

MANAGEMENT REVIEW

Management ohne Theater.

BETRIEB

DIE KPI-FALLE: WENN DIE ZAHLEN BESSER WERDEN, ABER DIE ARBEIT SCHLECHTER

Sechs Kapitel über Ziele, verzerrte Anreize, verborgene Zielkonflikte, statistische Streuung und die Kennzahlentafel, die den Kunden schützt.



2,3 %

kann ein Geschäftsziel von 1 % verfehlen, ohne eine hypothetische statistische obere Eingriffsgrenze zu überschreiten. Ein Signal ist kein Urteil.

P-KARTEN-MODELL, ERFUNDENES BEISPIEL

IN DIESEM BAND

TITELTHEMA

Die grüne Zahl, das gebrochene Versprechen

SEITE 05

DAS MODELL

Ziel gegen Eingriffsgrenze

SEITE 19

WERKZEUG VOR ORT

Das ehrliche KPI-Blatt

SEITE 28

EDITORIAL · BAND 6

Eine grüne Zahl kann einen schlechten Tag verbergen

Messung ist notwendig. Das Problem beginnt, wenn die Zahl leichter zu verteidigen ist als der Zweck, dem sie dienen sollte.

Betriebsleiter haben nicht den Luxus, nur nach Instinkt zu führen. Eine Abfahrt ist pünktlich oder verspätet. Eine Ladung ist vollständig oder nicht. Ein Zimmer ist bereit oder nicht. Jede Schicht erzeugt Fakten, die eine Aufzeichnung verdienen. Doch diese Aufzeichnungen werden gefährlich, wenn ein Indikator, der die Arbeit sichtbar machen soll, als vollständiges Urteil über Menschen und Service behandelt wird. Der Autor hat mit Last-Mile-Kennzahlen wie Verspätung, Schaden, Unvollständigkeit und Ladezeit gearbeitet und mit Berichten und Übergaben in der Hotellerie. Diese Kontexte erklären, warum ein KPI zugleich nützlich und unvollständig sein kann. Dieser sechste Band behauptet nicht, ein bestimmter Arbeitgeber habe ein Ziel manipuliert, einen Mangel verborgen oder Personal wegen eines irreführenden Dashboards bestraft. Seine Betriebsszenen sind konstruierte Illustrationen, klar gekennzeichnet, wo ein Ereignis oder eine Messung erfunden ist. Wir beginnen mit einer trügerisch positiven Zahl. Campbell und Goodhart erklären, warum Ziele unter hohem Druck die Kennzahlen schwächen können, die die Führung leiten sollen. Holmström und Milgrom zeigen das Anreizproblem, wenn manche Aufgaben messbar sind und andere weniger sichtbar. Bevan und Hood liefern einen dokumentierten Fall von Manipulationsrisiko im öffentlichen Dienst. Die statistische Prozesskontrolle unterscheidet ein Managementziel von einem Beleg für einen veränderten Prozess. Die letzten Kapitel fügen die Kennzahlen zu einer ausgewogenen, prüfbaren Entscheidung zusammen.

Stiven Janaqi HERAUSGEBER · STIVEN CATALYST

SO IST JEDES KAPITEL GEBAUT

- 1 Die lebendige Szene**
Eine Schicht, eine Rampe, eine Rezeption um 03:00: ein Stück Arbeit, verfolgt, wie es wirklich geschieht, mit der Übergabe, an der es bricht. Szenen sind aus echten Arbeitskategorien gebaut und als solche gekennzeichnet.
- 2 Die begriffliche und wissenschaftliche Aufschlüsselung**
Der Mechanismus hinter der Szene, benannt mit dem Modell, das ihn erklärt, und die Forschung, die ihn prüft: die Studie, ihre Population, ihr Jahr und ihre Grenzen, bevor sie anderswo angewandt wird.
- 3 Das operative Handlungsprotokoll**
Was ein Team am Montag ausprobieren kann: Felder, Fragen, einen Takt und eine Zählung, jeweils mit Verantwortlichem, nächstem Schritt und Rückmeldung an die Menschen, die die Arbeit tun.

DER LESEVERTRAG

Feld

Echte Kategorien betrieblicher Arbeit, keine erfundenen persönlichen Vorfälle.

Belege

Forschungsumfeld, Datum, Methode und Grenzen gemeinsam genannt.

Handeln

Eine KPI-Entscheidung muss eine Kunden- und Qualitätsprüfung überstehen.

INHALT

Die Kapitel,
in drei Ebenen

KAPITEL I Die grüne Zahl, das gebrochene Versprechen 05

1 SZENE	2 AUFSCHLÜSSELUNG	3 PROTOKOLL
Die grüne Zahl, das gebrochene Versprechen 05	Die Kennzahl ist nicht die Mission 06	Das Grün lesen, bevor Sie feiern 07

KAPITEL II Wenn eine Kennzahl zum Ziel wird 08

1 SZENE	2 AUFSCHLÜSSELUNG	3 PROTOKOLL
Das Ziel verändert den Raum 08	Goodhart ist eine Warnung, kein Schicksalsgesetz Was gemessen wird, bekommt den Einsatz Wenn Messung zum Spiel wird 10	Das Ziel von der Kennzahl fernhalten 12

KAPITEL III Die verborgenen Kosten eines KPI 13

1 SZENE	2 AUFSCHLÜSSELUNG	3 PROTOKOLL
Der schnelle Prozess und die Mehrarbeit 13	Ein ausgewogener Satz ist kein überfülltes Dashboard Eine Kennzahl wird besser. Eine andere ändert sich. 15	Eine Leitplanke neben das Ziel 16

KAPITEL IV Die rote Zahl und die falsche Schuld 17

1 SZENE	2 AUFSCHLÜSSELUNG	3 PROTOKOLL
Der Tag, an dem das Dashboard rot wird 17	Ein Geschäftsziel ist keine Eingriffsgrenze Der rote Punkt innerhalb des erwarteten Bereichs 19	Das Rot untersuchen, ohne einen Schuldigen zu erfinden 20

KAPITEL V Die Kennzahlentafel, die die Wahrheit sagt 21

1 SZENE	2 AUFSCHLÜSSELUNG	3 PROTOKOLL
Ein Durchschnitt, zwei verschiedene Tage 21	Wie beide Gruppen besser werden, während die Summe fällt Eine Kennzahlentafel mit Gewissen 23	Den KPI definieren, bevor Sie Erfolg definieren 24

KAPITEL VI Jenseits des Ziels 25

1 SZENE	2 AUFSCHLÜSSELUNG	3 PROTOKOLL
Die Besprechung, die eine Frage mehr stellt 25	Fünf Fragen vor einem Urteil 26	Jenseits des Ziels: eine Betriebsprüfung 27

AUSSERDEM IN DIESEM BAND

04 FORSCHUNGSNOTIZ
Was eine KPI-Studie sagen darf

28 WERKZEUG VOR ORT
Das ehrliche KPI-Blatt

29 SCHLUSSSESSAY
Verantwortlich für den Zweck

30 METHODE
Studien, Modelle und Methoden

31 QUELLEN
Quellen und Methode

FORSCHUNGSNOTIZ

Was eine KPI-Studie sagen darf

Eine Korrelation zwischen einem Ziel und einem Verhalten beweist nicht, dass jedes Ziel Manipulation verursacht.

Dieser Band verbindet mehrere Arten von Belegen: theoretische Arbeiten dazu, wie Kontrolldruck oder ungleiche Messung Verhalten verändern können, beobachtende Forschung im öffentlichen Dienst zu dokumentierten Risiken in einem institutionellen Umfeld, statistische Qualitätskontrolle zur Unterscheidung gewöhnlicher Streuung von ungewöhnlichen Signalen und Verbesserungsanleitung zu praktischen Kennzahlenfamilien. Sie liefern keinen universellen Prozentsatz für KPI-Manipulation, belegen kein gefälschtes Ergebnis von irgendjemandem in der Arbeit des Autors und zeigen nicht, dass ein Dashboard mit drei Kennzahlen schädliche Anreize immer verhindert. Sie rechtfertigen einen disziplinierten Satz von Fragen für die lokale Prüfung.

01**Beobachtet**

Befund mit Herausgeber, Population, Datum und Methode.

02**Modelliert**

Mathematisches Beispiel mit Annahmen und nachvollziehbaren Eingaben.

03**Vorgeschlagen**

Redaktionelle Kontrolle, Verantwortlicher und Test, die auf lokale Belege warten.

Eine grüne Zahl kann zugleich mathematisch richtig und betrieblich unzureichend sein, wenn die Definition eine relevante Kundenfolge ausschließt. Ebenso kann eine rote Zahl eine berechtigte Warnung sein, ohne Fehlverhalten oder eine ungewöhnliche Prozessverschiebung zu beweisen. Der Leser sollte fragen, was Zähler und Nenner bedeuten, wer ausgeschlossen ist, wann Ereignisse gezählt werden und was außerhalb der sichtbaren Grenze geschieht.

UNSERE LESART

Dies ist ein belegbasierter operativer Traktat, keine statistische Studie der Betriebe des Autors.

Quellen: Donald T. Campbell, Evaluation and Program Planning, 1979 (über ein Dartmouth Occasional Paper (1976), nachgedruckt im Journal of MultiDisciplinary Evaluation (2011)); Bengt Holmström & Paul Milgrom, Journal of Law, Economics, & Organization 7 (special issue), 1991; Gwyn Bevan & Christopher Hood, Public Administration, 2006; NIST & SEMATECH, 2026; Institute for Healthcare Improvement, 2026

Campbell (1979), Holmström und Milgrom (1991), Bevan und Hood (2006), das NIST/SEMATECH e-Handbook und das IHI; alle numerischen Betriebsfälle sind erfunden.

1 SZENE

2 AUFSCHLÜSSELUNG

3 PROTOKOLL

KAPITEL I · DIE GRÜNE ZAHL, DAS GEBROCHENE VERSPRECHEN

Die grüne Zahl, das gebrochene Versprechen

Das Dashboard ist zufrieden. Die Person, die die Arbeit erhält, nicht.

HYPOTHETISCHE LAST-MILE-SZENE, KEIN DOKUMENTIERTER VORFALL

An einer konstruierten Ladestation sieht eine Führungskraft, dass die Abfahrten schneller erfolgen als in der Vorwoche. Der Indikator Ladezeit ist grün. Das Team hat das Ergebnis auf seinem Bildschirm verdient: Das erfasste Intervall vom definierten Start bis zum definierten Ende ist kürzer. Später taucht ein Zustellproblem auf. Ein Teil eines Auftrags hat die Person, die ihn erwartete, nicht erreicht. Die Abfahrtsstatistik bleibt unverändert, weil der Mangel zu einer anderen Kennzahl gehört. Nichts in dieser Szene beweist, dass die Ladegeschwindigkeit den fehlenden Artikel verursacht hat. Der Fehlbestand kann in Vorbereitung, Erfassung, Sortierung, Beladung oder einer späteren Übergabe begonnen haben. Die Versuchung der Führung ist die einfache Geschichte: Eine Kennzahl wurde besser, und eine andere Person machte einen Fehler. Genau diese Geschichte muss geprüft werden, statt verkündet.

Die dokumentierte Last-Mile-Arbeit des Autors umfasste Ladezeit, Unvollständigkeit, fehlende Taschen und gemischte Wagen als Kategorien betrieblicher Aufmerksamkeit. Das Ereignis, seine Abfolge und sein Ausgang sind hier erfunden. Wir nutzen sie nur, um zu fragen, was eine Kennzahl bescheinigen kann und was nicht.

UNSERE LESART

Eine Kennzahl kann innerhalb ihrer Grenze stimmen und trotzdem zu wenig über das Erlebnis des Kunden sagen.

Redaktionelle Zusammensetzung echter betrieblicher Kategorien; begrifflicher Hintergrund in den Ergebnis-, Prozess- und Ausgleichskennzahlen des IHI. Der unten genannte Essay liest den Bericht einer Maschine genauso.

[Mehr im Essay](#)[Der Bericht, den eine Maschine geschrieben hat](#)

KAPITEL I · DIE GRÜNE ZAHL, DAS GEBROCHENE VERSPRECHEN

Die Kennzahl ist nicht die Mission

Jeder Indikator hat eine Definition; jeder Service hat einen größeren Zweck. Eine Ladezeitkennzahl betrifft ein Intervall eines Prozesses. Eine Kennzahl für unvollständige Zustellung betrifft eine andere Fehlerart, und ihr Nenner kann Aufträge, Taschen oder eine andere Einheit zählen. Beide können aussagekräftig sein, ohne austauschbar zu sein. Ein Prozentsatz mit unklarem Nenner ist noch kein entscheidungsreifer Beleg.

Das Institute for Healthcare Improvement unterscheidet Ergebnis-, Prozess- und Ausgleichskennzahlen. Das Ergebnis fragt, ob der Service die beabsichtigte Wirkung erzielt. Der Prozess fragt, ob die kritischen Schritte funktionieren. Der Ausgleich fragt, ob die Maßnahme anderswo neue Probleme schafft. Das ist auch außerhalb des Gesundheitswesens eine nützliche Gestaltungslinse, doch die tatsächlichen Kennzahlen und Annahmekriterien müssen im lokalen Betrieb definiert werden.

01	Prozess	Ladeintervall, Scan-Vollständigkeit oder eine definierte pünktliche Übergabe.
02	Ergebnis	Vollständiger, richtiger und pünktlicher Service für den Empfänger.
03	Ausgleich	Fehlende Artikel, Nacharbeit, Sicherheitsabweichungen oder neues Warten weiter unten.

Es ist verlockend, diese Zahlen zu einer einzigen Note zu verschmelzen. Doch ein gewichteter Durchschnitt kann ein schweres Ergebnis verbergen, wenn Gewichte, Vetobedingungen und Kundenprioritäten nicht ausdrücklich sind. Manche Fehler dürfen nie gegen schnelleren Durchsatz eingetauscht werden. Die Definition eines annehmbaren Ergebnisses kommt vor dem Entwurf des Dashboards.

UNSERE LESART

Fragen Sie, was sich für den Kunden geändert hat, bevor Sie ein internes Ziel für gewonnen erklären.

Quelle: Institute for Healthcare Improvement, 2026

IHI, Establishing Measures; die Übertragung auf die Logistik ist redaktionell, an keinem benannten Standort validiert.

1 SZENE

2 AUFSCHLÜSSELUNG

3 PROTOKOLL

KAPITEL I · DIE GRÜNE ZAHL, DAS GEBROCHENE VERSPRECHEN

Das Grün lesen, bevor Sie feiern

Eine Prüfung in fünf Schritten, die einen günstigen KPI am Service von Ende zu Ende misst.

1

Die Grenze des Indikators benennen

Betrieblicher Start und Ende, Ausschlüsse, Zeitfenster, Zähler, Nenner und Datenquelle.

2

Die zugrunde liegende Arbeit prüfen

Sind die als gut gezählten Ereignisse weiter unten vollständig und nutzbar? Anonymisierte Fälle stichprobenartig prüfen, wo erlaubt.

3

Das Kundenergebnis danebenlegen

Das Serviceergebnis und eine Ausgleichskennzahl neben dem verbesserten Prozess-KPI lesen.

4

Plausible Alternativen prüfen

Fallmix, Volumen, Stichtagsregeln, Personal und Ausnahmetage, bevor eine Ursache zugewiesen wird.

5

Die Entscheidung dokumentieren

Was bestätigt wurde, was noch unbekannt ist, wer es prüft und wann das Team Rückmeldung erhält.

Verbessert sich die Ladezeit, lesen Sie auch Vollständigkeit, pünktlichen Service und Sicherheitsaufzeichnungen. Zwei gemeinsam bewegte Zahlen beweisen keine Ursache; prüfen Sie, ob sie Zeitfenster und Einheiten teilen, denn eine Prozesszahl vom selben Tag und ein Kundenergebnis vom nächsten Tag müssen womöglich abgeglichen werden.

UNSERE LESART

Eine grüne Zahl sollte eine kurze Prüfung beginnen, nicht das Gespräch beenden.

Quelle: Institute for Healthcare Improvement, 2026

Redaktionelles Betriebsprotokoll nach der Messanleitung des IHI; keine Behauptung einer förmlichen Umsetzung bei einem Arbeitgeber.

1 SZENE

2 AUFSCHLÜSSELUNG

3 PROTOKOLL

KAPITEL II · WENN EINE KENNZAHL ZUM ZIEL WIRD

Das Ziel verändert den Raum

Ein Messprogramm kann genau die Tätigkeit verändern, die es beobachten soll.

KONSTRUIERTE TEAMBESPRECHUNG, KEINE AUFGEZEICHNETE DISKUSSION

In einer fiktiven Betriebsbesprechung kündigt eine Führungskraft an, dass in der Wochensitzung nur noch eine Zahl gezeigt wird: die Quote pünktlicher Abfahrten. Die Leute richten ihre Aufmerksamkeit natürlich auf das erfasste Abfahrtsereignis. Fragen, die früher zur Genauigkeit der Vorbereitung, zu verspäteter Information und zur Erholung nach einem späten Start gestellt wurden, bekommen weniger Zeit. Die Zahl bleibt nützlich, aber die Sitzung ist enger geworden. Für diese Verschiebung ist keine unehrliche Handlung nötig. Menschen reagieren auf das, was Aufmerksamkeit, Lob und Konsequenzen erhält. Das Personal tut womöglich genau die Arbeit, die die formale Kennzahl fördert, und lässt weniger gemessene Pflichten liegen. Das Risiko ist kein Rechenfehler; es ist ein Fehler im Entwurf des Anreizsystems.

Das ist kein Vorwurf gegen einen bestimmten Arbeitgeber oder eine Person. Es geht darum, ein dokumentiertes theoretisches Anliegen einzuführen: Wenn wir einen diagnostischen Indikator in ein starkes Kontrollziel verwandeln, kann sich sein Verhältnis zur Arbeit ändern.

UNSERE LESART

Je stärker die Konsequenz an einem Indikator hängt, desto sorgfältiger müssen seine blinden Flecken geprüft werden.

Konstruiertes Szenario; Theorie bei Goodhart (1975), Campbell (1979) und Holmström und Milgrom (1991). Der unten genannte Essay fragt, welchen KPIs ein Team vertraut.

[Mehr im Essay](#)

Kennzahlen, denen das Team vertraut

1 SZENE

2 AUFSCHLÜSSELUNG

3 PROTOKOLL

KAPITEL II · WENN EINE KENNZAHL ZUM ZIEL WIRD

Goodhart ist eine Warnung, kein Schicksalsgesetz

Der bekannte Slogan ist eine spätere Verallgemeinerung einer früheren Beobachtung über statistische Regelmäßigkeiten.

Charles Goodhart beobachtete in der geldpolitischen Analyse, dass eine zur Steuerung genutzte Regelmäßigkeit aufhören kann, sich so zu verhalten wie vor der Steuerung. Eine populäre Paraphrase, verbunden mit Marilyn Stratherns Diskussion der Auditkultur, sagt, eine Kennzahl könne aufhören, eine gute Kennzahl zu sein, wenn sie zum Ziel wird. Das sind verwandte Formulierungen, keine zwei Zitate aus demselben Originaltext. Keine davon bedeutet, dass jedes Ziel scheitern muss. In diesem Band dienen sie dazu, die Annahme zu hinterfragen, eine Kennzahl bleibe neutral, nachdem Belohnungen und Strafen daran hängen. Donald Campbells Arbeit von 1979 zu geplantem sozialem Wandel erinnert Evaluatoren zudem daran, dass beobachtete Indikatoren nicht zeigen müssen, ob eine Intervention die beabsichtigte Verbesserung wirklich verursacht hat. In komplexen Systemen können veränderte Bedingungen und andere Eingriffe dieselbe Kennzahl bewegen. Ein guter Trend kann echt sein, ein Artefakt geänderter Definitionen oder die Wirkung anderer Einflüsse.

01**Stellvertreter**

Die Zahl steht nur für einige Eigenschaften des Ergebnisses.

02**Druck**

Sichtbarkeit und Konsequenzen verändern Anreize und Aufmerksamkeit.

03**Gültigkeit**

Erneut prüfen, ob der Stellvertreter das Ergebnis noch beschreibt.

UNSERE LESART

Das Problem ist nicht, ein Ziel zu nutzen. Es ist zu vergessen, dass das Ziel den Kontext verändert, in dem die Kennzahl entsteht.

Quellen: Charles Goodhart, 1975 (über Bevan & Hood (2006), die den Nachdruck von 1984 zitieren); Marilyn Strathern, European Review, 1997; Donald T. Campbell, Evaluation and Program Planning, 1979 (über ein Dartmouth Occasional Paper (1976), nachgedruckt im Journal of MultiDisciplinary Evaluation (2011))

Goodhart (1975), zitiert nach Bevan und Hood (2006); Strathern (1997) für die Paraphrase; Campbell (1979). Die drei Karten sind redaktionell.

KAPITEL II · WENN EINE KENNZAHL ZUM ZIEL WIRD

Was gemessen wird, bekommt den Einsatz

Die Anreiztheorie mit mehreren Aufgaben erklärt, warum enge Kennzahlen schwer quantifizierbare Pflichten verdrängen können. Holmström und Milgroms Prinzipal-Agent-Analyse von 1991 untersucht, wie die Gestaltung von Belohnungen die Verteilung des Einsatzes auf verschiedene Pflichten beeinflusst. In einem Job mit mehreren Verantwortungen können starke Anreize, die nur an die messbaren geknüpft sind, die Aufmerksamkeit von den anderen ablenken. Das ist ein theoretischer Mechanismus, keine empirische Messung der Kollegen des Autors.

Ein Disponent tut mehr, als ein Dashboard zu beobachten: Er koordiniert, prüft Ausnahmen und kommuniziert. Ein Nachtauditor tut mehr, als einen Bericht zu erzeugen: Offene Arbeit muss der nächsten Schicht richtig übergeben werden. Ein Großhandels-Kundenbetreuer tut mehr, als Aufträge anzunehmen: Verfügbarkeit und Zusagen müssen glaubwürdig bleiben. In jedem Umfeld lässt sich wertvolle Tätigkeit schwerer in einen sauberen Tagesprozentsatz fassen.

- | | | |
|-----------|---|---|
| 01 | Letzte Meile: pünktliche Abfahrten sind leicht zu zählen | Nachgelagerte Vollständigkeit ist leicht zu übersehen. |
| 02 | Hotelübergabe: der Bericht wurde gesendet | Ob die nächste Schicht handeln kann, ist leicht zu übersehen. |
| 03 | Großhandel: angenommene Aufträge | Die Einlösung des Versprechens ist leicht zu übersehen. |

Die Zeilen illustrieren redaktionell, wie Aufgabendefinitionen sich unterscheiden können, und behaupten nicht, eine der drei Organisationen habe Anreize auf diese Kennzahlen gezahlt. Eine vernünftige Kennzahlentafel sollte die schwerer sichtbaren Pflichten einschließen, durch Stichproben oder qualitative Prüfung, wo numerische Stellvertreter nicht ausreichen.

UNSERE LESART

Wenn Einsatz mehreren Zwecken dient, beschreibt eine einzige belohnte Zahl selten alle.

Quelle: Bengt Holmström & Paul Milgrom, Journal of Law, Economics, & Organization 7 (special issue), 1991

Holmström und Milgrom, Journal of Law, Economics, & Organization 7 (Sonderheft), 1991; die Arbeitsbeispiele sind deuted.

KAPITEL II · WENN EINE KENNZAHL ZUM ZIEL WIRD

Wenn Messung zum Spiel wird

Veröffentlichte Belege aus öffentlichen Diensten zeigen Risiken, ohne identisches Verhalten in privaten Betrieben zu beweisen.

Gwyn Bevan und Christopher Hood untersuchten die zielbasierte Steuerung im englischen öffentlichen Gesundheitssystem und beschrieben berichteten Erfolg neben Messproblemen und Manipulationsrisiko. Ihre Arbeit von 2006 betrachtet ein bestimmtes politisches Umfeld mit öffentlichen Zielen von hoher Tragweite und Rechenschaftsregeln. Der Kontext zählt: Krankenhausdienste, politische Steuerung und öffentliche Berichterstattung sind nicht dasselbe wie eine privat betriebene Laderampe oder eine Hotelschicht. Sandra van Thiel und Frans Leeuw beschrieben das Leistungsparadox im öffentlichen Sektor: Wachsende Mengen an Leistungsinformation können mit abnehmender Einsicht in die tatsächliche Leistung, Tunnelblick oder Analyselähmung einhergehen. Auch das ist kein Beleg, dass der Autor Manipulation erlebt hätte. Es ist ein Grund zu prüfen, was der gewählte Indikator ausschließt und wie die Organisation darauf reagiert.

Die praktische Antwort ist nicht, Einzelne zu beschuldigen, wann immer eine Zahl ungewöhnlich gut aussieht. Prüfen Sie die Definition der Kennzahl, Änderungen an Zulässigkeit oder Zeitpunkt, die Qualität der Belegdaten, die an der Zahl hängenden Anreize und die unabhängigen Kundenergebnisse.

UNSERE LESART

Eine überraschende Zahl ist eine Frage für die Prüfung, keine Anklage gegen das Team.

Quellen: Gwyn Bevan & Christopher Hood, Public Administration, 2006; Sandra van Thiel & Frans L. Leeuw, Public Performance & Management Review 25(3), 2002

Bevan und Hood, Public Administration 84(3), 2006; van Thiel und Leeuw, Public Performance & Management Review 25(3), 2002.

1 SZENE

2 AUFSCHLÜSSELUNG

3 PROTOKOLL

KAPITEL II · WENN EINE KENNZAHL ZUM ZIEL WIRD

Das Ziel von der Kennzahl fernhalten

Fünf Schutzmaßnahmen, bevor ein Indikator an Lob, Rangfolge oder Konsequenzen geknüpft wird.

1

Eine diagnostische Kopie behalten

Die Kennzahl zum Lernen bleibt getrennt von der Zahl zum Urteilen.

2

Benennen, was der Stellvertreter nicht sieht

Die Pflichten ohne Zahl, und wie sie stichprobenartig geprüft werden.

3

Eine Gegenkennzahl hinzufügen

Ein Ausgleichsindikator, der sich verschlechtern würde, wenn das Ziel zu eng verfolgt wird.

4

Die Definition einfrieren

Zulässigkeit, Zeitpunkt und Ausschlüsse versioniert, vor dem ersten Vergleich.

5

Den Anreiz prüfen, nicht nur die Zahl

Regelmäßig fragen, was das Ziel inzwischen belohnt und was es nicht mehr bemerkt.

Eine ungewöhnlich gute Zahl verdient dieselbe Prüfung wie eine schlechte: die Definition, die Aufzeichnungen dahinter, Änderungen an Zulässigkeit oder Zeitpunkt und die unabhängigen Kundenergebnisse. Nichts davon unterstellt Fehlverhalten; es hält die Kennzahl vertrauenswürdig.

UNSERE LESART

Ein Ziel ist nur so lange sicher, wie noch jemand prüft, was es nicht mehr misst.

Quellen: Charles Goodhart, 1975 (über Bevan & Hood (2006), die den Nachdruck von 1984 zitieren); Donald T. Campbell, Evaluation and Program Planning, 1979 (über ein Dartmouth Occasional Paper (1976), nachgedruckt im Journal of MultiDisciplinary Evaluation (2011)); Bengt Holmström & Paul Milgrom, Journal of Law, Economics, & Organization 7 (special issue), 1991; Gwyn Bevan & Christopher Hood, Public Administration, 2006

Redaktionelles Protokoll nach Goodhart, Campbell, Holmström und Milgrom sowie Bevan und Hood; das PDF des Autors fasst diese Schutzmaßnahmen in seiner Forschungsseite zusammen.

1 SZENE

2 AUFSCHLÜSSELUNG

3 PROTOKOLL

KAPITEL III · DIE VERBORGENEN KOSTEN EINES KPI

Der schnelle Prozess und die Mehrarbeit

Lokale Geschwindigkeit kann sich verbessern, während eine benachbarte Aufgabe zusätzlichen Aufwand erbt.

HYPOTHETISCHER SERVICEABLAUF, KEIN GEMESSENES ARBEITGEBERERGEBNIS

Stellen Sie sich ein Team vor, das die Zeit zum Abschluss einer eingehenden Anfrage verkürzt. Das Dashboard zeigt mehr Anfragen, die vor der Frist als bearbeitet markiert sind. Ein Kollege weiter unten stellt fest, dass manche Anfragen nun ohne die Information ankommen, die zur Erledigung des Service nötig ist. Das erste Team erledigt den gemessenen Teil womöglich schneller, während das empfangende Team mehr Zeit damit verbringt, unvollständige Übergaben zu klären. Weder der Übergabemangel noch seine Ursache sind durch diese kurze Geschichte belegt. Eine ordentliche Untersuchung würde eine Stichprobe von Datensätzen vor und nach der Änderung vergleichen, prüfen, ob dieselben Anfragearten gezählt wurden, und ein echtes Prozessproblem von einem Zufall unterscheiden. Das Team braucht ein Maß für Arbeit, die im nächsten Schritt nutzbar ist, nicht nur das Datum, an dem der erste Schritt als fertig markiert wurde.

Dieses Szenario gilt im Prinzip für Aufträge, Schichtübergaben und Supportanfragen. Es ist bewusst keine Tatsachenbehauptung über bestimmte Gästebeschwerden, Transaktionen, Routenanpassungen oder eine dokumentierte Prozessmaßnahme.

UNSERE LESART

Schnellerer Abschluss am ersten Schalter ist nicht unbedingt schnellere Lieferung an den Kunden.

Hypothetischer Vergleich; begriffliche Quellen die Ausgleichskennzahlen des IHI und Kaplan und Norton (1992). Der unten genannte Essay verfolgt einen Teil eines Auftrags, der verloren ging.

[Mehr im Essay](#)

Wo ein Teil der Bestellung verloren geht

KAPITEL III · DIE VERBORGENEN KOSTEN EINES KPI

Ein ausgewogener Satz ist kein überfülltes Dashboard

Es geht darum, den Zweck der Arbeit mit ergänzenden Kennzahlen zu schützen, nicht darum, alles zu messen. Kaplan und Norton führten 1992 die Balanced Scorecard ein, um den Blick über traditionelle Finanzkennzahlen hinaus zu weiten und die Perspektiven von Kunden, internen Prozessen und organisatorischer Fähigkeit zu verbinden. Sie ist ein Ansatz der strategischen Führung, keine Regel, dass jedes operative Dashboard vier Kästen unverändert nachbilden muss. Eine nützliche Kennzahlentafel an der Front kann einfacher sein.

Die Unterscheidung des IHI zwischen Ergebnis, Prozess und Ausgleich liefert eine konkrete Prüfung. Soll ein Team einen Prozessindikator verbessern, muss die Prüfung ein vom Kunden erlebtes Ergebnis und eine Gegenmaßnahme einschließen, die unerwünschte Kosten anderswo erkennt. Messung kann auch eine qualitative Stichprobe verlangen, denn eine Zählung ausgefüllter Formulare sagt nicht, ob der empfangende Kollege den offenen Fall versteht.

01**Ergebnis**

Das Ergebnis, das ein Kunde oder empfangendes Team wirklich schätzt.

02**Prozess**

Ein steuerbarer Schritt, plausibel mit diesem Ergebnis verbunden.

03**Ausgleich**

Was sich verschlechtern könnte, wenn die Prozesszahl forciert wird.

Mehr Kennzahlen können auch mehr Lärm erzeugen. Wählen Sie eine kleine Familie mit verschiedenen Aufgaben; entfernen Sie Doppelungen; legen Sie Verantwortung fest; entscheiden Sie, wie oft jede geprüft wird. Jede sicherheitskritische Ausnahme sollte für sich sichtbar sein, statt von einer günstigeren Geschwindigkeitszahl weggemittelt zu werden.

UNSERE LESART

Eine Balanced Scorecard ist die Disziplin, das ganze Ergebnis zu betrachten, keine Ausrede, Ziele zu vervielfachen.

Quellen: Robert S. Kaplan & David P. Norton, Harvard Business Review, 1992; Institute for Healthcare Improvement, 2026

Kaplan und Norton, Harvard Business Review, Januar–Februar 1992; IHI, Establishing Measures.

KAPITEL III · DIE VERBORGENEN KOSTEN EINES KPI

Eine Kennzahl wird besser. Eine andere ändert sich.

Ein vollständig synthetisches Zahlenbeispiel zeigt, warum zwei Indikatoren zusammen betrachtet werden müssen: zwei Perioden mit je 1.000 Aufträgen.

SYNTHETISCHES BEISPIEL, FÜR DIESEN BAND VERFASST; KEINE ARBEITGEBERDATEN

HÄUFIGKEIT JE 1.000 AUFTRÄGE, ERFUNDEN

Verspätet, Periode A



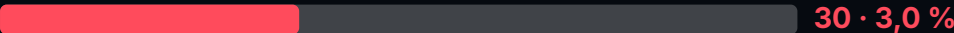
Verspätet, Periode B



Unvollständig, Periode A



Unvollständig, Periode B



Die Quote verspäteter Aufträge sinkt um zwei Prozentpunkte, während die Quote unvollständiger Aufträge um zwei Punkte steigt. Wir haben zwei Fakten über die künstliche Tabelle beschrieben. Wir haben nicht gezeigt, dass schnellere Abfahrten unvollständige Aufträge verursachen, und nicht entschieden, ob ein Prozentpunkt in beiden Ergebnissen denselben Wert hat. Außerdem können ein verspäteter und ein unvollständiger Auftrag derselbe Auftrag sein: Die Zahlen lassen sich nicht einfach addieren, um die Zahl der betroffenen Aufträge zu erhalten. Eine echte Analyse bräuchte Ereignisdefinitionen, Umgang mit Überschneidungen, Konfidenzüberlegungen, Schwere für den Kunden, vergleichbares Volumen und Belege zur Prozessänderung. Ohne sie ist der scheinbare Zielkonflikt nur ein Anlass zur Untersuchung.

UNSERE LESART

Addieren Sie keine Prozentsätze mit verschiedenen Bedeutungen, und folgern Sie keine Ursache aus zwei Linien in einem Diagramm.

Quelle: Institute for Healthcare Improvement, 2026

Jeder Wert auf dieser Seite ist synthetisch, für diesen Band verfasst; Kennzahlentaxonomie nach dem IHI.

KAPITEL III · DIE VERBORGENEN KOSTEN EINES KPI

Eine Leitplanke neben das Ziel

Ein praktisches Protokoll, um eine Verbesserung zu prüfen, ohne ihre Kosten zu verlagern.

1

Das Hauptziel festhalten

Die Kundenwirkung und die betriebliche Kennzahl, die dorthin führen soll.

2

Eine mögliche Nebenwirkung benennen

Ein oder zwei plausible nachgelagerte oder sicherheitsbezogene Folgen.

3

Den Messbereich abgleichen

Population, Zeitraum und Analyseeinheit abgleichen; bekannte Lücken festhalten.

4

Vor dem Pilot entscheiden

Prüfauslöser und nicht verhandelbare Stoppbedingungen, wo die Regelung sie verlangt.

5

Beide Seiten prüfen

Das Ergebnis, das Ausgleichssignal und eine kleine Stichprobe echter Arbeit.

6

Die Unterscheidung bewahren

Hat funktioniert, teilweise funktioniert oder noch nicht erkennbar: festhalten, was zutrifft.

Bei einer Hotelübergabe ist ein pünktlicher Bericht das eine; ob die nächste Schicht genug verwertbare Information hat, das andere. Bei der Last-Mile-Beladung verdienen Abfahrtszeit und Vollständigkeit womöglich gepaarte Aufmerksamkeit. Das sind Gestaltungsvorschläge, keine bereits geprüften kombinierten KPIs für die Betriebe des Autors.

UNSERE LESART

Sie brauchen kein perfektes Dashboard, um nicht ein Problem zu feiern, das an ein anderes Team verschoben wurde.

Quellen: Institute for Healthcare Improvement, 2026; Robert S. Kaplan & David P. Norton, Harvard Business Review, 1992

Redaktionelles Betriebsprotokoll aus den Kennzahlenfamilien des IHI und der Linse der ausgewogenen Leistung von Kaplan und Norton.

1 SZENE

2 AUFSCHLÜSSELUNG

3 PROTOKOLL

KAPITEL IV · DIE ROTE ZAHL UND DIE FALSCH E SCHULD

Der Tag, an dem das Dashboard rot wird

Eine schlechtere Zahl kann jetzt Aufmerksamkeit verlangen. Sie rechtfertigt nicht immer, den Prozess morgen zu ändern.

KONSTRUIERTE MANAGEMENTPRÜFUNG, KEIN BEHAUPTETES EREIGNIS

Eine Führungskraft sieht eine rote Tagesquote auf einem Leistungsbildschirm. Sie liegt über einem Ziel. Der Instinkt ist zu fragen, wer das verursacht hat, und dann sofort eine Besprechung einzuberufen. Doch ein einzelner Tagesanteil enthält mehrere Fragen, die die Farbe nicht beantworten kann: Wie viele Gelegenheiten wurden gezählt? Hat sich die Definition geändert? Schwankt der Prozess normalerweise? Ist das ein echtes Ausnahmeereignis oder eine Beobachtung in einem ohnehin instabilen System? Die Antwort hängt auch von der Art des Mangels ab. Ein ernster Sicherheitsvorfall, ein Compliance-Verstoß oder ein direktes Risiko für Kunden verlangt sofortiges Handeln nach Regelung, auch ohne statistischen Alarm. Bei gewöhnlicher Prozessverbesserung aber kann es unnötige Änderungen und mehr Streuung erzeugen, auf jeden roten Punkt als Beweis einer zuweisbaren Ursache zu reagieren.

Der Autor hat mit täglichen Leistungsgesprächen, Ladeabweichungen und Hotelberichten gearbeitet, doch es gibt keinen geprüften persönlichen Fall, in dem eine bestimmte rote Zelle genau die hier beschriebene Abfolge auslöste.

UNSERE LESART

Eine rote Farbe ist eine Einladung zur Prüfung. Sie ist keine vollständige Erklärung.

Konstruierte Szene; die statistische Unterscheidung folgt dem NIST/SEMATECH e-Handbook. Der unten genannte Essay fragt, ob es ein schlechter Tag oder ein Problem war.

[Mehr im Essay](#)[Ein schlechter Tag oder ein Problem?](#)

KAPITEL IV · DIE ROTE ZAHL UND DIE FALSCH E SCHULD

Ein Geschäftsziel ist keine Eingriffsgrenze

Zwei waagerechte Linien können in einem Diagramm gleich aussehen und verschiedene Fragen beantworten. Ein Geschäftsziel drückt ein gewünschtes Serviceniveau aus. Eine statistische Eingriffsgrenze leitet sich aus einem Modell ab, wie ein Prozess streut, wenn sein zugrunde liegender Mechanismus hinreichend stabil ist. Ein Tageswert kann das Ziel verfehlen und innerhalb der erwarteten Streuung bleiben; ein plötzliches Signal kann auftreten, während der Wert das Ziel noch erfüllt.

NIST beschreibt Regelkarten als über die Zeit aufgetragene Messungen mit einer Mittellinie und Eingriffsgrenzen. Die Grenzen werden nach statistischen Annahmen und dem Kartentyp gewählt. Für Anteile fehlerhafter Einheiten kann eine p-Karte geeignet sein, wenn die gezählten Gelegenheiten und die Unabhängigkeitsannahmen es tragen. Ein von der Führung gesetztes Ziel ist weder die Mittellinie noch die obere Eingriffsgrenze.

01**Ziel**

Welche Leistung wir wollen oder verlangen.

02**Eingriffsgrenze**

Was das Prozessmodell als gewöhnliche Streuung vorhersagt.

03**Sicherheitsauslöser**

Was die Regelung unabhängig vom Diagramm sofort verlangt.

Manche Prozesse rechtfertigen eine Shewhart-Karte nicht ohne weitere Arbeit. Wechselnder Fallmix, serielle Abhängigkeit, veränderliche Nenner oder eine instabile Ausgangslage können eine einfache Formel entwerfen. Regelkarten sind Werkzeuge für disziplinierte Untersuchung, keine sofortigen Urteile über Einzelne.

UNSERE LESART

Verfehltes Ziel, ungewöhnliches Signal und unsichere Lage sind drei getrennte Führungsfragen.

Quellen: NIST & SEMATECH, NIST/SEMATECH e-Handbook of Statistical Methods: What are Control Charts? (section 6.3.1), 2026; NIST & SEMATECH, NIST/SEMATECH e-Handbook of Statistical Methods: Proportions Control Charts (section 6.3.3.2), 2026

NIST/SEMATECH e-Handbook, Abschnitte 6.3.1 und 6.3.3.2.

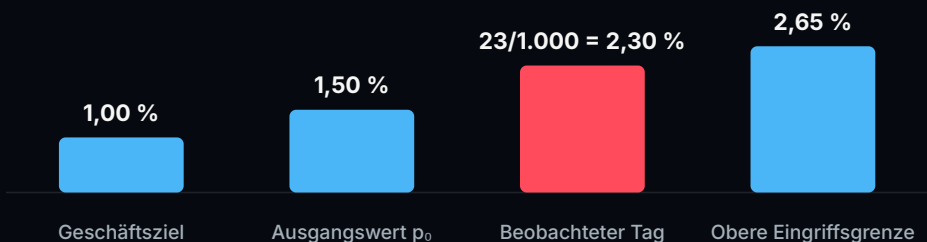
KAPITEL IV · DIE ROTE ZAHL UND DIE FALSCHESCHULD

Der rote Punkt innerhalb des erwarteten Bereichs

Eine illustrative p-Karten-Rechnung mit fester Stichprobengröße und ausdrücklichen Annahmen: $UCL = p_0 + 3\sqrt{[p_0(1 - p_0)/n]}$, mit $p_0 = 0,015$ und $n = 1.000$ Gelegenheiten je Tag. Die obere 3-Sigma-Grenze liegt bei etwa 0,0265 (2,65 %); die untere bei etwa 0,0035 (0,35 %).

SYNTHETISCHE AUSGANGSLAGE, UNTER ANNAHME UNABHÄNGIGER BINOMIALZÄHLUNGEN

TAGESANTEIL UNVOLLSTÄNDIGER AUFTRÄGE, IN PROZENT, ERFUNDEN



Der Tag mit 2,30 % verfehlt das Geschäftsziel, überschreitet unter dieser vereinfachten Karte aber nicht die obere Eingriffsgrenze. Das erklärt den Tag nicht für annehmbar: Der statistische Beleg eines veränderten Mechanismus ist eine andere Frage als die Angemessenheit des Service. Die Rechnung setzt einen stabilen historischen Anteil, unabhängige Gelegenheiten und einheitliche Definitionen voraus, Bedingungen, die reale Daten prüfen müssen. Die UCL ist eine konventionelle p-Karten-Rechnung, keine Arbeitgeberschwelle, und keine echte Ausgangslage wird offengelegt.

UNSERE LESART

Eine unerwünschte Quote kann für einen schlecht laufenden Prozess gewöhnlich sein. Gewöhnlich heißt nicht gut.

Quelle: NIST & SEMATECH, 2026

Synthetisches p-Karten-Beispiel nach der NIST-Formel für Anteilskarten: $3\sqrt{(0,015 \times 0,985 / 1.000)} \approx 0,0115$, also $UCL \approx 2,65 \%$ und $LCL = 0,35 \%$. Keine lokale Ausgangslage und kein Testergebnis wird behauptet.

1 SZENE

2 AUFSCHLÜSSELUNG

3 PROTOKOLL

KAPITEL IV · DIE ROTE ZAHL UND DIE FALSCHESCHULD

Das Rot untersuchen, ohne einen Schuldigen zu erfinden

Eine verhältnismäßige Antwort auf ein Datensignal, einen Serviceverstoß oder ein Sicherheitsrisiko.

1

Akutes Risiko prüfen

Zuerst Sicherheits-, Gästeschutz- und Compliance-Regeln, wenn eine Gefahr akut ist.

2

Den Eintrag prüfen

Quelle, Zeitstempel, Nenner, Einschlussregeln, ausstehende Korrekturen.

3

Die Frage einordnen

Verfehltes Ziel, ungewöhnliches Prozesssignal, wiederkehrender Kundenschieden oder mehrere?

4

Mit den Prozessverantwortlichen sprechen

Welche Bedingungen vorlagen, von Sendenden und Empfangenden; keine Schuld zuerst.

5

Eine Erklärung testen

Vergleichbare Perioden, Fallmix und direkte Beobachtung, bevor die Routine geändert wird.

6

Die Antwort abschließen

Geprüft, offen, nächster Schritt und wer Rückmeldung erhält.

Ein Team sollte sich nicht jeden Morgen gegen eine andere Ursachengeschichte verteidigen müssen; ebenso wenig entschuldigt statistische Vorsicht, einen gefährlichen Zustand oder ein chronisches Serviceversagen zu ignorieren. Die Entscheidung passt zum Risiko und zur Qualität der Belege.

UNSERE LESART

Die richtige Antwort ist nicht immer schneller. Sie passt besser zu dem, was das Signal tatsächlich sagt.

Quelle: NIST & SEMATECH, 2026

Redaktionelles Protokoll nach der NIST-Anleitung zu Regelkarten und üblichen betrieblichen Eskalationsgrundsätzen.

1 SZENE

2 AUFSCHLÜSSELUNG

3 PROTOKOLL

KAPITEL V · DIE KENNZAHLENTAFEL, DIE DIE WAHRHEIT SAGT

Ein Durchschnitt, zwei verschiedene Tage

Ändert sich der Arbeitsmix, kann ein aggregierter KPI das Gegenteil seiner Bestandteile erzählen.

REIN NUMERISCHE ILLUSTRATION, KEINE ARBEITGEBERDATEN

Eine Führungskraft prüft zwei Serviceperioden. In jeder Arbeitskategorie verbessert sich die Leistung. Die einfacheren Anfragen werden verlässlicher erledigt; die komplexen ebenfalls. Doch der Gesamtanteil pünktlicher Erledigungen sinkt. Niemand hat eine Zählung gefälscht. Das Dashboard hat sich geändert, weil der Anteil schwieriger Arbeit sich drastisch verschoben hat. Wer nur die Summe liest, mag schließen, das Team habe nachgelassen. Wer die Kategorien liest, mag schließen, der Prozess habe sich verbessert. Beide können auf richtige Arithmetik zeigen. Die fehlende Information ist die Zusammensetzung des Nenners. Die nächste Seite macht den Effekt mit synthetischen ganzen Zahlen sichtbar, die jeder nachrechnen kann.

Die redaktionelle Lehre reicht über die letzte Meile, Belegungsmuster im Hotel und Großhandelsaufträge: Verschiedene Arbeitsarten verlangen nicht zwangsläufig denselben Aufwand, schaffen nicht dasselbe Risiko und kommen nicht jeden Tag im selben Mix vor. Eine faire Analyse muss Definition und Fallmix jeder Kohorte bewahren.

UNSERE LESART

Ein Durchschnitt kann richtig sein und trotzdem die betriebliche Veränderung verfehlen, auf die es ankommt.

Begriffliche mathematische Illustration von Aggregationseffekten; keine implizierte Beobachtung aus den Betrieben des Autors. Der unten genannte Essay erklärt, warum KPIs sich nicht in Excel verbessern.

[Mehr im Essay](#)

Kennzahlen werden nicht in Excel besser

KAPITEL V · DIE KENNZAHLENTAFEL, DIE DIE WAHRHEIT SAGT

Wie beide Gruppen besser werden, während die Summe fällt

Ein nachrechenbares Beispiel einer Aggregationsumkehr durch veränderten Fallmix. Erfolg heißt pünktliche Erledigung.

SYNTHETISCHES BEISPIEL, FÜR DIESEN BAND VERFASST; KEINE ARBEITGEBERDATEN

PÜNKTLICHER ANTEIL JE KATEGORIE UND INSGESAMT, ERFUNDENE ZAHLEN

☐ Periode A ☒ Periode B

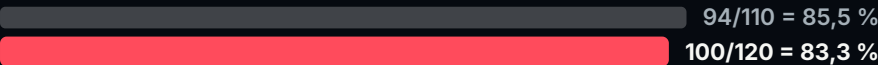
Standardarbeit



Komplexe Arbeit



Alle Arbeit



Beide Kategorien verbessern sich, doch der Gesamtanteil pünktlicher Erledigungen sinkt von etwa 85,5 % auf 83,3 %, weil die anspruchsvolle Kategorie in Periode B einen weit größeren Anteil hat. Die Summen sind keine Vergleiche gleich mit gleich: Die Umkehr ist eine arithmetische Eigenschaft der Gewichte, kein echter Trend. Eine Kennzahl, die Entscheidungen über Menschen stützt, sollte zeigen, ob sich der Mix änderte, und Volumen berichten; segmentieren Sie, wo Risiken sich unterscheiden, mit ausreichenden Stichproben und stabilen Definitionen, denn eine willkürliche Aufteilung erzeugt eigenen Lärm.

UNSERE LESART

Bevor Sie zwei Summen vergleichen, fragen Sie, ob sie dieselbe Art von Arbeit enthalten.

Alle Zahlen sind erfunden; die Rechnung wurde aus ganzzahligen Zählern und Nennern geprüft: $94/110 = 85,45 \%$, $100/120 = 83,33 \%$.

KAPITEL V · DIE KENNZAHLENTAFEL, DIE DIE WAHRHEIT SAGT

Eine Kennzahlentafel mit Gewissen

Die Kennzahlenfamilie muss betriebliches Handeln mit dem Kunden verbinden und ihre Ausschlüsse sichtbar machen. Eine nützliche Kennzahlentafel ist keine Tabelle, in der jeder Indikator denselben Rang hat. Manche belegen die Angemessenheit des Service; manche zeigen die Prozessgesundheit; manche warnen vor Nebenwirkungen; manche qualifizieren die Deutung, indem sie Volumen, Fallmix und Abdeckung zeigen. Eine Schwelle kann eine Prüfung anstoßen, ohne zur Schuldfeststellung zu werden.

01	Ergebnis	Vollständiger, pünktlicher Service; für den Kunden sichtbare Qualität.
02	Prozess	Abfahrts- oder Übergabeschritte mit kausaler Begründung.
03	Ausgleich	Unvollständige Aufträge, Nacharbeit, unsichere Abkürzungen oder wiederkehrende Beschwerden.
04	Kontext	Volumen, Fallmix, Ausnahmeexposition und Saisonalität.
05	Datenqualität	Abdeckung, Ausschlüsse, Verzögerungen, Korrekturen und Definitionsänderungen.

Balanced-Scorecard-Denken kann die Strategie leiten, während die Kennzahlenfamilien des IHI die lokale Verbesserung gestalten helfen. Die Führungskraft kann einen Hinweis zur Datenqualität ergänzen: Eine gut definierte Quote auf Basis unvollständiger Aufzeichnungen kann präzise sein, ohne repräsentativ zu sein. Der Anteil der Datensätze mit gültigem Status ist daher manchmal so wichtig wie die berichtete Servicequote. Persönliche Kennungen, Gästedaten und vertrauliche Betriebsaufzeichnungen gehören nicht in ein allgemeines Dashboard; aggregierte oder demonstrative Zahlen können die Methode erklären, ohne interne Information preiszugeben.

UNSERE LESART

Die Kennzahlentafel sollte eine strittige Zahl leichter untersuchbar machen, nicht schwerer anfechtbar.

Quellen: Robert S. Kaplan & David P. Norton, Harvard Business Review, 1992; Institute for Healthcare Improvement, 2026; NIST & SEMATECH, 2026

Begriffliche Synthese nach Kaplan und Norton (1992), der Messanleitung des IHI und den Datenannahmen von NIST.

KAPITEL V · DIE KENNZAHLENTAFEL, DIE DIE WAHRHEIT SAGT

Den KPI definieren, bevor Sie Erfolg definieren

Ein kompaktes Kennzahlenwörterbuch schützt Entscheidungen vor stillen Änderungen des Nenners.

1

Zweck und Nutzer

Wer braucht die Zahl, und welche Entscheidung könnte sie stützen?

2

Einheit und Formel

Zähler, Nenner, Einheiten, Ereignisgrenze und Zeitfenster.

3

Quelle und Aktualität

Die maßgebliche Quelle, Aktualisierungsrhythmus, bekannte Verzögerung und Verantwortung.

4

Zulässigkeit und Ausnahmen

Ausschlüsse, Stornierungen, Wiederholungen, späte Einträge und Segmentierung nach Fallmix.

5

Begleitkennzahlen

Die Prozesskennzahl mit einem Kundenergebnis, einer Ausgleichs-Leitplanke und einer Datenqualitätsprüfung paaren.

6

Änderungen steuern

Die Definition versionieren, vergleichbare Historie behalten und Änderungen ankündigen, bevor Perioden verglichen werden.

Wenden Sie keine erfundene Definition von Verspätung oder Unvollständigkeit auf ein Unternehmen an, dessen offizielles Kennzahlenwörterbuch andere Regeln hat. Ebenso wenig darf die Bedeutung eines Hotelberichts bloß aus seinem Titel abgeleitet werden. Eine Fachveröffentlichung kann allgemeine Strukturen beschreiben, doch eine echte Entscheidung verlangt die befugte lokale Definition.

UNSERE LESART

Der wichtigste Teil eines KPI ist womöglich das Kleingedruckte, das erklärt, was fehlt.

Quelle: Institute for Healthcare Improvement, 2026

Redaktionelles Protokoll für ein Kennzahlenwörterbuch, kein bestehender freigegebener Data-Governance-Standard eines Arbeitgebers.

1 SZENE

2 AUFSCHLÜSSELUNG

3 PROTOKOLL

KAPITEL VI · JENSEITS DES ZIELS

Die Besprechung, die eine Frage mehr stellt

Eine nützliche Leistungsprüfung fragt nach dem Serviceergebnis, nachdem sie die Zahl gesehen hat.

HYPOTHETISCHE MANAGEMENTPRÜFUNG, KEIN ANSPRUCH AUF DIREKTE ERFAHRUNG

Ein konstruiertes Führungsteam kommt zur Wochenprüfung. Ein KPI ist besser als in der Vorwoche. Die Folien sind fertig; der Jubel ist fast sofort da. Stattdessen fragt der Vorsitz, ob die Quote auf vergleichbarer Arbeit beruht, ob die Zählung vollständig ist und ob sich ein nachgelagerter Indikator verschlechtert hat. Der Analyst kann zwei Fragen beantworten, die dritte nicht. Eine kleine Stichprobe wird dem Prozessverantwortlichen zugewiesen, und das Ergebnis bleibt offen, bis es geprüft werden kann. Das ist keine Unentschlossenheit. Es ist eine verhältnismäßige Weigerung, eine unvollständige Statistik zu einer vollständigen Behauptung zu machen. Die Besprechung erkennt weiterhin an, was bekannt ist, und würdigt weiterhin die Menschen, die einen gemessenen Schritt verbessert haben. Sie vermeidet nur, ein Geschäftsergebnis zuzuschreiben, das die Belege nicht gestützt haben.

Die fiktive Besprechung beschreibt die Praxis, für die dieser Band wirbt. Sie ist kein rekonstruiertes Gespräch mit Vorgesetzten oder Kollegen aus der Laufbahn des Autors.

UNSERE LESART

Das stärkste Dashboard ist das, das ein ehrliches „wir wissen es noch nicht“ möglich macht.

Konstruiertes Szenario, angeregt von statistischer Prozesskontrolle, ausgewogener Messung und Belegdisziplin. Der unten genannte Essay handelt vom Gespräch mit jemandem, der einen Fehler gemacht hat.

[Mehr im Essay](#)

Wie man mit jemandem spricht, der einen Fehler gemacht hat

KAPITEL VI · JENSEITS DES ZIELS

Fünf Fragen vor einem Urteil

Eine Führungskraft kann Verantwortung für eine Entscheidung übernehmen, ohne vorzugeben, die Kennzahl enthalte die ganze Erklärung.

01	Ist sie definiert?	Kann ein anderer kompetenter Leser Formel, Population und Zeitraum nachvollziehen?
02	Ist sie vertrauenswürdig?	Sind die Aufzeichnungen vollständig, aktuell und stimmig mit dem physischen oder Serviceprozess?
03	Ist sie anders?	Ist die Veränderung praktisch bedeutsam und, wo angebracht, ein ungewöhnliches statistisches Signal?
04	Ist sie ausgewogen?	Hat sich das Kundenergebnis ohne unannehbare Sicherheits-, Qualitäts- oder Folgekosten verbessert?
05	Ist sie handlungsfähig?	Hat das verantwortliche Team die Befugnis, die Bedingung zu ändern, die das Ergebnis verursacht?

Die Reihenfolge zählt. Mit einem neuen Ziel zu beginnen, bevor der Nenner geprüft ist, erzeugt falsche Dringlichkeit. Ein Signal in eine Anklage zu verwandeln, bevor der Prozess verglichen ist, erzeugt Misstrauen. Zehn neue Kennzahlen hinzuzufügen, ohne veraltete zu entfernen, erzeugt Arbeit, aber keine Einsicht. Jede Kennzahlenprüfung sollte einen vereinbarten Weg für Korrekturen haben: Eine korrigierte Zahl darf die alte nicht still und ohne Prüfpfad ersetzen. Stimmt die Kennzahl, ist aber der Prozess schlecht, gehört das Problem zur Prozessverbesserung; ist der Indikator weder verlässlich noch nützlich, braucht er womöglich eine Neugestaltung.

UNSERE LESART

Ein KPI verdient Vertrauen, weil seine Bedeutung und seine Belege geprüft werden können, nicht weil er in Farbe angezeigt wird.

Quellen: Donald T. Campbell, Evaluation and Program Planning, 1979 (über ein Dartmouth Occasional Paper (1976), nachgedruckt im Journal of MultiDisciplinary Evaluation (2011)); NIST & SEMATECH, 2026; Institute for Healthcare Improvement, 2026

Redaktioneller Rahmen nach Campbell (1979), den Prozesskontrollmethoden von NIST und der Messanleitung des IHI.

1 SZENE

2 AUFSCHLÜSSELUNG

3 PROTOKOLL

KAPITEL VI · JENSEITS DES ZIELS

Jenseits des Ziels: eine Betriebsprüfung

Ein vorgeschlagener vierwöchiger Pilot, ohne die Annahme, dass sich jedes Problem in einem Monat lösen lässt.



- Woche 1: eine echte Entscheidung wählen**
Ein KPI hinter einer wiederkehrenden Führungsentscheidung, mit seiner offiziellen Definition.
- Woche 2: eine kleine Stichprobe prüfen**
Aktualität, Abdeckung, Dubletten, Zeitpunkt und eine begleitende Servicekennzahl.
- Woche 3: Abweichungen nachgehen**
Die Arbeit dort beobachten, wo eine irreführende Lesart entsteht; Privatsphäre schützen.
- Woche 4: die Entscheidungsregel vereinbaren**
Verantwortlicher, Eskalationsbedingungen, Folgetermin, eine Leitplanke gegen Nebenwirkungen.
- Nach dem Pilot: vergleichen und erklären**
Befunde, Unbekanntes, Änderungen; die Definition nur mit dokumentierter Steuerung geändert.

Vier Wochen sind ein redaktioneller Testtakt, keine belegte Dauer und kein Arbeitgeberverfahren: Ein sicherheitskritischer Befund wird sofort eskaliert, eine strukturelle Neugestaltung kann länger dauern. Das Ergebnis ist eine besser begründbare Entscheidung und ein Team, das weiß, was die Zahl bedeutet.

UNSERE LESART

Messung dient der Führung nur, solange die Führung für das Urteil verantwortlich bleibt.

Quellen: Institute for Healthcare Improvement, 2026; NIST & SEMATECH, 2026

Eigener redaktioneller Pilot nach IHI- und NIST-Anleitung; keine zertifizierte oder validierte Intervention.

WERKZEUG VOR ORT

Das ehrliche KPI-Blatt

Ein Indikator, eine Entscheidung und eine ausdrückliche Prüfung dessen, was die Kennzahl nicht sieht. In öffentlichen Beispielen leere oder synthetische Daten verwenden.

01

Zweck und Entscheidung

wer nutzt sie, und welche Wahl soll sie stützen?

02

Genaue Definition

Zähler, Nenner, Zeitfenster, Ausschlüsse und Verantwortlicher der Kennzahl

03

Quelle und Aktualität

maßgeblicher Datensatz, Aktualisierungsverzug und bekannte Lücken

04

Signal oder Ziel?

Geschäftsschwelle, Regelkartenannahmen oder dringender Risikoauslöser

05

Ausgleich und Kundenwirkung

welche nachgelagerte Qualitäts-, Sicherheits- oder Servicekennzahl könnte sich verschlechtern?

06

Handlung und Rückmeldung

nächste Prüfung, benannter Verantwortlicher, Prüfpunkt und Rückmeldung

Quellen: Institute for Healthcare Improvement, 2026; NIST & SEMATECH, 2026; Robert S. Kaplan & David P. Norton, Harvard Business Review, 1992

Keine Mitarbeiter-, Gast- oder Zahlungskennungen eintragen. Hat eine Frage keine gültige Antwort, „nicht gemessen“ oder „noch nicht geprüft“ festhalten, statt ein genaues Ergebnis zu erfinden. Eine Seite genügt, um den Unterschied zwischen einer Zahl und einer Entscheidung sichtbar zu machen.

SCHLUSSESSAY

Verantwortlich für den Zweck

Der Zweck einer Kennzahl ist, die Arbeit zu verbessern, nicht, die Arbeit durch eine ansprechende Beschreibung zu ersetzen.

Eine Führungskraft braucht Zahlen. Ohne sie bleiben wiederkehrende Fehler anekdotisch, Verbesserung lässt sich schwer prüfen, und die leiseren Teile der Organisation werden womöglich nie gesehen. Doch Messung schafft auch eine Versuchung. Sobald ein Ziel an der Wand hängt, sprechen die Leute darüber, als wäre es der Zweck des Betriebs. Eine grüne Zelle bringt Erleichterung. Eine rote Zelle verlangt eine Erklärung. Beide Gefühle kommen, bevor die Belege vollständig sind. Es zählt, ob die Zahl noch mit dem Service verbunden ist, den sie beschreiben sollte. Auf der letzten Meile heißt das, sich zu erinnern, dass Beladung, Abfahrt, Vollständigkeit und Kundenergebnis verknüpft, aber nicht austauschbar sind. Im Hotel müssen Bericht und Übergabe die Arbeit stützen, die über Schichten hinweg weitergeht. Im Großhandel sind ein angenommener Auftrag und ein verlässliches Versprechen getrennte Leistungen. Die passenden Kennzahlen unterscheiden sich, doch die Pflicht, ihre Bedeutung zu prüfen, ist dieselbe.

Campbell erinnert daran, dass geplante Verbesserung schwer zu bewerten ist. Goodhart warnt, dass Druck die Nützlichkeit einer Regelmäßigkeit verändern kann. Holmström und Milgrom erklären, warum unverhältnismäßige Anreize auf gemessene Aufgaben die Aufmerksamkeit verzerren können. Bevan und Hood beschreiben Risiken in einem dokumentierten öffentlichen Umfeld. IHI und NIST bieten Werkzeuge für eine ausgewogenere und statistisch ehrlichere Prüfung. Nichts davon nimmt der Führungskraft die Pflicht, den Menschen zuzuhören, den Prozess zu prüfen und den Kunden zu schützen. Eine Kennzahl muss nicht perfekt sein, um nützlich zu sein. Sie braucht einen definierten Geltungsbereich, eine vertrauenswürdige Quelle, ein Begleitsignal und eine Person, die bereit ist zu sagen, was noch nicht bekannt ist. Wenn sie aufhört, eine bessere Entscheidung zu leiten, wird ihre Farbe zur Dekoration.

UNSERE LESART

Ein Ziel kann erreicht werden, während der Zweck der Arbeit verloren geht. Die Führungskraft ist dafür verantwortlich, den Unterschied zu bemerken.

Redaktionelles Fazit. Kein Manipulation, versteckter Mangel oder Sanktion wird einem bestimmten Arbeitgeber zugeschrieben.

METHODE

Studien, Modelle und Methoden

Eine belegbasierte Synthese, kein Feldexperiment. Die Kontexte der letzten Meile und des Hotels spiegeln vom Autor beschriebene Arbeitskategorien; jede genannte Besprechung, jedes Einzelereignis und jeder numerische Betriebsdatensatz ist ausdrücklich synthetisch.

01	Theorie	Goodhart (1975, über Bevan und Hood), Strathern (1997), Campbell (1979) und Holmström und Milgrom (1991) beschreiben Mechanismen, keine gemessenen Quoten.
02	Beobachtung	Bevan und Hood (2006) sowie van Thiel und Leeuw (2002) betreffen öffentliche Dienste; ihre Befunde werden nicht als Ergebnisse der Logistik oder Hotellerie übertragen.
03	Methode	Kaplan und Norton (1992), die Kennzahlenfamilien des IHI und die Regelkartenabschnitte von NIST sind Gestaltungsanleitung; die p-Karte ist unter ihren genannten Annahmen berechnet.
04	Arithmetik	80, 60, 10 und 30 je 1.000; UCL $\approx$ 2,65 % und LCL $\approx$ 0,35 % für $p_0 = 0,015$ und $n = 1.000$; $94/110 = 85,5 \%$ und $100/120 = 83,3 \%$: alles nachgerechnet.

Das ehrliche KPI-Blatt und der vierwöchige Zyklus sind redaktionelle Vorschläge für lokale Tests, keine zertifizierten Interventionen. Redaktioneller Standard: Eine Theorie, eine Beobachtung, ein Modell und ein Vorschlag werden als solche benannt, und der Satz hört auf zu behaupten, wo die Belege enden.

Quellen: Charles Goodhart, 1975 (über Bevan & Hood (2006), die den Nachdruck von 1984 zitieren); Marilyn Strathern, European Review, 1997; Donald T. Campbell, Evaluation and Program Planning, 1979 (über ein Dartmouth Occasional Paper (1976), nachgedruckt im Journal of MultiDisciplinary Evaluation (2011)); Bengt Holmström & Paul Milgrom, Journal of Law, Economics, & Organization 7 (special issue), 1991; Gwyn Bevan & Christopher Hood, Public Administration, 2006; Sandra van Thiel & Frans L. Leeuw, Public Performance & Management Review 25(3), 2002; Robert S. Kaplan & David P. Norton, Harvard Business Review, 1992; Institute for Healthcare Improvement, 2026; NIST & SEMATECH, NIST/SEMATECH e-Handbook of Statistical Methods: What are Control Charts? (section 6.3.1), 2026; NIST & SEMATECH, NIST/SEMATECH e-Handbook of Statistical Methods: Proportions Control Charts (section 6.3.3.2), 2026

Jede Quelle ist auf der nächsten Seite mit ihrer DOI oder Seite zitiert; das PDF des Autors gibt die untere Grenze der p-Karte als auf null gekappt an, was die Website auf etwa 0,35 % korrigiert.

QUELLEN UND METHODE

Woher die Aussagen stammen

Die Zahlen dieses Bandes stammen aus den folgenden Quellen, jede mit ihrem Jahr. Die Szenen sind konstruiert; die Protokolle sind Vorschläge; was Forschung ist, wird Forschung genannt.

Assessing the impact of planned social change Donald T. Campbell, Evaluation and Program Planning, 1979 · über ein Dartmouth Occasional Paper (1976), nachgedruckt im Journal of MultiDisciplinary Evaluation (2011)	Problems of Monetary Management: The U.K. Experience Charles Goodhart, 1975 · über Bevan & Hood (2006), die den Nachdruck von 1984 zitieren
'Improving ratings': audit in the British University system Marilyn Strathern, European Review, 1997	Multitask Principal-Agent Analyses: Incentive Contracts, Asset Ownership, and Job Design Bengt Holmström & Paul Milgrom, Journal of Law, Economics, & Organization 7 (special issue), 1991
The Balanced Scorecard: Measures That Drive Performance Robert S. Kaplan & David P. Norton, Harvard Business Review, 1992	What's measured is what matters: targets and gaming in the English public health care system Gwyn Bevan & Christopher Hood, Public Administration, 2006
The Performance Paradox in the Public Sector Sandra van Thiel & Frans L. Leeuw, Public Performance & Management Review 25(3), 2002	Establishing Measures (Model for Improvement) Institute for Healthcare Improvement, 2026
NIST/SEMATECH e-Handbook of Statistical Methods: What are Control Charts? (section 6.3.1) NIST & SEMATECH, 2026	NIST/SEMATECH e-Handbook of Statistical Methods: Proportions Control Charts (section 6.3.3.2) NIST & SEMATECH, 2026



MANAGEMENT REVIEW

Management ohne Theater.

Jeder Band eine Frage zum Betrieb, verfolgt vom Betriebsalltag bis zum Protokoll
und an der Forschung geprüft.

ALLE BÄNDE

stivencatalyst.com/de/magazine/operational-treatise.html