LMU LUDWIG-MAXIMILIANS-UNIVERSITÄT MÜNCHEN

Evaluation im Spannungsfeld zwischen methodischen Ansprüchen, legitimatorischen Notwendigkeiten und pädagogischem Selbstverständnis

Anna Gieschen,
Prof. Dr. Thomas Eckert

Tagung des AK Methoden in der DeGEval
München, 11. Juli 2025

Bedeutung der Evaluation in Bezug auf Bildung und Erziehung

Verschiedene Tendenzen führen zur Bedeutungszunahme von Evaluationen in der Pädagogik, aber auch zur Diversifikation der **Stakeholder:innen**.

Exemplarisch hierzu folgendes:

- Die Bedeutung von Qualitätssicherung wird immer öfter gesetzlich fixiert und zur Förderbedingung (KJHG oder „Gute KiTa-Gesetz“, WB-Gesetze der Länder) – **Politik, Geldgeber**

- Stärkere steuerungspolitische Vorgaben im Zusammenhang mit Educational Governance: Bildungsmonitoring, Bildungsberichterstattung (national wie regional) – **Politik, Institutionen, Öffentlichkeit**

- Die Politik der EU über Ausschreibungen (Lernen vor Ort) oder eigene Konzepte (European Quality Assurance Framework) – **Projektverantwortliche**

- Projekte selbst führen zu Kooperationen mit Stiftungen (PPP) – **Stiftungen, pädagogisch Tätige**

Bedeutung der Evaluation in Bezug auf Bildung und Erziehung

Eine der **Konsequenzen** ist:

Die Ergebnisse von Evaluationen müssen einer breiten Palette an Adressat:innen vermittelt werden, die oft nicht über das methodische Wissen verfügen, die Ergebnisse einschätzen zu können.

Das führt zu **Entscheidungsproblemen** auf mehreren Ebenen:

1. Der Vermittlung (scheinbar) widersprüchlicher Befunde
2. Der Bezug auf eine (normative) Referenz
3. Die Herstellung der Akzeptanz für eine Evaluation und der Glaubwürdigkeit ihrer Befunde bei den pädagogisch Tätigen

1. Widersprüchliche Befunde: Die Studie ELEMENT

Die Studie „Erhebung zum Lese- und Mathematikverständnis in den Jahrgangsstufen 4 bis 6 in Berlin“ (ELEMENT) wurde 2003 vom Berliner Senat in Auftrag gegeben. Die Ergebnisse wurden publiziert zu einem Zeitpunkt, als in Hamburg die schwarz-grüne Koalition über eine sechsjährige Primarschule verhandelte.

Die sich daraus ergebenden Empfehlungen waren widersprüchlich. Nicht nur unterschiedliche Erziehungswissenschaftler:innen kamen zu gegensätzlichen Schlüssen, sondern auch der Autor selbst.

1. Die Leistungssteigerung war an beiden Schulformen vergleichbar, mit einem leichten Vorteil für Grundschulen (Lehmann & Lenkeit, 2008, S. 17):

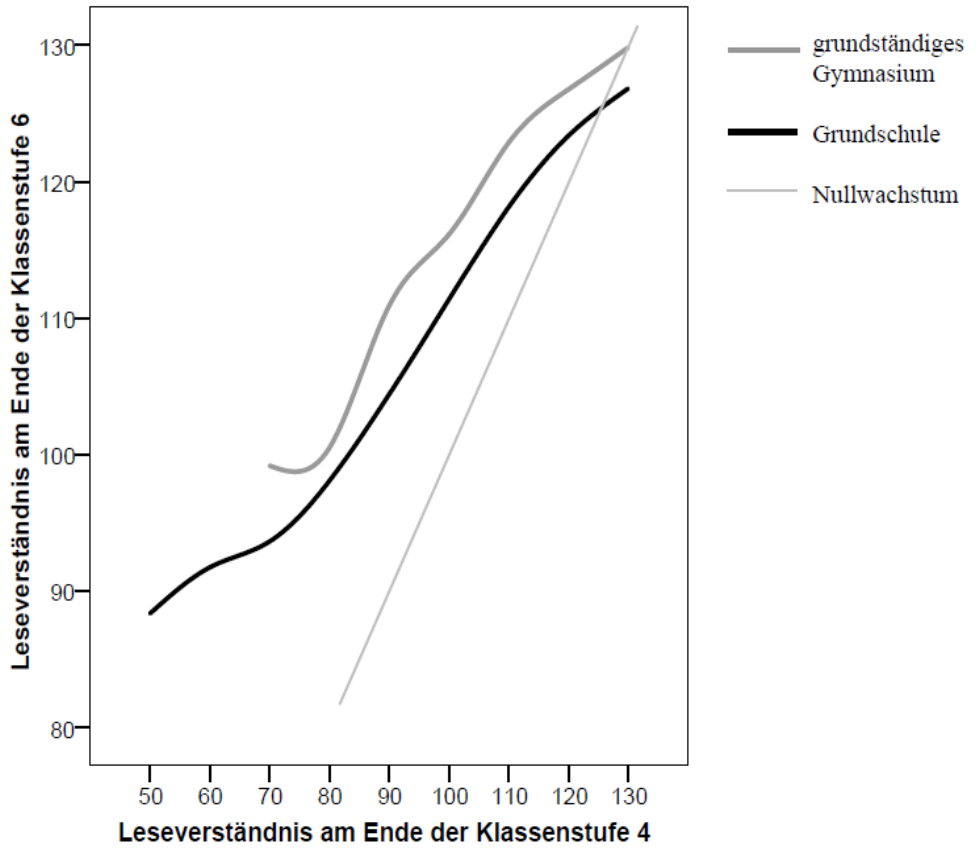
   Beispiel: Leseverständnis (Ende Klasse 5 vs. Ende Klasse 6)

Differenz Grundschule:	5,1 Punkte
Differenz Gymnasium:	4,0 Punkte

1. Widersprüchliche Befunde: Die Studie ELEMENT

2. Bei gleicher Lernausgangslage waren die Steigerungen an Gymnasien höher.

Abb. 3.3: Differenzielle Lernentwicklung im Leseverständnis zwischen dem Ende der Klasse 4 / Beginn Klasse 5 und dem Ende der Klasse 6 nach Schulform



(Lehmann & Lenkeit, 2008, S. 23)

1. Widersprüchliche Befunde: Die Studie ELEMENT

3. Eine eigene Analyse der Daten der ELEMENT-Studie zeigt einen bedeutsamen negativen Interaktionseffekt zwischen Schulart (Referenz: Grundschule) und Leseverständnis in der 5ten Klasse (eigene Darstellung).

Koeffizienten[a]

Modell		Nicht standardisierte Koeffizienten		Standardisierte Koeffizienten			Kollinearitätsstatistik	
		Regressionskoeffizient B	Std.-Fehler	Beta	T	Sig.	Toleranz	VIF
1	(Konstante)	30,775	,612		50,322	<,001		
	Bdl WLEskala Leseverständnis E5	,756	,006	,754	130,862	<,001	,559	1,790
	Gym	21,077	1,424	,764	14,807	<,001	,007	143,537
	bdl_Gym	-,155	,012	-,675	-12,746	<,001	,007	151,551

a. Abhängige Variable: Cdl WLEskala Leseverständnis E6

4. Eine Reanalyse der Daten durch Baumert et al. (2009) zeigte, dass die Unterschiede bei gleicher Ausgangslage verschwinden, wenn man mit Hilfe des Propensity Score Matchings abhängige Stichproben erzeugt (Baumert et al., 2009, S. 199-210).

2. Pädagogisch begründete Referenz?

Die ZEIT online vom 05.12.2017 berichtet über die IGLU-Studie folgendes:

Iglu-Studie

Ein Fünftel der Viertklässler kann nicht gut lesen

Der Anteil der Viertklässler in Deutschland, die sehr gut lesen, ist gestiegen. Ein Fünftel erreicht laut einer Studie jedoch nicht einmal ein mittleres Kompetenzniveau.

Später heißt es dann …

Allerdings erreichte im vergangenen Jahr knapp ein Fünftel (18,9 Prozent) nicht einmal die mittlere Kompetenzstufe III. Diese Kinder verfügen über ein nicht ausreichendes Leistungsniveau im Lesen. Sie sind mit erheblichen Schwierigkeiten beim Lernen in der Sekundarstufe I konfrontiert.

Ist das Ereignis nun wirklich besorgniserregend?
– Was ist dann der Maßstab dazu?

2. Pädagogisch begründete Referenz?

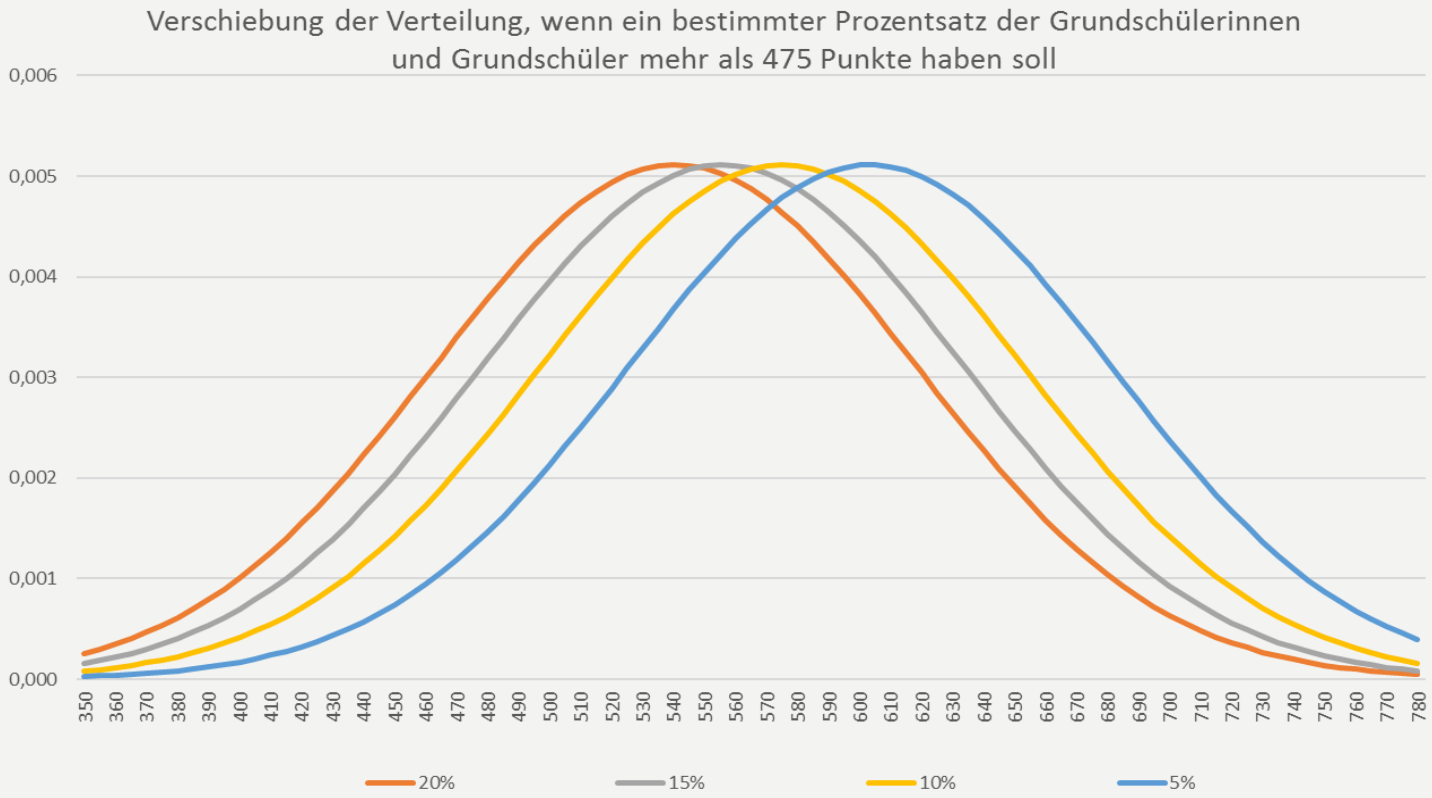
1. **Einfache statistische Referenz:** 50% der Schüler:innen sind schlechter als der Durchschnitt!

2. **Komplexere statistische Referenz:** Welche Verteilung der Schüler:innen auf die Kompetenzstufen von IGLU wäre bei vorgefundenem arithmet. Mittel (537) und Standardabweichung (78) zu erwarten?

Kompetenzstufe	Erwartet	Tatsächlich
1	3,84%	5,5%
2	17,38%	13,4%
3	34,91%	34,1%
4	29,92%	35,9%
5	12,96%	11,1%

2. Pädagogisch begründete Referenz?

3. **Pädagogische Machbarkeit:** Wie müsste sich die Verteilung verändern, wenn ein gegebener Anteil an Schüler:innen einen vorgegebenen Wert überschreitet?



	20%	15%	10%	5%
Mittelwerte:	540,6	555,8	575,0	603,3

3. Herstellung von Akzeptanz und Glaubwürdigkeit (übergreifend)

Ebene 1: Individuelle Ebene	**Ebene 2:** Institutionelle Ebene
• Professionelles (pädagogisches) Selbstverständnis • Haltung gegenüber Evaluation • Evaluationskompetenz und –Erfahrung	• Führungsverständnis und Steuerungslogik • Interne Qualitätsentwicklungs-prozesse und Reflexionsroutinen • Integration der Evaluation
Ebene 3: Systemebene	**Ebene 4:** Interaktionsebene
• Legitimationsbedarf durch Förderlogiken • Steuerungshandeln (evidenzbasiert vs. symbolisch) • Erwartungshaltung von Stakeholder:innen	• Transparenz des Evaluationsprozesses und der -Ergebnisse • Grad an Partizipation • Grad an Kokonstruktion

(Eigene Darstellung)

3. Herstellung von Akzeptanz und Glaubwürdigkeit (individuell)

Ebene 1: Individuelle Ebene

- Professionelles (pädagogisches) Selbstverständnis (Oevermann, 2008):
  - Beziehungsgestaltung im Fokus des Selbstverständnisses
  - Fundierung eigener Qualitätsurteile auf Erfahrung und Intuition
  - Ablehnung externer Beurteilungsmaßstäbe als Einschränkung professioneller Autonomie (Autonomie-Paritäts-Muster)

- Haltung gegenüber Evaluation (Ackeren et al., 2013):
  - Wahrnehmung von Evaluation als Kontrollinstrument
  - Skepsis gegenüber standardisierten Verfahren / quantitativen Indikatoren
  - Distanzierung bei fehlender Anschlussfähigkeit an pädagogische Praxis

3. Herstellung von Akzeptanz und Glaubwürdigkeit (individuell)

Ebene 1: Individuelle Ebene

- Evaluationskompetenz und –Erfahrung (Helmke, 2009):
  - Unterschiedliche Vertrautheit mit Evaluationskonzepten und -Verfahren
  - Wenig Routinen im Umgang mit Daten und Ergebnissen
  - Fehlende institutionalisierte Reflexionsmöglichkeiten

3. Herstellung von Akzeptanz und Glaubwürdigkeit (individuell)

- Professionelles (pädagogisches) Selbstverständnis: Fundierung eigener Qualitätsurteile auf Erfahrung und Intuition

  „Also ich bin so zwiegespalten, weil wie gesagt, ich erlebe die Kinder ja jede Woche, das heißt, ich sehe ja auch so, was passiert. [...]. Ich weiß es nicht, ob es am Ende des Tages so einen Mehrwert bringt.“

- Haltung gegenüber Evaluation: Skepsis gegenüber standardisierten Verfahren / quantitativen Indikatoren

  „An dem Tag haben die Kinder einfach null Bock gehabt. Da kann dann bei sowas auch nichts rauskommen. Da würde ich erst einmal verifizieren wollen, wer es ist und was wirklich dahintersteckt.“

- Evaluationskompetenz und –Erfahrung: Fehlende institutionalisierte Reflexionsmöglichkeiten

  „Es ist wichtiger als alles andere, dass wir in irgendeiner Form Nachweise liefern, dass das Programm wirkt.“

3. Herstellung von Akzeptanz und Glaubwürdigkeit (individuell)

- Professionelles (pädagogisches) Selbstverständnis (Terhart, 2013):
  - Anerkennung pädagogischer Expertise
  - Verknüpfung der Evaluation mit pädagogischen Zielen

- Haltung gegenüber Evaluation (Friedrich & Fischer, 2015):
  - Dialogische Klärung des Evaluationszwecks mit Bezug auf pädagogische Relevanz
  - Transparenz und Beteiligung bei Planung, Durchführung und Interpretation

- Evaluationskompetenz und –Erfahrung (Altrichter & Kemethofer, 2015):
  - Aufbau grundlegender Evaluationskompetenzen im pädagogischen Team
  - Sicherstellung von Nützlichkeit und Anwendung durch Rückbindung an pädagogische Ziele

VIELEN DANK FÜR IHRE AUFMERKSAMKEIT!

Literatur

Ackeren, I. V., Heinrich, M., & Thiel, F. (2013). *Evidenzbasierte Steuerung im Bildungssystem? Befunde aus dem BMBF-SteBis-Verbund*. Waxmann.

Altrichter, H., & Kemethofer, D. (2015). Neue Ansätze der Steuerung des Schulsystems und die Einstellung von Schulleitungen. *Bildung und Erziehung*, *68*(3), 291-310.

Baumert, J., Becker, M., Neumann, M., & Nikolova, R. (2009). Frühübergang in ein grundständiges Gymnasium: Übergang in ein privilegiertes Entwicklungsmilieu? Ein Vergleich von Regressionsanalyse und Propensity Score Matching. *Zeitschrift für Erziehungswissenschaft, 12*(2), 189–215.

Friedrich, V., & Fischer, A. (2015). *Wie nützt Evaluation (nicht)?* Zentrum für universitäre Weiterbildung.

Helmke, A. (2009). *Unterrichtsqualität und Lehrerprofessionalität. Diagnose, Evaluation und Verbesserung des Unterrichts*. Kallmeyer.

Lehmann, R., & Lenkeit, J. (2008). *Element. Erhebung zum Lese-und Mathematikverständnis. Entwicklung in den Jahrgangsstufen*, *4*. Humboldt Universität zu Berlin.

Oevermann, U. (2008). Profession contra Organisation? Strukturtheoretische Perspektiven zum Verhältnis von Organisation und Profession in der Schule. In W. Helsper, S. Busse, M. Hummrich & R. Kramer (Hrsg.), *Pädagogische Professionalität in Organisationen* (S. 55-77). VS Verlag für Sozialwissenschaften.

Terhart, E. (2013). *Erziehungswissenschaft und Lehrerbildung*. Waxmann.