

Automatische Aufsatzbewertung

Wissenschaftlich fundierte Bewertung
und Rückmeldung für den
Schreibunterricht



Warum wir die automatische Aufsatzbewertung entwickelt haben

Schreiben lernt man durch Schreiben. Doch häufiges Schreiben und individuelles Feedback sind im Schulalltag aufgrund des hohen Korrekturaufwands oft nur schwer umsetzbar.

Die automatische Aufsatzbewertung wurde entwickelt, um diesen Zielkonflikt zu entschärfen. Die wissenschaftlich entwickelte Lösung des Instituts für Bildungsevaluation (IBE) verbindet Psychometrie, künstliche Intelligenz und Sprachtechnologie, um Texte zuverlässig zu bewerten und differenzierte Rückmeldungen zu geben. Dabei stehen Fairness, Nachvollziehbarkeit und Datenschutz im Zentrum. Lehrpersonen gewinnen Zeit für individuelle Förderung und Schreibgespräche, während Lernende häufiger schreiben, ihre Texte gezielt überarbeiten und ihre Schreibkompetenz systematisch weiterentwickeln können.

Schreibkompetenz wächst durch Übung – und durch gutes Feedback.

Dieses Whitepaper erläutert die wissenschaftlichen Grundlagen, die technische Architektur sowie die empirischen Nachweise für die Fairness und Zuverlässigkeit der automatischen Aufsatzbewertung. Es zeigt, wie künstliche Intelligenz pädagogische Expertise sinnvoll ergänzt und einen wirksamen Schreibunterricht unterstützt.



Abb. 1 — Einsatz der automatischen Aufsatzbewertung im Unterricht

Was die automatische Aufsatzbewertung auszeichnet

Die automatische Aufsatzbewertung bewertet Schülertexte innerhalb von Sekunden anhand der vier Bereiche Inhalt, Textzusammenhang, Sprachrichtigkeit und Stil. Lernende erhalten zugleich verständliche, altersgerechte Rückmeldungen mit konkreten Verbesserungsvorschlägen. Das System unterstützt Deutsch, Englisch und Französisch, verarbeitet digitale und handschriftliche Texte und eignet sich ab der 3. Klasse bis in die Berufsbildung.

Die automatische Aufsatzbewertung entstand im Rahmen eines Inno-suisse-Forschungsprojekts gemeinsam mit der Universität Zürich und der EPFL Lausanne. Die Grundlage bilden rund 37'000 von Sprachexpertinnen und Sprachexperten bewertete Schülertexte.

Das System lernt nicht, einzelne Bewertungen von Fachpersonen möglichst genau nachzubilden. Stattdessen sagt es psychometrisch geschätzte Fähigkeitswerte voraus und erreicht eine hohe Übereinstimmung mit menschlichen Bewertungen (Korrelation ≈ 0.90).

Umfangreiche Analysen belegen eine vergleichbar zuverlässige Bewertung über Geschlecht, Erstsprache, Klassenstufen und Regionen hinweg. Sämtliche Daten werden ausschliesslich in der Schweiz verarbeitet. Die automatische Aufsatzbewertung nutzt ausschliesslich Open-Source-Sprachmodelle auf einer redundanten Schweizer Infrastruktur, alle Trainingsdaten sind anonymisiert und werden zu keinem Zeitpunkt an Dritte weitergegeben.

DIE PHILOSOPHIE

Wissenschaftlich fundiert

Gemeinsam mit der Universität Zürich und der EPFL entwickelt und evaluiert.

Der Mensch bleibt Massstab

Orientiert sich an den Beurteilungen geschulter Fachpersonen.

Datensicherheit aus der Schweiz

Gesamte Verarbeitung in der Schweiz, anonymisierte Trainingsdaten, keine Weitergabe an Dritte.

Die automatische Aufsatzbewertung verbindet **wissenschaftliche Fundierung** mit einem **praxistauglichen Einsatz** im Unterricht. So wird die automatische Aufsatzbewertung zu einem Werkzeug, das Lehrpersonen sinnvoll unterstützt.

So funktioniert die automatische Aufsatzbewertung

Die automatische Aufsatzbewertung begleitet den gesamten Prozess – von der Aufgabenstellung bis zum Feedback. Für Lehrpersonen bleibt der Ablauf vertraut: Schreibanlass auswählen, Texte schreiben lassen und

Rückmeldungen erhalten. Im Hintergrund übernehmen psychometrische Modelle, Sprachtechnologie und künstliche Intelligenz die Analyse und Bewertung.

1

Schreibanlass erstellen

Lehrpersonen wählen einen vorhandenen Schreibanlass oder erstellen einen eigenen. Die automatische Aufsatzbewertung unterstützt verschiedene Textsorten und mehrere Sprachen.

2

Texte verfassen

Die Schülerinnen und Schüler schreiben ihren Text digital oder handschriftlich. Handschriftliche Texte werden nach dem Einscannen automatisch transkribiert.

3

Automatische Analyse

Die automatische Aufsatzbewertung analysiert Inhalt, Textzusammenhang, Sprachrichtigkeit und Stil. Dabei werden mehrere spezialisierte Modelle kombiniert und die Ergebnisse zu einer aussagekräftigen Rückmeldung zusammengeführt.

4

Feedback erhalten

Lehrpersonen und Lernende erhalten innerhalb von Sekunden eine nachvollziehbare Bewertung mit verbalem Feedback sowie konkrete Verbesserungsvorschläge für die Überarbeitung.

Für Lehrpersonen bleibt der **Unterricht vertraut** – sie gewinnen jedoch **Zeit für individuelle Förderung** statt für Korrekturarbeit.

Bewertung und Feedback auf einen Blick

Die automatische Aufsatzbewertung bewertet jeden Text in vier Bereichen und ergänzt die Bewertung durch verständliches Feedback mit konkreten Verbesserungsvorschlägen. Die Rückmeldungen helfen Lernenden, ihre Schreibkompetenz Schritt für Schritt weiterzuentwickeln.

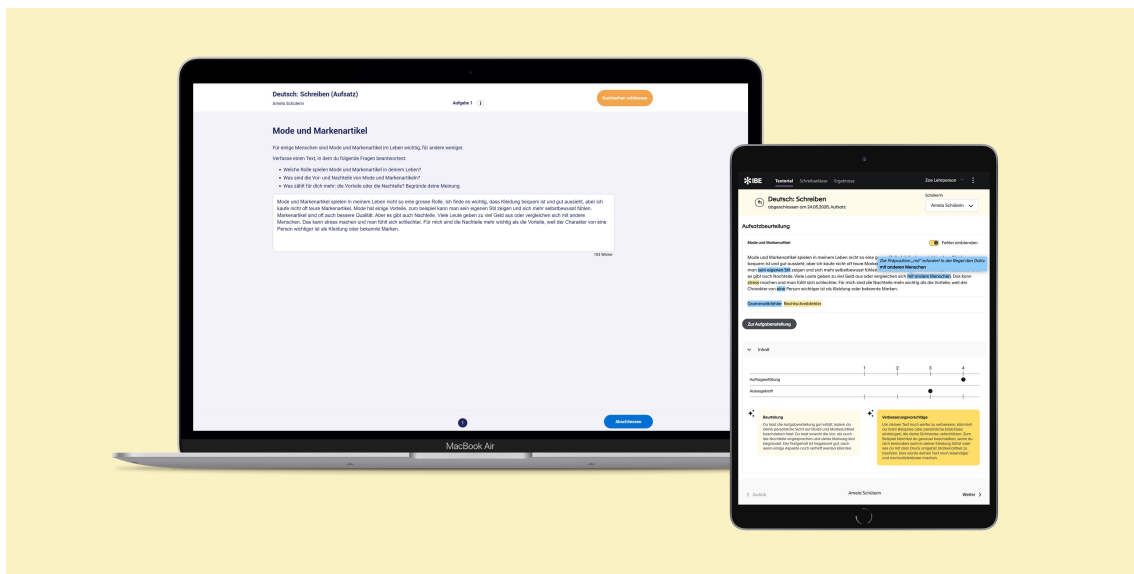


Abb. 2 — Beispiel einer Rückmeldung der automatischen Aufsatzbewertung.



Punktzahl

Punktzahl von 1 bis 4 in den Bereichen Inhalt, Textzusammenhang, Sprachrichtigkeit und Stil. Jeder der vier Bereiche umfasst zwei Kriterien. Orientiert sich an den Erwartungen der jeweiligen Klassenstufe.



Fehlermarkierung

Direkt im Text markierte Rechtschreib- und Grammatikfehler mit konkreten Verbesserungsvorschlägen und Hinweisen.



Vertikale Kompetenzskala

Jeder Text wird auf einer klassenübergreifenden Kompetenzskala eingeordnet. Das Ergebnis wird anschliessend mit Bezug zur jeweiligen Klassenstufe dargestellt.



Feedback

Verständliches Feedback zu den Stärken und Entwicklungsbereichen des Textes, ergänzt durch konkrete Verbesserungsvorschläge.

Mehr als ein Sprachmodell

Viele KI-Anwendungen überlassen Bewertung und Feedback einem einzelnen Sprachmodell. Die automatische Aufsatzbewertung verfolgt bewusst einen anderen Ansatz: Psychometrie, künstliche Intelligenz und regelbasierte Sprachtechnologie übernehmen unterschiedliche Aufgaben und bilden gemeinsam ein transparentes Bewertungssystem.



Psychometrische Bewertung

Das Bewertungsmodell bildet Schreibkompetenz statt Einzelbewertungen ab. Grundlage sind rund 37'000 mehrfach beurteilte Schülertexte.



Regelbasierte Sprachanalyse

Rechtschreibung und Grammatik werden mit einer erweiterten Version von LanguageTool sowie der NLP-Bibliothek spaCy zuverlässig und speziell für den Schulbereich analysiert.



KI-gestütztes Feedback

Ein Sprachmodell erstellt auf Grundlage der Bewertung, der Fehleranalyse und der Aufgabenstellung verständliches Feedback mit konkreten Verbesserungsvorschlägen.



Transparente Architektur

Psychometrie, Sprachtechnologie und KI übernehmen klar definierte Aufgaben und bilden gemeinsam ein nachvollziehbares Bewertungssystem.

Nicht das Sprachmodell macht den Unterschied, sondern das, worauf die automatische Aufsatzbewertung trainiert wurde: die zugrunde liegende Schreibkompetenz statt einzelner Bewertungen.

Zuverlässigkeit wissenschaftlich belegt

Die automatische Aufsatzbewertung ist kein gewöhnliches KI-System. Sie verbindet moderne Sprachmodelle mit wissenschaftlich etablierten psychometrischen Verfahren und schafft dadurch zuverlässige und faire Bewertungen.

Psychometrische Modelle erhöhen die Genauigkeit

Die automatische Aufsatzbewertung lernt nicht, einzelne Bewertungen nachzuahmen. Stattdessen wird das Modell auf wissenschaftlich berechnete Kompetenzwerte trainiert, die aus mehreren unabhängigen Bewer-

tungen abgeleitet werden. Diese Werte bilden den fachlich begründeten Konsens mehrerer geschulter Sprachexpertinnen und Sprachexperten ab und ermöglichen dadurch präzisere und fairere Bewertungen.

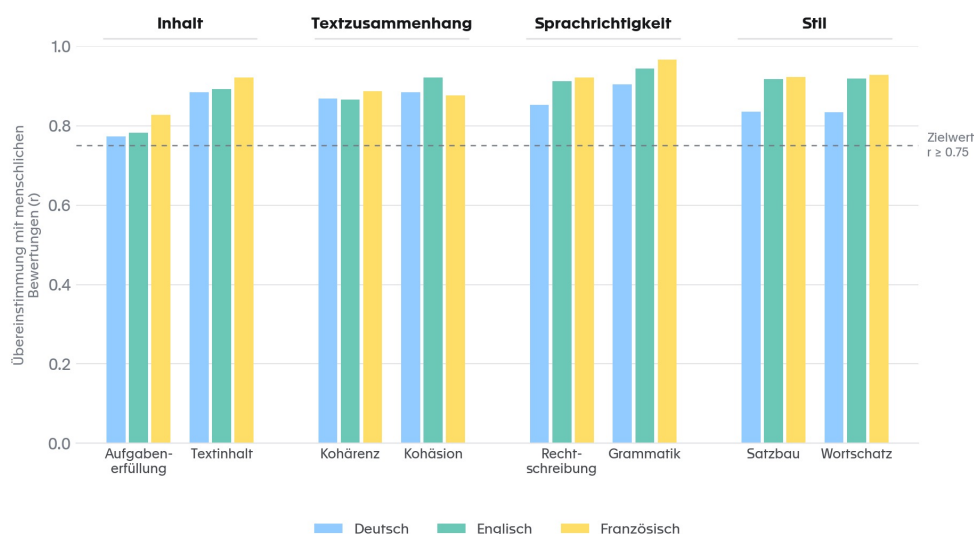


Abb. 3 — Übereinstimmung der automatischen Bewertung mit menschlichen Bewertungen nach den acht Kriterien und Sprache (Pearson Korrelationen). Quelle: IBE (2026). Projektberichte und Addenda zum Innosuisse-Projekt 119.542 «Innovative AI Approaches for Language Proficiency Testing in Switzerland». Institut für Bildungsevaluation, Zürich.

- * **Hohe Übereinstimmung** mit menschlichen Bewertungen ($r \approx 0.90$).
- * Alle acht Bewertungskriterien **übertreffen den definierten Qualitätswert** von $r \geq 0.75$.
- * Der psychometrische Trainingsansatz liefert **genauere und fairere Ergebnisse** als ein Modell, das direkt mit menschlichen Bewertungen trainiert wurde.

Einheitlicher Bewertungsmassstab für alle Lernenden

Faire Bewertungen setzen voraus, dass für alle Lernenden derselbe Massstab gilt. Deshalb wurde untersucht, ob die automatische Aufsatzbewertung die Schreibkompetenz unabhängig von Geschlecht, Erstsprache, Klassenstufe und Region gleich erfasst. Die Ergebnisse bestätigen,

dass das Bewertungsmodell in allen untersuchten Gruppen nach denselben Kriterien funktioniert. Unterschiede in den Bewertungen lassen sich somit auf Unterschiede in der Schreibkompetenz zurückführen – nicht auf Unterschiede im Bewertungsverfahren.

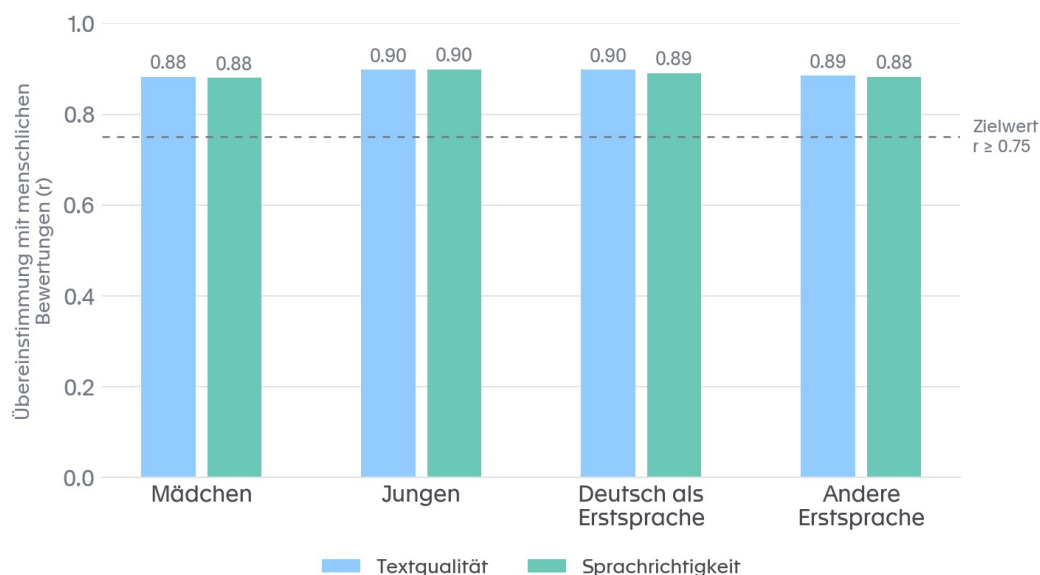


Abb. 4 – Übereinstimmung der automatischen Bewertung mit menschlicher Bewertung nach Schülergruppe. Keine systematischen Unterschiede nach Geschlecht und Erstsprache. Die Analysen zu Klassenstufe und Region finden sich im Projektbericht (IBE 2026). Quelle: IBE (2026). Projektberichte und Addenda zum Innosuisse-Projekt 119.542 «Innovative AI Approaches for Language Proficiency Testing in Switzerland». Institut für Bildungsevaluation, Zürich.

- * **Hohe Übereinstimmung** mit menschlichen Bewertungen in allen untersuchten Schülergruppen.
- * **Keine systematischen Unterschiede** nach Geschlecht, Erstsprache, Klassenstufe oder Region.
- * **Faire Vergleiche**, weil das Modell auf psychometrisch geschätzten Kompetenzwerten statt auf einzelnen Bewertungen trainiert wurde.

Kompetenzskala macht Lernfortschritte sichtbar

Durch die zugrunde liegende psychometrische Skalierung ist es möglich, Schreibkompetenzen auf einer klassenstufenübergreifenden vertikalen Kompetenzskala abzubilden.

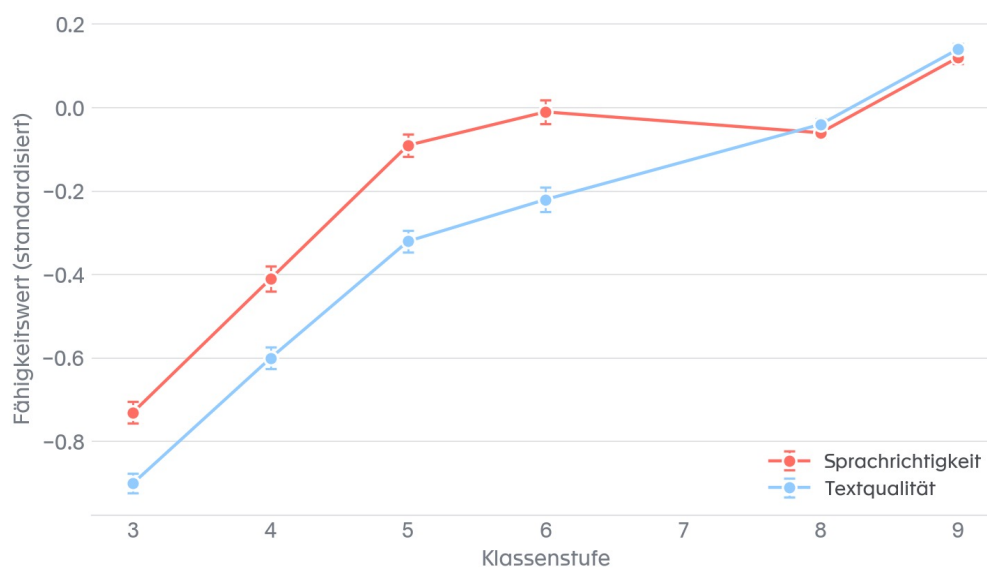


Abb. 5 — Entwicklung der Schreibkompetenz über die Klassenstufen 3–9 auf der klassenstufenübergreifenden vertikalen Kompetenzskala (Mittelwerte $\pm$ Standardfehler; für Klassenstufe 7 lagen keine Daten vor). Quelle: Debelak, R., Vignoud, J. & Tomasik, M. (2026). *Using Psychometric Modeling to Improve the Effectiveness of Automated Essay Scoring Systems*. Universität Zürich / EPFL Lausanne. [Eingereichtes Manuskript]

Erwartungsgemäss zeigen Schülerinnen und Schüler höherer Klassenstufen höhere Werte auf der Skala als jene tieferer Stufen.

Die Qualität der in der automatischen Aufsatzbewertung verwendeten Modelle wird kontinuierlich überwacht. Vor der Einführung werden sämtliche Modelle wissenschaftlich evaluiert – auf Genauigkeit, Fairness, Bias und Entwicklungsvalidität.

Datensicherheit nach Schweizer Standards

Datenschutz ist für die automatische Aufsatzbewertung kein Zusatz, sondern ein grundlegendes Entwicklungsprinzip. Sämtliche Schülertexte werden ausschliesslich in der Schweiz verarbeitet.



Verarbeitung in der Schweiz

Alle Daten werden ausschliesslich auf redundanten Servern in Schweizer Rechenzentren verarbeitet.



Keine Weitergabe an Dritte

Schülertexte werden ausschliesslich innerhalb der Schweizer Infrastruktur verarbeitet und nicht an externe KI-Dienste oder Modellanbieter weitergegeben.



Kontrollierte KI-Infrastruktur

Das IBE betreibt die eingesetzten Modelle innerhalb einer Schweizer Infrastruktur und behält die Kontrolle über Datenverarbeitung, Modelle und Systemkonfiguration.



Datenschutz ab der Entwicklung

Trainingsdaten werden anonymisiert und ausschliesslich für das Training von Open-Source-Modellen auf eigener Infrastruktur verwendet.

Digitale Souveränität bedeutet: vollständige Kontrolle über Daten, Modelle und Infrastruktur.

Die automatische Aufsatzbewertung entwickelt sich weiter

Mit der Version 1.0 wird die automatische Aufsatzbewertung ab August 2026 Schulen in der Schweiz, Deutschland und Österreich zur Verfügung stehen. Gleichzeitig markiert dieser Schritt nicht den Abschluss, sondern den Beginn einer kontinuierlichen wissenschaftlichen Weiterentwicklung. Neue Modelle und Verfahren werden laufend evaluiert und nur dann integriert, wenn sie nachweislich zu präziseren, faireren oder besser nachvollziehbaren Bewertungen führen.

Vier Entwicklungsschwerpunkte



Kontinuierliche Forschung

Die Datenbasis und die wissenschaftliche Evaluation werden laufend erweitert, um die Qualität der automatischen Bewertung kontinuierlich zu verbessern.



Modulare Architektur

Einzelne Komponenten können unabhängig voneinander weiterentwickelt werden, ohne die Stabilität und Vergleichbarkeit der Bewertungen zu beeinträchtigen.



Internationale Zusammenarbeit

Der Einsatz in weiteren Sprachregionen und Ländern schafft neue Möglichkeiten für Forschung und die Weiterentwicklung automatischer Aufsatzbewertung.



Wissenschaftlich begleitet

Auch zukünftige Modellversionen werden erst nach wissenschaftlicher Prüfung und klar definierten Qualitätskriterien produktiv eingesetzt.

Unser Ziel bleibt unverändert: Lehrpersonen sollen häufiger Schreibanlässe schaffen können – nicht häufiger korrigieren müssen.

Gute Schreibförderung beginnt dort, wo häufiges Schreiben und hochwertiges Feedback selbstverständlich werden.

Testen Sie die automatische Aufsatzbewertung während 30 Tagen kostenlos im Unterricht.



Zum 30-Tage-Test

KONTAKT

Nina König

Data Analyst & Software Manager

nina.koenig@ibe-edu.ch

www.ibe-edu.ch • 043 268 39 60

