

# Evidenzbasierter Unterricht in der Durchschnittsfall?

## *Typisierung von Lehrkräften nach diagnostischen Einstellungen und selbstberichteten Unterrichtsveränderungen*

Andrea Kallinger-Aufner<sup>1</sup>, Iris Arzberger<sup>2</sup>, Gundula Wagner<sup>1</sup>

---

DOI: <https://doi.org/10.53349/resource.2024.i3.a1288>

### **Zusammenfassung**

Evidenzbasierter Unterricht kennzeichnet sich dadurch, dass datenbasierte Rückmeldungen über Schüler\*innen in weitere Überlegungen des Unterrichts einfließen sollten. Im Rahmen dieser Studie wurden N = 242 Lehrkräfte sowohl zu ihren Einstellungen zu standardisierten Diagnoseverfahren als auch zu ihren selbstberichteten Unterrichtsveränderungen befragt und der Frage nachgegangen, welche Typen von Lehrkräften sich in Zusammenhang eines evidenzbasierten Unterrichts beschreiben lassen. Mittels Two-Step-Clusteranalyse konnten drei prototypische Profile von Lehrkräften hinsichtlich der Selbsteinschätzung diagnostischer Fähigkeiten sowie wahrgenommener Unterrichtsveränderung dargestellt werden, von denen sich zwei Profile um den Mittelwert gruppieren, während sich das dritte Profil durch deutliche Abweichungen kennzeichnet. Praktische Implikationen für Fortbildungen werden abgeleitet.

**Stichwörter:** Evidenzbasierung, datenbasierte Rückmeldungen, Kompetenzorientierung

---

## **1 Einleitung**

Seit dem Schuljahr 2021/22 erfolgt die Erhebung der Bildungsstandards auf der 3. Schulstufe jährlich im Rahmen der individuellen Kompetenzmessung PLUS (iKM<sup>PLUS</sup>), seit dem Schuljahr 2022/23 gilt dies auch für die 7. Schulstufe. Die Fortbildungen für Lehrkräfte wurden in

---

<sup>1</sup> Pädagogische Hochschule Wien, Grenzackerstraße 18, 1100 Wien.

E-Mail: [andrea.kallinger@phwien.ac.at](mailto:andrea.kallinger@phwien.ac.at)

<sup>2</sup> [www.zukunftsraeume.at](http://www.zukunftsraeume.at)

Zusammenarbeit des BMBWF mit dem IQS entwickelt, angeboten werden die Fortbildungen von entsprechend dafür geschulten Multiplikator\*innen u.a. an der PH Wien.

Mit der Einführung standardisierter Rückmeldeverfahren in Verbindung stand die Hoffnung, dass die Lehrkräfte diese zur Verbesserung ihres Unterrichts nutzen würden (Altricher & Maag Merki, 2016). Leider entspricht die tatsächliche Nutzung vielfach nicht den Erwartungen (Nachtigall & Hellrung, 2013; Westphal et al., 2018). Zudem nennen Lehrkräfte, werden sie nach dem Nutzen standardisierter Rückmeldeverfahren befragt, vorwiegend Nachteile und lassen eine gewünschte professionelle Einstellung vermissen (Müller & Wagner, 2024). Müller und Wagner fordern daher, die Professionalisierung im Bereich der Einstellungen und Werthaltungen von Lehrkräften voranzutreiben. Ziel der vorliegenden Studie ist nun, mehr über verschiedene Kategorien von Einstellungen zu evidenzbasiertem Unterricht zu erfahren, um in den Fortbildungen zur iKM<sup>PLUS</sup> gezielter auf den jeweiligen Typus an Lehrkräften eingehen zu können.

## 2 Evidenzbasierter Unterricht

### 2.1 Datenbasierte Rückmeldungen

Als Reaktion auf die enttäuschenden Ergebnisse von Pisa 2000 wurden in allen deutschsprachigen Ländern grundlegende Reformen im Schulsystem initiiert. In Österreich wurde mit der Einführung der Bildungsstandards 2008/09 eine Veränderung des Unterrichts von der Input- zur Outputorientierung mit dem Fokus auf die Kompetenzorientierung gestartet (Klieme et al.; 2003; Altrichter & Gamsjäger, 2019). Diese neue Sichtweise auf Lehren und Lernen sollte das Bewusstsein der Lehrkräfte für die Bedeutung eines kompetenzorientierten Unterrichts schärfen und sie bei der nachhaltigen Entwicklungsbegleitung unterstützen (Köller, 2010).

Zur Einschätzung von Schüler\*innen-Kompetenzen wurden wissenschaftliche Tools wie die Bildungsstandardüberprüfung, IKM und zuletzt die iKM<sup>PLUS</sup> entwickelt, die Lehrkräften zusätzliche Daten zur Reflexion ihres eigenen Unterrichts und der Kompetenzen ihrer Schüler\*innen bereitstellen sollen. Zum Umgang mit Ergebnissen wurden Zyklenmodelle (z.B. Helmke & Hosenfeld, 2005; Schildkamp & Ehren, 2013; Wiesner et al., 2015) entwickelt, die die Prozesse zu datenbasierter Unterrichtsentwicklung wissenschaftlich beschreiben.

In der wissenschaftlichen Auseinandersetzung werden sowohl konkret aufbereitete, belastbare Zahlen (Bewyl, 2017) als auch subjektive Wahrnehmungen (Mandinach, 2012) als „Daten“ bezeichnet. Beide Arten von Daten bedürfen der Analyse, Interpretation und Kontextualisierung durch die Lehrperson (Specht, 2002), um etwaige Handlungsfelder in der Praxis zu definieren. Gerade Daten in Form von Zahlen können den Eindruck erwecken, dass die Antworten bereits in den Zahlen enthalten sind (Klopsch & Silwka, 2020).

Eine datenbasierte Diagnose im Unterricht bezieht sich auf die Verwendung von verschiedenen Daten und Informationen (z.B. Leistungstests, Beobachtungen im Unterricht,

Schüler\*innenarbeiten und Feedback von Schüler\*innen und Eltern), um den Lernfortschritt der Schüler\*innen zu analysieren und geeignete Maßnahmen zur Verbesserung ihres Lernens zu ergreifen. Die Effektivität der Nutzung von datenbasierten Rückmeldungen als Feedback hängt davon ab, ob die Personen bereit und in der Lage sind, das Feedback zu lesen, zu verstehen und entsprechende Maßnahmen daraus abzuleiten. Dies gilt nicht nur für einzelne Rückmeldungen, sondern auch für die Zusammenführung verschiedener Datenquellen, die sich teilweise überschneiden oder widersprüchliche Informationen liefern können (Schulte et al., 2016).

Datenbasierte Entscheidungen werden als wichtige Kompetenz professionellen Handelns betrachtet, die dazu beitragen können, blinde Flecken aufzudecken und zu objektivieren (Honig & Coburn, 2008; Schildkamp & Lai, 2013) und damit Lehrkräften gezielte Hinweise zur Förderung von Schüler\*innen sowie zur Schul- und Unterrichtsentwicklung liefern können. Schulen und Lehrkräfte stehen vor der Herausforderung, die internen und externen Daten effektiv für ihren Unterrichtsalltag zu nutzen. Hierfür benötigen sie Strategien zur Datenerhebung, Kompetenzen für die Dateninterpretation und eine offene Haltung, um ihr berufliches Handeln kontinuierlich zu verbessern. Oft fehlt es jedoch an Unterstützung und Ressourcen, um Daten zu erheben, auszuwerten und für die Unterrichtsentwicklung zu nutzen. Zudem mangelt es oft an Wissen zur Interpretation statistischer Daten, wodurch die Potenziale der Daten ungenutzt bleiben (Krieg, 2020).

## 2.2 Unterrichtsveränderungen

Ausgangspunkt für die Fokussierung der Entwicklungsimpulse und -bemühungen hin zur Outputorientierung ist die Annahme eines kausalen Zusammenhangs zwischen kompetenzorientiertem Unterricht und der Verbesserung der Performanz von Lernenden (Asbrand, 2020).

Kompetenzorientierung als didaktisches Prinzip wird dabei charakterisiert durch die Merkmale guten Unterrichts, den Bildungsstandards als Grundlage für die Unterrichtsplanung und -gestaltung und kontinuierlichen, kompetenzorientierten Rückmeldungen (Harms et al., 2016). Damit eng verbunden ist ein besonderer Fokus auf die gezielte Diagnostik und die differenzierte Förderung individueller Lernprozesse (Beer & Benischek, 2011; Harms et al., 2016). Hier können datenbasierte Rückmeldungen wie die iKM<sup>PLUS</sup> ansetzen. Mit der verstärkten Nutzung von internen und externen Daten als Kern evidenzbasierten Unterrichts ist die Hoffnung verbunden, individuelles Handeln aufgrund von Erfahrungswissen durch eine Objektivierung zu ergänzen (Altrichter et al., 2016; Demski, 2017). Dabei wird vorausgesetzt, dass sich Lehrende selbst als Lernende verstehen und an der kontinuierlichen Weiterentwicklung und Steigerung der Wirksamkeit ihres Unterrichts interessiert sind (Altrichter et al., 2016; Helmke & Helmke, 2015).

Für die gezielte Auseinandersetzung mit Daten und der damit intendierten Steigerung der Adaptivität von Unterricht brauchen Lehrkräfte die Fähigkeit, diagnostische Expertise mit

didaktischem Wissen zum Kompetenzaufbau zusammenzuführen. Neben dieser Rekontextualisierungsleistung und ausreichend Zeit wird evidenzbasierter Unterricht durch positive Einstellungen in Bezug auf heterogene Lernvoraussetzungen und die Nutzung von Diagnoseinstrumenten begünstigt (Demski, 2017; Helmke, 2022; Westphal et al., 2018).

Der spürbare Mehrwert für die eigene Praxis, dass die Rückmeldungen nicht als Kontrolle empfunden werden und keine Einschränkung der professionellen Freiheit damit verbunden ist, sind wesentliche Aspekte für die Motivation, sich mit Daten auseinanderzusetzen und Unterrichtsveränderungen zu initiieren (Altrichter, 2016; Demski, 2017). Gelingt es nicht, diese Fähigkeiten und motivationalen Voraussetzungen ausreichend zu entwickeln, kann es zur Ausbildung der „Typologie der Verweigerer“ (Demski, 2017, S. 365) kommen.

Es zeigt sich, dass insbesondere die subjektiven Theorien und Überzeugungen äußerst stabil sind und die intendierte Entwicklung hin zu einem kompetenzorientierten Unterricht deutlich erschweren (Helmke, 2022). Rolff (2015) sieht daher eine Notwendigkeit, „die Überzeugungen und die Veränderung der Werte und Haltungen der in den Schulen handelnden Personen zum konstitutiven Bestandteil von UE [Unterrichtsentwicklung, Anm. d. Autorinnen] zu machen“ (Rolff, 2015, S. 17).

Kompetenzorientierter Unterricht und die damit verbundene verstärkte Nutzung von Evidenzen sind insgesamt äußerst voraussetzungsvoll und tragen damit unter Umständen zur Steigerung der individuellen Belastung von Lehrkräften bei. Das könnte Auswirkungen auf die nachhaltige Etablierung in der pädagogischen Praxis haben, die als erst teilweise gelungen beschrieben werden kann (Helmke, 2022; Lindner & Mayerhofer, 2018).

### 3 Methode

Im Rahmen dieser Studie wurden Lehrkräfte sowohl zu ihren Einstellungen zu standardisierten Diagnoseverfahren sowie ihren selbstberichteten Unterrichtsveränderungen befragt und im Anschluss der Frage nachgegangen, welche Typen von Lehrkräften sich im Zusammenhang mit den Anforderungen eines evidenzbasierten Unterrichts beschreiben lassen.

#### 3.1 Stichprobe

Zur Beantwortung der Forschungsfrage wurden Daten von Lehrkräften erhoben, welche im WS 2023 an den Schulungen der PH Wien zu iKM<sup>PLUS</sup> Modul 1 & 2 (Grundlagen und Ergebnisinterpretation) teilgenommen hatten. Insgesamt folgten N = 242 Lehrkräfte der Einladung zur Online-Befragung in Lime-Survey. Jedoch mussten im Anschluss eine Reihe von nicht vollständigen Datensätzen sowie Ausreißer im Zuge der Datenbereinigung ausgeschlossen werden, sodass schlussendlich n = 196 Datensets in die weiteren Berechnungen einfließen. In unserer bereinigten Stichprobe kamen 50,5 % der Lehrkräfte aus der Volksschule, 34,7 % unterrichten in der Mittelschule und 14,8 % in der AHS. Mit 83,7 % war ein Großteil der Lehrkräfte weiblich, 15,8 % waren männlich und 0,5 % gaben divers als

Geschlechtsbezeichnung an. 35,2 % der Lehrkräfte waren erst maximal fünf Jahre im Schuldienst.

### 3.2 Instrumente

Zur Anwendung kam ein Online-Fragebogen mit Skalen zu den diagnostischen Haltungen der Lehrkräfte sowie den wahrgenommenen Unterrichtsveränderungen. Die Skala „Diagnostische Selbstwirksamkeit“, „Einstellungen zur Diagnostik“ sowie die Skala „Diagnostische Aktivität“ stammen von Ohle et al. (2015). Die Skala zum „Nutzen von iKM<sup>PLUS</sup>“ wurde in Anlehnung an Emmrich et al. (2009) formuliert. Um die wahrgenommenen Unterrichtsveränderungen abzufragen, kamen die Skalen „Veränderung in der Kompetenzorientierung“, „Veränderung in der Binnendifferenzierung“, „Verengung des Lehrplans“ und „Keine Veränderung“ von Richter et al. (2014) zur Anwendung. Das Antwortformat aller Skalen war vierstufig und bewegte sich von 1 = „trifft nicht zu“ bis 4 = „trifft zu“.

Zur Überprüfung der Konstruktvalidität wurden alle Items einer explorativen Faktorenanalyse unterzogen. Dieses Verfahren führt zu einer Reduktion der Items auf wenige, den manifesten Variablen zugrunde liegende Faktoren (Backhaus et al., 2021, S. 414). Die aus der Literatur entnommenen Skalen wurden weitgehend bestätigt, zusätzlich wurde noch die Skala „Mehr Hausübungen“ ausgewiesen. Die verwendeten Skalen mit Beispielitem und unter Angabe der Reliabilität werden in Tab. 1 dargestellt.

Skala	Cronbach $\alpha$	Beispielitem
Diagnostische Selbstwirksamkeit	0.85	Ich bin mir sicher, dass ich die Fähigkeit meiner Schüler/innen richtig einschätzen kann.
Einstellungen zu Diagnostik	0.80	Diagnoseinstrumente können bei der Unterrichtsplanung hilfreich sein.
Diagnostische Aktivität	0.81	Ich verbringe Zeit damit, mir über die Anforderungen der Unterrichtsmaterialien für eine Unterrichtsstunde Gedanken zu machen.
Nutzen von iKM <sup>PLUS</sup>	0.86.	iKM <sup>PLUS</sup> ist ein guter Anhaltspunkt für die Einschätzung der Leistung einzelner Schüler/innen.
Kompetenzorientierung	0.62	Ich konzentriere mich stärker auf die Kompetenzen, die bei den Bildungsstandards erhoben werden.
Differenzierung	0.53	Ich konzentriere mich stärker auf Schüler/innen am oberen Ende des Leistungsspektrums.
Verengung des Lehrplans	0.61	Mein Unterricht ist thematisch enger geworden
Keine Veränderung	0.56	Ich halte es für falsch, wegen stand. Diagnoseverfahren Veränderungen in meinem Unterricht vorzunehmen.
Mehr Hausübung	0.87	Ich gebe mehr Hausaufgaben auf.

Tabelle 1: Übersicht der eingesetzten Skalen (Eigendarstellung)

Nachdem die Skalen als Vorbereitung für die folgende Clusteranalyse auf Korrelationen untereinander geprüft wurden, mussten die Skalen „Diagnostische Aktivität“, „Nutzen von iKM<sup>PLUS</sup>“ sowie „Verengung des Lehrplans“ wegen hoher Korrelationen mit anderen Skalen aus den weiteren Berechnungen ausgeschlossen werden.

### 3.3 Datenanalyse

Um Gemeinsamkeiten bzw. Unterschiede zwischen den Lehrkräften hinsichtlich ihrer Einstellungen zu evidenzbasiertem Unterricht zu ermitteln, führten wir eine Clusteranalyse durch. Sie ist ein Verfahren, in dem Personen oder Objekte zu Gruppen zusammengefasst werden, die innerhalb der Gruppen eine größtmögliche Ähnlichkeit aufweisen, die Gruppen zueinander jedoch möglichst hohe Unterschiedlichkeit (Backhaus et al., 2021, S. 490). Da sowohl kategoriale wie auch numerische Variablen im Datensatz vorhanden waren, griffen wir auf eine Two-Step Clusteranalyse und den Fusionierungsalgorithmus Partitioning Around Medoids (PAM) zurück (Shih et al., 2010). Im ersten Schritt dieser Methode werden die Daten in eine überschaubare Menge von Subclustern komprimiert, im zweiten Schritt werden die Subcluster mithilfe einer hierarchischen Clustermethode schrittweise zu immer größeren Clustern zusammengeführt (Backhaus et al., 2021, S. 570 f.). Als Distanzmaß zur Bestimmung der Ähnlichkeit zwischen den Clustern wählten wir im ersten Schritt die Gower Distanz (Höferl, 2022b) und im zweiten Schritt die Euklidische Distanz (Backhaus et al., 2021, S. 571). Um einer indirekten Gewichtung vorzubeugen, wurden die Variablen vorab z-standardisiert. Die Analyse wurden in R unter Verwendung des Pakets „cluster“ (Maechler et al., 2023) ausgeführt.

### 3.4 Ergebnisse

Als Entscheidungshilfe für die Bestimmung der Anzahl an Clustern wird für gewöhnlich das optische Elbow-Kriterium im Screeplot (Abb. 1) herangezogen (Backhaus et al., 2021, S. 529). Da der Sprung (Elbow) in unserem Fall nicht eindeutig zu erkennen war, wurde er direkt aus der Liste der Fehlerquadratsummen herausgelesen, die eine drei Cluster-Lösung nahelegte.

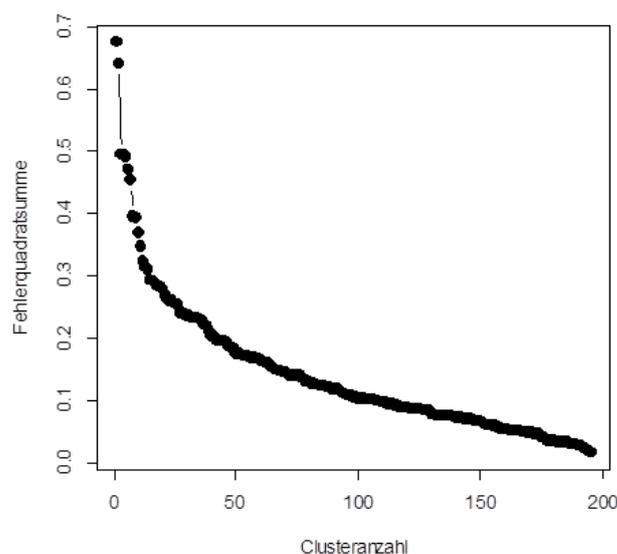


Abbildung 1: Screeplot zur Bestimmung der Clusteranzahl (Eigendarstellung)

Als weitere Entscheidungshilfe kann die Hauptkomponentenanalyse genutzt werden, mittels derer sich die Cluster visuell darstellen lassen (Höferl, 2022a). Wie im Clusplot (Abb. 2) ersichtlich ist, sind Cluster 1 und 2 nicht gänzlich trennscharf, sondern überlappen leicht. Zudem steckt Cluster 3 im Cluster 2, d. h. Personen des Clusters 3 gehören auch zu Cluster 2. Man spricht hier von geschachtelten Clustern. Was wie ein Spezialfall aussieht, ist vielmehr der Ansatz einer agglomerativen hierarchischen Clusterung, wie sie im zweiten Schritt der Analyse oben beschrieben wurde. „Beim hierarchischen Modell ist die Überlappung auf die Schachtelung begrenzt, gleichzeitig ist die Schachtelung zwingend“ (Uehlinger, 1988, S. 58).

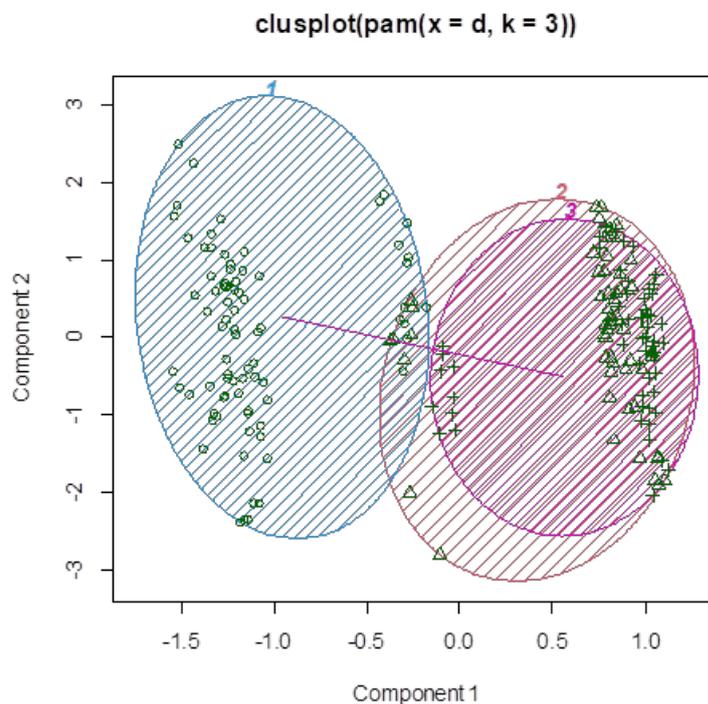


Abbildung 2: Clusplot zur Darstellung der Cluster (Eigendarstellung)

Für die Interpretation der Cluster sind die gestapelten Balkendiagrammen (Abb. 3) hilfreich, die genauen Werte sind Tab. 2 zu entnehmen. Ersichtlich ist, dass sich die z-standardisierten Variablen in Cluster 1 und 2 weitgehend um den Mittelwert gruppieren und damit als durchschnittlich bezeichnet werden müssen, während sich in Cluster 3 deutlichere Abweichungen vom Mittelwert zeigen. Allerdings zeigt sich nur in der Variable „Einstellung gegenüber standardisierter Diagnostik“ mit -2,17 eine signifikante Abweichung (siehe Tab. 2)<sup>1</sup>, die übrigen Variablen sind ebenfalls als durchschnittlich zu bezeichnen. Ihr Unterschied zu Variablen in Cluster 1 oder 2 kann daher nur deskriptiv beschrieben werden.

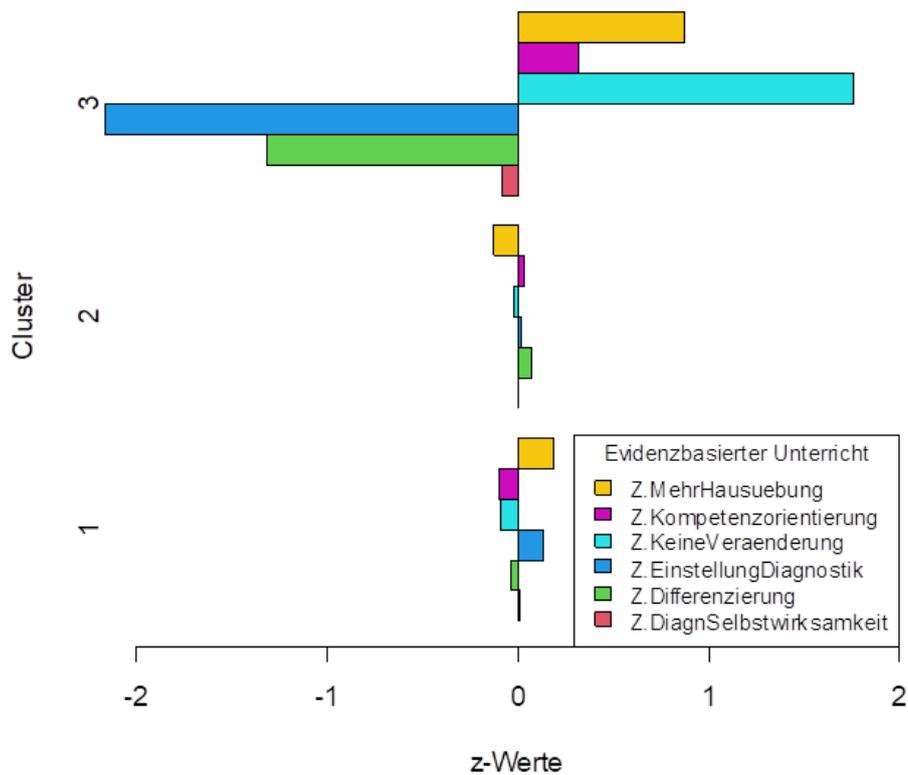


Abbildung 3: Gestapelte Balkendiagramme der Cluster (Eigendarstellung)

	Cluster 1	Cluster 2	Cluster 3
<b>Z.MehrHausübung</b>	0.18334138	-0.131571138	0.86917228
<b>Z.Kompetenzorientierung</b>	-0.09650049	0.038264306	0.31719887
<b>Z.KeineVeränderung</b>	-0.09540894	-0.019777212	1.75382827
<b>Z.EinstellungDiagnostik</b>	0.13443063	0.015697812	-2.16692966
<b>Z.Differenzierung</b>	-0.03341517	0.070441946	-1.31996846
<b>Z.DiagnSelbstwirksamkeit</b>	0.01049632	-0.002203731	-0.08345813

Tabelle 2: z-Werte der Cluster (Eigendarstellung)

Aufgrund der in Abb. 3 sowie in Tab. 2 dargestellten Ergebnisse der Two-Step Clusteranalyse beschreiben wir prototypische Profile von Lehrkräften, die als charakteristisch für die drei Cluster angesehen werden können.

Profil 1 – *Durchschnittliche Evidenzbasierung im Unterricht*: Die Lehrkräfte (n = 52) in diesem Cluster unterrichten ausschließlich in der Mittelschule. Sie kennzeichnen sich durch insgesamt durchschnittliche Werte in ihrer Einstellung gegenüber standardisierter Diagnostik, diagnostischer Selbstwirksamkeit sowie selbstberichteten Veränderungen im Unterricht wie mehr Hausübungen, Differenzierung und Kompetenzorientierung. Einem evidenzbasierten Unterricht stehen sie weder besonders ablehnend noch besonders befürwortend gegenüber.

Profil 2 – *Durchschnittliche Evidenzbasierung im Unterricht mit hartem Kern an Verweiger\*innen*: Es handelt sich hier um den größten Cluster (n = 126) bestehend aus Lehrkräften der Volksschule (n = 99) sowie Lehrkräften der AHS (n = 27). Diese unterscheiden sich hinsichtlich ihrer Einstellungen gegenüber einem evidenzbasierten Unterricht nicht wesentlich von Lehrkräften in Cluster 1. Auch sie berichten von Einstellungen, aus denen man einen durchschnittlich evidenzbasierten Unterricht ableiten kann. Allerdings enthält Cluster 2 auch Lehrkräfte, die einem eigenen Cluster, nämlich Cluster 3, zugeordnet werden können.

Profil 3 – *Harter Kern an Verweiger\*innen*: Hier gruppieren sich Lehrkräfte (n = 18) aus der Mittelschule und der AHS, die sich hinsichtlich ihrer diagnostischen Selbstwirksamkeit sowie dem Umfang an Hausübungen und der selbsteingeschätzten Differenzierung sowie Kompetenzorientierung im Unterricht etwas, aber nicht signifikant, von ihren Kolleg\*innen in Cluster 1 und 2 unterscheiden. Eine signifikante Abweichung vom Mittelwert gibt es lediglich in der Variable „Einstellung gegenüber standardisierter Diagnostik“ (-2,17). Das bedeutet, dass sie standardisierte Diagnostik, wie z.B. iKM<sup>PLUS</sup>, ablehnen. Sieht man sich die Werte nach Schultypen gelistet an (siehe Tab. 3), fällt auf, dass diese signifikante Ablehnung (-2,36) v.a. auf die AHS-Lehrkräfte (n = 2) zurückzuführen ist. Die Gruppe der Lehrkräfte aus Mittelschulen (n = 16) liegt mit ihrer Ablehnung knapp unter einem signifikanten Wert (- 1,88). Dafür kennzeichnet sich diese Gruppe durch eine signifikante Zustimmung zur Variable „Keine Veränderung im Unterricht“ (2,36), während die AHS-Lehrkräfte hier deutlich darunter liegen (1,35). Lehrkräfte in diesem Cluster haben also nicht nur eine signifikant negative Einstellung gegenüber standardisierter Diagnoseverfahren, ein Teil von ihnen gibt auch an, signifikant keine Veränderungen im Unterricht vorzunehmen. Die Zielsetzungen eines evidenzbasierten Unterrichts sind in diesem Cluster gescheitert, weshalb er auch mit der Bezeichnung „harter Kern“ versehen wurde.

Group. 1	Group. 2	Schul typ	Z. MehrH	Z. Komp	Z. KeineV	Z. Einstel	Z. Differ	Z.Diag Selbwi	Cluster
1	1	MS	0.18	-0.09	-0.09	0.13	-0.03	0.01	1
2	2	AHS	0.19	-0.15	-0.24	0.05	-0.00	-0.24	2
3	2	VS	-0.21	0.09	0.04	0.00	0.09	0.06	2
1	3	MS	1.88	1.74	2.36	-1.88	-0.74	0.37	3
2	3	AHS	0.19	-0.63	1.35	-2.36	-1.70	-0.39	3

Tabelle 3: z-Werte der Cluster nach Schultypen gelistet (Eigendarstellung)

## 4 Interpretation und Diskussion

Die vorliegende Studie verfolgte das Ziel, mehr über die Einstellungen von Lehrkräften zu evidenzbasiertem Unterricht zu erfahren. Dazu wurden Lehrkräfte sowohl zu ihren Einstellungen gegenüber standardisierten Rückmeldungen sowie ihren selbstberichteten

Unterrichtsveränderungen befragt. Mittels Clusteranalyse konnten drei prototypische Profile von Lehrkräften herausgearbeitet werden.

Vordergründig ins Auge sticht jenes dritte Profil, aus dem sich eine deutliche Ablehnung eines evidenzbasierten Unterrichts herauslesen lässt. Hier gruppieren sich Lehrkräfte der Sekundarstufe, die eine negative Einstellung gegenüber standardisierter Diagnostik aufweisen und von keinen Veränderungen im Unterricht berichten. Allerdings ist dieser *harte Kern an Verweiger\*innen* zahlenmäßig sehr klein. Laut Demski (2017) charakterisieren sich jene Verweiger\*innen dadurch, dass Lehrkräfte und Schulen in dieser Gruppe schlechte Ergebnisse häufig extern attribuieren, also auf ein als leistungsschwach angesehenes Lernendenklientel und strukturelle Rahmenbedingungen zurückführen. Das hat einerseits eine Entlastungsfunktion, gleichzeitig wird damit aber die intendierte Nutzung von Rückmeldungen zur Weiterentwicklung verhindert. Verweigerung als Entlastungsfunktion ist v.a. dann wichtig, wenn Bewältigungsstrategien zur Überwindung der Stressoren fehlen (Wenzel, 2017). Dass dieses Profil in der vorliegenden Studie ausschließlich Lehrkräfte der Sekundarstufe umfasst, könnte daran liegen, dass die Belastungen, wie aus einer Studie von Spenger et al. (2019) hervorgeht, in der Mittelschule deutlich über jenen der Volksschule liegen. Allerdings muss auch erwähnt werden, dass die Belastung der AHS-Lehrkräfte in eben dieser Studie wiederum unter jenen ihrer Kolleg\*innen in der Volksschule liegen. Insofern lässt sich erklären, weshalb die AHS-Lehrkräfte in dem betreffenden Profil aus Verweiger\*innen kaum vertreten waren, dafür umso mehr in jenem Profil der *durchschnittlichen Evidenzbasierung im Unterricht*, in dem die Lehrkräfte der Volksschule die überwiegende Mehrheit darstellten. Das nach Schularten unterschiedliche Belastungserleben von Lehrkräften könnte somit ein impliziter Faktor hinter den hier zur Clusterung verwendeten Merkmalen sein.

In den Profilen der *durchschnittlichen Evidenzbasierung im Unterricht* zu den Clustern 1 und 2 zeigte sich eine ähnliche Gruppierung von Lehrkräften, die sich durch insgesamt durchschnittliche Werte in ihrer Einstellung gegenüber standardisierter Diagnostik, diagnostischer Selbstwirksamkeit sowie selbstberichteten Veränderungen im Unterricht wie mehr Hausübungen, Differenzierung und Kompetenzorientierung kennzeichnen. Einem evidenzbasierten Unterricht stehen sie also weder besonders ablehnend noch besonders befürwortend gegenüber. Dahinter könnte sich jene Einstellung von Lehrkräften verbergen, die Reh (2005, S. 77) als „Homogenisierungssehnsucht“ bezeichnet. Gemeint ist der Wunsch, eine Normalität im Klassenzimmer zu konstruieren und daraus resultierend einen Unterricht zu halten, der allen Schüler\*innen eine Gleichbehandlung zukommen lässt. Überdurchschnittliche diagnostische Fähigkeiten wie auch besondere Expertise in der Differenzierung auf Seiten der Lehrkräfte sind in diesem auf Homogenisierung ausgerichteten Unterricht nicht gefragt. Die pädagogische Handlungskompetenz von Lehrkräften bezogen auf den evidenzbasierten Unterricht könnte sich hier also in der klassischen Durchschnittsfalle befinden, die sich durch das „Hervorbringen durchschnittlicher Allround-Köner“ (Hengstschläger, 2020, Klappentext) kennzeichnet. Durchschnittliche pädagogische Expertise reicht aber unter Umständen nicht aus, um Anforderungen in der Inklusion, in der

Begabtenförderung oder im Umgang mit Schüler\*innen mit Migrationshintergrund gerecht zu werden. Hier braucht es Lehrkräfte mit einer hohen diagnostischen Selbstwirksamkeit, einer überdurchschnittlich positiven Einstellung zu Diagnostik, hoher diagnostischer Aktivität, die sich dann in einer erhöhten Differenzierung und Kompetenzorientierung des Unterrichts niederschlägt.

Eine durchschnittliche Einstellung zu evidenzbasiertem Unterricht kann also, muss aber nicht, als Durchschnittsfalle gesehen werden. Bekanntlich ist iKM<sup>PLUS</sup> seit dem Schuljahr 2021/22 auf der Grundstufe und seit dem Schuljahr 2022/23 auf der Sekundarstufe verpflichtend durchzuführen. Richter et al. (2014) weisen darauf hin, dass Lehrkräfte ihren Unterricht vermutlich genau dann nicht verändern, wenn sie bildungspolitische Maßnahmen als Bevormundung erleben (siehe auch Demski, 2017). Aus dem Blickwinkel der „top-down“ Steuerung kann eine durchschnittliche Zustimmung zu standardisierten Rückmeldungen und Veränderungen im Unterricht daher auch als kleiner Erfolg gewertet werden. Der harte Kern an Verweiger\*innen hätte auch deutlich größer ausfallen können. Bemerkenswert erscheint auch, dass keine Gegensätzlichkeit zwischen der Einstellung zu standardisierter Diagnostik und Differenzierung, Kompetenzorientierung oder anderen Unterrichtsveränderungen aus den Profilen herausgelesen werden kann. Bemerkenswert ist dies deshalb, weil Kallenbach (2023) von zwei unvereinbaren Handlungslogiken zwischen der Standardisierung einerseits und der Individualisierung andererseits spricht und eben dieses Muster auch in den Interviews der Studie von Müller und Wagner (2024) aufgetaucht ist.

Dass Lehrkräfte generell eher nur dergestalt individualisieren und differenzieren, indem sie spontan auf die verschiedenen Bedürfnisse im Unterricht eingehen, ohne dies langwierig vorzubereiten, darauf haben Jennek et al. (2019) hingewiesen. Ob Lehrkräfte ihren Unterricht verändern, hängt also neben den Einstellungen zu standardisierten Rückmeldungen scheinbar auch vom Arbeitsaufwand und der damit verbundenen Arbeitsbelastung ab.

## 4.1 Limitation

Einschränkungen der hier berichteten Ergebnisse können darauf beruhen, dass die Auswahl der befragten Lehrkräfte nicht zufällig erfolgte, sondern Teilnehmer\*innen einer Fortbildungsveranstaltung befragt wurden. Eine positiv selektierte Stichprobe kann allerdings weitgehend ausgeschlossen werden, da die Durchführung von iKM<sup>PLUS</sup> verpflichtend und die Teilnahme am Modul mit hohem Empfehlungscharakter verbunden ist. In der vorliegenden Stichprobe könnten durch die Empfehlung somit eher negative Tendenzen im Antwortverhalten auftreten. Ebenfalls nachteilig ist, dass die unterschiedlichen Schularten nicht annähernd gleichverteilt in der Stichprobe vertreten waren. Nachfolgende Untersuchungen mit einer randomisierten und mehr repräsentativen Stichprobe wären daher wünschenswert.

Von Interesse wären auch tiefergehende Analysen, was den „harten Kern“ an Verweiger\*innen betrifft. Dem sind aber durch die Auflage einer anonymen Befragung

Grenzen gesetzt. Daher ist es nicht möglich, im Rahmen qualitativer Interviews nach weiteren Gründen für die Verweigerung der betreffenden Lehrkräfte zu fragen.

## 4.2 Pädagogische Implikation

Unabhängig davon, ob durchschnittliche Werte eines evidenzbasierten Unterrichts als Durchschnittsfalle oder als kleiner Erfolg gewertet werden, kann man sich aus Sicht der Fortbildung nicht damit zufriedengeben. Sowohl aus dem Blickwinkel der Inklusion wie auch der Begabtenförderung wäre es wünschenswert, die diagnostische Selbstwirksamkeit von Lehrkräften, ihre Einstellung gegenüber standardisierter Diagnostik sowie daraus resultierende Unterrichtsveränderungen in Richtung mehr Kompetenzorientierung und Differenzierung zu erhöhen. Dazu gilt es den Handlungsspielraum des Ministeriums bei Fortbildungen zu iKM<sup>PLUS</sup> stärker zu nutzen. Die Inhalte der Fortbildungsveranstaltungen werden weitgehend vom BMBWF gemeinsam mit dem IQS vorgegeben, die Umsetzung obliegt aber den Pädagogischen Hochschulen. Hier sollen an der PH Wien in Zukunft verstärkt Impulse gesetzt werden, bereits bei grundlegenden Erklärungen zur Handhabung auf mögliche unterrichtliche Konsequenzen hinzuweisen. Es gilt, ein noch besseres Zusammenspiel zwischen der „data literacy“, also der Fähigkeit Daten zu interpretieren und bewusst im Unterricht einzusetzen, und dem Kompetenzaufbau im Unterricht zu erreichen. Die Nutzung der Rückmeldungen sollte hier stärker mit individuellen Zielen der jeweiligen Lehrkraft verbunden werden und sich auf ganz konkrete Fragen der Unterrichtsplanung und -gestaltung beziehen. Fraglich ist, inwieweit dies in halb- oder eintägigen Fortbildungsveranstaltungen möglich ist. Sinnvoller erscheinen hier Angebote seitens der PH Wien zu gezielten externen Begleitungen von Schulstandorten, die Leitung, Fachgruppen, Teams und einzelne Lehrpersonen bei der Implementierung von datengestützter Unterrichtsentwicklung durch multiprofessionelle Teams (Rückmeldemoderation, Organisationsentwicklung, Fachdidaktik) unterstützen, um langfristig entsprechende Strukturen und Kompetenzen aufzubauen. Durch begleitete Schulentwicklungsprozesse könnte es auch gelingen, das Belastungserleben der Lehrkräfte gering zu halten.

## Literatur

- Altrichter, H. & Gamsjäger, M. (2019). Ein Wirkungsmodell für die Erforschung von Bildungsstandard-Politiken. In J. Zuber, H. Altrichter & M. Heinrich (Hrsg.), *Bildungsstandards zwischen Politik und schulischem Alltag* (S. 45–78). Springer VS.
- Altrichter, H. & Kanape-Willingshofer, A. (2012). Bildungsstandards und externe Überprüfung von Schülerkompetenzen: Mögliche Beiträge externer Messungen zur Erreichung der Qualitätsziele der Schule. In B. Herzog-Punzenberger (Hrsg.), *Nationaler Bildungsbericht Österreich 2012*. Band 2. Fokussierte Analysen bildungspolitischer Schwerpunktthemen (S.

- 355–394). Leykam. [https://www.bmbwf.gv.at/dam/jcr:4af4ffb7-0487-49f2-8be1-a0d4f334917a/nbb\\_2012\\_b02\\_kapitel09\\_23893.pdf](https://www.bmbwf.gv.at/dam/jcr:4af4ffb7-0487-49f2-8be1-a0d4f334917a/nbb_2012_b02_kapitel09_23893.pdf)
- Altrichter, H. & Maag Merki, K. (2016). Steuerung der Entwicklung des Schulwesens. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (S. 1–28). Springer VS.
- Altrichter, H., Moosbrugger, R., & Zuber, J. (2016). Schul- und Unterrichtsentwicklung durch Datenrückmeldung. In H. Altrichter & K. Maag Merki (Hrsg.), *Handbuch Neue Steuerung im Schulsystem* (S. 235–277). Springer VS.
- Asbrand, B. (2020). Der Umgang von Lehrpersonen mit der Bildungsstandardsreform. In U. Greiner, F. Hofmann, C. Schreiner & C. Wiesner (Hrsg.), *Bildungsstandards: Kompetenzorientierung, Aufgabenkultur und Qualitätsentwicklung im Schulsystem* (S. 52–546). Waxmann.
- Backhaus, K., Erichson, B., Gensler, S., Weiber, R. & Weiber, Th. (2021). *Multivariate Analysemethoden. Eine anwendungsorientierte Einführung*. 16. Aufl. Springer Gabler Verlag.
- Beer, R. & Benischeck, I. (2011): Aspekte kompetenzorientierten Lernens und Lehrens. In BIFIE (Hrsg.), *Kompetenzorientierter Unterricht in Theorie und Praxis* (S. 5–28). Leykam.
- Beywl, W. (2017). Datengestützte Schul- und Unterrichtsentwicklung aus wissenschaftlicher Sicht. In Ministerium für Kultus, Jugend und Sport BW (Hrsg.), *Dategestützte Schul- und Unterrichtsentwicklung an beruflichen Schulen. Konzept OES*, S. 5–9.
- Döring, N. (2023). *Forschungsmethoden und Evaluation in den Sozial- und Humanwissenschaften*. 6. Aufl. Springer.
- Demski, D. (2017). *Evidenzbasierte Schulentwicklung: Empirische Analyse eines Steuerungsparadigmas*. Springer VS.
- Emmrich, R., Krieg, S., & Wesselhöft, K. (2009). *Evaluation der Ermittlung der Lernausgangslage in Jahrgangsstufe 7 im Schuljahr 2008/09 in Brandenburg*. Unveröffentlichter Projektbericht.
- Harms, U., Schroeter, B. & Klüh, B. (2016). *Entwicklung kompetenzorientierten Unterrichts in Zusammenarbeit mit Forschung und Schulpraxis*. Waxmann.
- Hengstschläger, M. (2020). *Die Durchschnittsfalle. Gene-Talente-Chancen*. Ecwin.
- Heinrich, M., Kussau, J. (2010). Das Schulprogramm zwischen schulischer Selbstregelung und externer Steuerung. In Altrichter, H., & Maag Merki, K. (Hrsg.). *Handbuch Neue Steuerung im Schulsystem* (S. 17–194). VS Verlag für Sozialwissenschaften.
- Helmke, A. (2022). *Unterrichtsqualität und Professionalisierung: Diagnostik von Lehr-Lern-Prozessen und evidenzbasierte Unterrichtsentwicklung*. Kallmeyer in Verbindung mit Klett.
- Helmke, A. & Helmke, T. (2015). Unterrichtsdiagnostik als Ausgangspunkt für Unterrichtsentwicklung. In: H.-G. Rolff (Hrsg.). *Handbuch der Unterrichtsentwicklung* (S. 242–257). Beltz Verlag.
- Helmke, A. & Hosenfeld, I. (2005). Standardbezogene Unterrichtsevaluation. In G. Brägger, B., Bucher, N. Landwehr (Hrsg.), *Schlüsselfragen zur externen Schulevaluation* (S. 127–152). h.e.p.
- Honig, M. & Coburn, C. (2008). Evidence-Based Decision Making in School District Central Offices toward a Policy and Research Agenda. *Educational Policy*, 22, S. 578–608.
- Höferl, K. (2022a). *Clusteranalyse I: Grundlagen*.  
[https://kamihoeferl.at/lehre/vu\\_sozwiss\\_2/07\\_cluster\\_I.html](https://kamihoeferl.at/lehre/vu_sozwiss_2/07_cluster_I.html)
- Höferl, K. (2022b). *Clusteranalyse II: Gemischtskalige Modelle*.  
[https://kamihoeferl.at/lehre/vu\\_sozwiss\\_2/08\\_cluster\\_II\\_v2.html](https://kamihoeferl.at/lehre/vu_sozwiss_2/08_cluster_II_v2.html)

- Jennek, J., Gronostaj, A., & Vock, M. (2018). Wie Lehrkräfte im Englischunterricht differenzieren. Eine Re-Analyse der DESI-Videos. *Unterrichtswissenschaft*, 47, S. 99–116.
- Kallenbach, L. (2023). Evidenzbasierte Schulentwicklung als mehrdimensionale Spannungsbearbeitung. Ein übergeordneter Erklärungsansatz für anhaltende Wirksamkeitsprobleme. *Zeitschrift für Bildungsforschung*, 13, S. 109–126.
- Klieme, E., Avenarius, H., Blum, W., Döbrich, P., Gruber, H., Prenzel, M., Reiss, K., Riquarts, K., Rost, J., Tenorth, H., Vollmer, H., Bundesministerium für Bildung und Forschung (2003). *Zur Entwicklung nationaler Bildungsstandards. Eine Expertise*. BMBF 2003. [https://www.pedocs.de/volltexte/2020/20901/pdf/Klieme\\_et\\_al\\_2003\\_Zur\\_Entwicklung\\_Nationaler\\_Bildungsstandards\\_BMBF\\_A.pdf](https://www.pedocs.de/volltexte/2020/20901/pdf/Klieme_et_al_2003_Zur_Entwicklung_Nationaler_Bildungsstandards_BMBF_A.pdf)
- Klopsch, B. & Sliwka A. (2020). Schulqualität als Resultat eines komplexen adaptiven Systems: die Verschränkung von Systemebenen zur Verbesserung des Schülerlernens. Datengestützte Schulentwicklung in der Provinz Alberta, Kanada. In Fickermann, D., Veronika Manitus, V. Karcher, M. (Hrsg.). *Neue Steuerung – Renaissance der Kybernetik? Die Deutsche Schule*, Beiheft 15. (S. 58–74). Waxmann.
- Köller, O. (2010). Bildungsstandards. In R. Tippelt & Schmidt, B. (Hrsg.) *Handbuch Bildungsforschung*. VS Verlag für Sozialwissenschaften, S. 529–548.
- Krieg, Martina (2020). *Von Daten zu Taten in der Schulentwicklung*. Schulinfo Zug. <https://www.zg.ch/behoerden/direktion-fur-bildung-und-kultur/schulinfo/fokus/von-daten-lernen-professionelle-neugier-wecken>
- Lindner, G. & Mayerhofer, S. (2018). *Kompetenzorientiert guter Unterricht und bedarfsorientierte Lehrerfortbildung*. Waxmann.
- Maechler, M., Rousseeuw, P., Struyf, A., Hubert, M., & Hornik, K. (2023). "Finding Groups in Data": Cluster Analysis Extended Rousseeuw et al. <https://cran.r-project.org/web/packages/cluster/cluster.pdf>
- Müller, M.A., & Wagner, G. (2024). Einstellungen zur Diagnostik als Teil der professionellen Handlungskompetenz von Lehrpersonen der Sekundarstufe 1. *Pädagogische Horizonte*, 8 (Heft 1), S. 9–23.
- Nachtigall, C., & Hellrung, K. (2013). Zur zeitlichen Entwicklung der Rezeption von Vergleichsarbeiten. *Empirische Pädagogik*, 27 (Heft 4), S. 423–441.
- Ohle, A., McElvany, N., Horz, H., & Ullrich, M. (2015). Text-picture integration—Teachers' attitudes, motivation and self-related cognitions in diagnostics. *Journal for Educational Research Online*, 7 (Heft 2), S. 11–33.
- Richter, D., Böhm, K., Becker, M., Pant, H. A., & Stanat, P. (2014). Überzeugungen von Lehrkräften zu den Funktionen von Vergleichsarbeiten. *Zeitschrift für Pädagogik*, 60 (Heft 2), S. 225–244.
- Reh, S. (2005). Warum fällt es Lehrerinnen und Lehrern so schwer, mit Heterogenität umzugehen? Historische und empirische Deutungen. *Die Deutsche Schule*, 97 (Heft 1), S. 76–86.
- Rolff, H.-G. (2015). Formate der Unterrichtsentwicklung und Rolle der Schulleitung. In H.-G. Rolff (Hrsg.). *Handbuch der Unterrichtsentwicklung* (S. 12–32). Beltz Verlag.
- Schildkamp, K., & Ehren, M. (2013). The Netherlands: From "intuition"- to "data"-driven decision making in Dutch secondary schools? In K. Schildkamp, M. K. Lai, & L. Earl (Eds.), *Data-based decision making in education: Challenges and opportunities* (S. 49–68). Springer.

- Schildkamp, K & Lai, M. & Earl, L. (2013). *Data-based Decision Making in Education: Challenges and Opportunities*. [https://www.researchgate.net/publication/321554468\\_Data-based\\_Decision\\_Making\\_in\\_Education\\_Challenges\\_and\\_Opportunities](https://www.researchgate.net/publication/321554468_Data-based_Decision_Making_in_Education_Challenges_and_Opportunities)
- Schreiner, C. & Breit, S. (2019). Pädagogische Diagnostik als Transfer-Herausforderung. Instrumente pädagogischer Diagnostik im Spannungsfeld zwischen wissenschaftlichen und schulpraktischen Ansprüchen. In C. Schreiner, C. Wiesner, S. Breit, P. Döbelstein, M. Heinrich & U. Steffens (Hrsg.), *Praxistransfer Schul- und Unterrichtsentwicklung* (S. 171–188). Waxmann.
- Schulte, K., Fickermann, D. Lücken, M. (2016). Das Hamburger Prozessmodell datengestützter Schulentwicklung - In: *Die Deutsche Schule* 108. 2/16, S. 176–190.
- Shih, M.-Y., Jheng, J.-W., & Lai, L.-F. (2010). A Two-Step Method for Clustering Mixed Categorical and Numeric Data. *Tamkang Journal of Science and Engineering*, 13 (Heft 1), S. 11–19.
- Specht, W. (2002). Überlegungen zur Institutionalisierung zentraler Funktionen der Qualitätsentwicklung und Qualitätssicherung auf nationaler Ebene. In F. Eder (Hrsg.), *Qualitätsentwicklung und Qualitätssicherung im österreichischen Schulwesen* (Bildungsforschung des Bundesministeriums für Bildung, Wissenschaft und Kultur, Band 17. (S. 423–441). StudienVerlag.
- Spenger, J., Katschnig, T., Schrittmesser, I., & Wistermayer, L. (2019). Under pressure. Berufsvollzugsprobleme und Belastungen von Lehrpersonen. Eine empirische Studie. *Forschungsbericht der PH Niederösterreich und der Universität Wien*. [https://www.ph-noe.ac.at/fileadmin/root\\_phnoe/MitarbeiterInnen/Silvia\\_Zechmeister/Under\\_pressure\\_Projektbericht\\_Mai\\_2019.pdf](https://www.ph-noe.ac.at/fileadmin/root_phnoe/MitarbeiterInnen/Silvia_Zechmeister/Under_pressure_Projektbericht_Mai_2019.pdf)
- Uehlinger, H.-M. (1988). Überlappende Clusterstrukturen: ein Verfahren zur exploratorischen Datenanalyse. *ZUMA Nachrichten*, 12 (Heft 22), S. 58–73.
- Wenzel, B. (2017). *Belastungserleben von Lehrkräften und der Einfluss von Schulreformen am Beispiel der Bildungsstandards*. Diplomarbeit: Universität Linz. <https://epub.jku.at/obvulihs/download/pdf/1915379?originalFilename=true>
- Wiesner, C., Schreiner, C., Breit, S., Kemethofer, D., George, A. & Angerer, S. (2016). Die Bedeutsamkeit der professionellen Reflexion und Rückmeldekultur für eine evidenzorientierte Schulentwicklung durch Bildungsstandardüberprüfungen. *Journal für Schulentwicklung*. 4/16, S. 18–26.
- Westphal, A., Zuber, J., & Vock, M. (2018). Welche Rolle spielen Selbstwirksamkeit, Motivation und Einstellung zu Diagnostik für die Nutzung datenbasierter Rückmeldungen. *Zeitschrift für Bildungsforschung*, 8 (Heft 3), S. 289–307.

---

<sup>1</sup> Z-Werte die außerhalb von  $\pm 1,96$  liegen, werden als signifikant bezeichnet (Döring, 2023, S. 631).