

**Evidenz.Besser.Kommunizieren.: Wie Bildungswissenschaften und Fachdidaktiken ihre
Wissenschaftskommunikation weiterentwickeln können.**

Samuel Merk¹, Sarah Bez², and Kirstin Schmidt¹

¹Pädagogische Hochschule Karlsruhe

²Universität Tübingen

3

Author Note

Samuel Merk  <http://orcid.org/0000-0003-2594-5337>

Sarah Bez  <http://orcid.org/0000-0003-0726-6170>

Kirstin Schmidt  <http://orcid.org/0000-0002-8856-7380>

Die Daten und Code dieses Artikels sind unter

<https://github.com/sammerk/StudienergebnisseBesserKommunizieren> abrufbar. Die Autor:innen haben keine Interessenskonflikte zu berichten.

Author roles were classified using the Contributor Role Taxonomy (CRediT; <https://credit.niso.org/>) as follows: *Samuel Merk*: conceptualization, data curation, formal analysis, investigation, methodology, software, supervision, validation, visualization, writing initial draft, and writing – review & editing. *Sarah Bez*: conceptualization and writing – review & editing. *Kirstin Schmidt*: conceptualization and writing – review & editing

Correspondence concerning this article should be addressed to Samuel Merk,
Pädagogische Hochschule Karlsruhe, Bismarckstraße 10, Karlsruhe 76133, Germany, Email:
merk@ph-karlsruhe.de

Zusammenfassung

Lehrkräfte treffen tagtäglich unzählige Entscheidungen. Dabei rekurrieren sie vornehmlich auf persönliche Erfahrung, Konzeptwissen oder Heuristiken. Evidenz aus Bildungswissenschaften und Fachdidaktiken wird das Potenzial zugeschrieben, diese Entscheidungsprozesse ergänzend zu informieren und zu objektivieren. Dazu ist es jedoch notwendig, dass die betroffenen Lehrkräfte diese Evidenz nicht fehlinterpretieren, was wiederum entsprechende Kompetenzen der Lehrkräfte oder besonders geschickte Wissenschaftskommunikation voraussetzt. Der vorliegende Beitrag untersucht daher die Möglichkeiten und Grenzen der Kommunikation von Effektstärken an Lehramtsstudierende am Beispiel der Berichterstattung zu PISA 2022. Im Ergebnis zeigt sich, dass Lehramtsstudierende Effektstärken sehr ungenau (Noise) ein- und im Mittel drastisch überschätzen (Practical Significance Bias). Dieser Bias konnte durch die Verwendung alternativer Visualisierungen lediglich partiell reduziert werden. Im Lichte dieser Ergebnisse wird diskutiert, inwiefern eine konstruktive Entwicklung von Wissenschaftskommunikationsformaten evidenzinformierte Entscheidungen von Lehrkräften katalysieren kann.

Schlüsselwörter: Lehrpersonenprofessionalisierung, Evidenzinformierte Praxis, Wissenschaftskommunikation, Practical Significance Bias

Evidenz.Besser.Kommunizieren.: Wie Bildungswissenschaften und Fachdidaktiken ihre Wissenschaftskommunikation weiterentwickeln können.

Die bildungswissenschaftliche Literatur zu Schul- und Unterrichtsentwicklung bedient sich einer Vielzahl theoretischer Grundlegungen ([Bohl 2020](#)) und blickt daher aus ganz verschiedenen Winkeln auf diesen Gegenstand: Neben eher systemtheoretischen Perspektiven ([K.-O. Bauer und Rolff 1978](#)) finden sich u.a. Ansätze mit Entlehnungen aus der Lehr-Lern-
([Helmke 2022](#)) und Organisationspsychologie ([Holtappels 2007](#)) oder mit dem Leitgedanken der Praxisorientierung ([Brügelmann 2018](#)). Datenbasierte Schul- und Unterrichtsentwicklung hat im deutschsprachigen Raum erst in den vergangenen zwei Dekaden Verbreitung gefunden, wenngleich deren Grundidee des empirischen Einholens von Information über den Ist-Stand schon zuvor gefordert und auch umgesetzt wurde ([Altrichter und Rolff 2006](#)). In jüngerer Zeit ist jedoch von inner- wie außerwissenschaftlichen Stakeholdern vermehrt die Forderung nach einer Entwicklung von Schule und Unterricht hörbar geworden, die ihre Entscheidungen durch Evidenz informiert ([AERO 2023; J. Bauer und Prenzel 2012; Council of the European Union 2024; Pellegrini und Vivanet 2021; Slavin 2020](#)). Da jedoch einerseits die Genese und Interpretation von Evidenz nicht zu den professionellen Kernkompetenzen von Lehrkräften gehört und andererseits Bildungswissenschaftler- und Fachdidaktiker:innen keine Expert:innen für die Gestaltung von Schule und Unterricht sind, plädiert der vorliegende Beitrag dafür, Wissenschaftskommunikation erstens als wichtige Aufgabe von Bildungswissenschaftler:innen und Fachdidaktiker:innen aufzufassen, das Gelingen von Wissenschaftskommunikation zum Gegenstand empirischer Forschung zu machen und die Entwicklung von neuen Wissenschaftskommunikationsformaten als dialogischen Prozess zwischen Bildungswissenschaften/Fachdidaktiken und Lehrkräften aufzufassen.

Daher führt der folgende theoretische Hintergrund zunächst in Konzepte und Begriffe evidenzinformierter Praxis ein, bevor er auf Wissenschaftskommunikation in Bildungswissenschaften und Fachdidaktiken eingeht, um abschließend ein empirisches Beispiel zu skizzieren.

Theoretischer Hintergrund

Evidenzinformiertes Handeln

Was kann unter »Evidenz« verstanden werden?

Etymologisch kann »Evidenz« als Substantivierung des Adjektivs »evident« gesehen werden ([Kluge 2011, S. 263](#)), welches wiederum im 18. Jahrhundert dem lateinischen »evidens« (»ersichtlich, augenscheinlich«, [Hau et al. 2012](#)) entlehnt wurde ([Stark 2017](#)). Allerdings meinen Bildungswissenschaftler:innen und Fachdidaktiker:innen gerade nicht »das Augenscheinliche« oder »das direkt Ersichtliche«, wenn sie von Evidenz sprechen - vielmehr ist in Definitionsvorschlägen von »wissenschaftlichem Wissen« ([Stark 2017](#)), von einer »Funktion« von Daten für die Bestätigung oder Widerlegung von Hypothesen und Theorien ([Bromme et al. 2014](#)) oder von »warrants for making assertions or knowledge claims« ([Shavelson und Towne 2002](#)) die Rede. In einer aktuellen Systematisierung verschiedener Verständnisse des Evidenz-Begriffs in den Bildungswissenschaften hebt Schmidt ([2024](#)) hervor, dass nur wenige Definitionen ausschließlich quantitativer Empirie die Möglichkeit zuschreiben, Evidenz zu generieren, sondern meistens auch qualitative Empirie, Theorien sowie mathematische und logische Analysen als potenziell evidenzgenerierend definiert werden. Insbesondere die Inklusion nicht-empirischer Entitäten wie Theorien oder logischer Analysen mögen auf den ersten Blick widersprüchlich wirken, da der Begriff Evidenz insbesondere im deutschsprachigen Raum teils mit Ergebnissen explanativer quantitativer Studien assoziiert scheint. Dieser

scheinbare Widerspruch wirkt jedoch weniger stark, berücksichtigt man, dass insbesondere in der Lehr-Lernforschung mit »Theorien« wohl eher sogenannte »tried-and-tested theories« ([Renkl 2022](#)) gemeint sein dürften. Diese stellen eher Rahmenmodelle oder sogenannte »interventional models« (z.B. Cognitive Theory of Multi-Media Learning) dar (ebd.). Da solche Theorien wiederum meist stark von empirischen Ergebnissen beeinflusst sind, ist es plausibel, ihnen die Funktion als »warrant« für »knowledge claims« zuzuschreiben und sie also auch als Evidenz zu bezeichnen.

Evidenzinformiert, evidenzorientiert, evidenzbasiert

Im vorigen Abschnitt wurde deutlich, dass Evidenz ein uneinheitlich gebrauchter und gleichermaßen komplex wie unscharf definierter Begriff ist. Im Lichte dessen erscheint es nur konsequent, dass auch die Begriffe evidenzbasiert, evidenzinformiert, evidenzorientiert, datenbasiert, forschungsbasiert und forschungsinformiert als Jingle Jangle eingeordnet werden können ([Kelley 1927; Thorndike 1904](#)) darstellen - dass also unterschiedliche Begriffe für das Gleiche und gleiche Begriffe für Unterschiedliches gebraucht werden. Dabei speisen sich die Differenzierungen von **evidenzbasiert** und **evidenzinformiert** sowie **evidenzorientiert** aus recht verschiedenen ontologischen, epistemologischen und wissenschaftskritischen ([Schmid und Lutz 2007](#)) Überlegungen: Mit Evidenz**basierung** wird z.B. oft »the medical model« ([Jones 2024](#)) im Sinne von Evidenz aus Kontrollgruppenexperimenten als notwendige Voraussetzung für praktische Entscheidungen assoziiert und damit sowohl epistemologische (hier Kontrollgruppenexperiment) wie wissenschaftskritische (hier notwendige Voraussetzung) Kriterien zur Abgrenzung herangezogen. Den Begriffen »**evidenzorientiert**« und »**evidenzinformiert**« wird zum einen ein weniger enger Evidenzbegriff zugeordnet

(ontologischer & epistemologischer Aspekt, [Stark 2017](#)) und zum anderen der Evidenz in der praktischen Verwendung eher eine heuristische Funktion (wissenschaftskritischer Aspekt).

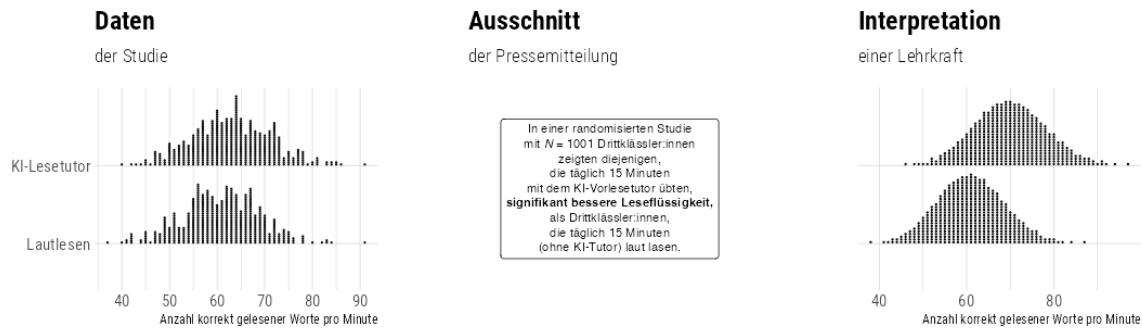
In der deutschsprachigen bildungswissenschaftlichen Diskussion sind nach Bromme et al.([2014](#)) zunächst zwei verschiedene Diskussionsstränge bzgl. evidenzinformierter Entscheidungen im Bildungskontext unterscheidbar: Ein Diskussionsstrang beschäftigt sich mit evidenzinformierten Entscheidungen in der Bildungspolitik und der andere mit evidenzinformierten Entscheidungen und Handlungen in der Bildungspraxis. In beiden Diskussionen werden der Evidenz verschiedene Funktionen zugeschrieben. Bromme et al. ([2014](#)) etwa sprechen davon, dass Evidenz über Zustände informieren, Mechanismen erklären oder Interventionen evaluieren kann. Groß Ophoff et al. ([2023](#)) wiederum unterscheiden konzeptuelle Nutzung (»evidence allows focussing attention, provides new insights, challenges beliefs or reframes thinking«, S. 2), instrumentelle Nutzung (»identify or develop concrete measures to be taken«, S. 2) und symbolische Nutzung (»justif[y] or support of existing positions or established procedures«, S. 2).

Potenzielle Wege zu einer gelingenden Wissenschaftskommunikation

Unabhängig vom Kontext und der Funktion evidenzinformierter Entscheidungen ist es plausibel anzunehmen, dass eine erfolgreiche Kommunikation im Sinne der Induktionen eines adäquaten Verständnisses von Evidenz zwischen Bildungswissenschaftler:innen/Fachdidaktiker:innen und den Akteuren im Bildungssystem eine notwendige Voraussetzung für das Gelingen evidenzinformierter Entscheidungen ist: Wird Evidenz fehlinterpretiert und erfolgt eine anschließende Entscheidung kohärent zu dieser Fehlinterpretation, wird die Wirkung dieser Entscheidung nicht die Erwünschte sein.

Abbildung 1

Daten einer (fiktiven) Studie, eine dazugehörige Pressemitteilung und die Vorstellung einer Lehrkraft von den Daten



Quelle: [Artikel-Notizbuch](#)

Liest eine Lehrkraft etwa die (fiktive) Pressemitteilung in Abbildung 1, stellt sich die Ergebnisse darauf basierend wie in Abbildung 1 rechts vor (Schmidt et al. 2023) und überzeugt anschließend ihre Schulleitung, diesen KI-Lesetutor zu beschaffen und schulweit einzusetzen, liegt höchstwahrscheinlich dysfunktionales evidenzinformiertes Handeln vor. Während die Forscher:innen mit »signifikant bessere Leseflüssigkeit« zum Ausdruck bringen, dass ihre Daten unter der Annahme eines Nulleffekts unwahrscheinlich sind (signifikanter p -Wert), interpretiert die Lehrkraft diese Formulierung als »Unterschied bedeutsamer Größe«. Folglich schlussfolgert sie, dass es Sinn macht Geld und Zeit in Anschaffung und Implementation des KI-Lesetutors zu investieren, weil sie den KI-Lesetutor für deutlich lernwirksamer hält als lautes Lesen, obwohl etwa die Implementation von Lautlesetandems lernwirksamer, kostengünstiger und weniger zeitaufwändig gewesen wäre.

Die Forschung zur Wissenschaftskommunikation hat eine Reihe solcher potenziellen Problematiken aufgezeigt: Z.B. das soeben beschriebene Verwechseln von Inferenzstatistik und

Effektstärke ([Schmidt et al. 2023](#)), das automatische Annehmen starker Effekte, wenn keine Effektstärken berichtet wurden (Practical Significance Bias, [Michal und Shah 2024](#)), Rückschaufehler ([Masnick und Zimmerman 2009](#)) oder die verzerrte Einschätzung der Belastbarkeit von Befunden (z.B. das Ergebnis einer Laborstudie mit $N = 56$ mit großem Effekt und daher hoher statistischer Power) durch irrelevante Zahlen (z.B. Stichprobengröße einer zuvor gelesenen Large-Scale-Studie, [Bohrer et al. 2025](#)).

Gleichzeitig liegt eine Reihe von Befunden vor, die implizieren, dass verbesserte Kommunikation von Evidenz an Lehrkräfte zu Zwecken evidenzinformierten Handelns vergleichsweise einfach umsetzbar ist (z.B. [Schneider et al. 2024](#)). Grundsätzlich lassen sich die bisherigen Befunde in angebotsseitige und nutzendenseitige Ansätze unterscheiden, also in Interventionen, die die Auswahl und Darstellung der Evidenz optimieren möchten und Ansätze, die bei der Scientific, Data und Statistical Literacy der Lehrkräfte ansetzen ([Brühwiler und Leutwyler 2020](#)).

Zu zweiterem gehören Programme wie »Data Teams« ([Schildkamp et al. 2018](#)), welche durch ein umfängliches Set an vordefinierten Leitlinien und Aktivitäten versucht, konkrete schulische Probleme mit Hilfe von (oft eigens dafür genierten) Daten zu lösen, wobei meist 4-6 Lehrkräfte und Schulleiter:innen mit Bildungswissenschaftler:innen und Fachdidaktiker:innen kooperieren. Hierzu gehören auch »Brokering-Ansätze« (teilweise auch als »research practice partnerships« bezeichnet), in welchen Wissenschaftler:innen und Lehrpersonen (insbesondere Schulleitungen) gemeinsam versuchen, konkrete schulischen Probleme unter Rückgriff auf wissenschaftliche Erkenntnisse zu lösen (z.B. [Sharples und Sheard 2015](#)). Auch Kurz- ([Merk et al. 2020](#)) oder längerfristig angelegte ([Karst et al. 2024](#)) Interventionen zur Anbahnung notwendiger Kompetenzen für evidenzinformiertes Handeln wie die Interpretation von grafisch

dargestellten Daten ([Friel et al. 2001](#)) oder Forschungskompetenz ([Neuenschwander 2005](#)) sowie die konkrete Unterstützung für evidenzinformiertes Handeln ([Clearing House Unterricht Academy 2025](#)), können diesem Ansatz zugerechnet werden.

Angebotsseitige Versuche die Kommunikation von Evidenz zu verbessern, stammen aus verschiedensten Disziplinen: So wird z.B. in der Psychologie untersucht ([Grice et al. 2020](#)), welche algebraisch äquivalenten Formulierungen zu standardisierten Effektstärken bei Rezeption durch Laien adäquatere Vorstellungen induzieren (siehe [Tabelle 1](#)). In der Human-Computer-Interaction-Forschung werden (teils dynamische) Visualisierungstechniken entwickelt, um Effektstärken und Inferential Uncertainty besser zu kommunizieren (z.B. [Hullman et al. 2015](#); [Zhang et al. 2023](#)). Die bildungswissenschaftliche Lehrerbildungsforschung sowie die Fachdidaktiken erproben innovative Formate für die Zielgruppe der Lehrkräfte (z.B. [Schneider et al. 2024](#)), was auch das Anliegen der vorliegenden Studie ist.

Tabelle 1

Beispiele für angebotsseitige Versuche verbesserter Kommunikation von Evidenz

	Unterschied	Zusammenhang
Standard-kommunikation	Die Leseleistung von Schülerinnen und Schülern sank in PISA 2022 um 28 Punkte und damit auf den Tiefststand.	Der sozioökonomische Status klärt 14% der Varianz der Mathematikleistung auf.
Verbesserte Kommunikation	Die Leseleistungen von Schülerinnen und Schülern in Deutschland aus PISA 2018 und aus PISA 2022 überlappen sich zu 88,9%, wobei der Mittelwert um 28 Punkte sank.	Von 100 Schülerinnen und Schülern, die einen überdurchschnittlichen sozioökonomischen Status haben, zeigen 69 eine überdurchschnittliche Mathematikleistung.

Quelle: [Artikel-Notizbuch](#)

Die vorliegende Studie

In diesem Kontext untersucht die vorliegende Studie, inwiefern verbreitete Standardgrafiken zur Kommunikation der Entwicklung der Lesekompetenz in den deutschen Kohorten des Programme of International Student Assessment (PISA) Practical Significance Bias induzieren und ob dieser mit Grafiken verringert werden kann, bei deren Gestaltung theoretische und empirische Erkenntnisse der Wissenschaftskommunikation (siehe [Kapitel 2.1.1](#)) berücksichtigt wurden.

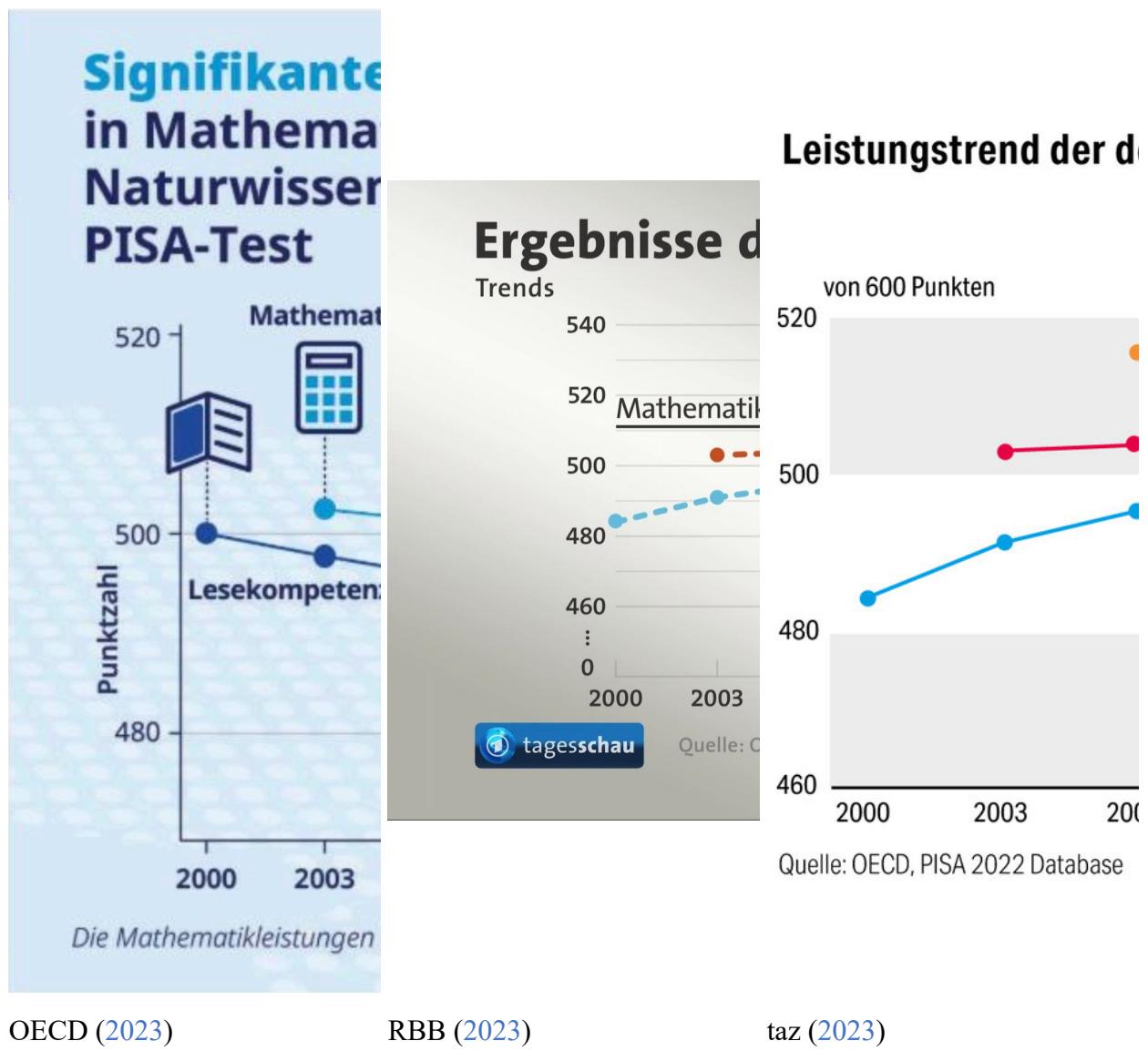
Methode

Materialien

In der wissenschaftlichen wie journalistischen Berichterstattung zu den Ergebnissen der PISA-2022-Kohorte wurden zahlreiche Darstellungsformate gewählt, insbesondere Liniendiagramme (siehe [Tabelle 2](#)), was angesichts der Anlage des PISA als Trendstudie ([Döring und Bortz 2016](#)) konsequent erscheint.

Tabelle 1

Verwendete Liniendiagramme in der Berichterstattung.

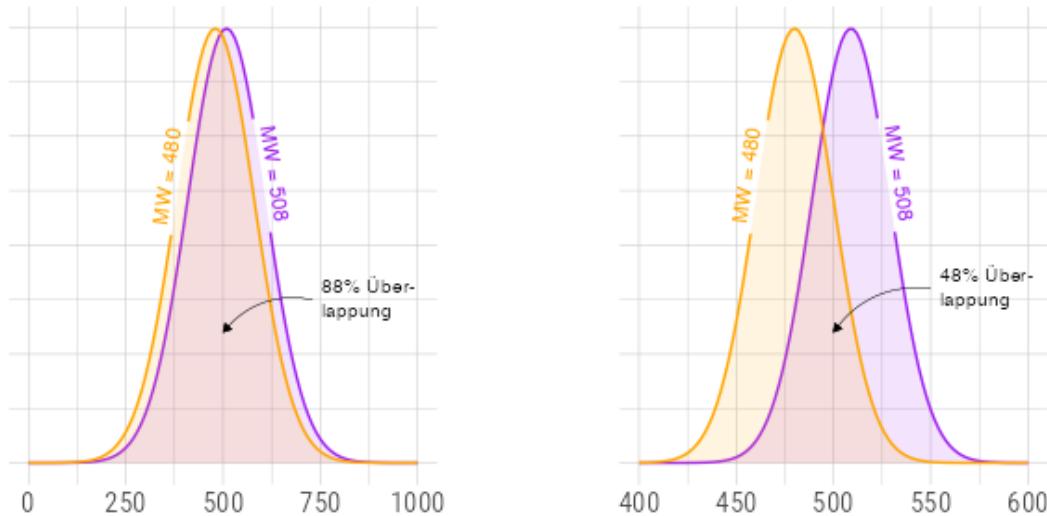


Diese Abbildungen erlauben einen effizienten Vergleich der Mittelwerte sowohl über die

Zeit als auch Variablen (hier: Fächer) hinweg. In solchen Grafiken ist jedoch die Bedeutsamkeit der Mittelwertsdifferenz nur bei bekannter Streuung interpretierbar: [Abbildung 2](#) zeigt jeweils die gleichen Mittelwerte von 508 (PISA Lesen 2015) und 480 (PISA Lesen 2022).

Abbildung 2

Illustration der Unabhängigkeit von Mittelwertsdifferenz und Größe des Effekts

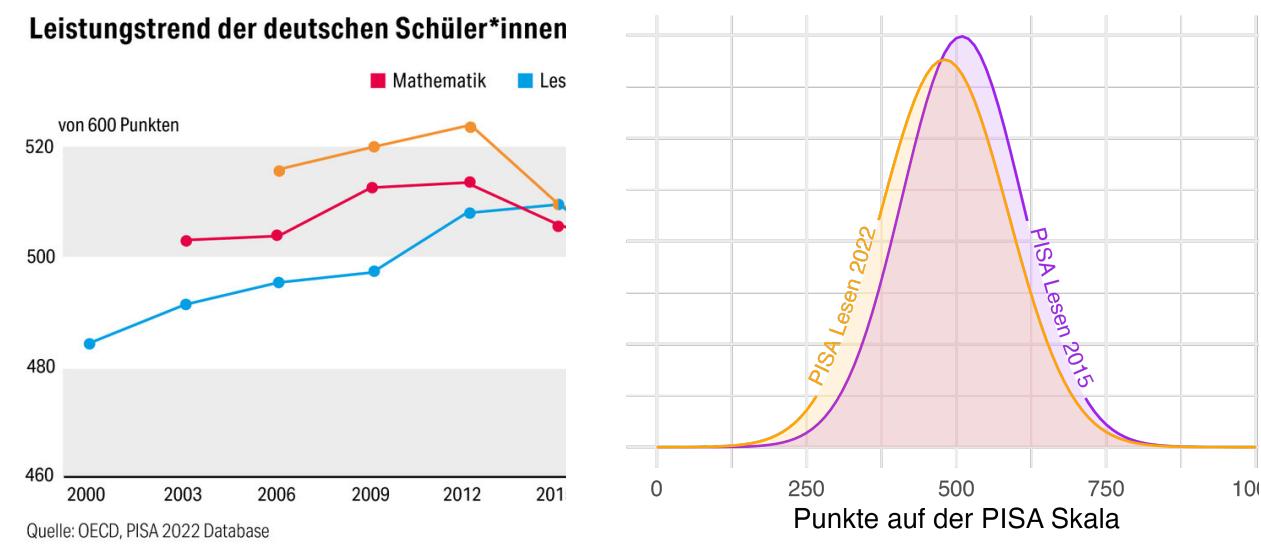


Quelle: [Artikel-Notizbuch](#)

Das Ausmaß der Bedeutsamkeit dieses (gleichen) Mittelwertsunterschiedes entsteht aber erst durch die Streuung der Daten um diesen Mittelwert herum. Weil die Variablen im rechten Teil der Abbildung weniger streuen, ist die Überlappung der beiden Gruppen geringer (48%, großer Effekt), während die große Überlappung im linken Teil (88%, kleiner Effekt) durch die große Streuung zustande kommt. Die Abbildungen in [Tabelle 2](#) sagen also nicht nur nichts über die Bedeutsamkeit der Mittelwertsunterschiede aus. Die nicht dargestellte Varianz induziert möglicherweise auch eine wahrgenommene große Bedeutsamkeit der Mittelwertsdifferenz ([Kale et al. 2021](#)).

Tabelle 2

Verwendete Stimuli



Daher wurden vorliegend neben Liniendiagrammen auch überlappende Verteilungskurven verwendet. Um diese barriereärmer zu gestalten wurde bei der Farbgebung auf hinreichenden Kontrast bei den prävalenten Sehbehinderungen geachtet (Garnier et al. 2023). Um unnötige Arbeitsgedächtnisbelastung zu vermeiden, wurde die Legende direkt in die Grafik integriert (Franconeri et al. 2021).

Design, Stichprobe und Instrument

Quelle: [Artikel-Notizbuch](#)

In einem Between-Person Design wurde $N = 195$ Studierenden in Bachelorstudiengängen des Primar- und Sekundarstufenlehramtes randomisiert eine der beiden in Tabelle 3 dargestellten Abbildungen gezeigt. Anschließend wurden sie mit folgenden Stimulus aufgefordert, die Effektstärke einzuschätzen: “*Basierend auf dieser Grafik: Wie hoch schätzen (exakte Antwort nicht möglich) Sie die Wahrscheinlichkeit ein, dass eine zufällig gezogene Schülerin oder ein*

zufällig gezogener Schüler aus dem Jahr 2022 im Lesen schlechter abschneidet als eine zufällig gezogene Schülerin oder Schüler aus dem Jahr 2015?". Beantwortet wurde diese Frage mit einem Schieberegler, dessen Enden mit "50% (beide Gruppen gleich)" und "100% (maximaler Effekt)" beschriftet waren. Diese Erfassung der wahrgenommenen Effektstärke als »Probability of Superiority« ist in der Human-Computer-Interaction-Forschung verbreitet und gilt als valide (Brooks et al. 2014; Kim et al. 2022), wenngleich die Operationalisierung als Schieberegler unklar lässt, inwiefern bei der Beantwortung tatsächlich eine Elaboration der Überlappung vorgenommen wird oder die Teilnehmenden eher intuitiv (etwa wie bei einem Likert-Item) vorgehen.

Statistische Analyse

Die abhängige Variable »Wahrgenommene Effektstärke« (operationalisiert als Probability of Superiority) ist per Design auf das geschlossene Intervall [0,5; 1] beschränkt und zeigt empirisch Bimodalität (siehe Abbildung 3). Um diesen Umständen in der inferenzstatistischen Modellierung Rechnung zu tragen, wurden bayesianische Mixture Regressionsmodelle für zwei trunkierte Normalverteilungen (Frischkorn und Popov 2023) in der probabilistischen Sprache Stan (Stan Development Team 2024) mithilfe des R-Pakets {brms} (Bürkner 2017) geschätzt.

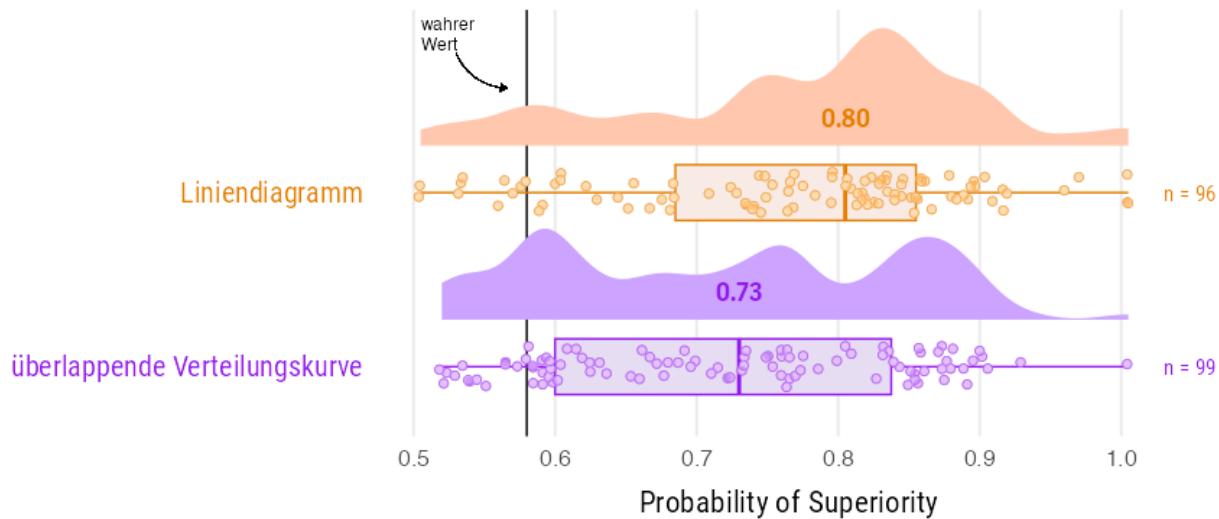
Quelle: Artikel-Notizbuch

Ergebnisse

Die Inspektion des Marcov-Chain-Monte-Carlo-Sampling-Prozesses zeigte eine zufriedenstellende Qualität bzgl. Konvergenz ($\hat{R} < 1.01$) und effektiver Sampling Size ($ESS_{Bulk} > 1000 < ESS_{Tail}$, Vehtari et al. 2021).

Abbildung 3

Geschätzte Effektstärke (Probability of Superiority) nach Stimulus. Beide Gruppen zeigen einen sehr deutlichen Practical Significance Bias (Abstand von Median und wahren Wert).



Quelle: [Artikel-Notizbuch](#)

Die Medianeinschätzung der Probability of Superiority lag in beiden Gruppen deutlich über dem wahren Wert (Liniendiagramm .80, überlappende Überlappungsgrafik .73). Dieser Unterschied in der Einschätzung entspricht einer Überlappung von 81.71% (Cliff's $d = 0.23$) oder anders ausgedrückt: Legt man 100 mal einem: einer Studierenden die Originalgrafik und einem: einer Studierenden die Überlappungsgrafik vor, schätzt 61mal die: der Studierende mit der Überlappungsgrafik den Effekt weniger verzerrt ein. Die Inferenzstatistik für diesen Unterschied ist mit einer Evidence Ratio von 14.8 klar konklusiv: Die Alternativhypothese einer kleineren Probability of Superiority für die Überlappungsgrafik ist gegeben die Daten 14,8-fach wahrscheinlicher als die Nullhypothese einer größeren Probability of Superiority.

Diskussion

Der vorliegende Beitrag zielt darauf ab, zu eruieren, inwiefern es nach dem Stand der Forschung gestaltete Wissenschaftskommunikation ermöglicht, bildungswissenschaftliche und fachdidaktische Evidenz »besser« an Lehrkräfte zu kommunizieren. Dabei wurde »besser« als »weniger biased« operationalisiert und gezeigt, dass die Wahl einer theoretisch fundierten grafischen Darstellung einen deutlich geringeren Bias induzierte als eine Standardgrafik. Allerdings war auch die Rezeption der verbesserten Darstellung immer noch erheblich verzerrt (siehe [Abbildung 3](#)).

Im Lichte dieser Ergebnisse werden im Folgenden drei Implikationen diskutiert: 1) Die Forderung, dass Lehrkräfte ihre professionelle Praxis evidenzinformiert gestalten sollen, setzt Anstrengungen in der Wissenschaftskommunikation seitens Bildungswissenschaften und Fachdidaktiken voraus. 2) Inwiefern diese Anstrengungen zielführend sind, sollte empirisch überprüft werden. 3) Erfolgreiche Wissenschaftskommunikation in den Bildungswissenschaften und Fachdidaktiken impliziert eine Passung von Angebots- und Nutzendenmerkmalen und damit einen dialogischen Prozess für die Entwicklung einer solchen Passung.

Sowohl Wissenschaftstheorie (z.B. [Mitchell und Jolley 2010](#)) als auch bildungswissenschaftliche Literatur (z.B. [Bohl et al. 2015; Dewe et al. 1992](#)) haben die Möglichkeiten und Limitationen der Abgrenzbarkeit von »Wissenschaft« und »Nicht-Wissenschaft« (bzw. in den Bildungswissenschaften von »Theorie« und »Praxis«) diskutiert und heben u.a. hervor, dass Entitäten und Aussagen in ihrer Bedeutung primär an den Herkunftskontext (z.B. »Wissenschaft« oder »Praxis«) gebunden sind. Also ist auch z.B. die »Evidenz« einer explanativen bildungswissenschaftlichen Studie per se zunächst bildungswissenschaftlich und muss für eine evidenzinformierte Handlung in der Praxis

reinterpretiert werden (z.B. [Groß Ophoff et al. 2023](#)). Damit liegt es auf der Hand, dass sich Bildungswissenschaftler:innen und Fachdidaktiker:innen fragen sollten, welche »wissenschaftlichen/theoretischen« Entitäten (z.B. Effektstärken oder inferenzstatistischen Maße) und Aussagen (z.B. kausale Effekte) sie wie in die Kommunikation ihrer Ergebnisse gegenüber der Praxis aufnehmen.

Dass diese Forderung selbst rein innerwissenschaftlich betrachtet nicht trivial ist, zeigt z.B. die Tatsache, dass Guidelines von Fachgesellschaften wie z.B. der American Psychological Association ([2019](#)) die Verwendung von Effektstärken fordern, diese aber in Pressemitteilungen (etwa der American Educational Research Association) und selbst in Fachzeitschriften selten sind, obwohl die Fachgesellschaften bzw. Zeitschriften diese in ihren Autor:innenrichtlinien verbindlich fordern ([McMillan und Jennifer 2011](#)). Für eine Wissenschaftskommunikation, die sich an die Praxis richtet scheint es also plausibel, zu schlussfolgern, dass es unter Forschenden noch nicht verbreitet scheint, sich literaturbasiert darüber Gedanken zu machen, inwiefern die eigene Wissenschaftskommunikation z.B. für Lehrkräfte günstig rezipierbar ist.

Doch selbst ein Bewusstsein für die Fallstricke der Kommunikation wissenschaftlicher Ergebnisse schützt nicht zwangsläufig vor der Induktion von Fehlvorstellungen: So fanden Schneider et al. ([2024](#)) etwa, dass selbst eine als leicht verständlich geltende Effektstärke für Mittelwertsvergleiche (wie etwa Cohen's U_3) bei einem erheblichen Anteil ($\geq 29\%$) der Rezipient:innen zu Fehlvorstellungen führte. Die erste Implikation, scheint also nicht hinreichend für eine gelingende Kommunikation von Evidenz an Lehrkräfte. Dies führt zur zweiten Implikation: Forschende sollten nicht nur den Stand der Forschung bei der Kommunikation von Evidenz berücksichtigen, sondern auch in intern und extern validen Studien untersuchen, inwieweit diese Berücksichtigung erfolgreich war. Denn statistische Informationen

werden nicht nur von unterschiedlichen Berufsgruppen (McDowell und Jacobs 2017), sondern auch in unterschiedlichen geografischen Regionen differentiell interpretiert (Gigerenzer et al. 2005). Inwiefern sich also generische Determinanten erfolgreicher Wissenschaftskommunikation auf die Kommunikation von Evidenz an Lehrkräfte etwa in einem bestimmten Teil eines Bildungssystems übertragen lassen, scheint nur schwer a priori bestimmbar.

Was aber, wenn Forschende ihre Wissenschaftskommunikation literaturbasiert verbessern, aber in empirischen Experimenten sehen, dass sie dennoch deutlich verzerrt, verrauscht oder konzeptuell falsch rezipiert wird? Der vorliegende Beitrag macht als dritte Implikation den Vorschlag, den Kommunikation von Evidenz an Lehrkräfte dialogisch weiterzuentwickeln und zu berücksichtigen, dass bei der Rezeption von vermutlich einer komplexen Interaktion von Angebots- und Nutzendenmerkmalen (Brühwiler und Leutwyler 2020) sowie Bottom-Up- bzw. Top-Down-Prozessen (Schmidt 2024) vorliegt: Man stelle sich eine Lehrkraft vor, die auf der Suche nach einer Entscheidungsgrundlage für oder gegen eine unterrichtsgestalterische Maßnahme A auf der Seite eines Clearing Houses landet. Dort liest sie, dass über viele Studien gemittelt Maßnahme A dazu geführt hat, dass 63% der Schülerinnen und Schüler bessere Leistungen zeigen als der Mittelwert der Schülerinnen und Schüler mit Maßnahme B. Dann können daraus manche Lehrkräfte möglicherweise anhand ihres Vorwissens unmittelbar eine korrekte/konsistente Vorstellung der Effektstärke dieses Unterschieds *schlussfolgern* (z.B. zwei Normalverteilungen mit 87% Überlappung). Hier läge ein Top-Down-Prozess vor, da die Merkmale der Kommunikation mit den im Langzeitgedächtnis der Rezipient:in vorhandenen Dispositionen wie Graph, Data oder Statistical Literacy dazu führen, dass in einem Schlussfolgerungsprozess ein korrektes mentales Modell erstellt wird. Umgekehrt kann es passieren, dass eine Lehrkraft auf diese Formulierung stößt und eben kein auf Wissen

basierendes mentales Modell abrufen kann - aber sich Stück für Stück mithilfe der gegebenen Informationen ein konsistentes mentales Modell erarbeitet. Dabei *lernt* sie, d.h. erwirbt Graph, Data, oder Statistical Literacy, was einem Bottom-Up Prozess entspricht. Da Lehrkräfte über sehr unterschiedliche Dispositionen zu Top-Down-Prozessen verfügen, aber auch Bottom-Up-Prozesse sehr individuell verlaufen dürften, liegt die dritte Implikation nahe: Die Kommunikation von Evidenz an Lehrkräfte sollte als dialogischer und differenzieller Prozess aufgefasst werden. Demnach würden zum einen Bildungswissenschaftler:innen und Fachdidaktiker:innen Kenntnis über Top-Down- und Bottom-Up-Prozesse ihrer Rezipienten:innen erwerben und deren Ausprägung und Entwicklung z.B. anhand von Think-Aloud-Studien wie z.B. Bez et al. (2021) beobachten und daraufhin ihre Angebote entsprechend differenzieren und anpassen. Zum anderen könnten Lehrkräfte in die Entwicklung von Kommunikationsprodukten anhand konstruktiver Verfahren eingebunden werden, in der Hoffnung, dass eine solche Kooperation von Akteuren aus den Systemen »Wissenschaft/Theorie« und »Nicht-Wissenschaft/Praxis« dazu führt, dass innerhalb dieser Systeme Ausdrucksweisen verfügbar werden, die zu verlustfreieren und damit erfolgreicheren Kommunikationsprozessen führen können (Leitz et al. 2024).

Literatur

- AERO. (2023). *Evidence-Based Teaching Practices*. Australian Education Research Organisation. <https://www.education.gov.au/quality-initial-teacher-education-review/resources/aero-evidence-based-teaching-practices>
- Altrichter, H., & Rolff, H.-G. (2006). Datenbasierte Schulentwicklung. Editorial. *Journal für Schulentwicklung*, 10(4), 4–6.

- Association, A. P. (2019). *Publication Manual of the American Psychological Association* (7. Aufl.). American Psychological Association.
- Bauer, J., & Prenzel, M. (2012). European teacher training reforms. *Science*, 336(6089), 1642–1643. <https://doi.org/10.1126/science.1218387>
- Bauer, K.-O., & Rolff, H.-G. (1978). Vorarbeiten zu einer Theorie der Schulentwicklung. In K.-O. Bauer & H.-G. Rolff (Hrsg.), (S. 219–263). Weinheim und Basel: Beltz.
- Bez, S., Poindl, S., Bohl, T., & Merk, S. (2021). Wie werden Rückmeldungen von Vergleichsarbeiten rezipiert? *Zeitschrift für Pädagogik*, 67(4), 551–572. <https://doi.org/10.3262/ZP2104551>
- Bohl, T. (2020). Theorien der Schulentwicklung. In M. Harant, P. Thomas, & U. Küchler (Hrsg.), (S. 97–109). Tübingen: Tübingen University Press. <https://doi.org/10.15496/publikation-45627>
- Bohl, T., Wacker, A., & Harant, M. (2015). *Schulpädagogik und Schultheorie* (1. Aufl.). Stuttgart: UTB GmbH. <https://doi.org/10.36198/9783838541808>
- Bohrer, K., Schmidt, K., & Merk, S. (2025). Zwei Studien, ein Ergebnis: Lehramtsstudierende unterliegen im Umgang mit Evidenz dem Ankereffekt. *Zeitschrift für Erziehungswissenschaft*.
- Bromme, R., Prenzel, M., & Jäger, M. (2014). Empirische Bildungsforschung und evidenzbasierte Bildungspolitik. *Zeitschrift für Erziehungswissenschaft*, 17(4), 3–54. <https://doi.org/10.1007/s11618-014-0514-5>
- Brooks, M. E., Dalal, D. K., & Nolan, K. P. (2014). Are common language effect sizes easier to understand than traditional effect sizes? *Journal of Applied Psychology*, 99(2), 332–340. <https://doi.org/10.1037/a0034745>

Brügelmann, H. (2018). Unterrichts- und Schulentwicklung in Communities of Practice. In H. Barz (Hrsg.), (S. 479–484). Wiesbaden: Springer Fachmedien.

https://doi.org/10.1007/978-3-658-07491-3_44

Brühwiler, C., & Leutwyler, B. (2020). Praxisrelevanz von Forschung als gemeinsame Aufgabe von Wissenschaft und Praxis: Entwurf eines Angebots-Nutzungs-Modells. *BzL - Beiträge zur Lehrerinnen- und Lehrerbildung*, 38(1), 21–36.

<https://doi.org/10.36950/bzl.38.2020.9309>

Bürkner, P.-C. (2017). brms: An R Package for Bayesian Multilevel Models Using Stan. *Journal of Statistical Software*, 80(1). <https://doi.org/10.18637/jss.v080.i01>

Clearing House Unterricht Academy. (2025). Clearing House Unterricht Academy.

<https://clearinghouse-academy.de/>. Zugegriffen: 23. Januar 2025

Council of the European Union. (2024). Council conclusions on promoting evidence-informed policy and practice in education and training to achieve the European Education Area.

https://eur-lex.europa.eu/legal-content/EN/TXT/PDF/?uri=OJ:C_202403642

Dewe, B., Ferchhoff, W., & Radtke, F.-O. (1992). Das „Professionswissen“ von Pädagogen. In B. Dewe, W. Ferchhoff, & F. Olaf-Radtke (Hrsg.), (S. 70–91). Wiesbaden: VS Verlag für Sozialwissenschaften. https://doi.org/10.1007/978-3-663-09988-8_5

Döring, N., & Bortz, J. (2016). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften* (5. Aufl.). Berlin, Heidelberg: Springer.

<http://dx.doi.org/10.1007/978-3-642-41089-5>

Franconeri, S. L., Padilla, L. M., Shah, P., Zacks, J. M., & Hullman, J. (2021). The science of visual data communication: What works. *Psychological Science in the Public Interest*, 22(3), 110–161. <https://doi.org/10.1177/15291006211051956>

- Friel, S. N., Curcio, F. R., & Bright, G. W. (2001). Making sense of graphs: Critical factors influencing comprehension and instructional implications. *Journal for Research in Mathematics Education*, 32(2), 124. <https://doi.org/10.2307/749671>
- Frischkorn, G. T., & Popov, V. (2023). A Tutorial for Estimating Bayesian Hierarchical Mixture Models for Visual Working Memory Tasks: Introducing the Bayesian Measurement Modeling (bmm) Package for R. <https://doi.org/10.31234/osf.io/umt57>
- Garnier, S., Ross, N., BoB Rudis, Filipovic-Pierucci, A., Galili, T., Timelyportfolio, et al. (2023). *sjmgarnier/viridis: CRAN release v0.6.3*. Zenodo. <https://doi.org/10.5281/ZENODO.4679423>
- Gigerenzer, G., Hertwig, R., Van Den Broek, E., Fasolo, B., & Katsikopoulos, K. V. (2005). “A 30. *Risk Analysis*, 25(3), 623–629. <https://doi.org/10.1111/j.1539-6924.2005.00608.x>
- Grice, J. W., Medellin, E., Jones, I., Horvath, S., McDaniel, H., O’lansen, C., & Baker, M. (2020). Persons as Effect Sizes. *Advances in Methods and Practices in Psychological Science*, 3(4), 443–455. <https://doi.org/10.1177/2515245920922982>
- Groß Ophoff, J., Brown, C., & Helm, C. (2023). Do pupils at research-informed schools actually perform better? Findings from a study at English schools. *Frontiers in Education*, 7, Artikel 1011241. <https://doi.org/10.3389/feduc.2022.1011241>
- Hau, R., Martini, U., & Dralle, A. (2012). *PONS Wörterbuch für Schule und Studium Latein-Deutsch*. PONS.
- Helmke, A. (2022). *Unterrichtsqualität und Professionalisierung: Diagnostik von Lehr-Lern-Prozessen und evidenzbasierte Unterrichtsentwicklung*. Hannover: Klett Kallmeyer.

- Holtappels, H. G. (2007). Schulentwicklungsprozesse und Change Management. Innovationstheoretische Reflexionen und Forschungsbefunde über Steuergruppen. In N. Berkemeyer (Hrsg.), (S. 11–39). Weinheim u.a.: Juventa.
- Hullman, J., Resnick, P., & Adar, E. (2015). Hypothetical outcome plots outperform error bars and violin plots for inferences about reliability of variable ordering. *PLOS ONE*, 10(11), Artikel e0142444. <https://doi.org/10.1371/journal.pone.0142444>
- Jones, A. (2024). Rethinking Evidence-Based Practice in Education: A Critical Literature Review of the ‘What Works’ Approach. *International Journal of Educational Researchers*, 15(2), 37–51. <https://doi.org/10.29329/ijer.2024.1041.3>
- Kale, A., Kay, M., & Hullman, J. (2021). Visual Reasoning Strategies for Effect Size Judgments and Decisions. *IEEE Transactions on Visualization and Computer Graphics*, 27(2), 272–282. <https://doi.org/10.1109/TVCG.2020.3030335>
- Karst, K., Yendell, O., Marx, A., Lettau, W.-D., & Hawlitschek, P. (2024). Die Etablierung von Evidenzteams in SchuMaS - Eine Strategie zur systematischen Nutzung von Daten für die Schul- und Unterrichtsentwicklung. In K. Maaz & A. Marx (Hrsg.), (S. 225–240). Münster: Waxmann.
- Kelley, T. L. (1927). *Interpretation of educational measurements*. World Book Company.
- Kim, Y.-S., Hofman, J. M., & Goldstein, D. G. (2022). CHI ’22: CHI Conference on Human Factors in Computing Systems. In (S. 1–14). New Orleans LA USA: ACM. <https://doi.org/10.1145/3491102.3502053>
- Kluge, F. (2011). *Etymologisches Wörterbuch der deutschen Sprache* (25. Aufl.). Berlin: De Gruyter.

- Leitz, A., Kleen, H., Hartmann, U., & Kunter, M. (2024). Tagung der Gesellschaft für Empirische Bildungsforschung. In. Potsdam.
- Masnick, A. M., & Zimmerman, C. (2009). Evaluating scientific research in the context of prior belief: Hindsight bias or confirmation bias? *Journal of Psychology of Science and Technology*, 2(1), 29–36. <https://doi.org/10.1891/1939-7054.2.1.29>
- McDowell, M., & Jacobs, P. (2017). Meta-analysis of the effect of natural frequencies on Bayesian reasoning. *Psychological Bulletin*, 143(12), 1273–1312.
<https://doi.org/10.1037/bul0000126>
- McMillan, J. H., & Jennifer, F. (2011). Reporting and Discussing Effect Size: Still the Road Less Traveled? *Practical Assessment, Research, and Evaluation*, 16(1).
<https://doi.org/10.7275/B6PZ-WS55>
- Merk, S., Poindl, S., Wurster, S., & Bohl, T. (2020). Fostering Aspects of Pre-Service Teachers' Data Literacy: Results of a Randomized Controlled Trial. *Teaching and Teacher Education*, 91, 103043. <https://doi.org/10.1016/j.tate.2020.103043>
- Michal, A. L., & Shah, P. (2024). A Practical Significance Bias in Laypeople's Evaluation of Scientific Findings. *Psychological Science*, 09567976241231506.
<https://doi.org/10.1177/09567976241231506>
- Mitchell, M. L., & Jolley, J. M. (2010). *Research design explained* (7. Aufl.). Belmont: Wadsworth.
- Neuenschwander, M. P. (2005). Forschungskompetenzen in der Lehrerinnen- und Lehrerbildung erweitern: Ein Weiterbildungskonzept. *BzL - Beiträge zur Lehrerinnen- und Lehrerbildung*, 23(2), 270–280. <https://doi.org/10.36950/bzl.23.2.2005.10132>

OECD (Hrsg.). (2023). *PISA 2022 Ergebnisse (Band I): Lernstände und Bildungsgerechtigkeit*. Bielefeld: wbv Media. <https://doi.org/10.3278/6004956w>

Pellegrini, M., & Vivanet, G. (2021). Evidence-Based Policies in Education: Initiatives and Challenges in Europe. *ECNU Review of Education*, 4(1), 25–45.

<https://doi.org/10.1177/2096531120924670>

RBB, M. K. (2023). Deutsche Schülerinnen und Schüler schneiden bei neuer PISA-Studie so schlecht ab wie nie zuvor. <https://www.tagesschau.de/multimedia/video/video-1280422.html>

Renkl, A. (2022). Meta-analyses as a privileged information source for informing teachers' practice? A plea for theories as primus inter pares. *Zeitschrift für Pädagogische Psychologie*, 36(4), 217–231. <https://doi.org/10.1024/1010-0652/a000345>

Schildkamp, K., Handelzalts, A., Poortman, C. L., Leusink, H., Meerdink, M., Smit, M., et al. (2018). The data team procedure: a systematic approach to school improvement. In K. Schildkamp, A. Handelzalts, C. L. Poortman, H. Leusink, M. Meerdink, M. Smit, et al. (Hrsg.), Springer International Publishing. https://doi.org/10.1007/978-3-319-58853-7_9

Schmid, S., & Lutz, A. (2007). Epistemologische Überzeugungen als Kohärente Laientheorien. *Zeitschrift für pädagogische Psychologie*, 21(1), 29–40. <https://doi.org/10.1024/1010-0652.21.1.29>

Schmidt, K. (2024). *Teachers' Engagement With Educational Science How to Communicate Findings From Educational Science in a User-Friendly Way to Teachers* (phdthesis). Karlsruhe.

Schmidt, K., Edelsbrunner, P. A., Rosman, T., Cramer, C., & Merk, S. (2023). When perceived informativity is not enough. How teachers perceive and interpret statistical results of

educational research. *Teaching and Teacher Education*, 130, Artikel 104134.

<https://doi.org/10.1016/j.tate.2023.104134>

Schneider, J., Schmidt, K., Bohrer, K., & Merk, S. (2024). Communicating Effect Sizes to Teachers. *Zeitschrift für Psychologie*. <https://econtent.hogrefe.com/doi/10.1027/2151-2604/a000573>

Sharples, J., & Sheard, M. (2015). Developing an Evidence-Informed Support Service for Schools – Reflections on a UK Model. *Evidence & Policy*, 11(4), 577–587.

<https://doi.org/10.1332/174426415X14222958889404>

Shavelson, R. J., & Towne, L. (2002). *Scientific Research in Education*. Washington: National Academies Press.

Slavin, R. E. (2020). How evidence-based reform will transform research and practice in education. *Educational Psychologist*, 55(1), 21–31.

<https://doi.org/10.1080/00461520.2019.1611432>

Stan Development Team. (2024). *Stan Modeling Language Users Guide and Reference Manual*.
<https://mc-stan.org>

Stark, R. (2017). Probleme evidenzbasierter bzw. -orientierter pädagogischer Praxis. *Zeitschrift für Pädagogische Psychologie*, 31(2), 99–110. <https://doi.org/10.1024/1010-0652/a000201>

taz.de. (2023). Pisa-Schock für deutsche Schüler:innen: Im freien Fall | taz.de. <https://taz.de/Pisa-Schock-fuer-deutsche-Schuelerinnen/!5974146/>

Thorndike, E. L. (1904). *Theory of mental and social measurements*. The Science Press.
<https://doi.org/10.1037/13283-000>

- Vehtari, A., Gelman, A., Simpson, D., Carpenter, B., & Bürkner, P.-C. (2021). Rank-
Normalization, Folding, and Localization: An Improved R⁺ for Assessing Convergence of
MCMC (with Discussion). *Bayesian Analysis*, 16(2). <https://doi.org/10.1214/20-BA1221>
- Zhang, S., Heck, P. R., Meyer, M. N., Chabris, C. F., Goldstein, D. G., & Hofman, J. M. (2023).
An illusion of predictability in scientific results: Even experts confuse inferential
uncertainty and outcome variability. *Proceedings of the National Academy of Sciences*,
120(33), Artikel e2302491120. <https://doi.org/10.1073/pnas.2302491120>