Bildungscampus

EMSE-Netzwerk
„Empiriegestützte Schulentwicklung"

SAARLAND

38. Tagung des EMSE-Netzwerkes:
KI als Instrument datengestützter Qualitätsentwicklung

PROGRAMM

Donnerstag, 19. November 2026

12:00 – 13:00 **Ankommen & Anmeldung**

13:00 – 13:15 **Begrüßung**
Christine Streichert-Clivot, Ministerin für Bildung und Kultur
Dr. Anika Limburg, Direktorin Bildungscampus Saarland

13:15 – 14:30 **Keynote I mit anschl. Diskussion**
»KI und Adaptivität: Hebel für Lernen oder Verlust von Selbstregulation?«
Prof. Dr. Andreas Lachner, Eberhard Karls Universität Tübingen,
Tübingen Center for Digital Education (TüCeDE), Co-Director

14:30 – 14:45 **Pause**

14:45 – 15:45 **Keynote II mit anschl. Diskussion**
»Bildungsmonitoring als Motor für Schulentwicklung? – Von Bildungsdaten zu gerechteren Lerngelegenheiten mit Künstlicher Intelligenz – das Beispiel Luxemburg«
Dr. Isabell Baumann, Université du Luxembourg

15:45 – 16:00 **Pause**

16:00 – 17:20 **Parallele Sessions**

	Aula	**Raum N-0.21**
16:00 – 16:40	»Von Daten zu Entwicklung: Das Funktionsraummodell als Analyseperspektive auf KI in schulischer Qualitätsentwicklung« Prof. Dr. Wolfram Rollett, Carl von Ossietzky Universität Oldenburg	»KI als Führungs- und Arbeitsassistent: Wie kann Künstliche Intelligenz Führungskräfte und Mitarbeitende bei anspruchsvollen Gesprächen, strategischen Entscheidungen und der täglichen Arbeitsorganisation konkret unterstützen?« (praxisorientierter Workshop) Thomas Bonerz, Bildungscampus Saarland (BCS)

Poststraße 6 · 66115 Saarbrücken
www.saarland.de/bildungscampus

16:40 – 17:20 »Wenn das Evaluationstool selbst nachfragt: KI-gestützte adaptive Interviews für datengestützte Qualitätsentwicklung«

Daniel Autenrieth, RWTH Aachen;
Nina Autenrieth, PH Schwäbisch Gmünd;
Dr. Lea Schulz, Europa-Universität Flensburg, Institut für Qualitätsentwicklung für Schulen in Schleswig-Holstein

»KI als Prozessmoderator schulischer Qualitätsentwicklung - Assistent für Berufsschulen im Startchancen Programm«

Daniel Hunold, Martina Hoffmann, Staatsinstitut für Schulqualität und Bildungsforschung (ISB) München

17:30 **Ende & Beginn Abendprogramm**
Treffpunkt 18:00 Uhr am Hauptportal Ludwigskirche Saarbrücken
(etwa 1 km Fußweg (ca. 15 Minuten) ab Veranstaltungsgebäude)

Freitag, 20. November 2026

08:00 – 08:30 **Ankommen**

08:30 – 09:15 **Vortrag**
»Schule.Macht.KI.– Evidenzbasierte Befunde zur Nutzung, Einstellungen und Entwicklung von KI in Österreich«

Priv.-Doz. Dr. Ann Cathrice George, IQS Salzburg; Dr. Martina Müller, PH Wien;
Prof. Dr. Christoph Helm, JKU Linz

09:15 – 10:45 **Parallele Sessions**

	Aula	Raum N-0.21	Raum N-0.19
09:15 – 10:00	»Nutzung von Künstlicher Intelligenz im Kontext interner Evaluationen in Schulen« Dirk Groß, Pädagogisches Landesinstitut Rheinland-Pfalz (PL); Dr. Jana Steinbacher, Institut für Bildungsanalysen Baden-Württemberg (IBBW)	»KI-gestützte Diagnose fachlichen Tiefenverständnisses am Beispiel Mathematik: Chancen und Grenzen für eine wirksame Förderung« Dr. Corinna Hankeln, TU Dortmund	»Ko-Design mit CoDe-Graph: eine KI-unterstütze, graphisch Sprache« Allison Kolling, Saroj Sharma, Prof. Dr. Armin Weinberger, Universität des Saarlandes
10:00 – 10:45	»LLMs-as-Judges zur Qualitätssicherung KI-generierter Erklärungs- und Interpretationshilfen in Schuldashboards« Dr. Jana Steinbacher, Dr. Paul Schröder-Pfeifer, Dr. Jessica Gnas, Institut für Bildungsanalysen Baden-Württemberg (IBBW)	»KI in der Schulentwicklung: Rahmenkonzeption von Verantwortung und Handlungsfähigkeit der Akteur*innen im Mehrebenensystem« (Posterbeitrag) PD Dr. Ramona Lorenz, TU Dortmund; Dr. Manuela Endberg, Universität Duisburg-Essen	»Fortbildungsbedarfe saarländischer Lehrkräfte: Erste Ergebnisse der landesweiten Bedarfserhebung mit Befunden zur Spannung zwischen Nutzen und Handlungssicherheit im Umgang mit KI.« Dr. Katharina Hary, Martin Ernst, Bildungscampus Saarland (BCS)

10:45 – 11:00 **Pause**

11:00 – 12:30	**Parallele Sessions** **Aula**	**Raum N-0.21**	**Raum N-0.19**
11:00 – 11:45	»Einsatz eines KI-basierten Tutorensystems zur Sprachförderung in der Primarstufe: Er-kenntnisse aus der begleitenden Evaluation« Dr. Svitlana Mokhonko, Dr. Alexandra Dehmel, Institut für Bildungsanalysen Baden-Württemberg (IBBW); Ines Pieronczyk, Zentrum für Schulqualität und Lehrerbildung (ZSL)	»Von qualitativen Evaluationsdaten zur organisationsdiagnostischen Analyse – Potenziale und Kontrollfragen generativer KI für datengestützte Schulentwicklung« (Vortrag)	»Implementierungsfragen und schulische Praxis am Beispiel der Implementation der Lernverlaufsdiagnostik quop in Baden-Württemberg« (Vortrag)
11:45 – 12:30		*anschließend* »Plausibel ist noch nicht belastbar – qualitative Schulentwicklungsdaten mit KI nutzen und beurteilen« (Workshop) Dr. Holger Gerlach, Anna-Victoria Kloth, Qualitätsentwicklung Thüringer Schulen (QThüS)	*anschließend* »Lernverlaufsdiagnostik quop – ein evidenzbasiertes Diagnosetool im Lesen und in Mathematik mit integrierter adaptiver Förderung« (Workshop) Dr. Ulrike Krug, Direktorin a. D. Hessisches Kultusministerium

12:30 – 13:00 **Abschlussrunde & EMSE-Perspektiven**

Hier gelangen Sie zur Tagungswebseite.

Anhang: Abstracts zu den Beiträgen

ABSTRACTS

Bildungsmonitoring als Motor für Schulentwicklung?

Von Bildungsdaten zu gerechteren Lerngelegenheiten mit Künstlicher Intelligenz – das Beispiel Luxemburg

Dr. Isabell Baumann, Universität Luxemburg

Mit den Épreuves standardisées (ÉpStan) verfügt Luxemburg über ein nationales Bildungsmonitoring, das jährlich die Kompetenzen vollständiger Jahrgangskohorten erfasst und die Ergebnisse bis auf Schul-, Klassen- und Individualebene zurückmeldet. Dabei werden auch Unterschiede zwischen Schülergruppen sichtbar. Gerade in Luxemburg hängen die Kompetenzen der Schüler:innen stark von ihrer sozialen Herkunft ab. Das zeigt auch PISA 2025: Der Leistungsunterschied zwischen Jugendlichen mit hohem und niedrigem sozioökonomischem Status beträgt in den Naturwissenschaften 130 Punkte; im OECD-Durchschnitt sind es 85 Punkte.

Der Vortrag geht der Frage nach, wie Bildungsdaten für Schul- und Unterrichtsentwicklung genutzt werden können. KI eröffnet in diesem Zusammenhang neue Möglichkeiten, Schüler:innen gezielt zu unterstützen und Lerngelegenheiten gerechter zu gestalten.

Von Daten zu Entwicklung: Das Funktionsraummodell als Analyseperspektive auf KI in schulischer Qualitätsentwicklung

Prof. Dr. Wolfram Rollett, Carl von Ossietzky Universität Oldenburg

Datengestützte schulische Qualitätsentwicklung verbindet Datengewinnung und -aufbereitung mit Deutung, Entscheidung, Umsetzung, Evaluation und der Rückbindung von Ergebnissen an weitere Entscheidungen und Entwicklungsprozesse. Gerade die Übergänge zwischen diesen Prozesslagen sind analytisch bedeutsam, weil sich hier entscheidet, ob und wie weitere Bearbeitung anschließen kann. Generative KI kann hier Informationen strukturieren, Befunde verdichten, alternative Deutungen sichtbar machen oder Handlungsoptionen formulieren. Damit stellt sich die Frage, welche Leistungen KI an solchen Übergängen übernehmen kann und unter welchen Bedingungen daraus tragfähige Anschlüsse für weitere Entwicklungsarbeit entstehen.

Das Funktionsraummodell (Rollett, 2026) bietet hierfür eine relationale Analyseperspektive. Es fasst Prozesse als Übergänge zwischen unterscheidbaren Prozesslagen und beschreibt Ermöglichung, Stabilisierung und Regulierung als allgemeine Leistungsarten. Für einen konkreten Übergang fragt das Modell, welche Leistungen erforderlich sind, welche Beiträge tatsächlich wirksam werden und ob deren Konfiguration den Anforderungen des Übergangs adäquat ist. Auf dieser Grundlage lässt sich bestimmen, welche funktionalen Rollen generative KI übernehmen könnte und welche weiteren Bedingungen für eine tragfähige Anschlusslage relevant sind.

Diese Perspektive wird anhand exemplarischer Störungs- und Risikolagen datengestützter schulischer Qualitätsentwicklung konkretisiert. Betrachtet werden Übergänge von Daten zu Deutungen, von Deutungen zu Entscheidungen sowie von Evaluation zu Revision. Dabei wird herausgearbeitet, wo Brüche, Drift oder Blockaden auftreten können und welche

Regulierungsrisiken bei fehlender Rückbindung, Prüfung oder Verantwortungszuordnung entstehen können. Untersucht wird, welche Leistungen generative KI bei der Rückbindung an Daten und Kriterien, der Prüfung konkurrierender Deutungen, der Explikation von Entscheidungsoptionen oder der Dokumentation von Revisionsbedarf übernehmen könnte. Zugleich werden Bedingungen betrachtet, unter denen der Einsatz generativer KI selbst Drift, Blockaden oder Regulierungsrisiken begünstigen kann.

Ziel ist, die Analyseperspektive für ausgewählte Übergänge schulischer Qualitätsentwicklung konzeptionell zu konkretisieren. Dabei werden Leistungsanforderungen präzisiert, mögliche funktionale Rollen generativer KI bestimmt und Ansatzpunkte für eine feldspezifische Operationalisierung und empirische Prüfung abgeleitet.

Literatur

Rollett, W. (2026). Der Modellkern des Funktionsraummodells als Theorie funktionaler Übergangsstabilisierung [Working Paper]. Oldenburger Online-Publikations-Server (OOPS), Carl von Ossietzky Universität Oldenburg. https://nbn-resolving.org/urn:nbn:de:gbv:715-oops-75199

Wenn das Evaluationstool selbst nachfragt: KI-gestützte adaptive Interviews für datengestützte Qualitätsentwicklung

Daniel Autenrieth, RWTH Aachen; Nina Autenrieth, PH Schwäbisch Gmünd; Dr. Lea Schulz, Europa-Universität Flensburg, Institut für Qualitätsentwicklung für Schulen in Schleswig-Holstein

Datengestützte Qualitätsentwicklung nutzt häufig standardisierte Befragungen und Kennzahlen. Diese erfassen Bewertungen vergleichbar, lassen aber wenig Raum, Begründungen, Bedingungen, Widersprüche und unerwartete Themen aus individuellen Antworten weiterzuverfolgen. Qualitative Interviews können solche Prozesse rekonstruieren, sind bei größeren Gruppen jedoch personal- und terminintensiv. Der Beitrag untersucht an einem Anwendungsfall den zusätzlichen Erkenntniswert adaptiver KI-Interviews und die Voraussetzungen ihres Einsatzes.

Das eingesetzte Instrument verbindet einen von Forschenden konfigurierten Leitfaden mit einer Gesprächsführung, die auf individuelle Antworten reagiert. Für jede Person führt eine eigene Instanz des KI-Interviewers das Gespräch. Sie kann Aussagen konkretisieren lassen, auf frühere Antworten zurückkommen und Wendepunkte vertiefen. Die Gespräche können parallel oder zeitlich unabhängig stattfinden. Erkenntnisinteresse, Vertiefungsdimensionen und Rahmenbedingungen werden vorab festgelegt. Die vollständigen Gesprächsverläufe bleiben für Prüfung und qualitative Analyse erhalten.

Untersucht wurde der Einsatz des Instruments in einem Projektseminar für angehende Grundschullehrkräfte. Die Studierenden setzten sich dort theoretisch und praktisch mit KI und ihren Potenzialen für Unterrichts- und Schulentwicklung auseinander. Zu Beginn erschlossen 15 adaptive Kurzinterviews den Problemraum: Die Studierenden beschrieben Herausforderungen des Grundschulalltags wie Heterogenität, fehlende Zeit für individuelle Förderung und Überlastung. Zum Abschluss folgten 15 ausführliche Reflexions- und Evaluationsinterviews. Sie rekonstruierten fallbezogen, wie die Teilnehmenden Veränderungen ihrer Handlungs- und Urteilsfähigkeit erklärten, wo Lernprozesse stockten und welchen Seminarbestandteilen sie Wirkung zuschrieben. Adaptive Nachfragen machten unerwartete Themen, Widersprüche und abweichende Verläufe sichtbar; der Fallvergleich zeigte gemeinsame Gelingensbedingungen und Unterschiede. Als wirksam beschrieben wurden

technische Begleitung, Zeit für iterative Überarbeitung, Zusammenarbeit, Praxisfeedback und öffentliche Präsentation. Technische Unsicherheit und Zeitaufwand blieben Belastungen.

Der Beitrag zeigt, wie Bildungseinrichtungen solche Daten für die Weiterentwicklung ihrer Angebote nutzen und mit standardisierten Evaluationen verbinden können. Ausgehend vom Fallbeispiel wird diskutiert, unter welchen Bedingungen sich adaptive KI-Interviews auf andere Bildungskontexte übertragen lassen, etwa auf Schüler:innen- und Elternfeedback oder die begleitende Evaluation von Schulentwicklungsvorhaben. Dabei werden auch Fragen einer verantwortlichen Nutzung behandelt, darunter die Grenzen automatisierter Gesprächsführung und die Zuständigkeit für Auswertung und Entscheidungen.

KI als Prozessmoderator schulischer Qualitätsentwicklung - Assistent für Berufsschulen im Startchancen Programm

Daniel Hunold, Martina Hoffmann, Staatsinstitut für Schulqualität und Bildungsforschung (ISB) München

Datenerhebungen allein verändern noch keine Schule. Entscheidend ist, wie schulische Teams Befragungsergebnisse gemeinsam interpretieren und daraus tragfähige Entwicklungsentscheidungen ableiten – ein komplexer Prozess mit zahlreichen Herausforderungen.[1] Hierbei stellt sich die Frage, auf welche Weise ein KI-Assistent unterstützend hinzugezogen werden kann: Schließlich liegt der spezifische Vorteil einer künstlichen Intelligenz im Erkennen von Mustern und Tendenzen in heterogenen umfangreichen Datenmengen, die anschließend gemeinsam eingeordnet werden können.

Im Startchancen-Programms an beruflichen Schulen in Bayern zeigt sich diese Herausforderung sehr konkret. Die schulische IST-Analyse verbindet standardisierte Befragungen des Personals und externer Beteiligter mit sehr unterschiedlichen analogen und digitalen Formen des Schülerfeedbacks. Diese Daten zunächst zusammenzuführen, lesbar aufzubereiten und anschließend sinnvoll aufeinander zu beziehen, stellt für die beteiligten Schulteams einen erheblichen Aufwand dar. Gleichzeitig hängt die zielgerichtete Nutzung der Daten wesentlich von schulorganisatorischen Bedingungen und der Zusammenarbeit der beteiligten Akteure ab, die zudem entsprechende Kompetenzen benötigen, um die erhobenen Daten sachgerecht zu analysieren und zu interpretieren.[2] Hier setzt die gemeinsam von der Abteilung Berufliche Schulen am ISB (Staatsinstitut für Schulqualität und Bildungsforschung) und KIKO (KI-Kompetenzzentrum der ALP Dillingen) entwickelte *IST-Analyse mit KI-Auswertung* an. Die Daten werden zunächst KI-gestützt aufbereitet. Darauf aufbauende Analyseassistenten machen deskriptiv Stärken, Auffälligkeiten und mögliche Entscheidungsfelder sichtbar, ohne Entscheidungen vorzugeben. Ausgewählte Handlungsfelder werden anschließend in einen Chancen-Chatbot übertragen, der das Schulteam im Chancengespräch bei der Priorisierung sowie bei der Entwicklung konkreter Ziele, Maßnahmen und überprüfbarer Indikatoren unterstützt. Nach der Umsetzung können diese Vereinbarungen in einen Evaluations-Chatbot übernommen werden, der passende neue Befragungsitems entwickelt, um Zielerreichung und angestrebte Veränderungen in einen nachhaltigen Entwicklungsprozess überführt.

---

[1] Solche Auswertungsprozesse weisen eine immense Komplexität auf und stellen hohe Anforderungen an die hermeneutische Kompetenz der Beteiligten (vgl. Schildkamp, Poortman & Handelzalts, 2016; Schildkamp et al., 2017. S. 246).
[2] vgl. Kim Schildkamp, Cindy Poortman, Hans Luyten & Johanna Ebbeler (2017) Factors promoting and hindering data-based decision making in schools, School Effectiveness and School Improvement, 28:2, 242-258, DOI: 10.1080/09243453.2016.1256901 S. 244f.

Konzeptionell folgt das aufeinander aufbauende KI-System der von Molenaar beschriebenen Augmentationsperspektive hybrider Systeme, bei der die Expertise des Menschen nicht ersetzt, sondern insbesondere bei der Verarbeitung und Analyse von Daten durch künstliche Intelligenz ergänzt und entlastet wird.[3] Menschliche und künstliche Intelligenz wirken in diesem System komplementär zusammen, indem die Verteilung von Aufgaben und Kontrolle zwischen Menschen und KI bewusst gestaltet wird. Die beteiligten Bildungsakteure werden im Rahmen einer strukturierten Prozessmoderation unterstützt. Die abschließende Priorisierung und Entscheidung über die Gestaltung und Umsetzung der Maßnahmen verbleibt bei den schulischen Akteuren. Im Mittelpunkt des Projekts steht somit die Frage, wie generative KI als Prozessmoderator in einen datensensiblen, partizipativen Schulentwicklungsprozess eingebunden werden kann.

Schule.Macht.KI.– Evidenzbasierte Befunde zur Nutzung, Einstellungen und Entwicklung von KI in Österreich

Priv.-Doz. Dr. Ann Cathrice George, IQS Salzburg; Dr. Martina Müller, PH Wien; Prof. Dr. Christoph Helm, JKU Linz

Schule steht vor der Herausforderung Lernende auf zum aktuellen Zeitpunkt nicht vorhersehbare und schnell wechselnde Anforderungen vorzubereiten. Eine dieser Entwicklungen ist der Einsatz Künstlicher Intelligenz (KI), den das gesellschaftliche Leben, die Wirtschaft und der Arbeitsmarkt zunehmend fordern.

Der Einzelvortrag geht der Frage nach, inwieweit Österreichs Schulen für die Nutzung von KI gerüstet sind (George, Helm & Müller, 2025). Zur Beantwortung dieser Frage werden Erkenntnisse aus aktuellen international-quantitativen Studien TIMSS, PIRLS, PISA und ICLIS auf der Systemebene genutzt und durch qualitative Interviews auf der Unterrichtsebene verfeinert. Diese evidenzbasierten Befunde können im Sinne einer datengestützten Schul- und Unterrichtsentwicklung Lehrende in ihrer Professionalisierung unterstützen und den Kompetenzerwerb der Lernenden vorantreiben (Maritzen, 2024, S. 405).

Aus den Kontextbefragungen der internationalen Studien wird abgeleitet, wie weit die Digitalisierung bereits vorangeschritten ist, welche Ausstattung Schulen aufweisen und welche Einstellungen Lehrpersonen zum Thema Digitalisierung und KI aufweisen. Dieser auf Systemebene berichtete Stand wird um die aktuellen bildungspolitischen Maßnahmen in Österreich ergänzt.

Zur Konkretisierung der Erkenntnisse werden Daten aus der quantitativen „öbv-Lehrkräftebefragung rund um Künstliche Intelligenz“ (Helm et al., 2024a) sowie eine repräsentative Elternbefragung zur KI-Nutzung (Helm et al., 2024b) herangezogen.
Ergänzend liefern qualitative Interviews aus Praxisbeispielen im Bereich der MS und AHS einen Einblick in den aktuellen Unterricht einzelner Fächer, um in einem weiteren Schritt mittels quantitativer Inhaltsanalyse die manifesten Inhalte zu erfassen, zu deuten, daraus Erkenntnisse zu gewinnen und Schlüsse zu ziehen (Schneijderberg et al., 2022, S. 18).
Die Erkenntnisse aus den unterschiedlichen Studientypen auf den unterschiedlichen Ebenen des Systems werden schließlich trianguliert, miteinander abgeglichen und diskutiert.

[3] vgl. Molenaar, I. (2022). Towards hybrid human-AI learning technologies. European Journal of Education, 57, 632–645. https://doi.org/10.1111/ejed.12527 S. 632f.

Literatur
George, A.C., Helm, C., & Müller, M. A. (2025). Zum Stand von Digitalisierung und Künstlicher Intelligenz im österreichischen Bildungssystem. *R&E Source*. 2025(3), 342 – 355.

Maritzen, N. (2024). Warum eigentlich mit Daten arbeiten und was die Alternative wäre. *Erziehung und Unterricht*. 174 (5-6).

Schneijderberg, Ch., Wieczorek, O., & Steinhardt, I. (2022). *Qualitative und quantitative Inhaltsanalyse: digital und automatisiert*. Beltz.

Helm, C., Grosse, C. S. & öbv (2024a). Einsatz künstlicher Intelligenz im Schulalltag – eine empirische Bestandsaufnahme. *Erziehung und Unterricht* (3-4), 370–381.

Helm, C., Dorfer, R., Aistleitner, T. & Postlbauer, A. (2024b). *Österreichweite Elternbefragung zum Lehrkräftemangel*. Forschungsbericht. Johannes Kepler Universität. Linz.

Nutzung von Künstlicher Intelligenz im Kontext interner Evaluationen in Schulen

Dirk Groß, Pädagogisches Landesinstitut Rheinland-Pfalz (PL); Dr. Jana Steinbacher, Institut für Bildungsanalysen Baden-Württemberg (IBBW)

Interne Evaluationen sind in Schulen bisher noch kaum entwickelt und können nicht als Stärken schulischer Arbeit beschrieben werden (Pietsch, 2011, S. 13). Bei Schulleitungen gehörten sie zu den „bei weitem unbeliebtesten und belastendsten“ Aufgaben (Huber et al., 2013, S. 261). Im Beitrag wird daher die Frage aufgegriffen, inwieweit die Nutzung von Künstlicher Intelligenz (KI) einen positiven Einfluss auf die Qualität von internen Evaluationen in Schulen haben kann. Die Qualitätseinrichtungen der Bundesländer beschäftigen sich im Rahmen ihrer „Länderkooperation Interne Evaluation“ bereits seit einiger Zeit mit Fragen der Professionalisierung interner Schulevaluationen durch den Einsatz von KI. Nach einer kurzen Vorstellung der Länderkooperation und ihrer Ziele werden im Beitrag zunächst einige Einsatzfelder, Chancen und Risiken der KI in der Evaluation skizziert (vgl. Hense, 2025). Evaluation wird dabei im Kontext datengestützter Schul- und Unterrichtsentwicklung ausdrücklich abgegrenzt von Lernstandserhebungen wie z.B. VERA und Verfahren der pädagogischen Diagnostik. Es wird überblicksartig dargelegt, welche Angebote für Schulen in den Bundesländern zum Thema bereits existieren und welche Erkenntnisse dazu vorliegen, wie Schulen bisher KI für datengestütztes Arbeiten verwenden. Im Hauptteil unseres Beitrages stellen wir vor, mit welcher Systematik und welchem Evaluationsverständnis in der Länderkooperation vorgegangen wird, um zielführende Produkte und Qualifizierungen für Schulen mit Blick auf das Thema zu generieren. Grundlage für diesen Arbeitsprozess sind die zehn Schritte einer Evaluation nach Balzer & Beywl (2026). Zum Schluss des Vortrags werden erste Anwendungsbeispiele und -ideen präsentiert und beispielhaft darauf eingegangen, welche Effekte auf die Qualität interner Evaluationen in Schulen denkbar sind (vgl. DeGEval, 2004, 2017). Bei dem Vortrag handelt es sich um einen Werkstattbericht, der keine fertigen Lösungen impliziert. Daher möchten wir in der sich anschließenden Diskussion mit den Teilnehmenden in einem offenen Format die Einsatzpotenziale von KI im Rahmen interner Evaluationen in Schulen reflektieren.

Literatur
Balzer, L. & Beywl, W. (2026, im Erscheinen). *evaluiert – erweitertes Planungsbuch für Evaluationen im Bildungsbereich*. hep-Verlag.

DeGEval – Gesellschaft für Evaluation (2004). *Empfehlungen zur Anwendung der Standards für Evaluation im Handlungsfeld der Selbstevaluation*. https://www.degeval.org/fileadmin/content/Z03_Publikationen/DeGEval_-_Empfehlungen_Selbstevaluation.pdf

DeGEval – Gesellschaft für Evaluation (2017). *Standards für Evaluation. Erste Revision 2016*. https://www.degeval.org/fileadmin/content/Z03_Publikationen/DeGEval-Standards/DeGEval_Standards_fuer_Evaluation_kurz_2016.pdf

Hense, J. U. (2025). *Anwendungen Künstlicher Intelligenz in der Evaluation: Stand von Forschung und Praxis* (PrEval Expertise 1/2025).DOI: 10.48809/PrEvalExp2501

Huber, S. G., Wolfgramm, C. & Kilic, S. (2013). Vorlieben und Belastungen im Schulleitungshandeln: Ausgewählte Ergebnisse aus der Schulleitungsstudie 2011/2012 in Deutschland, Österreich, Liechtenstein und der Schweiz. In S. G. Huber (Hrsg.), *Jahrbuch Schulleitung 2013* (S. 259-271).Wolters Kluwer.

Pietsch, M. (2011). Die Evaluationspraxis an deutschen Schulen. Ein Überblick aus empirischer Sicht. *Schulmanagement, 42*(4), 12-14.

LLMs-as-Judges zur Qualitätssicherung KI-generierter Erklärungs- und Interpretationshilfen in Schuldashboards

Dr. Jana Steinbacher, Dr. Paul Schröder-Pfeifer, Dr. Jessica Gnas, Institut für Bildungsanalysen Baden-Württemberg (IBBW)

Data Literacy gilt als zentrale Voraussetzung für die Nutzung schulischer Daten zur Qualitätsentwicklung (Lee et al., 2024). Dashboards können hierbei als Unterstützungsinstrumente fungieren, indem sie komplexe Schuldaten übersichtlich aufbereiten. Das „Schuldatenblatt“ ist ein solches Dashboard, welches die datengestützte Qualitätsentwicklung an Schulen in Baden-Württemberg unterstützt. Aufgrund unterschiedlicher Ausprägungen von Data Literacy besteht jedoch ein Bedarf an individualisierteren Hilfen beim Verständnis sowie bei der Interpretation schulbezogener Daten (Curran et al., 2024).

Large Language Models (LLMs) bieten das Potenzial, solche Erklärungs- und Interpretationshilfen im Kontext von Schuldashboards automatisiert bereitzustellen. Gleichzeitig erfordert ihr Einsatz eine systematische Qualitätssicherung. Ein hierfür zunehmend eingesetzter Ansatz ist *LLM-as-a-Judge*: Ein LLM bewertet die Qualität, Relevanz und Angemessenheit KI-generierter Outputs anhand definierter Kriterien (Li et al., 2024).

Der Beitrag geht am Beispiel des Schuldatenblatts auf zwei unterschiedliche LLM-basierte Bewertungssysteme („Judges“) ein, die als Single-LLM-Judges konfiguriert sind und aufgrund unterschiedlicher Bewertungsgegenstände jeweils eigene Kriterien und Bewertungslogiken verwenden. Der erste Judge fokussiert die faktische Genauigkeit KI-generierter Erklärungen (insbesondere die Wiedergabe korrekter Daten) anhand ausgewählter Prompting-Strategien (Schulhoff et al., 2025). Der zweite Judge bewertet die inhaltliche und kontextuelle Passung der Erklärungen. Hierzu werden definierte Bewertungskriterien und Dimensionen vorgegeben (z. B. Kohärenz und inhaltliche Genauigkeit, aber auch das Einhalten inhaltlicher Grenzen). Darüber

hinaus wird externes Wissen (z. B. zu Lernstandserhebungen) mittels Retrieval Augmented Generation (RAG) in den Bewertungsprozess integriert, um die Beurteilung an den relevanten fachlichen Kontext zu binden und Halluzinationen zu reduzieren (Chen et al., 2025; Lin & Chen, 2023).

Anhand konkreter Datendarstellungen des Schuldatenblatts wird untersucht, inwieweit die beiden Judges eine verlässliche und differenzierte Qualitätssicherung KI-generierter Erklärungs- und Interpretationshilfen ermöglichen. Der Beitrag zeigt damit auf, welche Anforderungen an LLMs-as-Judges in schulischen Dashboards zu stellen sind und wie solche Bewertungssysteme selbst einer kritischen Beurteilung unterzogen werden können (Meta-Validität).

Literatur

Chen, Y.-S., Jin, J., Kuo, P.-T., Huang, C.-W. & Chen, Y.-N. (2025). *LLMs are Biased Evaluators But Not Biased for Fact-Centric Retrieval Augmented Generation.* https://doi.org/10.48550/arXiv.2410.20833

Curran, F. C., Carlo, S. & Harris-Walls, K. (2024). Making the data visible: A systematic review of systems level data dashboards for leadership and policy in education. *Review of Educational Research, 96*(1), 3–52. https://doi.org/10.3102/00346543241288249

Lee, J., Alonzo, D., Beswick, K., Oo, C. Z., Anson, D. W. J. & Abril, J. M. V. (2024). Data literacy of principals in K–12 school contexts: A systematic review. *Educational Research Review, 45*, 100649. https://doi.org/10.1016/j.edurev.2024.100649

Li, H., Dong, Q., Chen, J., Su, H., Zhou, Y., Ai, Q., Ye, Z. & Liu, Y. (2024). *LLMs-as-Judges: A Comprehensive Survey on LLM-based Evaluation Methods.* https://doi.org/10.48550/arXiv.2412.05579

Lin, Y.-T. & Chen, Y.-N. (2023). *LLM-EVAL: Unified Multi-Dimensional Automatic Evaluation for Open-Domain Conversations with Large Language Models.* https://doi.org/10.48550/arXiv.2305.13711

Schulhoff, S., Ilie, M., Balepur, N., Kahadze, K., Liu, A., Si, C., Li, Y., Gupta, A., Han, H., Schulhoff, S., Dulepet, P. S., Vidyadhara, S., Ki, D., Agrawal, S., Pham, C., Kroiz, G., Li, F., Tao, H., Srivastava, A., Da Costa, H., Gupta, S., Rogers, M. L., Goncearenco, I., Sarli, G., Galynker, I., Peskoff, D., Carpuat, M., White, J., Anadkat, S., Hoyle, A. & Resnik, P. (2025). *The Prompt Report: A Systematic Survey of Prompt Engineering Techniques.* https://doi.org/10.48550/arXiv.2406.06608

KI-gestützte Diagnose fachlichen Tiefenverständnisses am Beispiel Mathematik: Chancen und Grenzen für eine wirksame Förderung

Dr. Corinna Hankeln, TU Dortmund

Im Zentrum von Unterrichts- und Schulentwicklung sollte die Frage stehen, ob Lernende fachliche Inhalte wirklich verstanden haben und nicht nur, ob sie richtige Antworten reproduzieren können. Datengestützte Qualitätsentwicklung greift sonst kurz, wenn sie sich nur auf leicht messbare, aber fachlich oberflächliche Indikatoren beschränkt. Der Beitrag plädiert dafür, dass Diagnose- und Fördersysteme gezielt auf fachlich relevante Aspekte wie tiefes Verständnis, Begriffsverständnis und Fehlvorstellungen ausgerichtet werden können, damit

Schul- und Unterrichtsentwicklung an den richtigen Stellen ansetzt.

Am Beispiel des Projekts „Assessment for Learning with AI“ (AIwAI) zeigen wir, wie KI-gestützte Verfahren dazu beitragen können, ein solches tieferes Verständnis in sprachlichen Antworten sichtbar zu machen. AIwAI erweitert dazu den Mathe-sicher-können-Online-Check (MSK-OC, entwickelt vom DZLM am IPN in Kooperation mit dem DIPF, Hankeln et al. 2025) unter anderem um Speech-to-Text und automatisierte Klassifikation offener Antworten: Während im MSK-OC geschlossene Items bereits vollautomatisch ausgewertet werden, erhalten Lehrkräfte bei offenen Items bislang vor allem Vorschläge für Auswertungskategorien, die noch manuell bestätigt werden müssen. Sprechen Lernende ihre Erklärungen stattdessen frei, statt sie zu verschriftlichen, werden ihre mündlichen Begründungen, Vorstellungen und Fehlvorstellungen deutlich detailreicher und authentischer erfasst als in kurzen schriftlichen Antworten (Hankeln et al., 2026). Eine KI kann die transkribierten, offenen Antworten anschließend automatisiert klassifizieren und sie typischen Verständnis- bzw. Fehlvorstellungsmustern zuordnen. So entsteht ein noch differenzierterer Einblick in das konzeptuelle Verständnis der Lernenden, als er mit dem bisherigen Verfahren möglich ist.

Gleichzeitig zeigt unsere bisherige Arbeit deutlich: Eine automatisierte Diagnose allein reicht für eine wirksame Förderung nicht aus. Der Beitrag diskutiert daher, wie KI-gestützte Diagnostik und lehrkraftgeführte Förderung sinnvoll zusammenwirken können, und benennt konzeptionelle, organisatorische und ethische Herausforderungen (u. a. Datenschutz, Transparenz, Bias), die bei der Einführung solcher Systeme in die Schulpraxis zu beachten sind.

Für die Teilnehmenden bietet der Beitrag konkrete Impulse, wie datengestützte Qualitätsentwicklung so gestaltet werden kann, dass fachliche Tiefe nicht der Messbarkeit geopfert wird und wie integrierte KI-Komponenten hier unterstützend, aber nicht ersetzend wirken sollte.

Literatur

Hankeln, C., Kroehne, U., Voss, L., Gross, S., & Prediger, S. (2025). Developing digital formative assessment for deep conceptual learning goals: Which topic-specific research gaps need to be closed? *Educational Technology Research and Development, 73*(4), 1953–1973. https://doi.org/10.1007/s11423-025-10486-x

Hankeln, C., Kroehne, U., Hahn, S., & Prediger, S. (2026). Fifth and sixth graders speak more richly than they write: Benefits of speech input for digitally assessing conceptual understanding. *Technology, Knowledge and Learning, 31*, 1099. https://doi.org/10.1007/s10758-025-09939-2

KI in der Schulentwicklung: Rahmenkonzeption von Verantwortung und Handlungsfähigkeit der Akteur*innen im Mehrebenensystem

PD Dr. Ramona Lorenz, TU Dortmund; Dr. Manuela Endberg, Universität Duisburg-Essen

Künstliche Intelligenz (KI) verändert zunehmend Prozesse der Schulentwicklung und eröffnet vielfältige Potenziale für datengestützte Entscheidungen, Organisation schulischer Prozesse, Unterricht, Kooperation und Professionalisierung. Gleichzeitig entstehen neue Herausforderungen hinsichtlich menschlicher Verantwortung, Transparenz von Entscheidungsgrundlagen und Gestaltungsoptionen. Während internationale Diskurse zu *Human-Centered AI* und *Responsible AI* die Notwendigkeit betonen, KI-Systeme so zu gestalten,

dass sie menschliche Autonomie und demokratische Werte stärken, fehlt bislang ein konzeptioneller Rahmen, der Verantwortung (*Accountability*) und Handlungsfähigkeit (*Agency*) der verschiedenen Akteur*innen im Schulsystem systematisch zusammendenkt.

Der Beitrag entwickelt ein theoretisches Mehrebenenmodell für die KI-gestützte Schulentwicklung, das Perspektiven der Schulentwicklungsforschung sowie Konzepte von *Human Agency* und *Responsible AI* auf Basis eines Forschungsreviews und zweier Workshops eines internationalen und interdisziplinären Expert*innennetzwerks des „School Development and AI Agency Research Lab“ (SD-AI) am College for Social Sciences and Humanities der UA Ruhr integriert. Im Zusammenspiel verschiedener Akteur*innen in Einzelschulen (Schulleitungen, Lehrkräfte, Lernende und Eltern), regionalen Bildungspartnerschaften, Bildungsadministration und Bildungsforschung entsteht ein komplexes soziotechnisches System, in dem Verantwortlichkeit beim Einsatz von KI ein zentraler Punkt ist.

Im Mittelpunkt stehen folgende Fragestellungen:

1. Wie lassen sich *Accountability* und *Human Agency* als Konzepte für eine KI-gestützte Schulentwicklung theoretisch integrieren?
2. Welche Verantwortlichkeiten und Handlungsoptionen ergeben sich daraus für die Akteur*innen auf den verschiedenen Ebenen des Schulsystems?
3. Wie können menschenzentrierte KI-Anwendungskontexte in der Schulentwicklung gestaltet werden, die Verantwortungsübernahme und Handlungsfähigkeit gleichermaßen fördern?

Als Ergebnis wird eine Rahmenkonzeption zu KI in der Schulentwicklung vorgestellt, die Verantwortlichkeiten, Entscheidungsspielräume und Wechselwirkungen zwischen den beteiligten Akteur*innen im Mehrebenensystem sichtbar macht. Das Modell versteht *Accountability* und *Agency* als Voraussetzungen menschenzentrierter KI-Anwendungskontexte in Schulentwicklungszusammenhängen.

Für die schulische Praxis bietet die Rahmenkonzeption eine Orientierungs- und Reflexionshilfe für Implementierungs- und Entwicklungsstrategien. Es unterstützt Schulleitungen, Bildungsadministration und weitere Akteur*innen dabei, Rollen, Zuständigkeiten und Kooperationsstrukturen im Kontext von KI explizit zu gestalten und so die Voraussetzungen für eine nachhaltige, partizipative und verantwortungsvolle KI-gestützte Schulentwicklung zu schaffen.

Ko-Design mit CoDe-Graph: eine KI-unterstütze, graphisch Sprache

Allison Kolling, Saroj Sharma, Prof. Dr. Armin Weinberger, Universität des Saarlandes

Gemeinsame Unterrichtsplanung im Sinn von „Ko-Design“ kann als eine Form der pädagogischen Mitgestaltung betrachtet werden. Der Unterrichtsplan wird nicht von einer einzelnen Person erstellt, sondern mehrere Beteiligte mit unterschiedlichen Fachkenntnissen diskutieren, verhandeln und überarbeiten den Plan gemeinsam (Bovill et. Al 2016)

Eine effektive Zusammenarbeit erfordert ein gemeinsames Verständnis davon, wie sich eine Unterrichtsstunde oder eine Unterrichtseinheit entwickeln wird, einschließlich der Abfolge der Aktivitäten, der Übergänge, der Zuständigkeiten und der Formen der Schülerunterstützung. Intuitive Werkzeuge können Ko-Design unterstützen, indem sie Unterrichtsideen sichtbar, diskutierbar und anpassbar machen (Sanders & Stappers 2008).

CoDe- Graph ist so ein Werkzeug, das die gemeinsame Planung vor Ort und aus der Distanz unterstützt (Kolling et al. 2022). Lehrkräfte und andere Beteiligte können den beabsichtigten Unterrichtsablauf mündlich oder schriftlich beschreiben, und das KI-unterstützte CoDe-Graph wandelt ihre Eingaben in eine erste visuelle Darstellung um (siehe Abbildung A1 im Anhang). Die Teilnehmer können die Darstellung auch jederzeit mit manuellen Werkzeugen erstellen oder modifizieren. Die Darstellung konzentriert sich auf den Lernweg der Lernenden durch (emergierende) Lernobjekte sowie auf Bewertung und Feedback. Darüber hinaus können die Teilnehmer Materialien direkt zur visuellen Darstellung hinzufügen oder mit ihr verknüpfen und so ein dynamisches Objekt erstellen.

Wichtig ist, dass die KI-generierte Darstellung nicht dazu gedacht ist, das fachliche Urteilsvermögen zu ersetzen oder pädagogische Entscheidungen anstelle der Lehrkräfte zu treffen. Sie dient als vorläufige Interpretation der Beschreibung der Gruppe und als Grundlage für weitere Diskussionen. Die Teilnehmer können das resultierende Diagramm prüfen, Unstimmigkeiten identifizieren, diskutieren, ob es ihre Absichten korrekt widerspiegelt, und es dann mit Hilfe von CoDe-Graph oder manuell erweitern, neu ordnen oder anpassen. Sollte KI-Feedback gewünscht sein, kann die Plattform Reflexionsfragen stellen und bewährte Designmuster sowie konkrete Verbesserungsvorschläge anbieten. Die Teilnehmer müssen entscheiden, inwieweit sie diese Vorschläge umsetzen. Der Zweck der Automatisierung besteht darin, die Teilnahme und die Externalisierung zu erleichtern, nicht darin, Lehrkräften und Beteiligten die Verantwortung für den Unterricht zu entziehen. Das resultierende Diagramm kann als Bild oder textbasierte Darstellung exportiert werden, um es zu teilen und als Referenz zu nutzen.

CoDe-Graph unterstützt somit die Zusammenarbeit auf mehreren miteinander verbundenen Ebenen. Erstens fördert es die Formulierung von Ideen, indem es mündliche oder schriftliche Beschreibungen akzeptiert. Zweitens unterstützt es die Visualisierung des Ablaufs von Anleitungen, indem es diese Beschreibungen in eine graphische Darstellung umwandelt. Drittens fördert es Verhandlungen und ein gemeinsames Verständnis, indem es den Beteiligten ein gemeinsames Objekt an die Hand gibt, das sie untersuchen und diskutieren können. Schließlich unterstützt es Iteration und professionelle Handlungsfähigkeit, indem es ermöglicht, die Darstellung manuell zu erweitern oder anzupassen.

Literatur

Bovill, C., Cook-Sather, A., Felten, P., Millard, L., & Moore-Cherry, N. (2016). *Addressing potential challenges in co-creating learning and teaching: Overcoming resistance, navigating institutional norms and ensuring inclusivity in student-staff partnerships*. Higher Education, 71(2), 195-208.

Kolling, A. L. G., Weinberger, A., & Niegemann, H. (2022). Co-designing MOOCs with CoDe-Graph. *Journal of Formative Design in Learning*. https://doi.org/10.1007/s41686-022-00064-2

Sanders, E. B.-N., & Stappers, P. J. (2008). *Co-creation and the new landscapes of design*. CoDesign: International Journal of CoCreation in Design and the Arts, 4(1), 5-18.

Einsatz eines KI-basierten Tutorensystems zur Sprachförderung in der Primarstufe: Er-kenntnisse aus der begleitenden Evaluation

Dr. Svitlana Mokhonko, Dr. Alexandra Dehmel, Institut für Bildungsanalysen Baden-Württemberg (IBBW); Ines Pieronczyk, Zentrum für Schulqualität und Lehrerbildung (ZSL)

KI-basierte Tools ermöglichen Diagnostik und individuelle Förderung von Schülerinnen und Schülern, wobei das adaptive Lernen im Fokus steht. Die Vorteile des Einsatzes solcher Tools werden häufig dokumentiert. Exemplarisch nennen lassen sich die Entlastung von Lehrkräften und die Möglichkeit, dass Schülerinnen und Schüler durch KI-basierte Tools individuelles Feedback zu ihren Lernprozessen erhalten und adaptiv gefördert werden können (Scheiter et al. 2025). Gleichzeitig werden Grenzen benannt, die insbesondere für Grundschulkinder relevant sind, wie z.B. fehlende Interaktionsprozesse, mangelnde Berücksichtigung individueller Bedürfnisse, Gefahr der Überforderung oder sinkende Motivation (Scheiter et al. 2025).

Die Lernwirksamkeit von KI-basierten Tools wird häufig untersucht und ist in zahlreichen Studien belegt (im Überblick z.B. Scheiter et al. 2025; Blanc et al. 2023). Es gibt allerdings noch wenig Erkenntnisse darüber, wie KI-basierte Tools in der schulischen Praxis und in verschiedenen Settings eingesetzt werden und welche Herausforderungen und Gelingensfaktoren dabei eine Rolle spielen.

Diesen Fragen wurde im Rahmen eines Pilotprojekts in Baden-Württemberg zum Einsatz eines Intelligenten Tutoriellen Systems (ITS) zur Förderung von Basiskompetenzen im Bereich Sprache in der Primarstufe nachgegangen. Im Schuljahr 2025/2026 wurde ein bereits etabliertes ITS zur Sprachförderung (Meister Cody Namagi) an 64 Grundschulen und Sonderpädagogischen Bildungs- und Beratungszentren (SBBZ) eingesetzt und erprobt.

Das Pilotprojekt wurde wissenschaftlich begleitet. Die begleitende Evaluation war als Längsschnittuntersuchung angelegt. Bei den Lehrkräften wurde eine Online-Befragung zu drei Messzeitpunkten durchgeführt. Es wurden verschiedene Aspekte erhoben wie z.B. die Integration des ITS in den Unterricht, Gestaltung des Unterrichts, Bewertung der Arbeit mit dem ITS, das Nutzungsverhalten sowie die wahrgenommene Lernförderlichkeit bei Schülerinnen und Schülern. Ergänzend dazu wurden zwei kurze Online-Befragungen bei Schülerinnen und Schülern der Klassenstufen drei und vier durchgeführt. Ziel war es zu erfassen, wie die Schülerinnen und Schüler die Arbeit mit dem ITS bewerten, beispielsweise hinsichtlich Häufigkeit und Einfachheit der Nutzung, sowie ihrer Motivation und des Spaßes bei der Arbeit mit dem ITS.

Im Rahmen des Vortrags werden die zentralen Erkenntnisse aus der wissenschaftlichen Begleitung des Pilotprojekts vorgestellt vor allem in Bezug auf die Fragen:

- Wie nutzen die Lehrkräfte das ITS in der Praxis?

- Wie wird der Einsatz des ITS aus Sicht der Lehrkräfte sowie der Schülerinnen und Schüler wahrgenommen und bewertet?
- Welche Gelingensbedingungen und Herausforderungen zeigen sich bei dem Einsatz des ITS?

Die im Rahmen des Vortrags präsentierten Ergebnisse liefern wichtige Erkenntnisse zum erfolgreichen Einsatz von KI-basierten Tools in der schulischen Praxis, zur Implementierung solcher Tools und daraus resultierenden Fortbildungsmaßnahmen.

Literatur

Blanc, B., Reichow, I. & Paaßen, B. (2023). Was wirkt? Eine Literaturstudie zur Wirksamkeit von Systemeigenschaften in Mathematik-Lernumgebungen. In: Röpke, R. & Schroeder, U. (Hrsg.): 21. Fachtagung Bildungstechnologien (DELFI), S. 293–294, Aachen, Deutschland.doi:10.18420/delfi2023-61.

Scheiter, K., Bauer, E., Omarchevska, Y., Schumacher, C. & Sailer M. (2025). Künstliche Intelligenz in der Schule: Eine Handreichung zum Stand in Wissenschaft und Praxis. Bonn: Herausgegeben im Rahmen des KI-Begleitprozesses im Rahmenprogramm empirische Bildungsforschung. https://www.empirische-bildungsforschung-bmbfsfj.de/img/KI_Review_20250318_Veroeffentlichung.pdf

---

Von qualitativen Evaluationsdaten zur organisationsdiagnostischen Analyse – Potenziale und Kontrollfragen generativer KI für datengestützte Schulentwicklung (Vortrag)

Dr. Holger Gerlach, Anna-Victoria Kloth, Qualitätsentwicklung Thüringer Schulen (QThüS)

Wie lassen sich umfangreiche qualitative Evaluationsdaten für datengestützte Schulentwicklung nutzen, ohne ihre Kontextgebundenheit durch Standardisierung oder automatisierte Verdichtung zu verlieren? Und welche Rolle kann generative KI dabei übernehmen?

Ausgangspunkt des Beitrags ist die Ausgangsevaluation Thüringer Schulen im Startchancen-Programm. In diesem Rahmen ist ein Bestand von inzwischen mehr als 90 leitfadengestützten Schulleitungsinterviews entstanden. In vertieften fallbezogenen und fallvergleichenden Analysen zeigen sich wiederkehrende organisationale Konstellationen, die darüber Auskunft geben, unter welchen Bedingungen schulische Entwicklungsanliegen im Organisationsalltag Haltekraft gewinnen (also in wiederholbare, verantwortete und auswertbare Praxis übersetzt werden) und unter welchen Bedingungen sie trotz hoher Aktivität, vorhandener Ressourcen oder Zustimmung fragil bleiben.

Daraus ergibt sich eine weiterführende methodische Frage: Generative KI eröffnet grundsätzlich die Möglichkeit, größere qualitative Fallbestände systematisch vergleichend zu bearbeiten. Gerade qualitative Organisationsdiagnostik stellt jedoch besondere Anforderungen an einen solchen Einsatz. Kontext und Aussagequalität müssen erhalten bleiben, unterschiedliche Befundreichweiten dürfen nicht nivelliert und Ähnlichkeiten nicht vorschnell zu fallübergreifenden Mustern verdichtet werden. Ebenso gilt es zu vermeiden, dass interpretative Setzungen durch scheinbare algorithmische Eindeutigkeit verdeckt werden.

Der Beitrag verbindet daher einen empirisch entwickelten organisationsdiagnostischen Zugang mit der Frage nach den Bedingungen einer kontrollierten KI-Unterstützung qualitativer Evaluation. Im Mittelpunkt stehen Materialbindung und Rückverfolgbarkeit, menschliche Interpretationskontrolle sowie Risiken analytischer Glättung, Überinterpretation und Drift.

An ausgewählten anonymisierten Praxisbeispielen soll gezeigt werden, welchen zusätzlichen Erkenntnisgewinn eine organisationsdiagnostische Lesart gegenüber der bloßen Erfassung von Maßnahmen und Entwicklungsaktivitäten ermöglichen kann. Abschließend wird diskutiert, wo generative KI qualitative Schulentwicklungsdaten sinnvoll erschließen kann – und an welchen Stellen gerade nicht automatisiert werden sollte.

Der Beitrag versteht KI damit nicht als automatisches Bewertungsinstrument, sondern als mögliche Unterstützung einer methodisch kontrollierten qualitativen Analyse für Evaluation und Schulentwicklung.

Plausibel ist noch nicht belastbar – qualitative Schulentwicklungsdaten mit KI nutzen und beurteilen (Workshop)

Im Mittelpunkt soll dabei die Frage stehen, welchen zusätzlichen Erkenntniswert qualitative Daten für Schulentwicklung bieten können und welche Möglichkeiten generative KI eröffnet, solche Daten systematisch auszuwerten und auch über einzelne Fälle hinaus nutzbar zu machen. An kontrastierenden KI-gestützten Auswertungen möchten wir gemeinsam mit den Teilnehmenden untersuchen, was solche Analysen tatsächlich leisten, wo zunächst plausible Ergebnisse über das Material hinausgehen und welche Formen der Kontrolle für eine belastbare Nutzung erforderlich sind. Die von Ihnen angesprochenen Perspektiven von Schulleitung und Instituten – gegebenenfalls auch von Schulaufsicht und Evaluation – lassen sich daran unmittelbar aufnehmen.

Implementierungsfragen und schulische Praxis am Beispiel der Implementation der Lernverlaufsdiagnostik quop in Baden-Württemberg (Vortrag)

Dr. Ulrike Krug, Direktorin a. D. Hessisches Kultusministerium

In meinem Vortrag möchte ich die Gelingensbedingung für die erfolgreiche Implementation von quop in Baden-Württemberg aufzeigen. Es ist eine systematische Implementation über alle Ebenen der Bildungsinstitutionen hinweg, gesteuert vom IBBW, Institut für Bildungsanalysen in BWB.

quop wird bundesweit von rund 350.000 Schülerinnen und Schülern und von 40.000 Lehrkräften genutzt, davon alleine in BWB rund 230.000 Schülerinnen und Schüler und 26.000 Lehrkräften.

Lernverlaufsdiagnostik quop – ein evidenzbasiertes Diagnosetool im Lesen und in Mathematik mit integrierter adaptiver Förderung (Workshop)

Im Gutachten der SWK „Datengestützte Entwicklung und Steuerung in Schulen und frühkindlicher Bildung“, 2026, wird quop als ein „lernprozessbegleitendes Diagnoseinstrument zur adaptiven Lernförderung“ (SWK, 2026) ausdrücklich genannt.

In meinem Workshop möchte ich die Lernverlaufsdiagnostik quop vorstellen. Mit quop wird mit zehn Messzeitpunkten über das Schuljahr hinweg im Lesen und in Mathematik eine differenzierte individuelle Diagnose plus adaptiver Förderangebote ermöglicht: für die Klassen 1 und 2 die Förderung des silbenbasierten Lesens mit **Silbenreich** (Prof. Richter/Uni Würzburg) und für die Förderung der Leseflüssigkeit in Klasse 2 bis 6 mit **FLINK** (DIPF/Frankfurt).
Mit quop wird zudem datengestützte Unterrichtsentwicklung angestoßen und bewirkt und Lehrkräfte dabei indirekt „an die Hand genommen“.