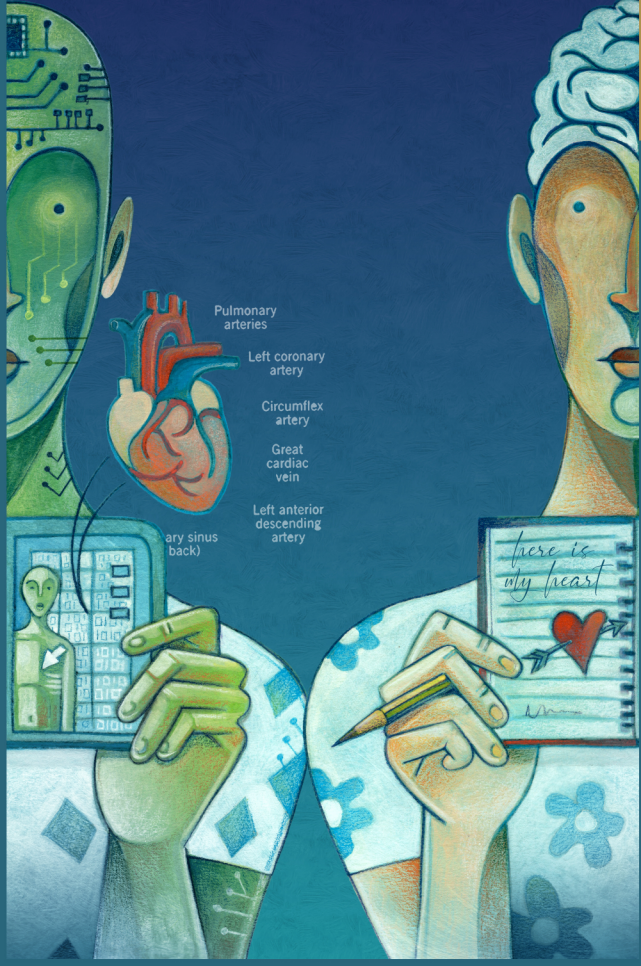
Lorella Giannandrea
Francesca Gratani
Lorenza Maria Capolla
[a cura di]



VALUTAZIONE E INTELLIGENZA ARTIFICIALE

Prospettive didattiche, etiche e metodologiche

sentieri scenari postmoderni

Pensa MULTIMEDIA

sentieri
scenari
postmoderni

Collana diretta da
Michele Corsi
Massimiliano Stramaglia

12

Comitato scientifico della collana

MATHIEU DEFLEM
Università del South Carolina - USA

MAURIZIO FABBRI
Università degli Studi di Bologna

TOMMASO FARINA
Università degli Studi di Macerata

ADRIAN-MARIO GELLEL
Università di Malta

CATIA GIACONI
Università degli Studi di Macerata

VANNA IORI
Università Cattolica del Sacro Cuore di Milano, sede di Piacenza

PIERLUIGI MALAVASI
Università Cattolica del Sacro Cuore Cuore, sede di Brescia

CONCEPCIÓN NAVAL
Università di Navarra - Spagna

LOREDANA PERLA
Università degli Studi di Bari "Aldo Moro"

MARIA GRAZIA RIVA
Università degli Studi di Milano Bicocca

GRAZIA ROMANAZZI
Università degli Studi di Macerata

MAURIZIO SIBILIO
Università degli Studi di Salerno

I volumi di questa collana sono sottoposti a un sistema di *double blind referee*

Lorella Giannandrea, Francesca Gratani
Lorenza Maria Capolla
[a cura di]

VALUTAZIONE E INTELLIGENZA ARTIFICIALE

PROSPETTIVE DIDATTICHE, ETICHE E METODOLOGICHE

ATTI DEL CONVEGNO PRIN "AI&F - ARTIFICIAL INTELLIGENCE & FEEDBACK FOR EFFECTIVE LEARNING".
22 E 23 GENNAIO 2026
UNIVERSITÀ DEGLI STUDI DI BARI ALDO MORO

Pensa MULTIMEDIA

Il presente lavoro è stato pubblicato con i fondi del Progetto PRIN 2022
AI&F - Artificial intelligence & feedback for effective learning
n. 2022ZMYYTH - CUP D53D23013110006 - PNRR M4.C2.1.1



L'opera, comprese tutte le sue parti, è tutelata dalla legge sul diritto d'autore ed è pubblicata in versionedigitale con licenza Creative Commons Attribuzione-Non Commerciale-Non opere derivate 4.0 Internazionale (CC-BY-NC-ND 4.0).
L'Utente, nel momento in cui effettua il download dell'opera, accetta tutte le condizioni della licenza d'uso dell'opera previste e comunicate sul sitohttps://creativecommons.org/licenses/by-nc-nd/4.0/deed.it

This work is licensed under a Creative Commons Attribution–NonCommercial–NoDerivatives 4.0 International License (CC BY-NC-ND 4.0). The license requires attribution, prohibits adaptation (altering or transforming the work or creating derivative works), and does not allow commercial use.

ISBN volume 979-12-5568-510-4
ISSN collana 2421-1133

2026 © by Pensa Evolution Srl
73100 Lecce • Via Arturo Maria Caprioli, 8 • Tel. 0832.230435
www.pensamultimedia.it

SOMMARIO

Introduzione IX

I parte
Interventi dei Componenti delle Unità di ricerca del PRIN AI&F

I.1 Pier Giuseppe Rossi, Lorella Giannandrea, Chiara Laici, Francesca Gratani, Lorenza Maria Capolla 3
Interazione tra agenti umani e artificiali per il Feedback Loop: evidenze sperimentali emerse nel PRIN "AI&F"

I.2 Valentina Grion, Laura Carlotta Foschi, Beatrice Doria, Giorgia Slaviero, Juliana Elisa Raffaghelli, Graziano Cecchinato, Corrado Petrucco 27
Pratiche di feedback e feedback automatizzato nell'istruzione superiore: analisi delle pratiche dei docenti e delle percezioni degli studenti nel progetto PRIN "AI&F"

I.3 Antonella Montone, Michele Giuliano Fiorentino 63
L'intelligenza artificiale per il feedback formativo in matematica: specificità epistemologiche, modelli e progettazione nei problemi aperti per la formazione universitaria

II parte
Area tematica 1. IA e valutazione universitaria

II.1 Angela Spinelli 79
Scrivere con l'IA generativa: usi, percezioni e nodi valutativi in ambito accademico

II.2 Nadia Sansone, Ilaria Bortolotti 91
Integrare l'IA nell'assessment: il modello rAIse

II.3 **Francesco Pio Sarcina, Anna Maria Cuzzi, Michele Baldassarre** **103**
Dalla valutazione tradizionale alla valutazione simbiotica: analisi comparativa tra docenti e sistemi di IA nella valutazione accademica

II.4 **Giovannina Albano, Anna Pierri, Agnese Telloni** **115**
Come gli studenti universitari valutano la risoluzione di un task generata da ChatGPT

II.5 **Claudia Baiata, Gabriele Biagini, Alice Roffi, Maria Ranieri** **129**
Chatbot e valutazione formativa nella didattica universitaria: evidenze da uno studio pilota di un corso blended

II.6 **Ilaria Fiore, Maria Clara Dicataldo, Mariasole Antonietta Guerriero, Alessia Scarinci, Anna Dipace** **139**
Ricostruire il compito, costruire la rubrica: esperienze di valutazione con il supporto dell'IA

II.7 **Maria Giulia Ballatore** **153**
From Challenge to Opportunity: Critical Use of GenAI in Mathematical Foundations for Architecture Students

II.8 **Rosamaria Nicassio** **163**
Ripensare la valutazione come processo nel percorso di apprendimento. Spunti per una riflessione di ricerca fenomenologica integrata con l'IA

Area tematica 2. IA e valutazione nella scuola

II.9 **Paolo Barabanti** **173**
Docenti e GenAI nella valutazione: tra entusiasmo e scetticismo. Evidenze dal Questionario Insegnante INVALSI 2024/25

II.10 **Irene Gianeselli, Dario Ianes, Rita Cosoli, Marco Cadavero, Gaetano Scianatico, Alessandro Oronzo Caffò, Andrea Bosco** **187**
Conceptual Domains of AI in Education for Neuroinclusive Measurement Using Topic Modelling

II.11 **Massimo Marcuccio, Maria Elena Tassinari** **197**
Le percezioni degli insegnanti circa l'uso dell'intelligenza artificiale nella valutazione della produzione scritta degli studenti: uno studio esplorativo

II.12 Biagio Di Liberto **207**
Rubriche IA-assistite per la valutazione formativa nella scuola secondaria di II grado: argomentazione di validità, feed-forward e audit di equità in chiave DD/UDL

II.13 Umberto Dello Iacono, Marco Picariello **225**
Interazioni con ChatGPT e impatto sulle attività metacognitive degli studenti: uno studio sulla risoluzione di problemi matematici nella scuola secondaria di secondo grado

II.14 Francesco Contel, Annalisa Cusi **233**
Potenzialità e limiti nell'uso dell'approccio strumentale nell'interpretazione delle interazioni tra studenti e IA generativa

II.15 Maria Lucia Bernardi, Ludovica Galiano, Roberto Capone, Eleonora Faggiano **243**
Chatbot come strumento di supporto alla "*Assessment Competency*" degli insegnanti: un'analisi esplorativa

Area tematica 3. IA e formazione insegnanti

II.16 Federica Troilo, Elisabetta Giuseppina Erione, Roberto Capone, Eleonora Faggiano **261**
Formare i docenti all'uso consapevole della GenAI nella valutazione: un'esperienza esplorativa basata sull'Instrumental Approach

II.17 Claudio Palazzi **273**
Formare i docenti per un uso consapevole dell'AI nella valutazione di una didattica ibrida

II.18 Alessandro Barca **283**
Valutazione dell'impatto dell'Intelligenza Artificiale nella formazione docente: una Scoping Review per un modello pedagogico-organizzativo

II.19 Silvia Artusi, Rosanna Pia Laccone, Marco Romano **301**
Docenti in formazione davanti all'IA: percezioni, pratiche e questioni etiche della valutazione

II.20 Marilena Di Padova, Angelo Basta, Andrea Tinterri, Anna Dipace **311**
RubricaBot: un supporto intelligente alla progettazione valutativa nella formazione dei docenti

II.21 **Viviana Vinci, Pierangelo Berardi** **325**
Progettare la valutazione con l'algoritmo. Uno studio sull'impatto dell'IA sulla competenza valutativa dei docenti pre-service

II.22 **Antonella Montone, Rosalba Romito, Michele Giuliano Fiorentino, Michele Romita** **337**
L'IA: un mediatore digitale nel processo di feedback per la formazione di insegnanti pre-service e a scuola

Area tematica 4. IA e feedback

II.23 **Manuela Fabbri, Chiara Laici, Maila Pentucci** **353**
L'AI generativa come supporto alla feedback literacy degli studenti

II.24 **Livia Petti, Marta De Angelis, Andrea Garavaglia** **365**
Efficacia del feedback docente e IA: un'analisi delle percezioni nel contesto accademico

II.25 **Umberto Barbieri, Lia Daniela Sasanelli, Raffaele Di Fuccio** **377**
Designing intelligent assessment systems: an MCP-based framework for UDL - oriented feedback

II.26 **Simona Michelon** **393**
Il feedback e l'apprendimento. Cosa cambia nell'era dell'intelligenza artificiale

II.27 **Sofia Di Crisci, Cristiana Bianchi** **405**
Dall'autovalutazione al feedback personalizzato: l'integrazione dell'intelligenza artificiale nella costruzione del bilancio di competenze nei percorsi di abilitazione dei docenti

II.28 **Elisabetta Lucia De Marco, Marilena Di Padova, Anna Dipace** **417**
Dall'uso delle rubriche all'IA generativa: un chatbot per il feedback intermedio dei Project Work

II.29 **Maila Pentucci, Giovanna Cioci** **427**
L'IA generativa per progettare dispositivi di feedback su artefatti complessi nella formazione degli insegnanti

II.11
Le percezioni degli insegnanti circa l'uso dell'intelligenza artificiale nella valutazione della produzione scritta degli studenti: uno studio esplorativo

Massimo Marcuccio
Professore ordinario di Pedagogia Sperimentale (SSD: PAED – 02/B)
Alma Mater Studiorum – Università di Bologna.
massimo.marcuccio@unibo.it
Maria Elena Tassinari
Assegnista di ricerca in Pedagogia Sperimentale (SSD: PAED – 02/B)
Alma Mater Studiorum – Università di Bologna.
mariaelena.tassinar3@unibo.it

Abstract

L'impiego dell'Intelligenza Artificiale (IA) nei contesti educativi è in rapida espansione, in particolare nelle pratiche di progettazione e valutazione, suscitando al contempo interesse e preoccupazioni tra i docenti. Questo studio esplorativo analizza le percezioni di due insegnanti di scuola secondaria di primo grado rispetto all'uso di ChatGPT nella valutazione della produzione scritta, mettendo a confronto i giudizi professionali con quelli generati dal chatbot. L'analisi delle interviste qualitative evidenzia opportunità, criticità e tensioni che emergono dall'incontro tra pratiche valutative consolidate e automazione intelligente, offrendo indicazioni utili per orientare future riflessioni e ricerche sul ruolo dell'IA nei processi valutativi scolastici.

Parole chiave: Intelligenza artificiale; valutazione degli apprendimenti; insegnanti; ChatGPT.

The use of Artificial Intelligence (AI) in educational contexts is rapidly expanding, particularly in instructional design and assessment practices, and is generating both interest and concern among teachers. This exploratory study examines the perceptions of two lower secondary school teachers regarding the use of ChatGPT in evaluating students' written work, comparing their professional judgments with those produced by the chatbot. The analysis of qualitative interviews highlights the opportunities, challenges, and tensions that arise when established

assessment practices encounter intelligent automation, offering insights to inform future reflections and research on the role of AI in school-based assessment processes.

Keywords: Artificial intelligence; learning assessment; teachers; ChatGPT.

1. Introduzione

L'intelligenza artificiale (IA) sta trasformando in modo profondo e trasversale numerosi ambiti della vita sociale, includendo sempre più stabilmente il settore educativo. La scuola diventa così un laboratorio privilegiato di sperimentazione, nel quale l'IA viene impiegata per potenziare l'apprendimento, personalizzare i percorsi formativi e automatizzare processi chiave quali la progettazione e la valutazione (Goralski et al., 2020). In quest'ultimo ambito, l'introduzione di strumenti basati su IA, soprattutto quelli che si avvalgono dell'IA generativa (IAGen), solleva tuttavia questioni teoriche e metodologiche rilevanti, che impongono un'analisi accurata delle percezioni, delle pratiche e delle dinamiche di interazione tra insegnanti e tecnologie emergenti. La letteratura sulle tecnologie educative evidenzia come l'adozione di nuovi strumenti digitali sia influenzata da un intreccio di fattori cognitivi, sociali e istituzionali. Il *Technology Acceptance Model* (TAM) di Davis (1989) e la *Unified Theory of Acceptance and Use of Technology* (UTAUT) di Venkatesh e colleghi (2003) forniscono utili cornici interpretative per comprendere le condizioni che favoriscono o ostacolano l'integrazione di sistemi automatizzati nella valutazione. A tali prospettive si affianca il contributo di Ertmer (1999), che distingue tra barriere di primo ordine, legate alle risorse infrastrutturali e alla formazione tecnica e barriere di secondo ordine, riconducibili alle convinzioni pedagogiche e alle resistenze epistemologiche degli insegnanti. Nell'ambito dell'*Artificial Intelligence in Education* (AIEd) (Luckin, 2018), queste dimensioni concorrono a delineare un continuum di atteggiamenti nei confronti dell'IA valutativa, che spazia dall'integrazione stabile dell'innovazione al suo uso condizionato, fino al rifiuto o alla limitazione del suo impiego.

All'interno di tale quadro si colloca l'interesse per la valutazione della produzione scritta, ambito che ha conosciuto una profonda evoluzione teorica. Come evidenziato da Huot (2002) e da Huot e O'Neill (2009), fino agli anni Settanta la valutazione della scrittura era inscritta nella tradizione della misurazione educativa, centrata su standardizzazione, affidabilità e validità, assumendo la scrittura come fenomeno quantificabile. A partire dagli anni Settanta, gli studi

sulla composizione e sulla retorica hanno invece messo in luce la natura sociale, situata e processuale della scrittura, criticando la riduzione a criteri oggettivi. Tra gli anni Ottanta e Novanta si consolida l'idea di una validità dinamica e contestuale e si diffondono pratiche valutative quali saggi, portfolio e compiti autentici, con attenzione alla validità costruttiva e consequenziale e al ruolo attivo dell'insegnante nel progettare pratiche coerenti con l'apprendimento. L'avvento delle tecnologie digitali e, più recentemente, dei modelli di IA ha introdotto ulteriori sviluppi, come i sistemi di *automated writing evaluation* (AWE) e le *computer-based assessments* (CBA), basati su algoritmi di NLP e tecniche di apprendimento automatico (Nardi, 2018; Tonelli et al., 2018). Tra questi strumenti, i chatbot conversazionali come ChatGPT, che si avvalgono di dispositivi di IAGen, stanno assumendo crescente rilevanza, sollevando un ampio dibattito scientifico sul loro potenziale e sui rischi connessi alla loro integrazione nei processi valutativi.

La letteratura mostra atteggiamenti ambivalenti degli insegnanti verso l'uso di ChatGPT nella valutazione. Alcuni studi ne evidenziano i vantaggi in termini di riduzione dei tempi di correzione e maggiore coerenza dei giudizi, soprattutto in contesti ad alto carico valutativo (Ruff et al., 2024; Zuckerman et al., 2023), nonché la possibilità di fornire feedback immediato agli studenti, permettendo ai docenti di concentrarsi su aspetti più complessi della scrittura (Pfau et al., 2023). Altri contributi segnalano invece criticità legate alla qualità del feedback e alla difficoltà dell'IA nel cogliere dimensioni soggettive della scrittura, quali originalità e costruzione argomentativa (Obata et al., 2024; Kim et al., 2024; Ljubojević et al., 2023), oltre al timore di una riduzione dell'autonomia professionale docente (Safrai et al., 2024). La letteratura sottolinea infine la necessità di una formazione specifica per un uso consapevole della GenAI, al fine di sviluppare competenze di mediazione tecnologica e ridurre disuguaglianze nelle pratiche valutative (Picciano et al., 2024; Jiang et al., 2024).

Muovendo da questa cornice teorica, il presente studio adotta un approccio esplorativo e interpretativo per indagare come gli insegnanti percepiscono l'inserimento di ChatGPT nella propria pratica valutativa. Il lavoro si colloca entro una concezione multifunzionale e formativa della valutazione, sviluppata in Italia da Giovannini (1994; 1995) e a livello internazionale da Broadfoot et al. (1999) attraverso il modello dell'*assessment for learning* (AfL). In questa prospettiva, la valutazione non è intesa come semplice misurazione degli esiti, ma come componente integrante del processo di apprendimento, orientata a generare feedback significativi per l'autoregolazione degli studenti e per la regolazione dell'agire didattico degli insegnanti (Shepard, 2000).

2. Metodo

Alla luce di tali riferimenti, l'indagine si propone di far emergere i punti di vista circa le opportunità, criticità, tensioni nonché i processi di rielaborazione soggettiva che si attivano quando le consolidate pratiche valutative degli insegnanti entrano in dialogo con dispositivi di intelligenza artificiale generativa (IAGen). L'obiettivo specifico è quello di esplorare i significati che gli insegnanti attribuiscono all'impiego di ChatGPT nella valutazione della produzione scritta degli studenti.

La ricerca adotta la forma dello *studio di caso* qualitativo a *orientamento esplorativo*, secondo l'impostazione proposta da Stake (1995). L'interesse quindi non è rivolto alla generalizzazione dei risultati ma alla comprensione contestuale delle esperienze individuali. I due casi presi in esame sono due docenti di italiano di scuola secondaria di primo grado selezionati in quanto portatori di posizioni iniziali differenti rispetto all'uso dell'IAGen: si tratta di un docente di circa quarant'anni, con cinque anni di esperienza e inizialmente più scettico rispetto all'impiego dell'IA (docente 1) e di una docente di circa cinquant'anni, con quindici anni di esperienza e una maggiore disponibilità a sperimentare nuove tecnologie (docente 2).

I dati sono stati raccolti attraverso *interviste in profondità* (Minichiello, Aroni & Hays, 2008) progettate per esplorare concezioni e percezioni in merito all'impiego di ChatGPT nella valutazione della produzione scritta degli studenti. L'intervista è stata preceduta da una fase preparatoria in cui ai docenti è stato richiesto, in primo luogo, di selezionare tre elaborati scritti rappresentativi di diversi livelli qualitativi (insufficiente, medio, ottimo) di una medesima classe, già corretti e valutati secondo i loro criteri abituali, e successivamente di compilare una scheda descrittiva per ciascuno studente. Gli elaborati selezionati, previo anonimizzazione rigorosa dei dati, sono stati successivamente analizzati con ChatGPT-4, dopo la trascrizione digitale dei testi e delle consegne. L'interazione con il chatbot è stata strutturata attraverso prompt specifici, elaborati sulla base della letteratura italiana sulla valutazione della scrittura con particolare riferimento alla griglia proposta da Calonghi (1972).

La scelta di adottare come riferimento la griglia di valutazione di Calonghi, anziché i criteri utilizzati individualmente dai docenti, è motivata dalla diversa formazione, esperienza professionale e modalità valutative degli insegnanti coinvolti. Al fine di garantire un quadro di riferimento condiviso e comparabile, si è pertanto optato per uno strumento standardizzato, conosciuto da entrambi i docenti.

Nello specifico, gli elementi presi in esame, così come articolati nella griglia

di Calonghi, hanno riguardato i contenuti, l'organizzazione del testo, la correttezza e la proprietà linguistica (punteggiatura, lessico, ortografia e grammatica). Il software, una volta forniti come riferimento la griglia di Calonghi (1972), il testo dell'elaborato, alcune informazioni socio-anagrafiche dello studente e la traccia proposta dal docente, ha prodotto le valutazioni articolandole secondo le macroaree previste dalla griglia stessa.

Terminata l'elaborazione da parte del software, i docenti sono stati intervistati separatamente mediante una traccia semistrutturata articolata in tre sezioni: (1) esplorazione delle conoscenze pregresse e degli atteggiamenti verso l'IA; (2) analisi degli output generati da ChatGPT e confronto con le loro valutazioni, avvenute invece in maniera tradizionale, come da consuetudine e abitudine dei docenti; (3) riflessioni sull'efficacia, sui limiti e sulle possibili implicazioni dell'uso del chatbot come supporto valutativo. Questa struttura ha permesso di cogliere sia i vissuti soggettivi sia le riflessioni critiche emerse dal confronto tra valutazione umana e automatizzata.

Le interviste sono state trascritte integralmente (verbatim) dopo aver ottenuto il consenso informato, garantendo il pieno rispetto della normativa vigente in materia di protezione dei dati personali. L'analisi ha seguito un approccio descrittivo-tematico per l'identificazione dei temi emergenti (ad esempio la conoscenza generale del software, l'approccio alla valutazione, i punti di forza e di criticità dell'impiego del software) attraverso una codifica iniziale mista (induttiva e deduttiva). Le categorie preliminari sono state integrate da codici generati durante la lettura approfondita delle trascrizioni. L'attenzione analitica è stata rivolta alla descrizione articolata dei punti di vista dei docenti, evitando eccessi di astrazione teorica o inferenze su strutture latenti non supportate dai dati.

3. Risultati

I risultati sono presentati seguendo l'articolazione dell'intervista. Per quanto riguarda la *familiarità con il software*, entrambi gli insegnanti hanno dichiarato di conoscere in modo generale strumenti come ChatGPT, pur senza averne fatto uso diretto. Il docente 1 ha espresso fin da subito una posizione più scettica, evidenziando il rischio di un impiego improprio da parte degli studenti: *«si usano a sproposito, per cose per cui non ci sarebbe bisogno… e si perdono occasioni per allenarsi e imparare»*. Il docente 2, pur riconoscendo la propria limitata esperienza, ha manifestato un atteggiamento più aperto, sostenendo che questi strumenti possano risultare utili se si conoscono le logiche d'uso.

Entrambi i docenti hanno inoltre descritto il proprio approccio valutativo

tradizionale, fondato su criteri strutturati quali correttezza ortografica, coerenza ed esattezza contenutistica e qualità della costruzione sintattica.

Il *confronto* tra le valutazioni dei docenti e quelle prodotte da ChatGPT ha evidenziato una serie di criticità negli output del chatbot. I docenti hanno rilevato una certa ripetitività nella rilevazione degli errori, la confusione tra diverse dimensioni linguistiche (morfologia, grammatica, sintassi) e la tendenza del software a classificare come problematici elementi perfettamente pertinenti al contesto, come nomi propri locali o espressioni idiomatiche. È stata inoltre segnalata la natura spesso generica dei feedback prodotti, percepiti come poco contestualizzati rispetto alle caratteristiche individuali degli studenti. Accanto a tali limiti, i docenti hanno riconosciuto alcuni *punti di forza*, in particolare, la capacità di ChatGPT di individuare errori "minimali", come ad esempio refusi, inversione di alcune lettere all'interno delle parole (pultroppo, sopratuto), che a volte sfuggono a una prima lettura umana e la rapidità con cui il chatbot valuta la pertinenza dell'elaborato rispetto alla traccia, contribuendo a ridurre il carico cognitivo richiesto al docente nelle fasi iniziali di analisi. Le criticità e le potenzialità individuate dai docenti possono essere ricondotte alle macroaree della griglia di Calonghi, che consentono di leggere in modo analitico il funzionamento del software. ChatGPT risulta maggiormente efficace nella dimensione della correttezza e proprietà linguistica, in cui l'analisi si fonda su elementi formalizzabili e ripetitivi, mentre incontra maggiori difficoltà nella valutazione dei contenuti, in particolare rispetto a originalità, stile ed espressività.

È emerso, infine, un disallineamento nei voti. Per gli elaborati di livello alto, i docenti hanno assegnato voti superiori rispetto all'IA; al contrario, per gli elaborati di livello basso, l'IA è stata più generosa dei docenti. Il "docente 1" ha interpretato il disallineamento come indice della mancanza, da parte dell'IA di "intelligenza pedagogica", evidenziando come essa conduca a una "correzione senza comprensione" incapace di cogliere stile, espressività e originalità e quindi poco sensibile agli elaborati di alto livello. Il "docente 2", invece, ha letto lo stesso fenomeno alla luce della funzione comunicativa del voto, sottolineando che *«Io il voto lo uso per comunicare qualcosa al ragazzo, non è solo una somma di errori»*, e spiegando così perché l'IA risulti più generosa con i testi deboli, standardizzando la valutazione verso il centro e non utilizzando il giudizio come segnale formativo intenzionale né in senso correttivo né valorizzante.

Nella parte conclusiva dell'intervista, entrambi gli insegnanti hanno espresso una *valutazione complessivamente* prudente ma interlocutoria dello strumento. ChatGPT è stato considerato potenzialmente utile come supporto, purché non sostituisca il giudizio professionale. Il docente 1 ha ipotizzato una possibile utilità *«nella formazione dei docenti inesperti»*, come occasione di confronto con un mo-

dello esterno che favorisca la riflessione critica sui propri criteri valutativi. Sono tuttavia emersi limiti significativi, tra cui la necessità, per gli insegnanti, di una formazione specifica per un uso consapevole dell'IAGen e l'incapacità del software di integrare variabili contestuali importanti, come il background culturale e linguistico degli studenti, elementi imprescindibili nel processo valutativo educativo.

4. Discussione e conclusione

In coerenza con l'approccio qualitativo e interpretativo adottato, l'attenzione non si è concentrata sull'efficacia tecnica dello strumento, ma sui processi di attribuzione di significato attivati dai partecipanti. Il confronto tra giudizi umani e valutazioni automatizzate ha evidenziato la coesistenza di atteggiamenti contrastanti – apertura e resistenza, curiosità e cautela, fiducia e diffidenza – che riflettono la complessità dell'incontro tra pratiche valutative consolidate e sistemi di nuova generazione. Una delle evidenze centrali emerse riguarda la difficoltà di riconoscere valore pedagogico a un dispositivo che, pur mostrando una certa coerenza nell'analisi formale, non è in grado di cogliere la dimensione semantica, stilistica e relazionale che caratterizza la scrittura scolastica. Si tratta di una resistenza che rimanda a quelle 'barriere di secondo ordine' (Ertmer, 1999), radicate nelle convinzioni profonde del docente sul senso dell'educare. La valutazione emerge così come un'operazione intrinsecamente interpretativa e contestuale, confermando la prospettiva situata descritta da Huot (2009), e appare difficilmente riducibile a un set di regole computazionali. Nel dialogo con i docenti, ChatGPT è apparso potenzialmente utile come supporto, ma solo se collocato all'interno di una cornice riflessiva che ne delimiti la funzione e ne problematizzi l'uso. Per gli insegnanti intervistati, dunque, l'obiettivo non è escludere a priori tali strumenti dalla pratica valutativa, bensì interrogare criticamente le condizioni e le finalità del loro impiego, alla luce di una concezione formativa e dialogica della valutazione (Giovannini 1994; 1995).

Dal punto di vista *metodologico*, lo studio conferma la pertinenza di un approccio qualitativo orientato all'ascolto delle voci degli insegnanti, capace di far emergere non solo dichiarazioni di convinzioni ma anche relative a resistenze, strategie di adattamento e tentativi di rielaborazione soggettiva, in cui i docenti rinegoziano il significato e il ruolo della tecnologia all'interno del proprio quadro professionale. Tale prospettiva si rivela particolarmente feconda in un ambito, quello dell'automazione della valutazione, dove dimensioni tecniche, etiche e pedagogiche si intrecciano in modo complesso e non eludibile.

I risultati indicano che gli insegnanti riconoscono a ChatGPT un'efficacia nell'analisi rapida e coerente degli aspetti formali del testo scritto, ma ne evidenziano i limiti nel cogliere le dimensioni semantica, pragmatica e contestuale proprie del giudizio professionale, inteso come processo interpretativo e relazionale. In questa prospettiva, l'IA può svolgere una funzione di supporto alla valutazione, senza sostituire il ruolo del docente come mediatore del significato, rendendo necessario un approccio ibrido che integri automazione e sensibilità pedagogica.

Lo studio presenta limiti legati al numero ridotto di partecipanti, alla necessità di trascrivere testi manoscritti per l'analisi automatizzata e al carico cognitivo richiesto dalla struttura dell'intervista. Tali elementi suggeriscono la necessità di perfezionare le strategie metodologiche e operative e di approfondire, in future ricerche, modalità di integrazione dell'IA più sostenibili e percorsi di formazione che ne favoriscano un uso critico e consapevole in ambito scolastico.

Riferimenti bibliografici

Broadfoot, P. M., Daugherty, R., Gardner, J., Gipp, C. V., Harlen, W., James, M., & Stobart, G. (1999). *Assessment for learning: Beyond the black box*. University of Cambridge School of Education.

Calonghi, L. (1972). *Valutazione delle composizioni scritte. Indicazioni docimologiche e psicometriche pratiche*. Roma: Armando.

Davis, F. D. (1989). Perceived usefulness, perceived ease of use, and user acceptance of information technology. *MIS Quarterly*, 13(3), p. 319–340.

Giovannini, M. L. (1994). *Valutazione sotto esame*. Milano: Ethel editoriale Giorgio Mondadori.

Giovannini, M. L. (1995). *La valutazione: ovvero, oltre il giudizio sull'alunno*. Milano: Ethel Editoriale Giorgio Mondadori.

Goralski, M., & Tan, T. (2020). Artificial intelligence and sustainable development. *The International journal of Managment Education*.

Ertmer, P. A. (1999). Addressing first-and second-order barriers to change: Strategies for technology integration. *Educational Technology Research and Development*, 47(4), p. 47–61.

Huot, B. (2002). *Rearticulating writing assessment for teaching and learning*. Logan: University Press of Colorado.

Huot, B. A., & O'Neill, P. (Eds.) (2009). *Assessing writing: A critical sourcebook*. Boston: Bedford/St. Martin's.

Jiang, Y., Hao, J., Fauss, M., & Li, C. (2024). Detecting ChatGPT-generated essays in a large-scale writing assessment: Is there a bias against non-native English speakers? *Computers and Education*, p. 1-14.

Kim, H., Yang, J., Chang, D., & Lenke, L. (2024). Assessing the Reproducibility of the Structured Abstracts Generated by ChatGPT and Bard Compared to Human-Written Abstracts in the Field of Spine Surgery: Comparative Analysis. *Journal of Medical Internet Research*, 26, e52001.

Ljubojevi , D., Kadijevich, D., & Gutvajn, N. (2023). Towards a Valid and Reliable Checklist to Evaluate Argumentative Essays Composed by ChatGPT. *CEUR Workshop Proceedings* , 3696, p. 82-90.

Luckin, R. (2018). *Machine Learning and Human Intelligence: The Future of Education for the 21st Century*. London, UCL IOE Press.

Minichiello, V., Aroni, R., & Hays, T. (2008). *In-depth Interviewing: Principles, Techniques, Analysis*. (Third ed.) Camberwell: Pearson Australia Group.

Nardi, A. (2018). Valutare l'apprendimento online: una rassegna degli studi sull'e-testing nel contesto universitario. *Form@ re-Open Journal per la formazione in rete*, 18(1), p. 179-191.

Obata, A., Tagawa, T., & Ono, Y. (2023). Assessment of ChatGPT's Validity in Scoring Essays by Foreign Language Learners of Japanese and English. *Proceedings - 2023 15th International Congress on Advanced Applied Informatics Winter, IIAI-AAI-Winter 2023*, p. 105-110.

Pfau, A., Polio, C., & Xu, Y. (2023). Exploring the potential of ChatGPT in assessing L2 writing accuracy for research purposes. *Research Methods in Applied Linguistics*, 2(3).

Picciano, A. (2024). Graduate Teacher Education Students Use and Evaluate ChatGPT as an Essay-Writing Tool. *Online Learning Journal*, 28(2).

Ruff, E., Engen, M., & Franz, J. (2024). ChatGPT Writing Assistance and Evaluation Assignments Across the Chemistry Curriculum. *Journal of Chemical Education*, 101(6), p. 2483-2492.

Safrai, M., & Orwig, K. (2024). Utilizing artificial intelligence in academic writing: an in-depth evaluation of a scientific review on fertility preservation written by ChatGPT-4. *Journal of Assisted Reproduction and Genetics*, 41(7), p. 1871–1880.

Shepard, L. A. (2000). The role of assessment in a learning culture. *Educational Researcher*, 29(7), p. 4-14.

Stake, R. E. (1995). *The art of case study research*. Thousand Oaks: SAGE Publications.

Tonelli, D., Grion, V., & Serbati, A. (2018). L'efficace interazione tra valutazione e tecnologie: evidenze da una rassegna sistematica della letteratura. *Italian Journal of Educational Technology*, p. 6-23.

Venkatesh, V., Morris, M. G., Davis, G. B., & Davis, F. D. (2003). User acceptance of information technology: Toward a unified view. *MIS Quarterly*, 27(3), p. 425–478.

Wiliam, D. (2011). *Embedded Formative Assessment*. Bloomington: Solution Tree Press.

Zuckerman, M., Flood, R., & Tan, R. (2023). ChatGPT for assessment writing. *Medical Teacher*, 45 (11), p. 1224-1227

Finito di stampare
nel mese di MAGGIO 2026 da



per conto di Pensa MultiMedia® • Lecce
www.pensamultimedia.it

Il volume raccoglie gli esiti del progetto PRIN *AI&F – Artificial Intelligence & Feedback for Effective Learning* e i contributi presentati al convegno finale, dedicato alle trasformazioni della valutazione e del feedback nell'era dell'intelligenza artificiale. Le ricerche documentano prospettive teoriche, pratiche e sperimentazioni sviluppate nei contesti dell'istruzione universitaria, della scuola e della formazione degli insegnanti, esplorando il rapporto tra progettazione della valutazione, processi di feedback e integrazione dell'intelligenza artificiale e le diverse implicazioni didattiche, metodologiche ed etiche che ne derivano. L'IA non viene considerata come una soluzione tecnica o come un sostituto del giudizio umano, ma come un agente con cui docenti e studenti possono collaborare nei processi di esplicitazione dei criteri, interpretazione degli elaborati e revisione delle pratiche valutative. La valutazione è concepita come un processo progettuale, dialogico e distribuito, integrato nelle attività di apprendimento e orientato a sostenere la riflessività e l'autoregolazione. Attraverso contributi teorici e ricerche empiriche, il volume analizza le opportunità offerte dall'IA e le questioni che il suo impiego solleva in termini di validità, equità, autorialità e responsabilità epistemica. Ne emerge un campo di ricerca ancora aperto, che invita ricercatori, docenti e formatori a ripensare il significato della valutazione e il ruolo dell'interazione tra agenti umani e artificiali nella costruzione di pratiche più consapevoli, formative e sostenibili.

Lorella Giannandrea è professoressa ordinaria di Didattica e Pedagogia speciale presso l'Università degli Studi di Macerata. I suoi principali ambiti di ricerca comprendono gli approcci critici e riflessivi alla didattica, la formazione degli insegnanti, i processi di progettazione e valutazione, il feedback e le relazioni tra educazione, tecnologie digitali e intelligenza artificiale.

Francesca Gratani è ricercatrice in *tenure track* in Didattica e Pedagogia speciale presso l'Università degli Studi di Macerata. I suoi principali ambiti di ricerca comprendono la *Maker Education*, le pratiche didattiche innovative, la valutazione formativa e la formazione degli insegnanti, con particolare attenzione all'integrazione delle tecnologie digitali e dell'intelligenza artificiale nei contesti educativi.

Lorenza Maria Capolla è assegnista di ricerca in Didattica e Pedagogia speciale presso l'Università degli Studi di Macerata. La sua attività di ricerca si concentra sulla progettazione didattica e sulla formazione degli insegnanti, con particolare attenzione all'educazione postdigitale, all'incertezza nei contesti di apprendimento e alle implicazioni educative e didattiche delle tecnologie digitali e dell'intelligenza artificiale.