

Messen, was in der Bildung zählt

Björn Rohles

24. März 2026

Bildung beeinflusst, wie wir denken, arbeiten und unseren Beitrag für die Gesellschaft leisten. Bildungseinrichtungen messen jedoch hauptsächlich, was einfach zu messen ist – nicht was zählt. Unsere Definition von Bildungsqualität bestimmt, worauf wir achten müssen und was wir riskieren zu übersehen. Deshalb sollten wir uns klar machen, was Qualität in der Bildung bedeutet und wie wir sie messen.



View online:

<https://rohles.net/de/artikel/messen-bildung>

Inhaltsverzeichnis

1	Was leicht zu messen ist	1
2	Worauf es ankommt: Qualitätsdimensionen	3
3	Evaluation von Bildungsqualität	5
3.1	Wahrgenommene und beobachtete Metriken	5
3.2	Zeitpunkt der Messung	6
3.3	Wer misst	8
3.4	Auf dem Weg zu einem Framework für die Messung von Bildungs- qualität	9
4	Fazit	10
5	Literatur	11

Evaluation ist wichtig für Management und Strategie, aber sie ist kein Selbstzweck: Sie muss mit ihrem strategischen Ziel übereinstimmen. In Bildungseinrichtungen ist das zentrale strategische Ziel, Bildungsqualität sicherzustellen. Aufbauend auf der Idee des **Menschenzentrierten Training-Management**¹ untersucht dieser Artikel die Messung von Bildungsqualität.

1 Was leicht zu messen ist

When a measure becomes a target, it ceases to be a good measure.

— Goodharts Gesetz, formuliert von Marilyn Strathern (1997)

In Bildungsorganisationen habe ich eine Reihe von Kennzahlen zur Definition von „Qualität“ beobachtet, zum Beispiel:

- Anmeldungen: wie viele Teilnehmer:innen sich eingeschrieben haben
- Raumauslastung: das Verhältnis von gebuchten Räumen zur Gesamtzahl der Räume

¹ <https://rohles.net/de/artikel/menschenzentriertes-training-management>

MENSCHENZENTRIERTER TRAINING-MANAGEMENT-LOOP

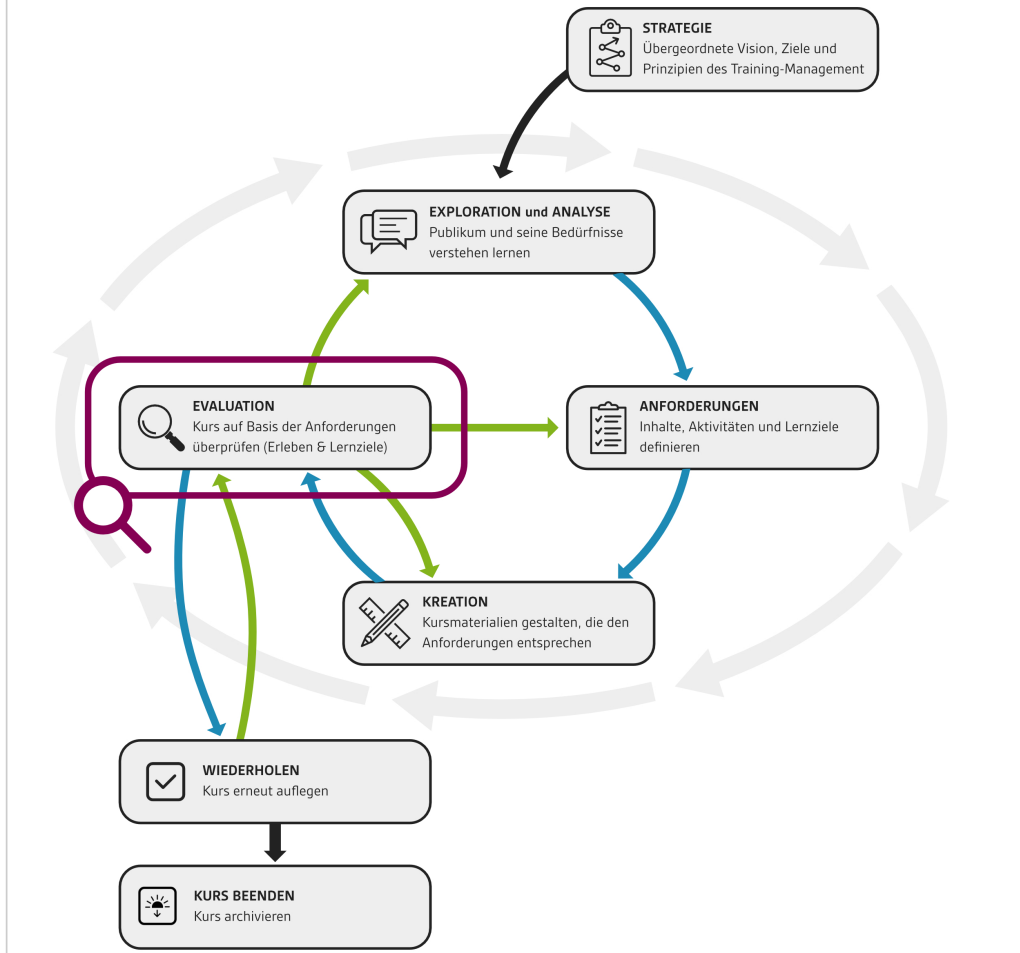


Abbildung 1: Evaluation im menschenzentrierten Training-Management-Loop

- Teilnahmequote: das Verhältnis von Teilnehmer:innen, die tatsächlich teilgenommen haben, zur Gesamtzahl der Angemeldeten
- Anzahl der Kurse oder Trainingsstunden: wie viele Kurse oder Trainingsstunden programmiert wurden

Diese Metriken würde ich als **Output-Variablen** bezeichnen. Sie sind einfach zu messen. Sie sind leicht verständlich. Und natürlich stehen sie zumindest irgendwie mit Qualität in Verbindung: Wir können im Training-Management kaum erfolgreich sein, wenn sich niemand anmeldet. Das Problem ist nur: Diese Metriken sind unvollständig. Es können sich viele Menschen anmelden, und wir können trotzdem kläglich scheitern. Output ist schlicht nicht dasselbe wie **Outcome**. Und wenn wir Outcomes messen wollen, müssen wir verstehen, worauf es in der Bildung wirklich ankommt.

2 Worauf es ankommt: Qualitätsdimensionen

However, not everything that can be counted counts, and not everything that counts can be counted.

— William Bruce Cameron (1966)

In der Bildung geht nicht um Anmeldezahlen, Raumauslastung oder Teilnahmequoten. Stattdessen geht es um **Bildungsqualität**². Bildungsqualität bedeutet letztlich, die Anforderungen eines Kurses zu erfüllen. Bildungsqualität zeigt sich in verschiedenen Dimensionen:

- **Funktionale Qualität** bezieht sich darauf, was Lernende erreichen müssen (die „Jobs-to-be-done“), etwa eine bestimmte Kompetenz entwickeln, spezifische Aufgaben ausführen oder an einem definierten Projekt arbeiten.
- **Hedonische Qualität** ist psychologisch und emotional. Sie schafft die notwendigen Voraussetzungen für Lernen und verwandelt Lernen in positive Erlebnisse.
- **Eudaimonische Qualität** handelt von Sinn – dem Gefühl, dass eine Lernerfahrung tief mit dem verbunden ist, was einer lernenden Person wirklich

² <https://rohles.net/de/artikel/qualitaet-menschzentriertes-training>

wichtig ist. Sie umfasst das wirksam zu sein, im Einklang mit den eigenen Werten zu handeln und zur besten Version seiner selbst zu werden.

- Die letzte Form von Qualität geht über die einzelne Person hinaus, weshalb ich sie „**Beyond-Self-Qualität**“ nennen möchte. Sie adressiert Auswirkungen auf die Gesellschaft, die Menschheit oder sogar alle Lebewesen – etwa Nachhaltigkeit, Gleichheit oder Demokratie.



Abbildung 2: Dimensionen von Bildungsqualität

Diese Qualitätsdimensionen zu kennen ist ein wichtiger erster Schritt im Menschenzentrierten Training-Management. Es schafft die Möglichkeit, Erwartungen von Lernenden für einen bestimmten Kurs oder eine andere Lernerfahrung zu definieren. Das Definieren allein reicht jedoch nicht: Irgendwann müssen wir überprüfen, ob ein Kurs diese Erwartungen tatsächlich erfüllt. Daher ist es notwendig, tiefer in die Evaluation dieser Qualitätsdimensionen einzutauchen.

3 Evaluation von Bildungsqualität

In meinem [Artikel zur Qualität im Human-Centered Training Management](#)³ habe ich einige mögliche Kriterien für die verschiedenen Qualitätsdimensionen identifiziert, bin aber nicht allzu tief darauf eingegangen, wie sie gemessen werden können. Drei Überlegungen sind für die Messung von Bildungsqualität besonders relevant.

3.1 Wahrgenommene und beobachtete Metriken

Erstens gibt es wahrgenommene und beobachtete Metriken:

- **Wahrgenommene Metriken** messen, was eine Person fühlt oder denkt. Wir sammeln diese Metriken, indem wir eine Person nach ihren Eindrücken fragen.
- **Beobachtete Metriken** messen, was von außen sichtbar oder testbar ist.

Subjektive Kriterien wie Zufriedenheit erfordern zwingend wahrgenommene Metriken. Andere Kriterien lassen jedoch verschiedene Optionen zu, etwa wahrgenommener Lernerfolg oder tatsächliche Testergebnisse. Wahrgenommene und beobachtete Metriken müssen nicht übereinstimmen: Man kann den subjektiven Eindruck haben, alles verstanden zu haben, und dennoch Fehler in einem Test machen. Andererseits ist die Evidenz stärker, wenn wahrgenommene und beobachtete Metriken übereinstimmen – etwa wenn wir eine hohe Anzahl von Interaktionen in einem Kurs beobachten und Lernende tatsächlich von einem starken Gefühl der Verbundenheit berichten.

Es gibt auch unterschiedliche Vorstellungen davon, was als valide Evidenz gilt. Manche betrachten beobachtete Metriken als „objektiver“ und bevorzugen daher eine numerische Zufriedenheitsbewertung wie 4,2 von 5 gegenüber anekdotischen Aussagen wie „Dieser Kurs hat mir geholfen, ein Projekt zu verwirklichen, das ich schon immer machen wollte“. Aber gerade eine solche Aussage ist für eudaimonische Bildungsqualität weitaus aussagekräftiger. Das macht eine Metrik nicht besser als die andere: Jede Qualitätsdimension baut auf einem anderen Verständnis davon auf, was Lernen bedeutet:

³ <https://rohles.net/de/artikel/qualitaet-menschzentriertes-training>

- **Funktionale Qualität** definiert Lernen letztlich als *Wissenserwerb*. Lernende haben eine Kompetenz oder nicht. Und ein Test oder Projekt kann dies objektiv zeigen.
- **Hedonische Qualität** definiert Lernen als *Erlebnis*. Aber ein Erlebnis ist per Definition subjektiv. Wir können zwar beobachtbare Hinweise haben, etwa eine nervös wirkende Person. Aber Erfahrungen müssen letztlich berichtet werden.
- **Eudaimonische Qualität** ist weit *transformativer* und *reflektiver*. Sie entfaltet sich im Laufe der Zeit. Wir können sie erst im Nachhinein oder nach einer tiefgehenden Reflexion erkennen.
- Funktionale, hedonische und eudaimonische Qualität beziehen sich alle auf die einzelne Person. **Beyond-Self-Qualität** per Definition nicht. Sie ist daher deutlich schwerer zu evaluieren. Anstatt im eigentlichen Sinne zu messen, müssen wir Fragen zur *Ethik* und zur *Rolle von Bildung* insgesamt stellen – und zu unseren Trainings im Besonderen, bezogen auf unsere *Werte* und das *Gemeinwohl*. Wie gehen wir zum Beispiel mit dem enormen Energieverbrauch der KI-Tools um, die wir in unseren Kursen besprechen? Oder wie sprechen wir über soziale Medien, die Menschen zweifellos verbinden und ihnen eine Stimme geben, aber auch Fake News und Ghost Work (Gray & Suri, 2019) hervorgebracht haben? Diese Fragen sind nicht einfach, aber sie sind notwendig.

Jede Qualitätsdimension impliziert also eine unterschiedliche Erkenntnistheorie der Evidenz. Und in der Praxis benötigen wir eine gute Balance von Metriken für diese Qualitätsdimensionen. Das Wissen über die verschiedenen Metriken ist jedoch nur ein erster Schritt hin zu einer soliden Messung. Wir müssen auch berücksichtigen, *wann* diese Messung stattfindet und *wer* sie durchführt.

3.2 Zeitpunkt der Messung

Zweitens machen der **Zeitpunkt** und die **Häufigkeit** der Messung einen Unterschied. Meiner Erfahrung nach messen die meisten Bildungseinrichtungen nur einmal – unmittelbar am *Ende* einer Lernerfahrung, etwa in einer Umfrage nach dem Kurs. Das ist nützlich, um Feedback zu sammeln, solange es noch frisch im Gedächtnis der Lernenden ist. Der richtige Zeitpunkt der Messung ist jedoch ebenso wichtig wie die richtige Metrik.

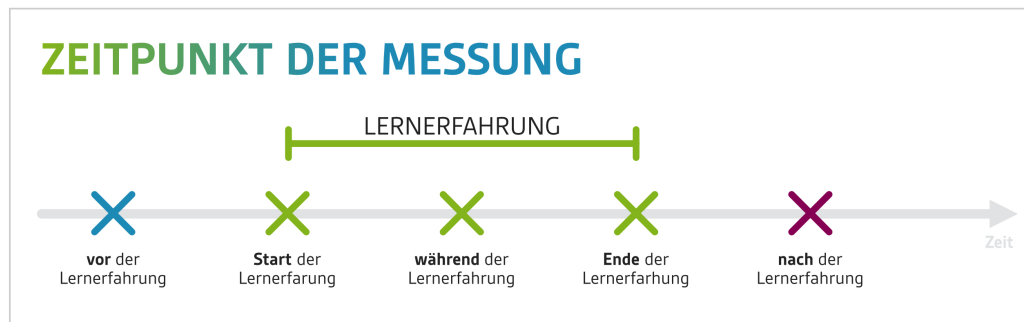


Abbildung 3: Verschiedene Zeitpunkte der Datenerhebung über Lernerfahrungen

Allgemeine Erkenntnisse über Lernende und ihre Bedürfnisse entstehen typischerweise *vor* einer Lernerfahrung, etwa um Lernziele bei der Definition von Trainingsanforderungen zu identifizieren.

Wenn Erwartungen von Lernenden oder Einblicke in ihr Vorwissen erfasst werden müssen, geschieht dies typischerweise direkt zu *Beginn* einer Lernerfahrung. Trainer:innen können diese Erkenntnisse nutzen, um Kursinhalt oder -tempo anzupassen.

Einige Metriken sind am nützlichsten *während* eines Trainings. Beispielsweise kann das Erfragen der mentalen Belastung während eines Kurses Trainer:innen helfen, den Kurs an die Bedürfnisse der Lernenden anzupassen: Sie können langsamer werden und komplizierte Inhalte erneut erklären oder schneller vorgehen, wenn der Inhalt leicht ist. Diese Messungen im Moment sind auch für Training-Manager:innen wichtig. Wenn wir beispielsweise wissen, dass die hedonische Qualität zu Beginn von Kursen tendenziell niedrig ist, wäre es sinnvoll, die erste Stunde eines Kurses zu beobachten, um herauszufinden, was vor sich geht. Kommen die Lernenden in einen Raum voller fremder Menschen und fühlen sich unwohl? Müssen sie sich in ein System mit Übungen einloggen und finden ihr Passwort nicht? Solche Daten und Beobachtungen helfen Training-Manager:innen, diese entscheidenden Momente systematisch zu verbessern.

Andere Metriken ergeben erst Sinn, wenn sie einige Zeit *nach* einer Lernerfahrung oder *wiederholt* über einen längeren Zeitraum gemessen werden – etwa der Transfer von Wissen in den eigenen Kontext oder die Arbeitsumgebung. Eudaimonische Qualität ist ein gutes Beispiel: Wir haben vielleicht während eines Kurses ein vages Gefühl, dass er uns hilft, unser Potenzial zu entfalten. Aber eudaimonische Qua-

lität bedeutet letztlich Transformation. Ob der Kurs das tatsächlich erreicht hat, wird erst im Nachhinein sichtbar – indem wir beobachten, wie wir uns danach entwickelt haben.

3.3 Wer misst

Drittens ist eine wichtige Frage, **wer** etwas misst. Wir neigen dazu, die einfache, leicht verfügbare Option zu wählen. Aber jede Entscheidung beeinflusst, welche Daten wir erheben und wie wir sie interpretieren:

- Lernende können sich selbst einschätzen, unterliegen aber verschiedenen Verzerrungen. Ein Beispiel ist der Dunning-Kruger-Effekt, eine kognitive Tendenz von Personen mit relativ geringer Kompetenz, ihre Fähigkeiten zu überschätzen.
- Trainer:innen können Teilnehmer:innen hauptsächlich im Moment beobachten, haben aber begrenzte Kapazität, auf spezifische individuelle Bedürfnisse einzugehen.
- Mentor:innen können Aussagen analysieren und Kompetenznachweise beobachten, etwa ein Portfolio. Das ermöglicht spezifische Begleitung, erfordert aber erheblichen Zeit- und Energieaufwand.
- Eine externe Person, etwa die Führungskraft einer Mitarbeiterin oder eines Mitarbeiters, die an einem Kurs teilgenommen haben, weiß möglicherweise wenig über das Training selbst, könnte aber eine Verhaltensänderung bemerken.
- Training-Manager:innen erheben systematisch Daten über Trainingsbedarfe, Lernende, Trainer:innen, Kurse und Zeit. Besonders im **Menschenzentrierten Training-Management**⁴ ist Datentriangulation entscheidend. Training-Manager:innen können nur durch Berücksichtigung einer großen Vielfalt dieser Perspektiven zu einer validen Evaluation gelangen.

Die Entscheidung, wer misst, ist auch mit Fragen von Macht und Vertrauen verbunden. Messen gibt der messenden Person Autorität. Wenn wir beispielsweise einen On-Demand-Kurs organisieren und eine HR-Manager:in die Kursqualität bewerten lassen, legen wir weniger Wert auf die Perspektive der eigentlichen Lernenden,

⁴ <https://rohles.net/de/artikel/menschenzentrierte-digitale-skills>

die möglicherweise sehr unterschiedliche Lernziele und Prioritäten haben – es sei denn, wir gleichen dies aus, indem wir auch mit den Lernenden sprechen. Machtfragen sind heikel, und es gibt kaum eine einzig richtige Antwort. Aber es ist wichtig zu verstehen, dass die Wahl nicht neutral ist und daher bewusst getroffen werden muss.

Vertrauen ist besonders wichtig für wahrgenommene Metriken. Lernende müssen beispielsweise das Gefühl psychologischer Sicherheit haben, um ihre ehrlichen Eindrücke zu berichten. Anonyme oder pseudonyme Datenerhebung kann sie dazu ermutigen.

3.4 Auf dem Weg zu einem Framework für die Messung von Bildungsqualität

In meinem früheren [Artikel zur Bildungsqualität](#)⁵ habe ich bereits einige mögliche Kriterien für funktionale, hedonische, eudaimonische und Beyond-Self-Qualität identifiziert. Diese gegen die drei hier untersuchten Evaluationsdimensionen zu mappen ist ein erster Schritt hin zu einem Framework für die Messung von Bildungsqualität.

- Als Teil der funktionalen Qualität bezieht sich **Effektivität** darauf, ob Menschen die erforderliche Kompetenz in einem Kurs erwerben. Wir könnten sie messen, indem wir Lernende bei der Ausführung von Aufgaben oder beim Beantworten von Fragen beobachten (beobachtete Metriken durch Trainer:innen), aber wir könnten auch fragen, ob sie sich in der neuen Kompetenz sicher fühlen (wahrgenommene Metrik, berichtet von den Lernenden). Die Messung kann während einer Lernerfahrung oder an ihrem Ende stattfinden.
- Effektivität erzählt jedoch nicht die ganze Geschichte. Kurzfristig, direkt nach dem Training, mag man effektiv sein; aber zwei Wochen später ist man nicht in der Lage, irgendetwas im realen Kontext anzuwenden. Hier kommt **Transfer** ins Spiel. Sie ist ebenfalls Teil der funktionalen Qualität, muss aber wiederholt über einen längeren Zeitraum bewertet werden – entweder durch die Lernenden selbst oder durch eine externe beobachtende Person

⁵ <https://rohles.net/de/artikel/qualitaet-menschzentriertes-training>

wie eine Führungskraft oder Mentor:in. Sowohl wahrgenommene als auch beobachtete Metriken könnten verwendet werden.

- **Stimulation** ist Teil der hedonischen Qualität und bezieht sich darauf, wie engagiert sich eine lernende Person fühlt. Sie wird typischerweise als wahrgenommene Metrik von den Lernenden während oder nach einer Lernerfahrung gemessen.
- Eudaimonische Qualität umfasst **Selbstverwirklichung**, also ob die Lernerfahrung Lernenden hilft, ihr volles Potenzial zu entfalten. Die Evaluation erfordert wiederholte, langfristige Reflexionen durch eine lernende Person (wahrgenommene Metrik) und potenziell eine Mentor:in.
- **Gesellschaftliche Wirkung** ist Teil der Beyond-Self-Qualität – etwa ob eine Lernerfahrung Demokratie unterstützt. Das geht über klassische Messung hinaus, aber ethische Reflexionen aller Beteiligten können dennoch helfen, sie fortlaufend zu bewerten.

Die folgende Tabelle fasst diese Beispiele zusammen.

Qualitätsdimension	Kriterium	wahrgenommen		Wer
		/ beobachtet	Zeitpunkt	
funktional	Effektivität	beides	während oder danach	Trainer:in, Lernende:r
funktional	Transfer	beides	langfristig	Lernende:r, extern
hedonisch	Stimulation	wahrgenommen	während oder danach	Lernende:r
eudaimonisch	Selbstverwirklichung	wahrgenommen	langfristig	Lernende:r, Mentor:in
Beyond-Self	gesellschaftliche Wirkung	ethische Reflexion	fortlaufend	alle

4 Fazit

Da Messung für Organisationen wichtig ist, greifen Bildungsanbieter:innen zu oft auf einfache Kennzahlen wie Anmeldungen und Raumauslastung als Indikatoren ihres „Erfolgs“ zurück. Diese einfachen Metriken können sogar zu Zielen werden,

die es zu optimieren gilt. Das ist Goodharts Gesetz in der Praxis. Und es bedeutet, zu ignorieren, dass es bei Bildungsqualität letztlich um etwas anderes geht: Sie hat funktionale („Jobs-to-be-done“), hedonische (wie Bildung sich anfühlt), eudaimonische (wie Bildung eine Person zu einem besseren Selbst macht) und Beyond-Self-Dimensionen (wie Bildung zur Gesellschaft, Menschheit und Lebewesen als Ganzes beiträgt). Bildungsanbieter:innen brauchen eine gute Balance von Metriken, um diese Dimensionen zu bewerten. Deren Definition erfordert bewusste und informierte Entscheidungen darüber, *was wann* und durch *wen* gemessen wird. Gleichzeitig müssen wir uns bewusst sein, dass nicht alles messbar ist – aber nur weil wir etwas nicht messen können, bedeutet das nicht, dass es nicht zählt.

5 Literatur

Cameron, W. B. (1966). *Informal sociology: A casual introduction to sociological thinking* (Bd. 21). Random House.

Gray, M. L., & Suri, S. (2019). *Ghost work: how to stop Silicon Valley from building a new global underclass*. Houghton Mifflin Harcourt.

Strathern, M. (1997). 'Improving ratings': Audit in the British University system. *European Review*, 5(3), 305–321. [https://doi.org/10.1002/\(SICI\)1234-981X\(199707\)5:3%253C305::AID-EURO184%253E3.0.CO;2-4](https://doi.org/10.1002/(SICI)1234-981X(199707)5:3%253C305::AID-EURO184%253E3.0.CO;2-4)