

1.2 Lehren über KI vs. Lehren mit KI

Kernbotschaften

- **KI ist ein Navigationsgerät.** Es rettet die Fahrerin, die die Karte bereits im Kopf hat, und hindert diejenige, die sie nicht hat, am Lernen. Die Wirkung hängt von der Person ab, nicht vom Produkt.
- **Lehren über KI ist die Voraussetzung für Lehren mit KI.** Man kann eine Antwort nicht überprüfen, wenn man nicht weiß, wie das Werkzeug, das sie erzeugt hat, versagen kann.
- **Wissen ist nicht Tun.** Lehrkräfte bewerten ihr KI-Wissen höher als ihre Fähigkeit, damit zu unterrichten, und Fortbildung allein schließt diese Lücke nicht.
- **Lehrkräfte nutzen KI hinter den Kulissen, nicht im Unterricht.** Die Hälfte der Lehrkräfte nutzt KI regelmäßig, weniger als ein Drittel setzt sie mit Schüler:innen ein.
- **Pädagogik schlägt Technologie.** Unterricht zu transformieren schlägt das bloße Chatten mit KI; die Arbeit zu zweit schlägt Alleinarbeit; Feedback schlägt Gamification.
- **Besseres Feedback bedeutet nicht besseres Lernen.** Als eindeutig besser bewertete KI-Kommentare führten zu Überarbeitungen, die nicht besser waren. Entscheidend ist, ob das Feedback *aufgenommen* wird.
- **KI steigert den Output, nicht das selbstgesteuerte Lernen.**
- **Klare Regeln und intensive Nutzung gehen Hand in Hand.** Lehrkräfte an Schulen mit einer schriftlichen KI-Richtlinie berichten, dass sie mehr Zeit sparen, nicht weniger – das Gegenteil dessen, was die meisten erwarten.
- **Bewahren Sie Ihre eigene Landkarte.** Entlastung wird zu Dequalifizierung, wenn KI die Arbeit übernimmt, die Urteilsvermögen aufbaut und erhält.

“ Wenn Sie sich nur einen Satz von dieser Seite merken: **KI ist ein Navigationsgerät. Es rettet die Fahrerin, die bereits eine Karte im Kopf hat, und hindert die Fahrerin, die keine hat, still und leise daran, jemals eine aufzubauen.**

Das Navi-Problem

Um ein Londoner Taxi zu fahren, muss man zunächst eine Wissensprüfung (den „Knowledge“) bestehen: 320 festgelegte Routen über die 113 Quadratmeilen im Umkreis von sechs Meilen um Charing Cross, wofür Transport for London bis zu zwei Jahre Vorbereitungszeit einräumt, bevor die Prüfungen überhaupt beginnen.

Auch geprüfte Fahrer:innen nutzen ein Navi, und niemand denkt deswegen schlechter von ihnen. Die Wissensprüfung wurde nicht wertlos, als das Gerät aufkam. Das Navi wurde zu dem, was die Fahrt sicher macht. Eine Fahrerin mit der Stadt im Kopf nutzt das Navi für das, was sie nicht wissen kann: den Unfall auf der Brücke oder den heutigen Verkehr. Und sie bemerkt, wenn es sie durch eine Straße lenkt, die seit April aufgerissen ist.

Setzen Sie nun eine Neuling in dasselbe Taxi mit demselben Navi. Sie erreicht jede Adresse. Eine Zeit lang ist sie von der erfahrenen Fahrerin nicht zu unterscheiden. Doch indem sie den Anweisungen des Navis folgt, baut sie nichts auf. Nach einem Jahr, in dem sie auf Anweisung links abgebogen ist, kann sie die Stadt immer noch nicht ohne das Gerät durchqueren, und sie kann nicht erkennen, wann es sich irrt.

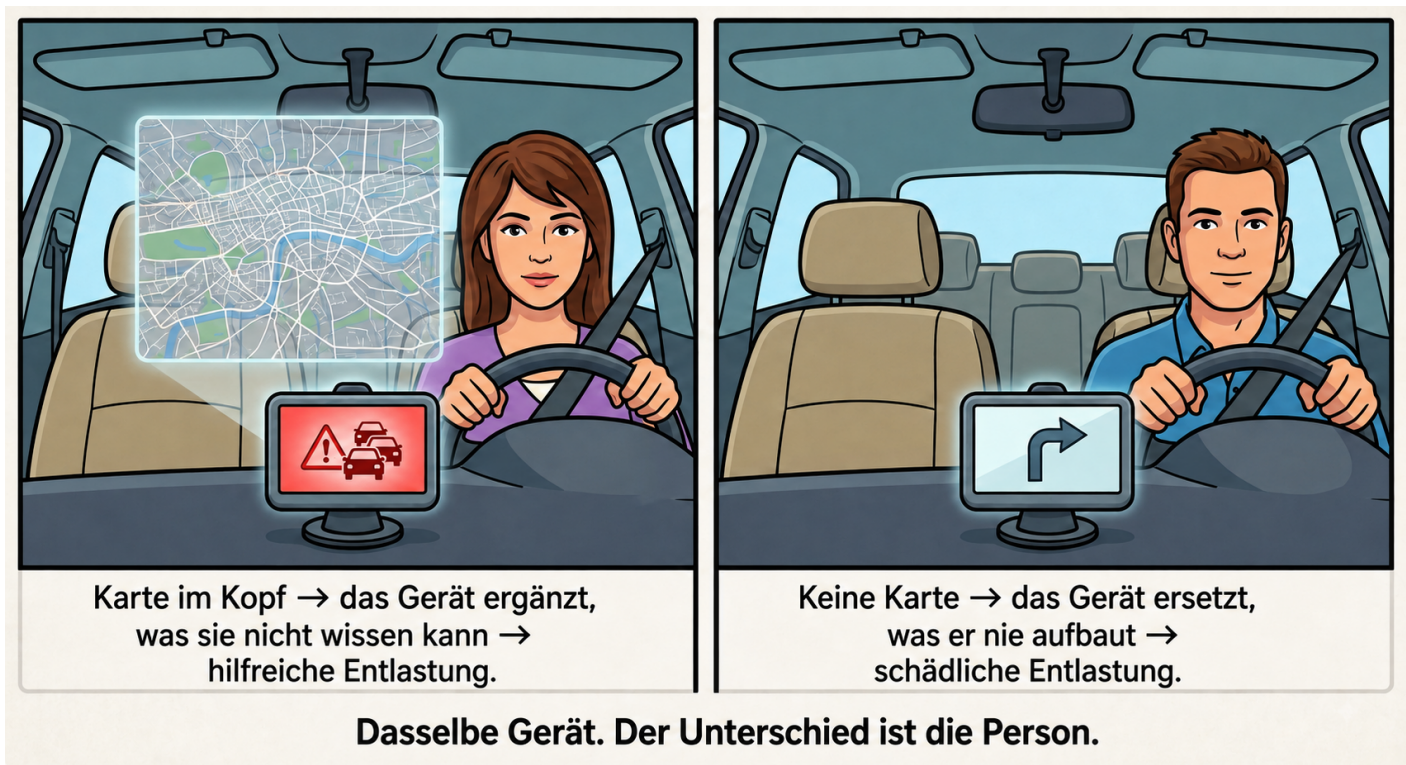


Abb. 1. Das Ergebnis hängt vom Fahrer ab. Bild erstellt mit Google Gemini.

Dasselbe Navi, zwei Fahrerinnen, gegensätzliche Ergebnisse. **Der Unterschied lag nie im Gerät.**

Das ist ein Jahrzehnt an KI-Forschung in einem Bild – und es gilt auch ohne das Bild. Die Wirkung eines KI-Werkzeugs hängt weit stärker davon ab, was die nutzende Person bereits weiß, als vom Werkzeug selbst. Forschende nennen dies ein [Schema](#), das organisierte Vorwissen einer Person. Ist

es vorhanden, entlastet die Abgabe von Arbeit an eine Maschine. Forschende nennen dies nützliches [kognitives Auslagern](#). Muss Wissen jedoch erst aufgebaut werden, kann derselbe Vorgang genau das verhindern. Das nennt man schädliches [kognitives Auslagern](#).

Eine Studie mit rund tausend weiterführenden Schüler:innen zeigt genau das in einem Mathematikunterricht. Eine Gruppe übte mit einem uneingeschränkten Chatbot, eine andere ohne. Während der Nutzung löste die Chatbot-Gruppe **48 % mehr Übungsaufgaben!** Dann kam eine Prüfung ohne Chatbot. Nun schnitten dieselben Schüler:innen, die zuvor so gut Übungsaufgaben gelöst hatten, **17 % schlechter** ab als die Gruppe, die nie einen KI-Chatbot genutzt hatte. Eine dritte Gruppe nutzte eine KI-Version, die so gebaut war, dass sie Antworten zurückhielt und stattdessen Rückfragen stellte. Diese Gruppe verlor nichts.

Diese dritte Gruppe ist die eigentliche Gestaltungsfrage. Ein Navi, das das Ziel nennt und Ihnen hilft, die Route selbst zu finden, bringt Ihnen die Stadt bei. Ein Navi, das sagt „*In 200 Metern links abbiegen*“, tut das nicht, so gut seine Daten auch sein mögen. Forschende nennen diesen Unterschied [Leitplanken \(Guardrails\)](#), und **die Gestaltung entscheidet, ob Sie Ihren Schüler:innen helfen, die Karte der Stadt zu lernen, oder ob sie einfach nur das Ziel finden.**

Bitte denken Sie daran: **Sie haben das Wissen. Ihre Schüler:innen haben es, in dem Fach, das Sie unterrichten, noch nicht.** Diese Asymmetrie ist das ganze Kapitel.

Zwei Stockwerke, nicht zwei Themen

In der Fachliteratur gelten diese beiden Bereiche als getrennte Felder. Man kann sie sich als zwei Stockwerke eines Gebäudes vorstellen, wobei sich die meisten Lehrkräfte im Erdgeschoss befinden und keine Treppe ins erste Stockwerk führt.

Auf die Bitte, ihr KI-Wissen einzuschätzen, gaben sich Lehrkräfte in Deutschland im Schnitt 3,4 von 5 Punkten. Ihre Fähigkeit, dieses Wissen im Unterricht anzuwenden, bewerteten sie hingegen deutlich niedriger, mit 2,6 von 5. Wissen und Handeln hatten sich voneinander gelöst. Eine Studie in fünf europäischen Ländern untersuchte daraufhin, ob Fortbildung diese Lücke schließen kann. In der Studie erhielten 736 Lehrkräfte in Frankreich, Irland, Italien, Luxemburg und Slowenien einen Kurs samt Begleitung. Die Ergebnisse zeigten, dass ihr Wissen und ihre Fähigkeit, die Leistungsfähigkeit eines Werkzeugs einzuschätzen, zunahmen. Ihr Unterricht änderte sich jedoch nicht. Sie probierten Dinge aus und kehrten dann zu dem zurück, was sie vorher getan hatten.

Das ist die Wissen-Handlungs-Lücke. Kurse erzeugen Wissen, aber die Veränderung der Praxis braucht Begleitung, Ausprobieren und Kolleg:innen zum gemeinsamen Nachdenken. **Transfer muss gestaltet werden, nicht erhofft.**

Was Lehrkräfte tatsächlich tun

Der robusteste Befund über alle Länder hinweg ist auf leise Weise überraschend: **Lehrkräfte nutzen KI intensiv für sich selbst und kaum mit ihren Schüler:innen**. In Deutschland nutzen 52 % KI mindestens mehrmals im Monat. Nur 29 % setzen sie im Unterricht vor der Klasse ein. Die öffentliche Debatte über „KI im Klassenzimmer“ dreht sich also größtenteils um etwas, das in der Fläche noch gar nicht stattfindet.

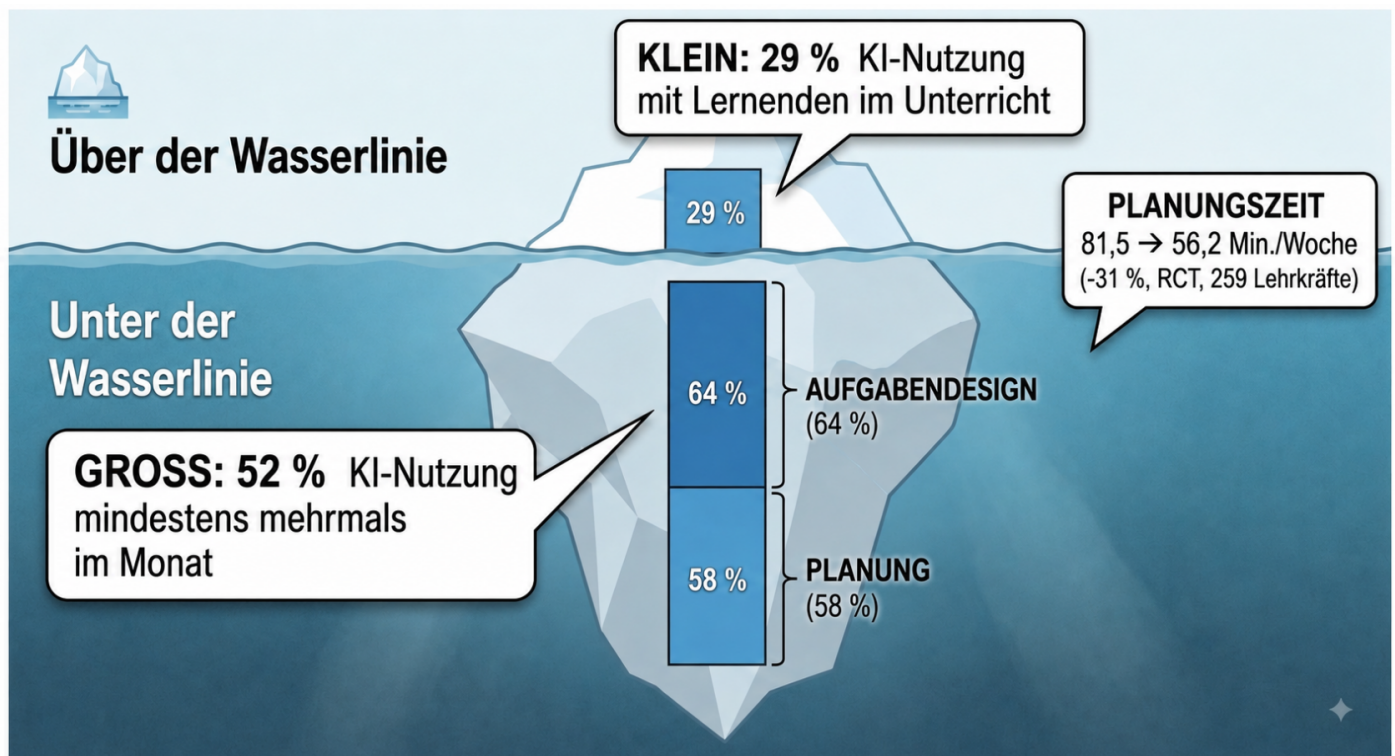


Abb. 2: Unterschied zwischen der KI-Nutzung zur Vorbereitung und der Nutzung im Unterricht. Bild erstellt mit Google Gemini.

Was **tatsächlich** passiert, ist Entlastung von der Arbeitsbelastung, und hier sind die Nachrichten gut. In einer britischen Studie erhielten 259 Lehrkräfte an 68 Schulen KI-Unterstützung für die Planung von Naturwissenschaftsunterricht für 11- bis 13-Jährige. Die Planungszeit sank **von 81 auf 56 Minuten pro Woche**. Als unabhängige Gutachter:innen die entstandenen Materialien beurteilten, ohne zu wissen, welches welches war, stellten sie keinen Qualitätsabfall fest. Die eigenen Schätzungen der Lehrkräfte zur eingesparten Zeit über alle Fächer hinweg liegen sogar noch höher, bei rund sechs Stunden pro Woche für regelmäßige Nutzer:innen.

Und ein Muster verdient einen Moment Aufmerksamkeit. In einer großen amerikanischen Befragung **berichteten Lehrkräfte an Schulen mit einer klaren schriftlichen KI-Richtlinie, dass sie etwa ein Viertel mehr Zeit sparten als Lehrkräfte an Schulen ohne eine solche**. Das ist die eigene Einschätzung der Lehrkräfte und keine Messung, und eine Befragung kann nicht zeigen, in welche Richtung der Zusammenhang tatsächlich verläuft. Gut organisierte Schulen machen möglicherweise einfach beides gut. Aber es ist das Gegenteil dessen, was die meisten erwarten, und es ist bemerkenswert, dass niemand das Umgekehrte gefunden hat: Wo immer dies untersucht wurde, **gehen klare Regeln und intensive Nutzung Hand in Hand**. -> [1.4](#)

[Konformer Einsatz von KI](#)

Warum es nicht auf das Werkzeug ankommt

Wer sich ein Ranking der besten KI-Werkzeuge wünscht, wird von der Forschung zunächst enttäuscht – und beim zweiten Lesen davon befreit. Wenn Forschende viele einzelne Studien zusammenführen und fragen, was Klassenzimmer, in denen KI stark geholfen hat, von solchen unterscheidet, in denen sie kaum geholfen hat, lautet die Antwort nie die Marke der Software. Es ist das, was die Schüler:innen damit *tun* sollten.

Es gibt drei Vergleiche, die diesen Punkt wirklich verdeutlichen.

- Schüler:innen, die KI nutzten, um etwas zu transformieren – etwa einen Entwurf zu überarbeiten oder auf dem eigenen Versuch aufzubauen –, erzielten etwa zweieinhalbmal so viel Lernzuwachs wie Schüler:innen, die einfach nur mit einem Chatbot ein Gespräch führten.
- Schüler:innen, die KI zu zweit nutzten, erreichten fast doppelt so viel wie diejenigen, die sie allein nutzten.
- KI wurde eingesetzt, um Feedback zu Schülerarbeiten zu geben, und es ist fair zu sagen, dass dies zu einem der größten Lernzuwächse in diesem Bereich führte. Der Einsatz von KI zur Einbindung von Spielelementen in den Unterricht brachte dagegen keinerlei Nutzen.

Pädagogik schlägt Technologie, und das deutlich. Das zeigt sich jedes Mal, wenn jemand die Studien danach sortiert, was die Schüler:innen tatsächlich getan haben. Zwei Einschränkungen sollten dabei gleichermaßen bedacht werden.

Gutes Feedback ist nicht dasselbe wie wirksames Feedback. In einem Experiment schrieben 70 Studierende einen argumentativen Aufsatz und überarbeiteten ihn eine Woche später. Sie wurden zufällig drei Gruppen zugeteilt: Eine erhielt Feedback von einer erfahrenen Lehrkraft, eine von einem Chatbot mit einfacher Anweisung und eine von demselben Chatbot, der angewiesen wurde, den Aufsatz Schritt für Schritt zu durchdenken. Bewertet nach demselben Raster, lieferte der schrittweise vorgehende Chatbot das **beste Feedback der drei**. Es war besser als das des einfachen Chatbots und besser als das der Lehrkraft. Dann kamen die Überarbeitungen, und **alle drei Gruppen verbesserten sich um etwa denselben Betrag**. Das beste Feedback der Studie brachte keine größere Verbesserung als das am schlechtesten bewertete. **Bessere Kommentare führten nicht zu einem besseren Aufsatz.**

Das verdient einen Moment Aufmerksamkeit, denn es stellt die verbreitete Annahme infrage, besseres Feedback führe zu besserem Lernen. Für sich genommen tut es das nicht. Feedback bewirkt nur dann etwas, wenn es **aufgenommen** wird: gelesen, verstanden, umgesetzt – und so gut verstanden, dass die Schülerin oder der Schüler es beim nächsten Mal auch ohne Hilfe richtig machen könnte. **Die Aufnahme ist der Engpass, nicht die Qualität.**

Zwei Dinge scheinen dabei im Weg zu stehen, und beide lassen sich nutzen.

Mehr ist nicht immer besser. In einer zweiten Studie mit 60 Schüler:innen der Jahrgangsstufe 11 und 240 Textstücken produzierte die KI weit mehr inhaltsbezogene Kommentare als die Lehrkräfte – 333 gegenüber 241. Es waren die Kommentare der Lehrkräfte, die die Schüler:innen

häufiger umsetzten. Konfrontiert mit dreißig präzisen Korrekturen, hört eine Schülerin oder ein Schüler womöglich auf zu diagnostizieren und beginnt stattdessen zu sortieren – oder fügt sich einfach. Wer den Regeln folgt, sagt damit im Grunde, dass die Überarbeitung unnötig war. Es ist wieder das Navi, nur auf einer anderen Ebene. Es gibt einem wirklich detailliertes Feedback, wie „In 200 Metern links abbiegen“ – nur für den eigenen Text.

Wer es sagt, spielt eine Rolle. Wir wissen nicht genau, warum Kommentare von Lehrkräften häufiger aufgenommen werden. Das ist Interpretation, keine Messung, aber die wahrscheinlichste Erklärung ist, dass Feedback als etwas zwischen Menschen wahrgenommen wird. Was jemand sagt, der weiß, was man letztes Halbjahr gemacht hat, und sehen wird, was man damit macht, ist etwas anderes als derselbe Satz, auf Anfrage generiert.

Was das für die Praxis bedeutet. Alle Belege, die wir zu automatisiertem Feedback gesehen haben, zeigen, dass es am besten wirkt, wenn Schüler:innen das in einer Aufgabe Gelernte auf eine neue Aufgabe übertragen. Am wenigsten wirksam ist es, wenn sie das Gelernte einfach direkt aus dem vorliegenden Text übernehmen. Und es ist deutlich wirksamer, wenn man es über einen längeren Zeitraum einsetzt, etwa ein ganzes Halbjahr, statt nur einmalig. **Wir sollten also nicht fragen, ob das Feedback der KI gut genug ist. Wir sollten fragen, ob diese Schülerin oder dieser Schüler es je wieder brauchen wird. Lautet die Antwort Nein, helfen bessere Kommentare vermutlich wenig.**

Und dieselbe Aufspaltung zeigt sich auch im großen Maßstab. Ein Team führte 228 Studien zusammen: KI verbesserte klar und durchgängig, was Schüler:innen produzieren konnten, während jeder Zuwachs bei der Fähigkeit, das eigene Lernen zu planen, zu überwachen und zu korrigieren, klein genug war, um Zufall sein zu können. Am deutlichsten zeigt sich das beim Programmieren. Schüler:innen mit einer KI-Assistenz waren messbar schneller, und am Ende nicht besser als jene ohne. **Mehr zu schaffen und besser zu werden sind nicht dasselbe, und KI trennt die beiden voneinander.**

Europa als Spiegel

Der Anteil der KI nutzenden Lehrkräfte reicht europaweit **von 89 % in Tschechien bis 14 % in Frankreich.** Bedenken Sie, dass sich die Erhebungen in Methode und Zeitpunkt unterscheiden – lesen Sie daher die Spannweite, nicht die Einzelwerte.

Frankreich ist der aufschlussreiche Fall. Das Land hat eines der ambitioniertesten staatlichen KI-Programme des Kontinents aufgebaut, ein KI-Tutoring-Programm in Mathematik und Französisch für das erste Jahr der Oberstufe, erprobt an 150 Schulen. Dennoch steht Frankreich weiterhin am Ende dieser Tabelle, da nur 9 % der französischen Lehrkräfte überhaupt eine KI-Fortbildung erhalten haben, verglichen mit rund 38 % in vergleichbaren Ländern. **Werkzeuge bereitzustellen ist die leichte Hälfte, aber die Qualifizierung ist die Hälfte, die den Unterschied macht.**

Zwei Länder haben das Navi-Problem so in ihre Politik überführt, wie es eine Fahrschule tun würde: erst die Karte, dann das Gerät. Norwegen empfahl im Juni 2026, generative KI im Wesentlichen aus

der Grundschule herauszuhalten und sie erst in der Sekundarstufe I zu unterrichten, wo Lehrkräfte entsprechend geschult wurden, und KI dabei als bewusst zu vermittelnde Fähigkeit zu behandeln. Luxemburg geht ähnlich stufenweise vor. Beachten Sie genau: Norwegen sprach eine **Empfehlung** aus, kein verbindliches Verbot – auch wenn die internationale Presse es so darstellte.

Der blinde Fleck

Ein Risiko von KI wird selten benannt, und auch das Bild führt dorthin: Das Wissen über die Karte der Stadt London **ist nicht von Dauer**. Eine FahrerIn, die fünf Jahre lang den Pfeilen folgt, behält die Stadt nicht im Kopf. Es verblasst, und sie bemerkt die aufgerissene Straße nicht mehr. Ihre eigene Expertise erhalten Sie, indem Sie genau die Dinge selbst tun, die KI am besten übernimmt: die Aufgabe formulieren, das Missverständnis vorausahnen und die Erklärung entwerfen. **Wo KI die Arbeit übernimmt, durch die sich Urteilsvermögen bildet, wird aus Entlastung Dequalifizierung. In einem deutschen Expertengremium benannten 65 % der Bildungsexpert:innen dieses Risiko für den Lehrberuf – ein Anstieg gegenüber 44 % im Jahr zuvor.**

Nutzen Sie das Navi. Bewahren Sie sich nur die Karte.

Revision #12

Created 2026-09-01 09:16:52 UTC by Alexander

Updated 2026-09-02 07:12:57 UTC by Jannik