

Veronica Mobilio (a cura di)

La scuola che legge sé stessa

Il valore dei dati e della ricerca educativa per l'analisi e il contrasto alla dispersione scolastica

# Capitolo quattordicesimo Dati per il futuro: analisi predittiva e prevenzione della dispersione scolastica

**di Arianna Beri, dottoranda, Dipartimento di Scienze della persona e del nuovo welfare, Università di Bergamo; Laura Sara Agrati, professoressa ordinaria di Pedagogia sperimentale, Università telematica Pegaso; Marco Lazzari, professore ordinario di Didattica e Pedagogia speciale, Università di Bergamo**

## Abstract

Il capitolo evidenzia la complessità e la non univocità della definizione di dispersione scolastica (dropout), sottolineando come questa eterogeneità comprometta la misurazione e l'efficacia delle politiche. Una revisione sistematica della letteratura (2021-2024) ha messo in luce criticità quali l'uso di strumenti non specifici, la mancanza di dati longitudinali accurati (fondamentali per l'identificazione precoce), e la necessità di una definizione condivisa del costrutto. Viene esplorata la potenziale efficacia dell'Educational Data Mining (EDM) per sviluppare modelli predittivi di allerta precoce, ma si sollevano le problematiche etiche connesse, in particolare il rischio di bias algoritmico e di etichettamento degli studenti. Si raccomanda la collaborazione inter-agenzia per l'integrazione e la condivisione dei dati e la necessità di bilanciare innovazione e tutela etica e pedagogica.



## 1. Introduzione: cos'è la dispersione scolastica?

L'istruzione è essenziale per lo sviluppo e l'accesso universale, eppure circa il 16% della popolazione mondiale in età scolastica non frequenta la scuola [UNESCO 2020]. Questo fenomeno, noto come dispersione scolastica, è un problema globale [Nakajima, Kijima e Otsuka 2018], particolarmente critico nei paesi in via di sviluppo, dove la crisi dell'apprendimento è una priorità [Boucha, Youness e Berrouyne 2024].

La dispersione è stata ampiamente studiata [Zapata-Medina, Bedoya e Jiménez-Builes 2024] a causa dell'alto tasso di abbandono degli ultimi 20 anni [Guzmán et al. 2021] ed è un indicatore utile per chi voglia misurare la continuità scolastica, monitorare il completamento degli studi, valutare la qualità educativa e migliorare i sistemi educativi.

Tuttavia, la dispersione è fenomeno complesso e multidimensionale [Corica e Vicente 2024], influenzato da variabili sociali, economiche e sanitarie, che richiede analisi statistiche accurate per evitare correlazioni spurie [Cruz, Cozman e Souza 2021].

Le ricerche tematiche condotte negli anni si sono concentrate su numerose variabili ricorrenti, quali genere, status socioeconomico e risultati scolastici. Per classificarle, Rumberger [2004] ha individuato due prospettive di analisi: a) individuale, legata a fattori e caratteristiche degli studenti

{p. 248}

(ad es., impegno, atteggiamenti nei confronti della scuola e dello studio e rendimento scolastico); b) istituzionale, incentrata sul contesto educativo e le politiche scolastiche (ad es., assegnazione dei voti e pratiche didattiche).

A questa complessità si aggiunge la difficoltà di definire univocamente il fenomeno della dispersione. Il dibattito scientifico oscilla tra due poli: alcuni studiosi considerano la dispersione come un concetto onnicomprensivo che include qualsiasi irregolarità o deviazione rispetto alla linearità del percorso educativo [Besozzi 2017]; altri la limitano all'abbandono del sistema di istruzione e formazione [Morgagni 1998].

Alla luce di queste problematiche, all'inizio del processo che ha condotto a questo contributo è stata condotta una preliminare analisi comparativa e documentale delle definizioni e degli indicatori adottati per lo studio della dispersione a livello locale e globale [Beri 2024].

L'analisi ha coinvolto: a) agenzie educative internazionali (UNICEF, UNESCO, Eurostat) e 19 Stati membri UE); b) istituzioni educative italiane (Ministero dell'istruzione e INVALSI); c) ulteriori definizioni emerse dalla letteratura più recente. I risultati evidenziano l'assenza di una definizione condivisa di dispersione scolastica, con significative discrepanze tra enti internazionali, europei e nazionali. I criteri adottati per identificare gli studenti dispersi variano in base a fattori quali età, iscrizione, qualifiche e misure di successo scolastico, determinando incongruenze nella misurazione del fenomeno. A livello europeo, il concetto predominante è quello di abbandono scolastico, mentre nella letteratura accademica il termine più utilizzato è dropout. Questa eterogeneità compromette l'affidabilità dei dati e ne limita l'utilità per l'elaborazione di politiche educative efficaci, laddove l'adozione di indicatori più coerenti e predittivi potrebbe migliorare le strategie di prevenzione e intervento [ibidem].

Il fenomeno della dispersione scolastica è studiato da un secolo [Fuller 1927], eppure poche ricerche si sono concentrate sulle tecniche di raccolta e analisi dei dati. L'uso di dati accurati consente di identificare con oltre il 70%

{p. 249}

di precisione gli studenti a rischio di abbandono [Guzman Rincón et al. 2023].

L'uso di dati e informazioni storiche consente di comprendere gli eventi passati, ma anche di prevedere scenari futuri e adottare le misure necessarie per affrontarli [Ferreira e Andrade 2016].

In questo contesto, la dispersione non rappresenta una decisione improvvisa, ma è spesso il risultato di un processo graduale di disimpegno [Haimovich, Vazquez e Adelman 2021], accompagnato da segnali evidenti già nei periodi precedenti, come il rendimento scolastico degli studenti in relazione al loro contesto sociale, familiare ed economico [bin Mat et al. 2013] e i loro stati affettivi [Pardos et al. 2014].

Pertanto, lo sviluppo di un modello predittivo che identifichi con ragionevole accuratezza i bambini a rischio dispersione, con largo anticipo rispetto all'evento, potrebbe facilitare interventi mirati per aumentarne le probabilità di completamento degli studi: la letteratura suggerisce che prevenzione precoce e valutazione del rischio di dispersione dovrebbero iniziare già a partire dai 10 anni d'età [Florin-Bogdan, Teodorescu e Cioclu 2024].

Nonostante i dati sulla dispersione siano fondamentali per definire di politiche educative, allocare risorse e progettare interventi, manca uno standard metodologico consolidato per misurare il fenomeno [Lindhart et al. 2022]. L'assenza di una definizione condivisa e di una metodologia uniforme compromette l'efficacia delle strategie educative [Balfanz et al. 2020]: le interpretazioni dei tassi di dispersione variano in base alle definizioni adottate [Dekkers e Classen 2001] e la loro eterogeneità ostacola la comprensione precisa della gravità e portata del problema, rappresentando un limite significativo per l'attuazione di interventi efficaci.

Un'ulteriore criticità riguarda la percezione del fenomeno da parte di insegnanti e dirigenti scolastici. Studi condotti negli USA evidenziano discrepanze tra i tassi di abbandono riportati a livello nazionale e i dati reali, suggerendo come la mancanza di chiarezza metodologica possa influenzare la rappresentazione del problema [Balfanz et al. 2020].

{p. 250}

## 2. Indagine sui dati della dispersione

Questa ricerca mira a esaminare sistematicamente gli aspetti relativi ai dati della dispersione scolastica, per contribuire a una maggiore chiarezza metodologica e supportare politiche educative più efficaci.

A partire dalle riflessioni riportate nell'introduzione, è stata condotta una revisione sistematica [Pellegrini e Vivinet 2018] della letteratura sui dati della dispersione usando come parola chiave di riferimento e di interrogazione delle banche dati il termine dropout, con un duplice proposito:

- 1) indagare fonti, metodologie e criteri adottati per la raccolta dei dati sul fenomeno;
- 2) mettere a fuoco il significato del termine, lasciando aperta la possibilità di approfondire in futuro altre forme di dispersione non intercettate dall'uso della parola chiave.

Le domande di ricerca, basate sulle linee guida PRISMA [1] [Page et al. 2021], sono:

- 1) Quali sono le fonti dati e le metodologie usate a livello internazionale per la raccolta dei dati sulla dispersione scolastica?
- 2) Quali sono i temi emergenti dalla letteratura?

Alla luce delle problematiche emerse dal quadro teorico, la ricerca vuole analizzare con sistematicità i metodi di raccolta e analisi dei dati sulla dispersione scolastica, evidenziandone il ruolo cruciale nel monitorare, prevedere e ridurre il fenomeno del dropout. Un approccio metodologico chiaro e condiviso permetterebbe infatti di individuare tempestivamente gli studenti a rischio, ottimizzare le strategie di intervento e migliorare l'efficacia delle politiche educative. L'obiettivo è quindi fornire una base solida per sviluppare strategie di prevenzione mirate ed efficaci.

Per selezionare gli studi sono stati adottati i seguenti criteri [Cooper e Hedges 2019]: 1) studi quantitativi, qualitativi, misti o multimetodo; 2) che includono studenti fino

{p. 251}

alla scuola secondaria di II grado, escludendo formazione adulti e universitari; 3) condotti in contesti scolastici o relativi a dati sulla dispersione scolastica; 4) pubblicati tra il 2021 e il 2024 (fino al 10 luglio); e 5) condotti in qualsiasi paese, ma scritti in inglese.

{p. 252}

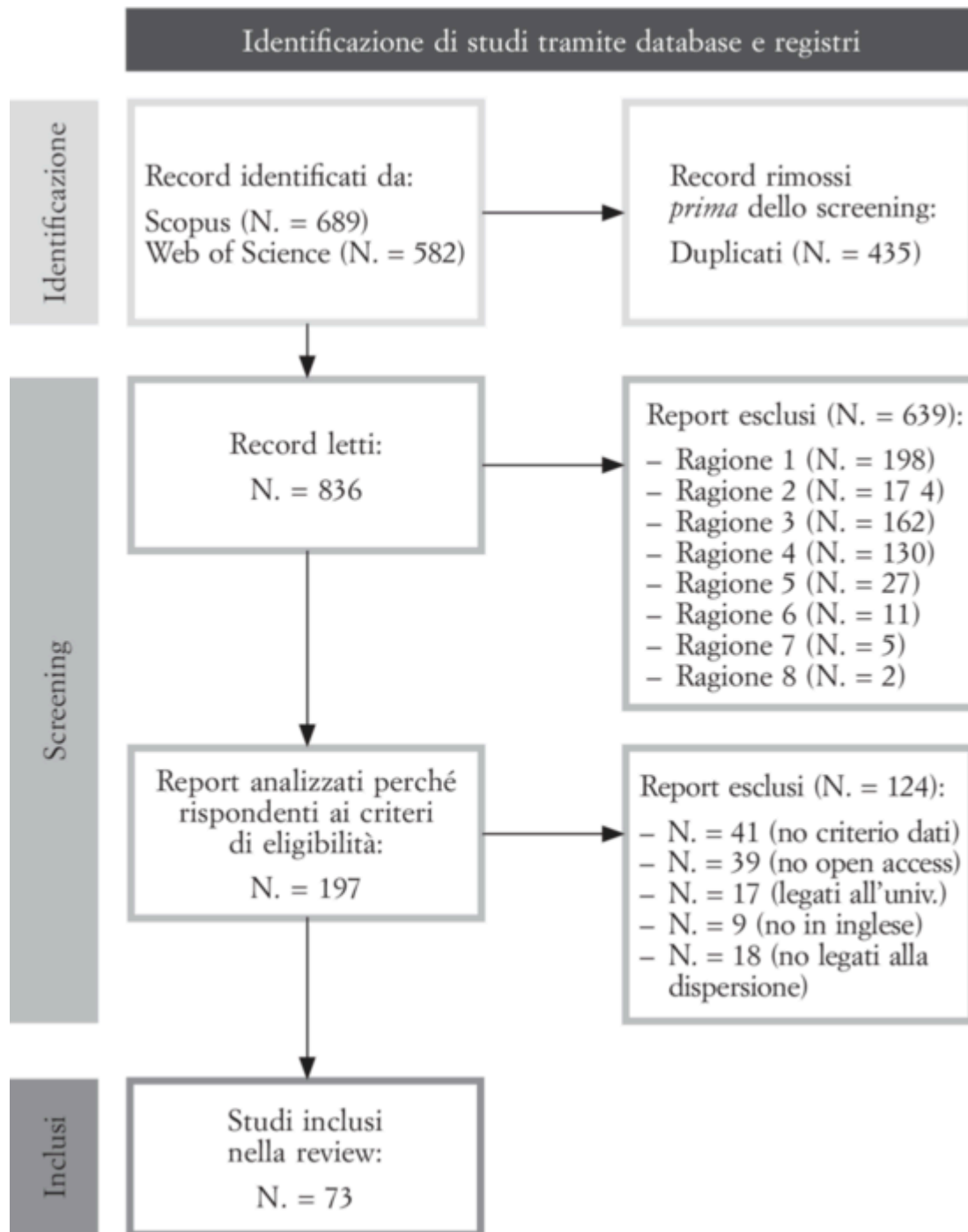


Fig. 14.1. Processo di selezione degli studi.  
Fonte: Adattamento da Page et al. [2021].

La ricerca, condotta con un approccio sistematico, ha fatto riferimento a due database scientifici generali (Scopus e Web of Science), escludendo la letteratura grigia, prevista per future indagini. I termini di ricerca usati sono stati: dropout, school, data. Sulla base delle riflessioni precedenti, il termine dropout è stato privilegiato rispetto a early school leaving. Successivamente gli studi sono stati esportati nel software Rayyan [2] per rimuovere i duplicati e creare l'elenco degli studi da esaminare. Lo screening ha previsto la lettura di titoli e abstract, e l'eliminazione degli articoli che non soddisfacevano i primi tre criteri di ammissibilità. Nella seconda fase, i testi completi degli articoli inclusi sono stati letti, effettuando la selezione finale sulla base dei restanti criteri e della disponibilità open access (fig. 14.1).

Infine, abbiamo sviluppato un modulo di estrazione e indicizzazione dei dati, per registrare le seguenti informazioni: paese di riferimento, tipologia dei dati e database usato, anni di raccolta, campione, problematiche emerse e definizione di dispersione scolastica. Gli studi sono stati classificati in categorie di esito sulla base dei risultati.

## 3. Risultati

### 3.1. Selezione degli studi

Dai due database bibliografici sono stati individuati 1.271 record. Dopo la rimozione dei duplicati, 836 sono stati valutati tramite titolo e abstract. Di questi, 639 sono stati esclusi poiché non trattavano dispersione scolastica, dati o contesti scolastici, o perché relativi ad ambiti accademici e universitari. Dopo l'analisi dei 197 articoli selezionati, 73 studi hanno soddisfatto pienamente i criteri di ammissibilità.

{p. 253}

(fig. 14.1). Il numero è elevato rispetto ad altre rassegne sull'argomento poiché nessuno degli studi analizzati approfondisce specificatamente le problematiche relative ai dati e, per costruire un quadro complessivo, è stato necessario estrapolare informazioni rilevanti attingendo anche a saggi che affrontano solo in parte il tema.



Fig. 14.2. Provenienza dei dati dello studio.

### 3.2. Caratteristiche degli studi

Le principali caratteristiche degli studi scelti sono sintetizzate e supportate da grafici riassuntivi.

\* Paese di riferimento. Il continente più rappresentato è l'America (30,9%), seguito da Europa (24,7), Africa (22,2), Asia (21) e Oceania (1,2) [3] (fig. 14.2).

\* Provenienza dei dati e campione. Vengono illustrati nella tabella 14.1.

{p. 254}

Tab. 14.1. Provenienza dei dati e campione

	N. di studi	%
Provenienza degli studi		
Sistema informatico del ministero	3	4,5
Uso di diversi dataset	2	2,9
Registro scolastico	3	4,5
Dati raccolti nella ricerca tramite strumenti ad hoc	4	5,9
Altri dataset educativi	6	8,8
Sondaggio nazionale	11	16,2
Dati amministrativi	19	27,9
Dataset longitudinale	20	29,4
Campione		
Basato sugli studenti	47	64,4
Basato sul numero di scuole	11	15,1
Nessuna indicazione	15	20,5
Campione basato sugli studenti		
Classe 1: 0-10.000 studenti	30	63,8
Classe 2: 10.001-100.000 studenti	9	19,1
Classe 3: 100.001-1.000.000 studenti	2	4,2
Classe 4: oltre 1.000.000 di studenti	6	12,8
Campione basato sulle scuole		
Sotto le 100 scuole	5	45,5
Tra 100 e 500 scuole	2	18,2
Oltre le 500 scuole	4	36,4
Età del campione		

Scuola dell'infanzia	1	1,8
Scuola primaria	10	17,5
Scuola secondaria di I grado	9	15,8
14-19 anni	29	49,0
Tutti i cicli scolastici	5	8,8
Adulti (dati rif. a quando erano nel sistema scolastico)	4	7,0

\* Anno di conduzione dello studio. Gli studi analizzati sono stati pubblicati tra il 2021 e il 2024, sebbene molti usino dati raccolti in periodi precedenti (fig. 14.3).

{p. 255}

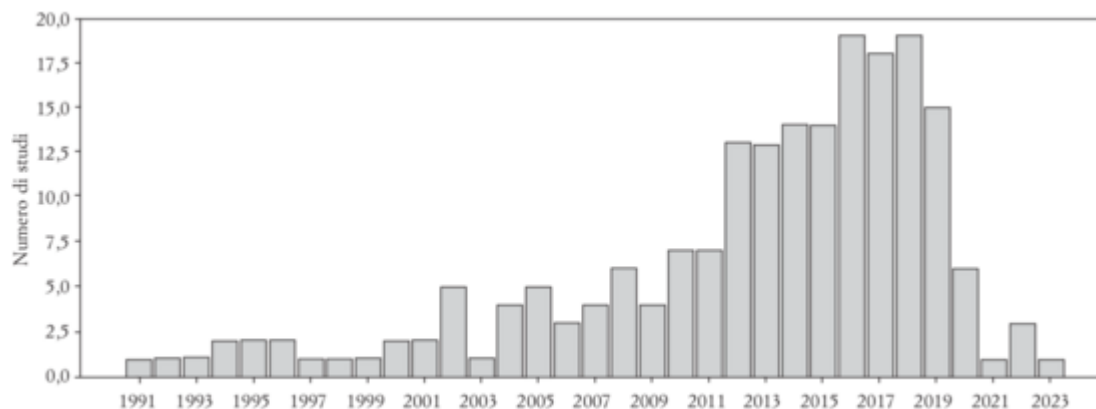


Fig. 14.3. Anno di conduzione dello studio.

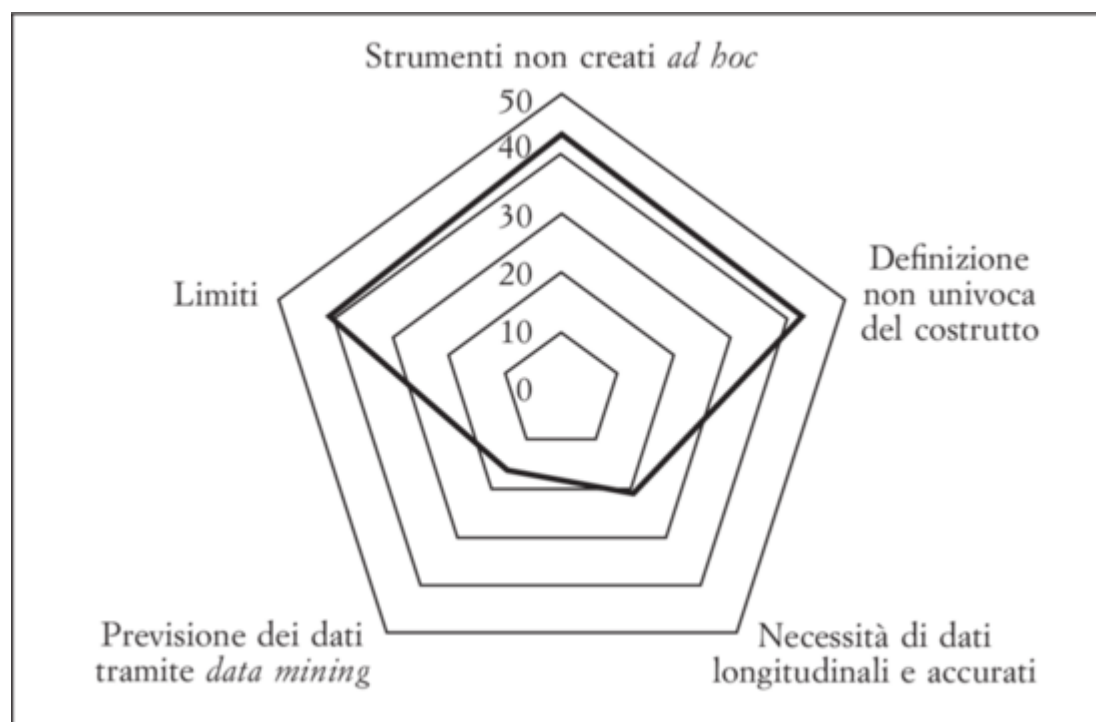


Fig. 14.4. Frequenze riscontrate.

### 3.3. Sintesi dei risultati

Nessuno studio si focalizza esplicitamente sulla dispersione scolastica e l'uso dei dati. Tuttavia, emergono elementi significativi: uso di strumenti non creati *ad hoc* (43 studi), definizione non univoca del costrutto (42), necessità di dati longitudinali e accurati (20), previsione tramite *data mining* (16), limitazioni identificate (41) (fig. 14.4).

#### 3.3.1. Uso di strumenti non creati «*ad hoc*»

Gli studi analizzati non indagavano direttamente la gestione e l'analisi dei dati relativi alla dispersione scolastica, ma piuttosto si concentravano sugli effetti del fenomeno in altri ambiti, come per esempio criminalità [Fu, Grau e Rivera 2022], abuso di alcol [Mitra e Sengupta 2023], chiusura delle scuole durante la pandemia da Covid-19 [Zulaika et al. 2022] o sintomi depressivi [Odoardi et al. 2022]. Ciò ha portato all'uso, nella maggior parte degli studi, di strumenti non progettati appositamente per la dispersione, spesso adattati da scale già esistenti o da questionari ideati

{p. 257}

per altri scopi [Cho e Bae 2024]. In altri casi, la dispersione veniva indagata tramite una singola domanda inclusa in un questionario più ampio, riducendo il fenomeno a un aspetto parziale, come l'abbandono scolastico [Mbebi 2023].

Tali approcci limitati compromettono la possibilità di avere un quadro completo e dettagliato del fenomeno, e si evidenzia la necessità di strumenti specifici per raccogliere, analizzare e interpretare i dati sulla dispersione.

### 3.3.2. Necessità di dati longitudinali accurati

Il 30% degli studi usa modelli longitudinali per raccogliere dati: questa metodologia offre una panoramica approfondita della situazione di iscrizione degli studenti, e coglie fenomeni come la promozione, la ripetizione e l'abbandono [Cho e Bae 2024]. I dati longitudinali, di conseguenza, consentono di monitorare il percorso scolastico e facilitano l'identificazione precoce di chi è a rischio di abbandono [Galvis-Restepo 2022].

Inoltre, una raccolta sistematica di dati longitudinali durante l'anno scolastico, piuttosto che solo alla sua conclusione, potrebbe intercettare per tempo un fenomeno cruciale come l'assenteismo. Per esempio, si pensi a un momento critico come la transizione dalla scuola secondaria inferiore alla superiore, che è caratterizzato da elevato rischio di abbandono scolastico: Markussen, Daus e Bedre [2020] segnalano che l'assenteismo nella scuola inferiore è un segnale precoce di disimpegno che può culminare con l'abbandono scolastico. In questi casi pratiche di monitoraggio potrebbero consentire di intervenire con tempestività.

Studi condotti nei paesi più sviluppati evidenziano forte correlazione tra la ripetizione e l'abbandono scolastico, oltre a mostrare come anche assenze limitate possano compromettere il rendimento [Gershenson, Jackowitz e Brannegan 2016; Klein, Sosu e Dare 2022]. Tuttavia, l'UNESCO [2012] segnala che nei paesi in via di sviluppo la relazione tra ripetizione e abbandono non è sempre chiara, per difficoltà di rilevazione [Cho e Bae 2024].

Un altro aspetto rilevante riguarda l'uso avanzato dei dati longitudinali in alcuni studi (14%). In Danimarca, ad

{p. 258}

esempio, ogni studente ha un identificativo personale univoco, che collega i dati autoriportati con quelli dei registri ufficiali [Lindhart et al. 2022].

Dove l'identificativo unico non era disponibile, sono state usate combinazioni di informazioni come data di nascita, sesso e codice postale per collegare i dati [Putrik et al. 2021]. Questo approccio è stato inoltre impiegato per integrare i dati sugli esiti scolastici con le risposte a sondaggi somministrati agli studenti [Tvedt e Bru 2023]. Questi metodi, consolidati soprattutto in Danimarca, rappresentano un modello efficace per la raccolta accurata di dati sulla dispersione.

I set di dati collegati permettono di ottenere informazioni fondamentali circa background familiare, caratteristiche scolastiche, rendimento e altre variabili, e offrono una visione più completa delle condizioni degli studenti [Fu, Grau e Rivera 2022].

### 3.3.3. Definizione non univoca del costrutto

L'indicatore prevalente usato per definire la dispersione scolastica è stato il termine dropout (89%), con solo uno studio che ha impiegato il termine early school dropout (1,4%) e sette studi che hanno adottato early school leaving (9,6%).

Dei 73 studi, circa il 58% ha fornito una definizione esplicita di dispersione scolastica; tuttavia, un'analisi preliminare ha rivelato che le definizioni adottate non erano sempre coerenti tra loro. Nella maggior parte dei casi (35,7%), il termine è stato ridotto al concetto di abbandono, definito come: «chi lascia la scuola e non si iscrive nuovamente durante quel periodo» [Smith Uldall e Gutierrez Rojas 2022] o «persona che smette di andare a scuola prima di aver terminato» [Mbogo et al. 2022].

Altri studi (28,6%) hanno legato la dispersione al conseguimento del diploma, come nel caso della definizione di «conseguimento del diploma entro 5 anni» [Askeland et al. 2022] o «ottenimento di qualsiasi titolo di studio secondario superiore, sia professionale che accademico, entro i 22 anni» [McMullin et al. 2021].

{p. 259}

Alcuni studi (19%) si sono basati solo sulla differenza tra iscritti all'inizio e alla fine dell'anno scolastico, senza analizzare i flussi interni degli studenti [Ronak e Rashmini 2021]. Un gruppo ristretto (4,8%) ha invece considerato la dispersione legata al fallimento nell'ottenere risultati scolastici adeguati, ossia «chi rimane, ma non riesce a superare tutte le materie» [Sletten, Tøge e Malmberg-Heimonen 2022], o a parametri di rendimento, ritardo e impegno scolastico come il Dropout Prediction Index [Janosz et al. 2007].

Altri studi (7,1%) hanno approfondito la problematica della definizione [Alexa e Baciu 2021; Ndour 2024], mentre il 4,7% ha adottato una definizione più ampia, come ad esempio: «ultima fase nel processo di disimpegno che si manifesta precocemente con un cattivo rendimento nell'impegno scolastico» [Zanoni 2023] o «ultima fase di un processo dinamico e cumulativo di disimpegno dalla scuola» [Putrik 2021].

Un aspetto interessante emerso da uno studio è che gli studenti che «hanno cambiato istituto scolastico non sono stati definiti come dropouts, a meno che non abbiano intrapreso un nuovo tipo di programma educativo o abbiano ripreso l'anno scolastico in una nuova scuola» [Pisinger et al. 2023]. Questo aspetto, talvolta trascurato negli studi sulla dispersione, sottolinea l'importanza di analizzare non solo i flussi in uscita, ma anche i cambiamenti interni nel percorso educativo degli studenti.

### 3.3.4. Previsione del fenomeno

Sedici studi hanno usato i dati in ottica predittiva, avvalendosi dell'Educational Data Mining (EDM), inteso come l'uso di tecniche di data mining per analizzare ed estrarre conoscenze nascoste dai dati educativi [Oppong 2023].

I modelli di apprendimento automatico, addestrati su dati storici, identificano schemi e relazioni tra fattori di rischio, come la dispersione scolastica, e metriche comportamentali o di performance a essa correlate [Xu e Wilson 2021]. Diversi studi hanno tentato di prevedere la dispersione usando l'apprendimento automatico. Molti di questi approcci si

{p. 260}

basano su indagini estese e complesse [Márquez-Vera et al. 2016], che generano dataset con un elevato numero di caratteristiche, ma risultano applicabili solo a un numero limitato di studenti. Altri studi, seppur meno frequenti, sfruttano dati estratti direttamente dai sistemi informativi accademici [Sales, Balby e Cajueiro 2016].

Le ricerche hanno sviluppato o sfruttato piattaforme digitali, esistenti o create ad hoc, per individuare gli studenti a rischio dispersione, promuovere il loro coinvolgimento e fornire supporto personalizzato [Ebabu et al. 2024]. In questi casi, una volta identificato uno studente a rischio (per assenze o presenze irregolari), le istituzioni vengono avvisate della situazione e supportate nella gestione del caso. Un sistema di prevenzione ben implementato può ridurre significativamente il tasso di abbandono. Gli insegnanti, in particolare, giocano un ruolo cruciale nell'uso di questi strumenti, monitorando regolarmente la frequenza scolastica e intervenendo tempestivamente [Márquez-Vera et al. 2016].

Tuttavia, la previsione della dispersione non può basarsi su un modello unico, poiché gli studenti a rischio non presentano necessariamente le stesse caratteristiche [Ebabu et al. 2024]. Le ricerche analizzate hanno considerato un'ampia gamma di variabili per la costruzione dei modelli predittivi, tra cui lo status sociale [Melo et al. 2022], il rendimento scolastico, il genere dello studente e il livello di istruzione dei genitori [Karhina et al. 2024].

### 3.3.5. Limitazioni emerse

Alcuni studi (22%) riportano tra le principali limitazioni l'adozione di definizioni di dispersione non corrette o non allineate a quelle ufficiali, un problema che rischia di compromettere l'accuratezza della rappresentazione del fenomeno. Ad esempio, in alcuni casi venivano computati erroneamente studenti trasferiti in altri comuni senza informare la scuola, sovrastimando così il tasso di dispersione [Serban 2024].

Connesso a ciò, il 9,7% degli studi evidenzia problemi legati al campione, come l'inclusione di studenti non in

{p. 261}

età scolastica secondo le definizioni ufficiali di dispersione o l'uso di campioni troppo piccoli. Inoltre, il 4,9% ha segnalato l'impiego di strumenti inadeguati, spesso progettati per altri scopi [Askeland et al. 2022].

La maggior parte delle limitazioni (39%) è risultata connessa alla difficoltà di ottenere dati accurati, con il rischio di distorsioni in caso di dati approssimativi [Mbogo et al. 2022] o di analisi che forniscono visioni parziali del problema [Cho e Bae 2024]. Un ulteriore ostacolo evidenziato dal 12,2% degli studi è la complessità della raccolta di dati longitudinali, una difficoltà particolarmente accentuata nei paesi in via di sviluppo [Jack e Oster 2023].

Alcuni studi (4,9%) sottolineano le sfide legate all'integrazione dei dati derivati da diverse agenzie. Questi dati, spesso in formati diversi e provenienti da fonti eterogenee, richiederebbero una comunicazione efficace tra i vari sistemi, ma le difficoltà tecniche nella loro armonizzazione costituiscono un ulteriore problema [Cho e Bae 2024].

Infine, il 12,2% delle limitazioni è risultato legato ai processi di data mining. Da un lato, approcci complessi come quelli descritti da Márquez-Vera e colleghi [2016] richiedono indagini estese che producono insiemi di dati caratterizzati da molteplici variabili, ma sono frequentemente applicati a campioni di dimensioni limitate. Dall'altro lato, i dati raccolti possono essere intrinsecamente distorti, e alcuni fattori chiave, come l'accesso alle risorse educative, potrebbero non essere stati adeguatamente considerati nei modelli algoritmici [Crawford 2017].

## 4. Discussione

Secondo Delen [2011] esistono due principali approcci per affrontare il problema della dispersione scolastica: 1) basato su sondaggi; 2) guidato dai dati (analitico). Nel primo approccio, vengono sviluppati modelli teorici come quello di Tinto [4], mentre nel secondo metodo i dati istituzionali

{p. 262}

sono analizzati attraverso tecniche di data mining [Gustian e Hundayani 2017].

La disamina degli studi mostra che entrambi i metodi trovano applicazione, e mette in luce la varietà delle strategie adottate. Alcuni studi hanno avviato la creazione di sistemi di dati longitudinali a livello statale (come Cile, Filippine, Norvegia e Danimarca), mentre altri (come Canada, Stati Uniti e India) usano queste informazioni per fornire allerta precoce a distretti e scuole locali, o pianificano di farlo [Balfanz et al. 2020].

In generale, molti studi concordano sulla necessità di sistemi di dati integrati e interconnessi per migliorare il processo decisionale. I ministeri e le agenzie educative dovrebbero collaborare per garantire collegamenti tra i sistemi e fornire un accesso adeguato alle parti interessate, inclusi educatori, decisori, studenti, genitori, ricercatori, organizzazioni non profit e comunità locali, nel rispetto della privacy degli studenti, una questione particolarmente dibattuta [ibidem].

Tuttavia, disporre di dati accurati non è sufficiente: i leader statali devono anche potenziare la capacità delle parti interessate di usare le informazioni. Ciò richiede formazione e sviluppo professionale, per garantire che gli educatori comprendano e sfruttino efficacemente i dati per le proprie decisioni [ibidem].

Inoltre, sarebbe opportuno promuovere la condivisione dei dati oltre le singole scuole, considerando l'elevata mobilità della popolazione studentesca. Questo permetterebbe di generare report più accurati, evitando di classificare erroneamente come abbandono il trasferimento degli studenti tra segmenti educativi, paesi o scuole [Bhat e Khan 2021].

Un tema rilevante riguarda la creazione di sistemi di dati longitudinali che non solo segnalino i tassi aggregati di dispersione, ma permettano di monitorare il percorso dei singoli studenti dalla prima infanzia fino all'istruzione

{p. 263}

superiore, post-secondaria e al mondo del lavoro [Carson et al. 2010]. Alcuni paesi, come la Norvegia, hanno già implementato sistemi efficaci in tal senso [Tvedt e Bru 2023], offrendo modelli utili per lo sviluppo di future politiche educative. I meccanismi di tracciamento degli studenti dovrebbero essere resi obbligatori, includendo sondaggi annuali per identificare chi ha abbandonato la scuola e analizzare le motivazioni dietro tale decisione. Ciò consentirebbe di adottare interventi tempestivi e mirati [Putrik et al. 2021].

L'uso di sistemi di allerta precoce è cruciale per affrontare la dispersione, poiché consente interventi preventivi piuttosto che reazioni tardive. Quando i fattori di rischio vengono identificati in una fase iniziale, è possibile attuare misure mitigative, come la fornitura di supporto educativo supplementare [Jiménez-Gutiérrez et al. 2024]; nel contempo, l'effetto dei fattori non modificabili può essere attenuato, per esempio coinvolgendo studenti provenienti da contesti familiari svantaggiati in attività sociali e sportive strutturate [Audas e Willms 2001].

Diversi studi evidenziano inoltre il ruolo predittivo del rendimento scolastico pregresso nel completamento del percorso educativo [Gubbels, van der Put e Assink 2019]. Il rendimento è stato ripetutamente identificato come il fattore determinante per il completamento

della scuola [Markussen, Frøseth e Sandberg 2011].

Queste evidenze giustificano l'impegno per una previsione precoce della dispersione [Chung e Lee 2019], sottolineando l'importanza di adottare strategie innovative di prevenzione e contrasto.

## 5. Implicazioni e conclusioni

È stata condotta una systematic review per colmare le lacune nella ricerca sulla dispersione scolastica, prestando particolare attenzione all'accesso ai dati. Malgrado il riconosciuto valore della scuola nel processo di sviluppo umano [Stanica 2019] e la vasta quantità di ricerche disponibili, i

{p. 264}

tassi di abbandono scolastico continuano a essere elevati, con ripercussioni sia a livello individuale che sociale.

L'analisi ha evidenziato diverse criticità, tra cui: a) uso di strumenti non ad hoc, che ostacolano il confronto tra studi diversi; b) mancanza di una definizione chiara di dispersione scolastica, con differenze nelle metriche adottate; c) necessità di dati longitudinali per tracciare i percorsi educativi degli studenti nel tempo e per evitare errori nell'etichettare come «disperso» chi segue percorsi alternativi; e d) potenzialità del data mining, che offre nuove prospettive di analisi e intervento, ma richiede protocolli etici e metodologici solidi.

I risultati ottenuti indicano che le istituzioni educative dovrebbero implementare algoritmi predittivi per il monitoraggio dei dati, per migliorare la capacità di tracciamento degli studenti, identificare chi ha abbandonato il percorso scolastico e analizzare le motivazioni sottese a tale scelta. Ciò consentirebbe di adottare interventi tempestivi e mirati [Putrik et al. 2021].

Perché questi strumenti siano realmente efficaci, è fondamentale migliorare raccolta, archiviazione e aggiornamento dei dati, includere variabili come benessere psicologico e supporto familiare, che influenzano il rendimento e il rischio di abbandono [Mbogo et al. 2022], e definire standard condivisi tra le diverse agenzie educative.

È essenziale che le istituzioni collaborino nella definizione e condivisione di costrutti e indicatori per misurare la dispersione, ampliandone la definizione per includere anche altre forme di difficoltà educative e promuovere così un approccio coordinato ed efficace.

Tuttavia, l'adozione di sistemi predittivi solleva problematiche etiche e legate alla privacy. Il monitoraggio sistematico degli studenti richiede una gestione trasparente dei dati per prevenire bias algoritmici, che possono svantaggiare gruppi sociali già in difficoltà qualora i dati storici riflettano disuguaglianze preesistenti. Esiste poi il rischio di etichettamento precoce, che potrebbe dar vita a profezie autoavveranti. È infine fondamentale considerare i rischi per la privacy, nel caso in cui i dati raccolti non venissero gestiti secondo

{p. 265}

protocolli di sicurezza adeguati. Per affrontare queste sfide, è fondamentale sviluppare linee guida etiche per l'uso del data mining in ambito educativo, bilanciando innovazione e tutela dei diritti degli studenti.

In conclusione, l'analisi dei dati fornisce prospettive importanti per l'esame, la comprensione e la modellazione dei processi pedagogici, oltre a sostenere un processo decisionale fondato su evidenze, orientato a interventi mirati in risposta alle tendenze globali. D'altro lato, l'efficacia dei programmi di contrasto dipende strettamente dalla chiarezza delle definizioni adottate e dalla precisione dei tassi di dispersione. Una definizione non adeguata o un uso impreciso degli indicatori possono indurre significativi sprechi di risorse in termini di energie e di fondi da allocare, ma anche compromettere la capacità dei programmi stessi di raggiungere i propri obiettivi.

Per il futuro, sarà quindi necessario promuovere la collaborazione tra agenzie nella raccolta e analisi dei dati per sviluppare strategie basate su evidenze; bisognerà anche migliorare la formazione di insegnanti e operatori scolastici, perché usino e interpretino i dati consapevolmente. Inoltre, sarà opportuno integrare metodi di ricerca qualitativi e quantitativi, per comprendere meglio le cause della dispersione. Un approccio basato sui dati, ma attento agli aspetti etici e pedagogici, può contribuire a costruire sistemi educativi più inclusivi ed efficaci, a ridurre il fenomeno del dropout e a garantire pari opportunità di apprendimento a tutti.

## Appendice. Gli studi inclusi

Tab. A1.

Articolo	Autore e anno	Paese	Principali caratteristiche
Chronic school absenteeism for health-related reasons among children ages 5-17 years: United States, 2022	Black e Elgadda 2024	Stati Uniti	– sondaggio nazionale – dispersione individuata con una sola domanda
Machine learning predicts upper secondary education dropout as early as the end of primary school	Psyridou et al. 2024	Fin-landia	– dati longitudinali – mancanza di dati – previsione della dispersione
An early warning system for school dropout in the State of Espírito Santo: A machine learning approach with variable selection methods	Pereira et al. 2024	Brasile	– dati educativi – previsione della dispersione
Determinants of school dropouts and the impact on youth unemployment: Evidence from Ethiopia	Sileshi et al. 2024	Etiopia	– sondaggio nazionale – dispersione definita sulla base del diploma
Unwinding the tangle of adolescent pregnancy and socio-economic functioning: Leveraging administrative data from Manitoba, Canada	Jakubowski et al. 2023	Canada	– dati amministrativi – mancanza di dati – dispersione definita sulla base del diploma
School dropout in Cameroon and its determinants	Mbebi 2023	Camerun	– necessità di dati longitudinali – sondaggio nazionale – solo due domande relative alla dispersione
A methodology to design, develop, and evaluate machine learning models for predicting dropout in	Rodríguez	Cile	– dati amministrativi gestiti dal ministero – incongruenze tra i dati

school system: The case of Chile

et al. 2023

– previsione

Alcohol intake and academic performance and dropout in high school: A prospective cohort study in 65,233 adolescents

Hijarnaa et al. 2023

Danimarca

– sondaggio nazionale  
– dispersione come raggiungimento del diploma  
– correlazione tra variabili

(segue)

Tab. A1. Segue

Articolo

Autore e anno

Paese

Principali caratteristiche

An explainable machine learning approach for student dropout prediction

Kruger et al. 2023

Brasile

– database scolastico  
– dati mancanti  
– previsione della dispersione

Birthweight, gestational age, and early school trajectory

Gansa-onré et al. 2023

Burkina Faso

– dati longitudinali  
– dispersione come abbandono

Development of an early warning system to support educational planning process by identifying at-risk students

Skittou et al. 2023

Marocco

– unione di diversi dataset  
– analisi del «ritardo accademico»

School integration of Syrian refugee children in Turkey

Kirdar et al. 2023

Turchia

– indagine nazionale  
– impossibile studio longitudinale

Does socioeconomic status impact the relationship between school absence and outcomes?

Mooney et al. 2022

Australia

– dati longitudinali  
– dispersione come assenza

The role of learning in school persistence and dropout: A longitudinal mixed methods study in four countries

Kaffenberger et al. 2023

Etiopia, India, Perù, Vietnam

– dati longitudinali  
– dispersione collegata ad altre variabili

Longitudinal trajectories of the career barriers of school dropouts using latent class growth analysis

Cho et al. 2023

Corea del Sud

– dati panel  
– adattamento di un precedente questionario

Determinants of school dropouts among adolescents: Evidence from a longitudinal study in India

Kumar et al. 2023

India

– dati longitudinali  
– possibilità di sovrastima del dato

School dropouts in the Philippines: Causes, changes and statistics

Parreno 2023

Filip-pine

– dati amministrativi  
– dispersione come termine dell'anno

A scientific approach to addressing social issues using administrative data

Green et al. 2023

Canada

– dati amministrativi  
– unione di diversi dataset

(segue)

{p. 268}

Tab. A1. Segue

Articolo

Autore e anno

Paese

Principali caratteristiche

School dropout prediction and feature importance exploration in Malawi using household panel data: Machine learning approach

Colak Oz 2023

Malawi

– indagine nazionale  
– previsione

Self-reported mental health in adolescents attending school and its association with later school dropout: A prospective 2.5-year follow-up study

Lind-hardt et al. 2022

Danimarca

– utilizzo del numero unico  
– dispersione come non completamento

Effects of an early warning system on student absence and completion in norwegian upper secondary schools: A cluster-randomised study

Sletten et al. 2021

Nor-vegia

– dati amministrativi  
– le scuole non hanno lo stesso metodo di registrazione

On the use of explainable artificial intelligence to evaluate school dropout

Melo et al. 2022

Brasile

– addestrare i dataset può causare la reiterazione di pregiudizi

Detection of hotspots of school dropouts in India: A spatial clustering approach

Venkatesan e Mappillairaju 2023

India

– difficoltà di ottenere dati  
– dati aggregati

Expanding the understanding of high school non-graduates through a comparison of high school dropouts and persisters

Uretsky e Henneberger 2023

Stati Uniti

– dati amministrativi

Early prediction of at-risk students in secondary education: A countrywide k-12 learning analytics initiative in Uruguay

Queiro-ga et al. 2022

Uru-guay

– dati amministrativi  
– unione di diversi database  
– previsione

Parental separation and school dropout in adolescence

Karhina et al. 2024

Nor-vegia

– sondaggio nazionale  
– unione con stato socioeconomico  
– uso di definizione statale

(segue)

{p. 269}

## Tab. A1. Segue

Articolo	Autore e anno	Paese	Principali caratteristiche
Genetically optimized ensemble classifiers for multiclass student performance prediction	Begum e Padmannavar 2022	Portogallo	– uso di diverse fonti – problematiche di dati inesistenti e sovrapposti
Educational trajectories, traceability, and self-perception of grades in post-compulsory transitions	Gracia et al. 2022	Spagna	– dati longitudinali – creazione dello strumento
A dataset of dropout rates and other school-level variables in Louisiana public high schools	Stein et al. 2022	Stati Uniti	– dataset scolastici
EduBot: A proof-of-concept for a high school motivational agent	Faria et al. 2022	Portogallo	– dispersione come fallimento – previsione
Gifting relationships and school dropout in rural Malawi: Examining differences by gender and poverty level	Pike e Grant 2022	Malawi	– dati longitudinali
Comparative analysis of schooling situations using longitudinal data: Improving enrollment patterns of individual children in Honduras	Ashida 2021	Honduras	– dati longitudinali
Does lack of parental involvement affect school dropout among Indian adolescents? Evidence from a panel study	Ronak et al. 2021	India	– studio panel – indagini nazionali (questionario famiglie)
Consequences of school closure on access to education: Lessons from the 2013-2016 Ebola pandemic	Smith 2021	Sierra Leone, Guinea, Costa d'Avorio	– indagini nazionali – dispersione come mancata frequenza
Accounting for repetition and dropout in contemporaneous cross-section learning profiles: Evidence from Rwanda	Crawford 2021	Ruanda	– sondaggio nazionale rappresentativo – dati del ministero

(segue)

{p. 270}

## Tab. A1. Segue

Articolo	Autore e anno	Paese	Principali caratteristiche
Strength and Difficulties Questionnaire (SDQ) score as a predictor of early school dropout: Results from a longitudinal study in 24,988 Dutch children	Putrik et al. 2021	Paesi Bassi	– studio longitudinale – la definizione usata non è in linea con quelle formali
School dropout in satellite towns around Bucharest, Romania	Petre et al. 2024	Romania	– dispersione come differenza entrata/uscita
The effect of free tuition in private high schools on the high school dropout rate	Nomoto 2024	Giappone	– dati ministeriali – dispersione come abbandono
Beyond education dollars: Does social safety net spending affect high school graduation rates?	Kukla-Acevedo e Acevedo-Polakovich 2024	Stati Uniti	– database costruito per lo studio
Violence and children's education: Evidence from administrative data	Duque et al. 2024	Colombia	– dati amministrativi – informazioni di tutte le scuole
Prediction of school dropout outside school setting: Potential for early risk stratification by youth health care services in the Netherlands. Results from a retrospective cohort study	Putrik et al. 2024	Paesi Bassi	– registri nazionali – definizione non in linea con quelle formali
Determinants of school dropouts in India: A study through survival analysis approach	Garg et al. 2024	India	– indagine sulle famiglie – necessità di dati longitudinali
The effects of grade retention on high school outcomes: Evidence from New York City school	Mariano et al. 2024	Stati Uniti	– dati amministrativi – unione di diverse fonti dati
Application of the performance of machine learning techniques as support in the prediction of school dropout	Jiménez-Gutiérrez et al. 2024	Messico	– analisi dei dati di dispersione con altre variabili

(segue)

{p. 271}

## Tab. A1. Segue

Articolo	Autore e anno	Paese	Principali caratteristiche
----------	---------------	-------	----------------------------

Poorer self-reported mental health and general health among first year upper secondary school students do not predict school dropout: A five-year prospective study	Goll et al. 2024	Nor-vegia	– progetto longitudinale – dati autoriportati quindi a rischio correttezza
Prospective associations between adolescent risky substance use and school dropout and the role of externalising and internalising problems	Heradstveit et al. 2024	Nor-vegia	– sondaggio e registro dati educativi – definizione statale
Spatially-informed insights into early marriage and school dropout: An advanced bivariate binary multilevel model for understanding Ethiopia's contex	Ebabu et al. 2024	Etiopia	– disallineamento di definizioni – indagine nazionale
School dropout report-behind the numbers of dropouts from compulsory education in the Kosovo	Kamberi 2024	Kosovo	– sistema informatico del ministero – mancanza di dati
Understanding school dropout in Senegal: Exploring the impact of governance factors	Ndour 2024	Senegal	– indagini nazionali – riflessioni sulle definizioni
Analysing the dropout patterns of individual children in Nepal: Relationship between school dropout and grade repetition/entrance age	Ezaki 2024	Nepal	– dati longitudinali – raccolta difficile dei dati
Completion of upper secondary education: Predictions of the psychosocial learning environment and academic achievement	Tvedt e Bru 2023	Nor-vegia	– dati di registro e autosegnalazioni degli studenti – dispersione come abbandono
Do slum upgrading impact school attendance?	Zanoni et al. 2021	Uruguay	– dati amministrativi – necessità di dati più dettagliati

(segue)

{p. 272}

Tab. A1. Segue

Articolo	Autore e anno	Paese	Principali caratteristiche
Local food price volatility and school dropout in Sub-Saharan Africa	Brawn et al. 2023	14 paesi africani	– indagini nazionali – dati non accurati
Absenteeism, dropout, and on-time school completion of vulnerable primary school students in Ethiopia: Exploring the role of adolescent decision-making	Woldehanna et al. 2021	Etiopia	– dati longitudinali – problematiche relative ai costi
School education and tribal children of Jammu and Kashmir: Analysing trends in enrolment, gender parity and dropout rate	Bhat e Khan 2021	India	– dati provenienti da diversi periodi
Student engagement in vocational education and training: Differential analysis in the province of Valencia	Crespo et al. 2022	Spagna	– strumento creato ad hoc – dispersione come abbandono
School dropout: The role of childhood conduct problems and depressive symptoms	Lau et al. 2021	Canada	– dati longitudinali – dispersione come abbandono prima del diploma
School victimization, immigration, dropping out, and gender disparities	Peguero et al. 2021	Stati Uniti	– studio longitudinale – monitoraggio di studenti nel tempo
Sleep and dropout from upper secondary school: A register-linked study	Hysing et al. 2023	Nor-vegia	– dati educativi ufficiali – dispersione come non raggiungimento del diploma nei 5 anni
Online credit recovery school-level enrollment: Intended and unintended consequences	Viano 2023	Stati Uniti	– dati amministrativi – inclusione dei voti degli studenti
Is educational wellbeing associated with grade repetition and school dropout rates among Indian students? Evidence from a panel study	Ronak e Rashimi 2021	India	– dataset longitudinale – dati incompleti (morte, trasferimento o altro)

(segue)

{p. 273}

Tab. A1. Segue

Articolo	Autore e anno	Paese	Principali caratteristiche
Mental health groups in high school students and later school dropout: A latent class and register-based follow-up analysis of the Danish National Youth Study	Andersen et al. 2021	Danimarca	– dati amministrativi – questionario creato per un altro studio

Geographical mobility and children's non-completion of upper secondary education in Finland and Germany: Do parental resources matter?	McMullin et al. 2021	Germania e Finlandia	– dati amministrativi e registri nazionali – campione adulto ma dati del periodo scolastico
Investigation of high school students' dropout risk level.	Zergin 2021	Turchia	– utilizzo di una scala Likert – dispersione come abbandono
Grad nation: Building a grad nation, progress and challenge in ending the high school dropout epidemic, 2020-2021 annual update	Balfanz et al. 2021	Stati Uniti	– diverse problematiche legate ai dati
Childcare attendance and academic achievement at age 16 years	Larose et al. 2021	Regno Unito	– studio longitudinale – molti dati mancanti
School dropout and early school leaving in Romania: Tendencies and risk factors	Alexa e Baciu 2021	Romania	– differenza tra dispersione e abbandono
Association of depressive symptoms in late adolescence and school dropout	Askeland et al. 2022	Nor-vegia	– utilizzo identificativo personale
Impact of Covid-19 lockdowns on adolescent pregnancy and school dropout among secondary schoolgirls in Kenya	Zulaika et al. 2022	Kenya	– dati del registro scolastico – studenti dispersi contattati personalmente
Stopped by the police: The end of «stop-and-frisk» on neighborhood crime and high school dropout rates	Tebes e Fagan 2022	Stati Uniti	– registri amministrativi – dati sulle iscrizioni

(segue)

{p. 274}

Tab. A1. Segue

Articolo	Autore e anno	Paese	Principali caratteristiche
An application of machine learning in public policy: Early warning prediction of school dropout in the Chilean public education system	Smith Uldall e Gutierrez Rojas 2022	Cile	– dati ministeriali – studio longitudinale
Prediction of school dropout in rural Antioquia, Colombia, using machine learning: Improving targeting and identifying important predictors	Galvis-Restepo 2022	Colombia	– dati ministeriali – dati mancanti
Coping with access difficulties and absenteeism through data visualization: A case study from a rural vocational school in Northern Greece	Samaras et al. 2022	Grecia	– dati scolastici uniti ad altre variabili – dispersione come abbandono

## Note

[1] PRISMA è l'acronimo di Preferred Reporting Items for Systematic reviews and Meta-Analyses. Si tratta della guida più rigorosa nel reporting delle revisioni sistematiche [Maraolo 2021] (<https://bit.ly/ariprisma>).

[2] Rayyan è un software online utilizzato per organizzare, gestire e accelerare il processo di revisione della letteratura (<https://www.rayyan.ai>).

[3] La somma dei paesi citati non corrisponde al numero totale degli studi inclusi, perché alcuni studi non indicano il paese di riferimento, mentre altri hanno coinvolto più paesi. Questo aspetto ricorre anche per le successive caratteristiche.

[4] La teoria dell'abbandono degli studenti di Tinto [1993] è la più comunemente citata per spiegare il processo di abbandono scolastico. Nel tempo, questa teoria ha assunto quasi uno status paradigmatico nel campo dell'istruzione superiore ed è tuttora alla base degli obiettivi istituzionali della maggior parte delle università negli Stati Uniti [Barnett 2010].

Copyright © 2025 by Società editrice il Mulino - [Legal notice](#)

Per le opere presenti in questo sito si sono assolti gli obblighi dalla [normativa sul diritto d'autore](#) e sui diritti connessi.