Report di novembre 2026

I numeri vanno dal 1° ottobre al 5 novembre 2026. Pubblicato il 6 novembre 2026.

Ed è un tutor educativo con intelligenza artificiale. Dal 1° ottobre al 5 novembre 2026 abbiamo lavorato ogni giorno sul modo in cui risponde, ricorda, prepara le attività e dice se le risposte degli studenti sono giuste. Questo report racconta che cosa è cambiato, coi numeri del periodo.

Un cambiamento entra in Ed solo se supera la regola che abbiamo deciso prima di provarlo: di solito deve essere meglio di quello che c'era, a volte basta che non sia peggio. Come ci siamo arrivati resta nostro; i risultati no.

Che cosa è migliorato

10 ottobre 2026

Il voto dell'interrogazione è vicino a quello di riferimento



Entro un punto dal voto di riferimento: il 70% delle volte prima, sempre dopo. Le interrogazioni del campione sono poche: è una direzione, non una certezza.

dal 14 al 23 ottobre 2026

Nei test, i vero/falso non sono più quasi sempre veri



I test coi vero/falso tutti veri, che si indovinano senza studiare: da uno su cinque a meno di uno su duecento. Contati su tutti i test, dal 1° al 13 ottobre e dal 24 ottobre al 5 novembre.

29 ottobre 2026

Ed ricorda la conversazione più a lungo



Le domande che arrivavano a Ed senza l'inizio della conversazione. E nelle chat lunghe confrontate, i riferimenti persi a un esercizio di prima sono spariti.

30 ottobre 2026

Meno errori gravi nelle domande delle partite a quiz



Le domande con un errore grave ($p = 0,03$). Lo stesso giorno, in chat, la risposta nuova è preferita nel 64% dei confronti ($p = 0,0001$).

30 ottobre 2026

Meno errori gravi nelle risposte in chat



I turni con un errore grave: quasi dimezzati, ma gli errori sono pochi e non è ancora una certezza statistica (p = 0,12).

31 ottobre 2026

Le mappe rispondono meglio alle loro domande



Le informazioni che non rispondono alla loro domanda (p = 0,038): meglio, ma di poco.

Non tutto è migliorato: nelle misure qui sotto ci sono anche quello che è rimasto alla pari, quello che sbaglia ancora e quello che abbiamo spento.

Quando Ed risponde in chat

Gli errori, la memoria, il controllo di ogni giorno.

Gli errori gravi nelle risposte

30 ottobre 2026

Su turni di chat veri, la versione nuova e quella di prima, confrontate senza sapere quale fosse quale.



- Errori gravi: dal 4,7% al 2,6% dei turni.
- La risposta preferita: la nuova nel 64% dei confronti (p = 0,0001).
- Nelle domande delle partite a quiz, errori gravi dall'8,2% all'1,9% (p = 0,03).
- I titoli che Ed dà alle chat: sbagliati dal 23% al 3%; il nuovo preferito nel 79% dei confronti (p < 0,01).
- Quando lo studente chiede un'attività sulle foto mandate in chat, le foto giuste scelte dal 76% al 93% delle volte.
- Il prezzo è l'attesa: circa 2 secondi in più in chat, 14 per preparare una partita a quiz.

Gli errori gravi calano di quasi la metà, ma sono pochi: da soli non bastano a una certezza statistica (p = 0,12).

Quanto Ed ricorda della conversazione

29 ottobre 2026

Nelle chat lunghe, una parte delle domande arrivava a Ed senza l'inizio della conversazione: l'esercizio di cui si parlava, la risposta di prima. Abbiamo misurato quante, dal 15 al 28 ottobre, e confrontato turni veri di chat lunghe, con la memoria di prima e con quella nuova, senza sapere quale fosse quale.



- Le domande con la conversazione tagliata: dal 20,1% all'8,4%, cioè da una su cinque a meno di una su dieci.
- Nei turni confrontati: la versione nuova meglio una volta su tre, peggio una volta su sei, pari nella metà.

  I riferimenti persi a un esercizio o a una risposta di prima: spariti. Gli errori nella spiegazione: meno

della metà.

I turni confrontati sono pochi. E ricordare di più ha un prezzo: in un caso Ed è tornato su un esercizio già chiuso.

Il controllo di ogni giorno

4 e 5 novembre 2026

Ogni giorno una parte delle risposte di Ed viene controllata, in cerca degli errori gravi: un fatto sbagliato, un conto sbagliato, una risposta che non c'entra. Quello che troviamo diventa una correzione.

- In chat: una risposta con un errore grave nel 3,2% dei casi il 4 novembre, nel 3,0% il 5 novembre.
- Nei test: il 2,2% delle domande il 4 novembre e l'1,7% il 5; più della metà, in tutte e due le giornate, veniva da un difetto solo, poi corretto.

Il controllo di ogni giorno non confronta due versioni: è un allarme, non una misura da mettere accanto alle altre. Serve a trovare gli errori, non a dire di quanto siamo migliorati. Per questo il suo 3% non si legge accanto al 2,6% degli errori gravi, più sopra: sono risposte diverse, guardate per uno scopo diverso.

Test, mappe e interrogazioni

Quello che Ed prepara e quello che corregge.

Le mappe che rispondono alle loro domande

31 ottobre 2026

Le mappe di Ed sono fatte di domande, e ogni informazione dovrebbe rispondere alla sua. Sugli argomenti e sulle pagine di mappe vere, la versione di prima e quella nuova, confrontate senza sapere quale fosse quale.



- Le informazioni che non rispondono alla loro domanda: dal 5,7% al 4,1% ($p = 0,038$).
- Le domande che hanno davvero una risposta: dall'89,9% al 93,1% ($p = 0,021$).
- Il guadagno sta tutto nelle mappe fatte dalle foto delle pagine (dal 7,5% al 5,5%); da un argomento scritto non cambia niente. Il prezzo: 2,7 secondi di attesa in più.

Meglio, ma di poco: senza l'ultimo gruppo di mappe la differenza non sarebbe stata sicura ($p = 0,22$). L'abbiamo tenuta perché la regola decisa prima era quella.

Il voto dell'interrogazione

10 ottobre 2026

Interrogazioni a voce, ognuna con un voto di riferimento. Poi il voto di Ed, più volte sulla stessa interrogazione, prima e dopo il cambiamento.



- Prima: entro un punto dal voto di riferimento il 70% delle volte.
- Dopo: entro un punto sempre, e uguale più di una volta su tre; lo scarto medio è di 0,64 punti.

Le interrogazioni del campione sono poche: è una direzione, non una certezza. Il voto lo vede lo studente alla fine; durante l'interrogazione Ed non lo dice.

–

I vero/falso dei test

dal 1° ottobre al 5 novembre 2026

Un test coi vero/falso tutti veri si indovina senza studiare. Abbiamo contato le risposte giuste dei vero/falso in tutti i test, prima e dopo il cambiamento, con la stessa regola per i due periodi.



- Le affermazioni vere: dal 63,6% (dal 1° al 13 ottobre) al 42,1% (dal 24 ottobre al 5 novembre).
- I test con almeno due vero/falso, tutti veri: dal 20,3% allo 0,4%, cioè da uno su cinque a meno di uno su duecento.

Adesso pende un poco dall'altra parte: più falsi che veri.

La correzione di esercizi e test

15 ottobre 2026

Prima di cambiare il modo in cui Ed corregge abbiamo scritto la regola del passaggio: la versione nuova non doveva sbagliare più di quella di prima. Poi correzioni vere, con le due versioni, e le righe in cui non erano d'accordo valutate senza sapere quale fosse quale.

- Il risultato: lo stesso numero di errori con le due versioni.

Non meglio: alla pari. Lo scriviamo così.

Quando lo studente chiede «è giusto?»

I conti, le risposte in foto, l'analisi della frase.

Il giudizio sui conti

29 ottobre 2026

Quando Ed doveva dire se il tentativo di uno studente era giusto, nella prima giornata di prova ha sbagliato il 15% dei giudizi. Un tutor che dà per giusto un conto sbagliato insegna l'errore.

- Da quel giorno, sui conti che lo studente scrive in chat, il giudizio è controllato con certezza ogni volta.

Il giudizio sulle risposte mandate in foto

5 novembre 2026

Sui tentativi veri che gli studenti hanno mandato in foto, dal 1° ottobre al 5 novembre, abbiamo controllato se il tentativo era giusto e se Ed l'aveva detto bene, dove si poteva dire con certezza.

- Ed sbaglia il giudizio il 15% delle volte (a ottobre era il 17%).
- Soprattutto per troppa indulgenza: a una risposta sbagliata dice «ci sei quasi», o anche «giusto», circa una volta su tre; «giusto» pieno, circa una volta su nove. Troppo severo, invece, solo sul 4% delle risposte giuste.
- Gli errori stanno nelle lingue straniere, nei problemi di geometria scritti a parole e nelle foto lette male.

Il 15% dei conti, sopra, era il primo giorno di prova: lì oggi il giudizio è controllato con certezza; qui no, ed è il numero che vogliamo far scendere. Due correzioni provate finora non l'hanno spostato in modo misurabile, e non le abbiamo messe in Ed.

L'analisi grammaticale e logica

8 ottobre 2026

L'analisi grammaticale e logica che Ed propone la misuriamo parola per parola, su frasi di cui si conosce già l'analisi giusta.

- Analisi grammaticale: quando Ed si dichiara sicuro, è giusta il 97,9% delle volte.

– Analisi logica, quando Ed si dichiara sicuro: il soggetto è giusto il 92,3% delle volte, il predicato nominale il 91,2%, il complemento oggetto l’84,7% (contando come errori anche quelli trovati dove non c’erano).
– Quando non è sicuro, non finge di sapere.

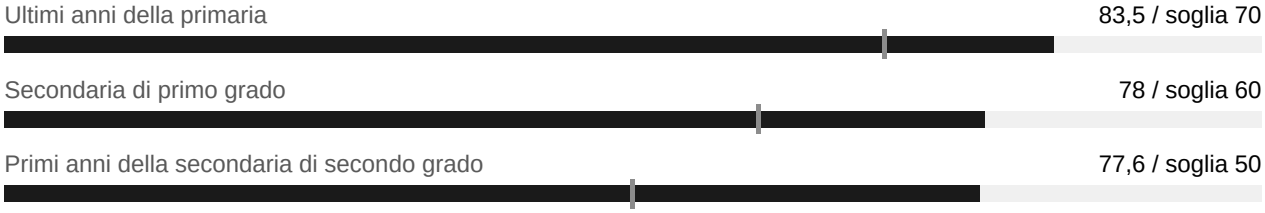
Quanto si leggono le risposte

Ogni giorno, su ogni risposta lunga in italiano.

L’indice di leggibilità per grado di scuola

dal 4 ottobre al 5 novembre 2026

Ogni risposta in italiano di almeno 40 parole a uno studente riceve l’indice Gulpease, la formula di leggibilità pensata per l’italiano e per la scuola (Lucisano e Piemontese, 1988), da 0 a 100: più è alto, più il testo è facile. La confrontiamo con la soglia adatta al suo grado di scuola: sopra 80 un testo è facile per chi ha finito la primaria, sopra 60 per chi ha finito la secondaria di primo grado, sopra 40 per chi ha finito la secondaria di secondo grado.



La barra arriva all’indice medio; il segno è la soglia del grado di scuola.

– Secondaria di primo grado: indice medio 78, soglia 60. Il 98,5% delle risposte è sopra.
– Primi anni della secondaria di secondo grado: indice medio 77,6, soglia 50. Il 100% è sopra.
– Ultimi anni della primaria: indice medio 83,5, soglia 70. L’85,4% è sopra.
– Primi anni della primaria e ultimi anni della secondaria di secondo grado: troppo poche risposte per dirlo.
– Di ogni misura si salvano il giorno, il grado e il numero: nessun nome e nessun testo.

Alla primaria una risposta su sette resta sotto la soglia: è il grado su cui lavoriamo. E la leggibilità dice quanto una risposta è facile da leggere, non se è giusta né se aiuta a capire.

Quello che abbiamo spento o scartato

Anche questo è una misura.

Una funzione che i ragazzi non usavano

dal 26 al 29 ottobre 2026

Per settimane abbiamo costruito un modo di fare i compiti con Ed passo per passo, e il 26 ottobre l’abbiamo acceso per tutti. Poi abbiamo guardato che cosa facevano gli studenti.

– Degli studenti che in quei giorni hanno usato Ed, la metà l’ha provato, e uno su sei ha finito un compito.
– Il 29 ottobre l’abbiamo spento.

Una prova che non è passata

29 ottobre 2026

Per le mappe fatte dalle foto delle pagine abbiamo provato un modo diverso di prepararle: mappe vere, con la versione di prima e con quella nuova, confrontate con le pagine davanti, senza sapere quale fosse

quale.

- Errori gravi: non meno di prima, anzi uno in più con la nuova.
- Non l’abbiamo messa in Ed.

Come citare questo report

ChiarAmEnte (2026). Report di novembre 2026. «Come misuriamo Ed», n. 1, pubblicato il 6 novembre 2026. https://ed.lume.study/come-misuriamo/novembre-2026

Le regole con cui misuriamo valgono per ogni report e stanno in «Come misuriamo Ed», ed.lume.study/come-misuriamo.