formativo della valutazione tra pari e dell'autovalutazione (e.g., Lynch, et al., 2012; Poon et al., 2009; Vickerman, 2009; Liu & Carless, 2006; Nicol & Macfarlane-Dick, 2006), un buon grado di accordo tra queste ultime e quelle dei docenti del corso ci permetterebbe di sostenere che il percorso di formazione ha contribuito a far maturare negli insegnanti coinvolti buone competenze valutative nell'ambito dell'approccio metodologico proposto. Relativamente al secondo aspetto, la questione inerente alla validità della valutazione tra pari e dell'autovalutazione (Falchikov & Goldfinch, 2000; Li et al., 2016), soprattutto per quanto concerne la formazione degli insegnanti in servizio, rimane aperta. Infatti, mentre lo sviluppo di sofisticati algoritmi consente nei Massive Open Online Courses (MOOC) di utilizzare la valutazione tra pari per fornire *feedback* affidabili e del tutto sovrapponibili a quelli dei docenti (Piech et al., 2013), non si conoscono procedure che possano rendere altrettanto affidabili e validi i processi di valutazione tra pari nei tradizionali contesti della formazione insegnanti. Negli studi condotti sono stati infatti riscontrati bassi (e.g., Kovach, Resch, & Verhulst, 2009), moderati (e.g., Cho, Schunn, & Wilson, 2006) e solo talvolta

elevati (e.g., Harris, 2011) livelli di accordo fra le valutazioni dei pari e quelle degli esperti. Inoltre, nella maggior parte di questi studi viene analizzata la valutazione tra pari e l'autovalutazione tra studenti dai gradi inferiori fino all'istruzione superiore, mentre un minor numero di ricerche riguarda insegnanti in formazione e ancora più rari sono gli studi relativi a insegnanti in servizio.

5. CONTESTO

L'intervento di formazione è stato condotto in modalità *blended* e ha coinvolto 76 insegnanti di un Istituto Comprensivo di Verona. Per sostenere gli insegnanti mentre erano impegnati nello svolgimento delle attività online del corso e fornire loro un *feedback* costante è stato utilizzato, come già indicato, l'ambiente Moodle, mentre per le pratiche di valutazione tra pari e di autovalutazione ci si è avvalsi dell'ambiente Peergrade. Al termine del corso, infatti, agli insegnanti era richiesta la progettazione di UdA basate sull'approccio metodologico proposto. Queste sono state oggetto di valutazione da parte nostra (in qualità di docenti del corso), nonché dei processi di valutazione tra pari e autovalutazione.

Proprio questi ultimi processi sono stati presi in esame per valutare se gli insegnanti hanno maturato buone competenze nell'ambito dell'approccio metodologico proposto con il corso. Tenendo conto che la competenza valutativa è ritenuta una competenza di alto livello (Anderson et al., 2001; Krathwohl, 2002), si è ipotizzato di poter indagare se il percorso di formazione abbia contribuito a far maturare negli insegnanti coinvolti buone competenze valutative nell'ambito dell'approccio metodologico proposto, analizzando il grado di concordanza tra le valutazioni da loro stessi fornite e le corrispondenti valutazioni dei docenti del corso.

6. METODO

La consegna dell'UdA avveniva tramite la compilazione di un documento con campi predefiniti che richiedevano di dettagliare, oltre a informazioni di carattere generale come il grado scolare, la materia, la classe e l'argomento curricolare, le tre fasi previste dall'approccio metodologico proposto (*Lancio*, *Conduzione* e *Chiusura* della *Sfida*).

La valutazione delle UdA è avvenuta utilizzando 12 criteri da noi appositamente predisposti in forma di domanda che consentivano una risposta basata su una scala a tre punti (0: no; 1: in parte; 2: sì). I criteri non sono stati condivisi con gli insegnanti prima della fase di valutazione per evitare che potessero influenzarli nella realizzazione delle UdA stesse. I 12 criteri erano articolati in 3 sezioni, ognuna riferita a uno dei tre momenti dell'UdA: *Lancio della sfida* 5 criteri, *Conduzione della sfida* 4 criteri, *Chiusura della sfida* 3 criteri. Ogni criterio faceva riferimento a specifiche caratteristiche che ogni fase dovrebbe avere per corrispondere all'approccio metodologico proposto (es. induttività nel *Lancio*, apprendimento attivo nella *Conduzione*, valutazione formativa nella *Chiusura*). Il confronto fra le valutazioni, quindi, ai fini della presente ricerca, è stato svolto prendendo in esame le somme dei punteggi derivanti dalla valutazione tramite questi criteri.

Dei 76 insegnanti coinvolti nella formazione hanno partecipato alle attività su Peergrade in 66, di cui 51 di scuola Primaria e 15 di scuola Secondaria di I grado. Le attività di valutazione, che hanno mantenuto distinti gli insegnanti dei due gradi scolari, si componevano di due attività: l'attività di valutazione tra pari - che richiedeva a ogni insegnante di valutare le UdA di altri tre colleghi in forma anonima, assegnate in modo casuale da Peergrade - e l'attività di autovalutazione attraverso le stesse modalità. Delle 51 UdA della Primaria e delle 15 della Secondaria di I grado hanno ricevuto tutte e quattro le valutazioni (tre pari e l'autovalutazione), rispettivamente, 37 e 12 UdA. Tra le UdA della Primaria, 43 hanno ricevuto le tre valutazioni

dei pari ma non l'autovalutazione. Le restanti 8 UdA della Primaria e 3 della Secondaria di I grado hanno ricevuto solamente una o due valutazioni dei pari e per questo sono state escluse dalle analisi.
Si è proceduto quindi ad analizzare l'accordo inter-osservatori, ovvero «*la misura in cui diversi giudici tendono a esprimere esattamente gli stessi giudizi sull'oggetto di valutazione*» (Tinsley & Weiss, 1975, p. 359), sui seguenti confronti:

- tra pari / docenti del corso;
- autovalutazioni / docenti del corso;
- tra pari e autovalutazioni / docenti del corso.

Assumendo che le valutazioni dei docenti del corso rappresentino i "gold standard" (Falchikov & Goldfinch, 2000) e che pertanto un buon grado di concordanza tra queste e quelle degli insegnanti attesti la validità di queste ultime, le ipotesi prevedono che le valutazioni dei docenti del corso e quelle tra pari dimostrino un buon grado di concordanza e che lo stesso si verifichi per l'autovalutazione e per la valutazione "complessiva" (valutazione tra pari e autovalutazione).
Per verificare le ipotesi è stato utilizzato il coefficiente di concordanza di Kendall (*W*) (Kendall & Babington Smith, 1939). Il test di Kendall, appartenente alla famiglia dei test non parametrici, raccoglie i punteggi dati dai valutatori e genera ranghi su cui si formula per correlazione l'indice *W*, i cui valori sono 1 = massima concordanza e 0 = nessuna concordanza. L'interpretazione del *W* di Kendall si è basata sulle linee guida relative alla valutazione del κ di Cohen (Landis & Koch, 1977) tali per cui valori di *W* minori di .20 indicano un pessimo livello di accordo, compresi tra .20 e .40 un modesto livello di accordo, compresi tra .40 e .60 un moderato livello di accordo, compresi tra .60 e .80 un buon livello di accordo e maggiori di .80 un ottimo livello di accordo.

7. RISULTATI

Per quanto concerne la scuola Primaria, il confronto tra le valutazioni dei pari e quelle dei docenti del corso è risultato significativo e ha dato come esito un buon livello di accordo tra i valutatori (Tabella 1).

N	2
W di Kendall	,693*
Chi-quadrato	58,244
gl	42
Sign. asint.	,049

Tabella 1. Indici relativi al confronto tra le valutazioni dei pari e quelle dei docenti del corso per la scuola Primaria.

Sempre per quanto concerne la scuola Primaria, il confronto tra le autovalutazioni degli insegnanti e quelle dei docenti del corso è risultato significativo e ha dato come esito un buon livello di accordo tra i valutatori (Tabella 2).

N	2
W di Kendall	,734*
Chi-quadrato	52,82
gl	36
Sign. asint.	,035

Tabella 2. Indici relativi al confronto tra le autovalutazioni e quelle dei docenti del corso per la scuola Primaria.

Il confronto tra le valutazioni "complessive" (valutazione tra pari e autovalutazione) e quelle dei docenti del corso è risultato significativo e ha dato come esito un buon livello di accordo tra i valutatori (Tabella 3).

N	2
W di Kendall	,743*
Chi-quadrato	53,523
gl	36
Sign. asint.	,03

Tabella 3. Indici relativi al confronto tra le valutazioni "complessive" e quelle dei docenti del corso per la scuola Primaria.

Per quanto riguarda invece la scuola Secondaria di I grado, il confronto tra le valutazioni dei pari e quelle dei docenti del corso è risultato significativo e ha dato come esito un ottimo livello di accordo tra i valutatori (Tabella 4).

N	2
W di Kendall	,898*
Chi-quadrato	19,761
gl	11
Sign. asint.	,049

Tabella 4. Indici relativi al confronto tra le valutazioni dei pari e quelle dei docenti del corso per la scuola Secondaria di I grado.

Sempre per quanto riguarda la scuola Secondaria di I grado, il confronto tra le autovalutazioni degli insegnanti e quelle dei docenti del corso è risultato significativo e ha dato come esito un ottimo livello di accordo tra i valutatori (Tabella 5).

N	2
W di Kendall	,934*
Chi-quadrato	20,548
gl	11
Sign. asint.	,038

Tabella 5. Indici relativi al confronto tra le autovalutazioni e quelle dei docenti del corso per la scuola Secondaria di I grado.

Il confronto tra le valutazioni "complessive" (valutazione tra pari e autovalutazione) e quelle dei docenti del corso è risultato significativo e ha dato come esito un ottimo livello di accordo tra i valutatori (Tabella 6).

N	2
W di Kendall	,926*
Chi-quadrato	20,379
gl	11
Sign. asint.	,04

Tabella 6. Indici relativi al confronto tra le valutazioni "complessive" e quelle dei docenti del corso per la scuola Secondaria di I grado.

8. DISCUSSIONE

Le analisi condotte e i risultati ottenuti dal confronto fra le valutazioni tra pari e le valutazioni dei docenti del corso hanno evidenziato una buona concordanza per la Primaria e un'ottima concordanza per la Secondaria di I grado. Questi stessi gradi di concordanza sono stati ottenuti anche nei confronti effettuati per le autovalutazioni e per le valutazioni "complessive". In merito alle autovalutazioni possiamo notare come, sia nella Primaria che nella Secondaria di I grado, la concordanza sia risultata, seppur leggermente, superiore a quella relativa alle valutazioni tra pari. Su questo dato potrebbe aver influito positivamente, come messo in luce in letteratura (e.g., Nicol et al., 2014; Cho & MacArthur, 2011), il fatto che le valutazioni tra pari fossero anteposte all'autovalutazione, consentendo a quest'ultima di divenire più consapevole e competente.

I risultati ottenuti ci permettono di affermare che gli insegnanti coinvolti nel percorso di formazione mostrano buone competenze valutative nell'ambito dell'approccio metodologico proposto. Questo risultato è in linea con quanto riscontrato da Lynch e colleghi (2012) e Sluijsmans e colleghi (2002) nei loro studi sugli insegnanti in formazione (*pre-service* e *student teachers*). Sebbene non si possa concludere che gli

insegnanti coinvolti nella formazione abbiamo sviluppato effettive competenze realizzative nell'ambito dell'approccio metodologico proposto, possiamo però desumere che questi risultati rappresentino un indicatore della buona comprensione della metodologia proposta. In letteratura infatti si sottolinea il potenziale formativo delle pratiche di autovalutazione e valutazione tra pari (e.g., Vickerman, 2009; Liu & Carless, 2006; Nicol & Macfarlane-Dick, 2006), soprattutto nel promuovere un apprendimento significativo (e.g., Poon et al., 2009), nonché come l'atto di revisionare favorisca una comprensione più profonda dell'oggetto di indagine (e.g., Nicol et al., 2014).

Le analisi condotte, inoltre, ci hanno permesso di indagare la validità delle procedure di valutazione tra pari e autovalutazione in un ambito in cui sono poco esplorate, come quello della formazione degli insegnanti in servizio. La maggior parte degli studi sono infatti relativi a studenti o insegnanti in formazione e i loro esiti evidenziano risultati contrastanti. Si riscontrano infatti bassi (e.g., Kovach et al., 2009), moderati (e.g., Cho et al., 2006) e talvolta elevati (e.g., Harris, 2011) livelli di accordo. Assumendo, come anticipato, che le valutazioni dei docenti del corso rappresentino i "gold standard" (Falchikov & Goldfinch, 2000) e che pertanto un buon grado di concordanza fra queste e quelle degli insegnanti (valutazione tra pari e autovalutazione) attesti la validità di queste ultime, i risultati ottenuti consentono di sostenere la validità delle procedure di valutazione tra pari e autovalutazione anche nell'ambito della formazione degli insegnanti in servizio.

9. CONCLUSIONI

I risultati ottenuti indicano la possibilità di utilizzare proficuamente pratiche di valutazione tra pari e di autovalutazione nell'ambito della formazione degli insegnanti in servizio. Richiedere agli insegnanti la valutazione degli elaborati dei pari e l'autovalutazione dei propri favorisce lo sviluppo di competenze valutative nell'ambito dell'approccio metodologico proposto e offre inoltre la possibilità di sperimentare queste pratiche di valutazione in vista di un loro uso con i propri studenti.

Prospettive di ricerca future potrebbero includere analisi pre e post intervento di formazione volte a indagare se quest'ultimo abbia prodotto uno sviluppo di competenze anche di carattere progettuale e realizzativo nell'ambito dell'approccio metodologico proposto. I dati così prodotti potrebbero inoltre essere corroborati da analisi qualitative che entrino nel merito, per esempio, della qualità degli elaborati prodotti dagli insegnanti.

Queste analisi potrebbero fornire validi elementi anche per valutare l'efficacia di interventi formativi rivolti agli insegnanti in servizio. In genere questi interventi si concludono con questionari di gradimento, che non indagano l'effettivo sviluppo di competenze. Strumenti di indagine più specifici, somministrati prima e dopo l'intervento e anche attraverso il confronto con un gruppo di controllo, possono valutare l'effettivo cambiamento prodotto nelle pratiche didattiche o nelle convinzioni educative degli insegnanti coinvolti ma, oltre a richiedere un certo impegno, possono incorrere in *bias*, come la desiderabilità sociale, oppure incontrare le difficoltà tipiche dell'analisi di processi che richiedono tempi di maturazione più lunghi di quelli solitamente previsti dagli interventi di formazione. Altre modalità, come quelle che si basano su una valutazione formale di prove, sono chiaramente difficilmente applicabili nel contesto della formazione degli insegnanti in servizio.

L'analisi qui proposta potrebbe pertanto essere utilizzata, insieme ad altri indicatori, per valutare l'efficacia formativa di un intervento di formazione. Presenta infatti il vantaggio di non richiedere, né agli insegnanti partecipanti, né ai responsabili dell'intervento di formazione, attività *ad hoc*, ma di utilizzare le attività stesse proposte durante il corso e i dati generati dalle pratiche di valutazione tra pari e di autovalutazione.

10. RICONOSCIMENTI

Questo contributo, pur essendo frutto dell'intensa collaborazione fra gli autori, è stato così redatto: i paragrafi "Valutazione tra pari e autovalutazione", "La scelta di Peergrade per la valutazione tra pari e l'autovalutazione", "Obiettivi della ricerca", "Metodo" e "Discussione" sono stati redatti da Laura Carlotta Foschi; i paragrafi "Introduzione", "Contesto" e "Conclusioni" sono stati redatti da Graziano Cecchinato; il paragrafo "Risultati" è stato redatto da Filippo Say e Laura Carlotta Foschi.

11. BIBLIOGRAFIA

Anderson, L. W., Krathwohl, D. R., Airasian, P. W., Cruikshank, K. A., Mayer, R. E., Pintrich, ... Wittrock, M. C. (2001). *A taxonomy for learning, teaching, and assessing: A revision of Bloom's taxonomy of educational objectives.* New York, NY, USA: Longman.

Anolli, L., & Mantovani, F. (2011). *Come funziona la nostra mente*. Bologna, IT: Il Mulino.

Ausubel, D. P., Novak, J. D., & Hanessian, H. (1968). *Educational psychology: A cognitive view*. New York, NY, USA: Holt Rinehart and Wilson.

Baker, W. J. (2000). The "classroom flip": Using web course management tools to become the guide by the side. *Cedarville University: Communication Faculty Publication*, 9-17.

Berlyne, D. E. (1960). *Conflict, arousal, and curiosity*. New York, NY, USA: McGraw-Hill.

Bishop, J. L., & Verleger, M. A. (2013). The flipped classroom: A survey of the research. *Proceedings – 120*th *ASEE Annual Conference & Exposition. American Society for Engineering Education,* Atlanta, GA, USA.

Black, P., & Wiliam, D. (1998). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice, 5*(1), 7-74. doi: 10.1080/0969595980050102

Bloxham, S., & West, A. (2004). Understanding the rules of the game: Marking peer assessment as a medium for developing students' conceptions of assessment. *Assessment and Evaluation in Higher Education, 29*(6), 721-733. doi: 10.1080/0260293042000227254

Brown, G. T. L., & Harris, L. R. (2013). Student self-assessment. In J. H. McMillan (Ed.), *The SAGE handbook of research on classroom assessment* (pp. 367–393). Thousand Oaks, CA, USA: Sage.

Carless, D. (2007). Learning-oriented assessment: Conceptual bases and practical implications. *Innovations in Education and Teaching International, 44*(1), 57-66. doi: 10.1080/14703290601081332

Carless, D., Salter, D., Yang, M., & Lam, J. (2011). Developing sustainable feedback practices. *Studies in Higher Education, 36*(4), 395-407. doi: 10.1080/03075071003642449

Cecchinato, G. (2014). Flipped classroom, innovare la scuola con le tecnologie digitali. *TD Tecnologie didattiche*, *22*(1), 11-20.

Cecchinato, G., & Papa, R. (2016). *Flipped classroom: un nuovo modo di insegnare e apprendere*. Torino, IT: UTET.

Cecchinato, G., & Papa, R. (2018). Gaming experience a scuola: innovare la didattica nell'ecosistema digitale. In *La formazione nell'era delle Smart City. Esperienze ed Orizzonti*. Milano, IT: Monduzzi Editoriale.

Cheng, M. M. H., Cheng, A. Y. N., & Tang, S. Y. F. (2010). Closing the gap between the theory and practice of teaching: Implications for teacher education programmes in Hong Kong. *Journal of Education for Teaching, 36*(1), 91-104. doi: 10.1080/02607470903462222

Cho, K., & MacArthur, C. (2010). Student revision with peer and expert reviewing. *Learning and Instruction, 20*(4), 328-338. doi: 10.1016/j.learninstruc.2009.08.006

Cho, K., & MacArthur, C. (2011). Learning by reviewing. *Journal of Educational Psychology, 103*(1), 73-84. doi: 10.1037/a0021950

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. doi: 10.1037/0022-0663.98.4.891

Davies, P. (2006). Peer assessment: Judging the quality of students work by comments rather than marks. *Innovations in Education and Teaching International, 43*(1), 69-82. doi: 10.1080/14703290500467566

Dewey, J. (2007). *Experience and education*. New York, NY, USA: Simon & Schuster.

Falchikov, N. (1995). Peer feedback marking: Developing peer assessment. *Programmed Learning, 32*(2), 175-187. doi: 10.1080/1355800950320212

Falchikov, N., & Goldfinch, J. (2000). Student peer assessment in higher education: A meta-analysis comparing peer and teacher marks. *Review of Educational Research, 70*(3), 287-322. doi: 10.3102/00346543070003287

Festinger, L. (1962). *A theory of cognitive dissonance*. Stanford, CA, USA: Stanford University Press.

Gee, J. P. (2003). *What video games have to teach us about learning and literacy*. New York, NY, USA: Palgrave Macmillan.

Harris, J. R. (2011). Peer assessment in large undergraduate classes: An evaluation of a procedure for marking laboratory reports and a review of related practices. *American Journal of Physiology - Advances in Physiology Education, 35*(2), 178-187. doi: 10.1152/advan.00115.2010

Hattie, J. (2016). *Apprendimento visibile, insegnamento efficace. Metodi e strategie di successo dalla ricerca evidence-based*. Trento, IT: Erickson.

Jenkins, H., Purushotma, R., Weigel, M., Clinton, K., & Robison, A. J. (2009). *Confronting the challenges of participatory culture: Media education for the 21st century*. Cambridge, MA, USA: Mit Press.

Jonassen, D. H. (2008). *Meaningful learning with technology*. Upper Saddle River, NJ, USA: Prentice Hall.

Kaufman, J. H., & Schunn, C. D. (2011). Students' perceptions about peer assessment for writing: Their origin and impact on revision work. *Instructional Science, 39*(3), 387-406. doi: 10.1007/s11251-010-9133-6

Keengwe, J., Onchwari, G., & Oigara, J. (2014). *Promoting active learning through the flipped classroom model*. Hershey, PA, USA: IGI Global.

Kendall, M. G., & Babington Smith, B. (1939). The Problem of m Rankings. *The Annals of Mathematical Statistics*, *10*(3), 275-287. doi:10.1214/aoms/1177732186

Kovach, R. A., Resch, D. S., & Verhulst, S. J. (2009). Peer assessment of professionalism: A five-year experience in medical clerkship. *Journal of General Internal Medicine, 24*(6), 742-746. doi: 10.1007/s11606-009-0961-5

Krathwohl, D. R. (2002). A revision of bloom's taxonomy: An overview. *Theory into Practice, 41*(4), 212-218. doi: 10.1207/s15430421tip4104_2

Lage, M. J., Platt, G. J., & Treglia, M. (2000). Inverting the classroom: A gateway to creating an inclusive learning environment. *Journal of Economic Education, 31*(1), 30-43. doi: 10.2307/1183338

Landis, J. R., & Koch, G. G. (1977). The measurement of observer agreement for categorical data. *Biometrics, 33*(1), 159-174. doi: 10.2307/2529310

Li, H., Xiong, Y., Zang, X., Kornhaber, M. L., Lyu, Y., Chung, K. S., & Suen, H. K. (2016). Peer assessment in the digital age: A meta-analysis comparing peer and teacher ratings. *Assessment and Evaluation in Higher Education, 41*(2), 245-264. doi: 10.1080/02602938.2014.999746

Liu, N., & Carless, D. (2006). Peer feedback: The learning element of peer assessment. *Teaching in Higher Education, 11*(3), 279-290. https://dx.doi.org/10.1080/13562510600680582

Lynch, R., McNamara, P. M., & Seery, N. (2012). Promoting deep learning in a teacher education programme through self- and peer-assessment and feedback. *European Journal of Teacher Education, 35*(2), 179-197. doi: 10.1080/02619768.2011.643396

Magin, D. (2001). Reciprocity as a source of bias in multiple peer assessment of group work. *Studies in Higher Education, 26*(1), 52-63. doi: 10.1080/03075070020030715

Mayer, R. E. (2002). Rote versus meaningful learning. *Theory into Practice, 41*(4), 226-232. doi: 10.1207/s15430421tip4104_4

Mazur, E. (1997). *Peer instruction: A user's manual.* Upper Saddle River, NJ, USA: Prentice Hall.

Miller, P. J. (2003). The effect of scoring criteria specificity on peer and self-assessment. *Assessment and Evaluation in Higher Education, 28*(4), 383-394. doi: 10.1080/0260293032000066218

Nicol, D. (2010). From monologue to dialogue: Improving written feedback processes in mass higher education. *Assessment and Evaluation in Higher Education, 35*(5), 501-517. doi: 10.1080/02602931003786559

Nicol, D., & MacFarlane-Dick, D. (2006). Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education, 31*(2), 199-218. doi: 10.1080/03075070600572090

Nicol, D., Thomson, A., & Breslin, C. (2014). Rethinking feedback practices in higher education: A peer review perspective. *Assessment and Evaluation in Higher Education, 39*(1), 102-122. doi: 10.1080/02602938.2013.795518

Novak, J. D. (2002). Meaningful learning: The essential factor for conceptual change in limited or inappropriate propositional hierarchies leading to empowerment of learners. *Science Education, 86*(4), 548-571. doi: 10.1002/sce.10032

O'Mahony, T. K., Vye, N. J., Bransford, J. D., Sanders, E. A., Stevens, R., Stephens, ... Soleiman, M. K. (2012). A comparison of lecture-based and challenge-based learning in a workplace setting: Course designs, patterns of interactivity, and learning outcomes. *Journal of the Learning Sciences, 21*(1), 182-206. doi: 10.1080/10508406.2011.611775

Orsmond, P., Merry, S., & Reiling, K. (2002). The use of exemplars and formative feedback when using student derived marking criteria in peer and self-assessment. *Assessment and Evaluation in Higher Education, 27*(4), 309-323. doi: 10.1080/0260293022000001337

Panadero, E., Brown, G. T. L., & Strijbos, J. (2016). The future of student self-assessment: A review of known unknowns and potential directions. *Educational Psychology Review, 28*(4), 803-830. doi: 10.1007/s10648-015-9350-2

Panadero, E., Tapia, J. A., & Huertas, J. A. (2012). Rubrics and self-assessment scripts effects on self-regulation, learning and self-efficacy in secondary education. *Learning and Individual Differences, 22*(6), 806-813. doi: 10.1016/j.lindif.2012.04.007

Piaget, J. (1974). *La prise de conscience.* Paris, FR: PUF.

Piech, C., Huang, J., Chen, Z., Do, C., Ng, A., & Koller, D. (2013). Tuned models of peer assessment in MOOCs. *Proceedings of the 6th International Conference on Educational Data Mining*, Memphis, Tennessee, USA.

Poon, W., McNaught, C., Lam, P., & Kwan, H. S. (2009). Improving assessment methods in university science education with negotiated self-and peer-assessment. *Assessment in Education: Principles, Policy & Practice, 16*(3), 331-346. doi: 10.1080/09695940903319729

Price, M., Handley, K., & Millar, J. (2011). Feedback: Focusing attention on engagement. *Studies in Higher Education, 36*(8), 879-896. doi: 10.1080/03075079.2010.483513

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology, 25*(1), 54-67. doi: 10.1006/ceps.1999.1020

Sadler, D. R. (1987). Specifying and promulgating achievement standards. *Oxford Review of Education, 13*(2), 191-209. doi: 10.1080/0305498870130207

Sadler, D. R. (1989). Formative assessment and the design of instructional systems. *Instructional Science, 18*(2), 119-144. doi: 10.1007/BF00117714

Sadler, P. M., & Good, E. (2006). The impact of self- and peer-grading on student learning. *Educational Assessment, 11*(1), 1-31.

Schwartz, D. L., Lin, X., Brophy, S., & Bransford, J. D. (1999). *Toward the development of flexibly adaptive instructional designs*. Hillsdale, MI: Erlbaum.

Sluijsmans, D. M. A., Brand-Gruwel, S., & van Merriënboer, J. J. G. (2002). Peer assessment training in teacher education: Effects on performance and perceptions. *Assessment and Evaluation in Higher Education, 27*(5), 443-454. doi: 10.1080/0260293022000009311

Somervell, H. (1993). Issues in assessment, enterprise and higher education: The case for self-, peer and collaborative assessment. *Assessment & Evaluation in Higher Education, 18*(3), 221-233. doi: 10.1080/0260293930180306

Talbert, R. (2017). *Flipped learning: A guide for higher education faculty*. Sterling, VA, USA: Stylus Publishing.

Tinsley, H. E., & Weiss, D. J. (1975). Interrater reliability and agreement of subjective judgments. *Journal of Counseling Psychology, 22*(4), 358-376. doi: 10.1037/h0076640

Topping, K. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research, 68*(3), 249-276. doi: 10.3102/00346543068003249

Topping, K. J. (2005). Trends in peer learning. *Educational Psychology, 25*(6), 631-645. doi: 10.1080/01443410500345172

Vickerman, P. (2009). Student perspectives on formative peer assessment: An attempt to deepen learning? *Assessment & Evaluation in Higher Education, 34*(2), 221-230. doi: 10.1080/02602930801955986

Wen, M. L., & Tsai, C. (2006). University students' perceptions of and attitudes toward (online) peer assessment. *Higher Education, 51*(1), 27-44. doi: 10.1007/s10734-004-6375-8

Yan, Z., & Brown, G. T. L. (2017). A cyclical self-assessment process: Towards a model of how students engage in self-assessment. *Assessment and Evaluation in Higher Education, 42*(8), 1247-1262. doi: 10.1080/02602938.2016.1260091

Yilmaz, F. M. (2017). Reliability of scores obtained from self-, peer-, and teacher-assessments on teaching materials prepared by teacher candidates. *Educational Sciences: Theory & Practice, 17*(2), 395-409. doi: 10.12738/estp.2017.2.0098