

Startup Alignment Sprint

Die Evidenz hinter der Methode

Welche Befunde meine Arbeit tragen, wie belastbar sie sind – und wo ihre Grenzen liegen.

Levi Neumann

M.Sc. Psychologie · IHK-zertifizierter Trainer · Aachen

Warum es dieses Dokument gibt

Auf meiner Website steht, dass ich psychologisch fundiert arbeite. Das ist schnell behauptet und schwer zu überprüfen. Deshalb dieses Dokument.

Es enthält jede Studie, auf die sich mein Vorgehen stützt, mit vollständiger Fundstelle. Du kannst sie nachschlagen. Du kannst sie jemandem vorlegen, der Psychologie studiert hat. Und du kannst mich darauf festnageln.

Es enthält außerdem einen Abschnitt darüber, was die Forschung **nicht** hergibt. Der ist mir der wichtigste. Wer dir bei Teamentwicklung Umsatzsteigerungen in Prozent verspricht, hat die Studien entweder nicht gelesen oder hofft, dass du es nicht tust.

Wie ich Belege einordne

Meta-Analyse: fasst viele Einzelstudien rechnerisch zusammen und gleicht Zufallsbefunde aus. Der stärkste Beleg, den die Sozialwissenschaft anzubieten hat.

Einzelstudie: ein Hinweis, kein Beweis. Kann in einem anderen Kontext anders ausfallen.

Unternehmensinterne Untersuchung oder Praxisbericht: macht etwas plausibel, belegt es aber nicht. Wo ich mich darauf stütze, schreibe ich es dazu.

Die Kurzfassung

Was mein Angebot behauptet, und wie gut es belegt ist. Die letzte Zeile ist die ehrlichste.

Aussage	Beleg	Belastbarkeit
Teams, in denen man Fehler ansprechen kann, lernen schneller und leisten mehr	Edmondson (1999); Frazier et al. (2017), Meta-Analyse über 136 Stichproben	hoch
Strukturierte, moderierte Rückschauen verbessern die Leistung	Tannenbaum & Cerasoli (2013), Meta-Analyse über 46 Studien	hoch
Entscheidendes Wissen bleibt in Gruppen systematisch ungeteilt	Lu et al. (2012), Meta-Analyse über 65 Studien, 3.189 Gruppen	hoch
Kontakt zwischen verfeindeten Gruppen wirkt, aber nur unter Bedingungen	Pettigrew & Tropp (2006), Meta-Analyse über 515 Studien	hoch
Wenn-dann-Vorsätze werden häufiger umgesetzt als Absichtserklärungen	Gollwitzer & Sheeran (2006), Meta-Analyse über 94 Tests	hoch
Namentlich zugeordnete Beiträge verhindern Trittbrettfahren	Karau & Williams (1993), Meta-Analyse über 78 Studien	hoch
Teamentwicklung wirkt, am stärksten auf Zusammenarbeit und Prozesse	Klein et al. (2009), 60 Korrelationen; Salas et al. (2008), 52 Effektstärken, 1.563 Teams	hoch
Bestehende Teams profitieren stärker als kurzfristig zusammengestellte	Salas et al. (2008), Vergleich intakter und ad hoc gebildeter Teams	mittel
Teams korrigieren ihren Kurs zur Halbzeit eines Projekts	Gersick (1988), qualitative Studie an acht Teams	eingeschränkt
Teamentwicklung steigert Umsatz oder Entwicklungsgeschwindigkeit	keine belastbare Grundlage, wird von mir nicht behauptet	nicht belegt

Was ich tue – und warum

Für jede Phase: was passiert, welcher Befund dahintersteht und wo dessen Grenze liegt.

Fünf Phasen, sechs Wochen von Kickoff bis Übergabe. Die angegebenen Zeiträume sind die Regeltaktung. Wenn Urlaub, ein Release oder eine Finanzierungsrunde dazwischenkommt, verschieben wir sie. Der Umfang bleibt derselbe. Das ist keine Unverbindlichkeit, sondern Voraussetzung dafür, dass die Termine tatsächlich mit den richtigen Leuten stattfinden.

Phase 1

Alignment Audit

ca. 1 Woche

Was passiert. Ein 90-minütiger Deep-Dive mit Gründung und Schlüsselrollen. Ergebnis sind Hypothesen, ausdrücklich keine Diagnose.

Warum es nur Hypothesen sind. Das Bild der Führung ist an dieser Stelle systematisch unvollständig, und zwar nicht, weil jemand lügt. In einer Interviewstudie berichteten 85 % der Befragten von Situationen, in denen sie ein aus ihrer Sicht wichtiges Thema gegenüber Vorgesetzten nicht angesprochen haben (Milliken et al., 2003). Detert und Edmondson (2011) haben gezeigt, woran das liegt: An ungeschriebenen Regeln, die niemand ausspricht und alle befolgen. Man widerspricht dem Chef nicht vor anderen, man braucht harte Daten, bevor man den Mund aufmacht, man geht nicht am direkten Vorgesetzten vorbei.

Konsequenz für das Vorgehen. Schweigen ist für Mitarbeitende die rationale Wahl. Wer nur die Führung befragt, bekommt ein sauberes, plausibles und unvollständiges Bild. Deshalb steht am Ende dieser Phase eine Hypothesenliste, die in Phase 2 gegen die Team-Realität geprüft wird.

Grenze: Die 85 % stammen aus 40 qualitativen Interviews. Das ist eine Größenordnung, kein Messwert, und so verwende ich sie auch.

Phase 2

Diagnostic Retrospektiven

ca. 2 Wochen

Was passiert. Zwei moderierte, abteilungsspezifische Retrospektiven, typischerweise Tech und Customer-Facing getrennt.

Der wichtigste Befund dieses Dokuments. Tannenbaum und Cerasoli (2013) haben 46 Studien zu strukturierten Rückschauen zusammengefasst. Teams, die sie durchführen, übertreffen vergleichbare Teams im Mittel um rund 25 %. Entscheidend ist der Zusatzbefund: **Der Effekt fällt deutlich größer aus, wenn die Rückschau strukturiert ist, moderiert wird und auf konkrete Verbesserungen zielt.**

Genau deshalb moderiere ich selbst, statt dem Team eine Vorlage zu geben. Der Unterschied zwischen einer moderierten Retrospektive und einer Feedbackrunde ist nicht Geschmackssache, sondern der Unterschied zwischen Wirkung und Zeitverlust.

8x

seltener

So viel seltener finden Gruppen die beste Lösung, wenn entscheidendes Wissen nur bei einzelnen Personen liegt. Nicht weil die Leute schlecht sind, sondern weil Gruppen überwiegend das besprechen, was ohnehin schon alle wissen.

Lu, Yuan & McLeod (2012): Meta-Analyse über 65 Studien, 101 Effekte und 3.189 Gruppen.

Warum der geschützte Rahmen keine Nettigkeit ist. Damit ungeteiltes Wissen überhaupt auf den Tisch kommt, braucht es psychologische Sicherheit: die geteilte Überzeugung, dass man ohne Nachteil einen Fehler zugeben oder widersprechen kann. Edmondson (1999) hat an 51 Arbeitsteams gezeigt, dass dieser Faktor über Lernverhalten auf die Teamleistung wirkt. Die Meta-Analyse von Frazier et al. (2017) über 136 Stichproben mit mehr als 22.000 Personen hat den Befund bestätigt.

Grenze: Die 25 % sind ein Mittelwert über sehr unterschiedliche Einsatzfelder, darunter viel Medizin und Militär. Für ein Software-Team ist das eine begründete Erwartung, keine Vorhersage.

Phase 3

Management Insight Report

zur Halbzeit

Was passiert. Die Befunde aus den Retrospektiven werden zu einem Bericht an die Geschäftsführung verdichtet: über Muster und Strukturen, nicht über Personen.

Warum der Bericht in die Mitte gehört. Gersick (1988) hat acht Projektteams über ihre gesamte Laufzeit begleitet, von sieben Tagen bis zu sechs Monaten. Teams richten sich nicht kontinuierlich neu aus, sondern in einem Sprung, und zwar ziemlich genau zur Halbzeit, **unabhängig von der absoluten Dauer**. Entscheidend ist also die relative Lage, nicht das Datum: Der Bericht kommt zur Halbzeit, bei einem Sechs-Wochen-Sprint um Woche drei bis vier. Verschiebt sich der Zeitplan, verschiebt sich der Bericht mit. Er kommt nicht irgendwann, sondern dann, wenn Kurskorrekturen erfahrungsgemäß ankommen.

Warum über Muster und nicht über Personen. Kluger und DeNisi (1996) haben 607 Effektstärken zu Feedback ausgewertet. Im Mittel wirkt Feedback positiv, aber **rund 38 % der Interventionen verschlechtern die Leistung**. Der Unterschied liegt darin, ob Feedback auf die Aufgabe zielt oder auf die Person. Ein Bericht, der Namen nennt, richtet mit hoher Wahrscheinlichkeit Schaden an.

Grenze: Gersick ist eine qualitative Studie mit acht Teams. Als Timing-Heuristik gut begründet, als Gesetzmäßigkeit nicht belastbar.

Phase 4

Cross-Functional Action Sessions

ca. 1 Woche

Was passiert. Zwei Sessions mit bewusst gemischt besetzten Gruppen, an deren Ende verbindliche Vereinbarungen stehen.

Warum gemischt und nicht getrennt. Silos zwischen Tech und Business sind kein Kommunikationsproblem, sondern ein Intergruppenproblem: zwei Gruppen, die sich wechselseitig typisieren. Pettigrew und Tropp (2006) haben 515 Studien mit 713 Stichproben und 1.383 Einzeltests ausgewertet: Kontakt zwischen Gruppen geht durchgängig mit geringeren Vorbehalten einher, im Mittel mit einer mittleren Korrelation von $r = .21$ in die erwartete Richtung. Wirksam ist Kontakt allerdings nur unter Bedingungen: gleicher Status in der Situation, ein gemeinsames übergeordnetes Ziel, Kooperation statt Wettbewerb. Diese Bedingungen stellen sich nicht von selbst her. Sie sind der eigentliche Inhalt der Moderation.

Warum Vereinbarungen im Wenn-dann-Format formuliert werden. Gollwitzer und Sheeran (2006) haben 94 Tests mit über 8.000 Personen zusammengefasst: Vorsätze, die festlegen, **wann, wo und wie** gehandelt wird, werden erheblich häufiger umgesetzt als allgemeine Absichten, mit einem mittleren bis großen Effekt ($d = 0,65$). „Wir wollen besser kommunizieren“ ist deshalb wertlos. „Wenn ein Ticket länger als zwei Tage blockiert ist, schreibt der Bearbeiter es in den Kanal“ ist es nicht.

Warum jede Zusage einen Namen bekommt. Karau und Williams (1993) haben in 78 Studien gezeigt, dass individuelle Anstrengung in Gruppen nachlässt, sobald der eigene Beitrag nicht mehr identifizierbar ist (g von rund 0,44). Eine Maßnahme ohne Zuständigen ist eine Absichtserklärung.

Grenze: Die Kontaktforschung stammt überwiegend aus gesellschaftlichen Intergruppenkontexten, nicht aus Unternehmen. Der Mechanismus ist übertragbar, die Effektstärke nicht eins zu eins.

Phase 5

Performance Roadmap und Übergabe

ca. 1 Woche

Was passiert. Abschlusspräsentation vor Führung und Teams, die psychologische Roadmap für eure Skalierungsphase, das Retro-Playbook mit allen Vorlagen, eine Einschätzung dazu, wer bei euch die Retrospektiven künftig moderieren könnte, und ein Follow-up-Check nach drei Monaten.

Warum die Übergabe der wichtigste Teil ist. Leblanc et al. (2024) finden für Team-Reflexivität einen positiven Zusammenhang mit der Teamleistung. Er hängt allerdings von Randbedingungen ab, unter anderem von Teamgröße und Zugehörigkeitsdauer. Reflexion wirkt also nicht von allein. Den zweiten Teil liefert die Lernforschung: Donovan und Radosovich (1999) werten 63 Studien aus und finden verteiltes Üben dem geballten überlegen, wobei der Vorteil bei komplexen Aufgaben kleiner ausfällt. Ein einmaliger Workshop bleibt ein Ereignis, kein Mechanismus. Deshalb endet der Sprint nicht mit einer Präsentation, sondern mit einem Verfahren, das ihr selbst weiterführen könnt.

Zwei unabhängige Meta-Analysen kommen zum selben Ergebnis. Klein et al. (2009) werten 60 Korrelationen aus, Salas et al. (2008) 52 Effektstärken aus 1.563 Teams. Beide finden einen moderaten Effekt, und beide finden ihn **am stärksten bei Prozessergebnissen**. Bei Salas et al. ist Prozess mit Abstand die am besten belegte Ergebnisart. Insgesamt erklärt Teamtraining dort 12 bis 19 % der Unterschiede zwischen Teams; die Autoren halten fest, dass dies die praktische Bedeutung eher unterschätzt.

Ehrliche Einordnung: Der Transfer in den Alltag ist die bekannte Schwachstelle jeder Personalentwicklungsmaßnahme. Playbook, Host-Empfehlung und der Termin nach drei Monaten federn das ab, lösen es aber nicht vollständig. Wer die Retrospektiven dauerhaft in eigener Hand führen will, braucht Menschen, die moderieren können – und das lernt man nicht aus einem Dokument. Dafür gibt es ein eigenes Format; es ist bewusst nicht im Sprint enthalten, damit hier niemand für etwas zahlt, das er vielleicht gar nicht braucht.

Zuschnitt

Welche Phase gehört zu welchem Angebot

Ihr müsst nicht mit allem anfangen. Die Bausteine sind dieselben, sie werden nur unterschiedlich weit zugeschnitten.

	Alignment Diagnose	Startup Alignment Sprint	Retro Host Ausbildung
Phase 1 – Alignment Audit	•	•	–
Phase 2 – Diagnostic Retrospektiven	•	•	–
Phase 3 – Management Insight Report	•	•	–
Phase 4 – Cross-Functional Action Sessions	–	•	–
Phase 5 – Roadmap und Übergabe	–	•	–
Ausbildung interner Moderatoren	–	–	•
Wirkungsmessung vorher und nachher	•	•	–
Follow-up-Check nach drei Monaten	–	•	–
Dauer	2–3 Wochen	6 Wochen	2 Tage

Die Diagnose ist der Anfang des Sprints. Wer später aufstockt, zahlt insgesamt nicht mehr. Die Ausbildung ist kein Teil des Sprints, sondern das, was danach kommt: Im Sprint bekommt ihr das Verfahren, in der Ausbildung Menschen, die es können.

Gilt das überhaupt für euer Team?

Die berechtigte Rückfrage zu jeder Studie lautet: schön – aber wurde das an Teams wie unserem untersucht? Die ehrliche Antwort hat eine gute und eine unangenehme Hälfte.

Die große Meta-Analyse von Salas et al. (2008) erlaubt hier eine genauere Auskunft, weil sie nicht nur fragt, **ob** Teamtraining wirkt, sondern **wo besser und wo schlechter**.

Drei Ergebnisse sprechen für euren Fall

- **Gewachsene Teams profitieren mehr als kurzfristig zusammengestellte.** Bei Leistungsergebnissen liegt der Effekt für bestehende Teams deutlich über dem für ad hoc gebildete Gruppen, und zwar mit klar getrennten Konfidenzintervallen. Das ist bemerkenswert, weil ein großer Teil der Teamforschung mit kurzfristig gebildeten Gruppen arbeitet. Euer Fall ist der, in dem es besser funktioniert.
- **Training, das an der Zusammenarbeit ansetzt, wirkt stärker als Training, das an der Aufgabe ansetzt,** jedenfalls für Prozess- und Klimaergebnisse. Das ist der Unterschied zwischen meiner Arbeit und einer weiteren Runde Prozessoptimierung.
- **Mehr Inhalt ist nicht besser.** Programme, die Aufgaben- und Zusammenarbeitsinhalte mischen, waren den fokussierten nicht überlegen. Die Autoren vermuten Überfrachtung. Das ist der Grund, warum der Sprint wenige Dinge tut und nicht viele.

Und die unangenehme Hälfte

Die Autoren schreiben selbst, dass die überwiegende Mehrheit der ausgewerteten Studien mit Studierenden in simulierten Umgebungen durchgeführt wurde, eine Folge jahrzehntelanger Förderung durch militärische Institutionen. Militärische Settings erreichen die höchsten Werte, Laborstudien mittlere, medizinische die niedrigsten.

Die Zahl, die ich nicht verwende

Dieselbe Meta-Analyse berichtet für Studien aus **Wirtschaftskontexten** einen Zusammenhang von **.85** – gegenüber **.34** für den Gesamtdatensatz. Das ist mit Abstand der höchste Wert der ganzen Untersuchung, und er beruht auf fünf Studien. Die Autoren warnen ausdrücklich davor, Effektstärken aus so wenigen Studien zu interpretieren.

Das wäre die schmeichelhafteste Zahl, die ich zitieren könnte – ausgerechnet aus dem Kontext, in dem du arbeitest. Ich nenne sie trotzdem hier und nicht im Verkaufsgespräch, weil ein Dossier, das nur die passenden Befunde zeigt, keines wäre. Verwenden werde ich sie nicht.

Was daraus folgt: Die Mechanismen sind gut belegt und übertragbar, also strukturierte Rückschauen, psychologische Sicherheit, geteiltes Wissen, verbindliche Vorsätze. Die genauen Zahlen sind es nicht. Für ein 40-köpfiges Software-Unternehmen in Aachen gibt es keine Studie. Es gibt gut begründete Erwartungen, und es gibt die Bereitschaft, nach drei Monaten nachzusehen, ob sie eingetroffen sind.

Warum ich nicht „offen ausdiskutieren“ lasse

Eine verbreitete Managementweisheit lautet: Streit über die Sache ist gut, Streit über Personen ist schlecht. Der erste sei sogar produktiv. Diese Unterscheidung steht in vielen Lehrbüchern.

Sie hält der Prüfung nicht stand. De Dreu und Weingart (2003) haben die vorliegenden Studien zusammengefasst und finden, dass **beide** Konflikttypen negativ mit Leistung und Zufriedenheit zusammenhängen, bei komplexen Aufgaben sogar stärker. Die Autoren halten ausdrücklich fest, dass ihr Ergebnis dem widerspricht, was sowohl die Forschung als auch einführende Lehrbücher nahelegen.

Was das praktisch bedeutet

„Lasst uns das mal offen ausdiskutieren“ ist keine Methode, sondern der Verzicht auf eine. Sachkonflikte werden nicht dadurch produktiv, dass man sie zulässt, sondern dadurch, dass Reihenfolge, Rederecht und Ergebnisformat vorher feststehen. Das ist der Grund, warum ich mit einer festen Struktur arbeite und nicht mit einer offenen Runde.

Und die Grenze meiner Methode

Friedrich Glasl hat Konflikteskalation in neun Stufen beschrieben. Etwa ab der vierten Stufe geht es den Beteiligten nicht mehr um die Sache, sondern darum, dass die Gegenseite verliert. Ab diesem Punkt ist Moderation strukturell überfordert. Dann braucht es Mediation, und in manchen Fällen eine Personalentscheidung.

Ich prüfe das im Erstgespräch. Wenn ich zu dem Schluss komme, dass ihr etwas anderes braucht als das, was ich anbiete, sage ich das, bevor Geld fließt.

Die Gegenprobe

Was die Forschung nicht hergibt

Ein Evidenzdossier ist nur so viel wert wie der Abschnitt, in dem der Absender gegen sich selbst argumentiert. Hier ist er.

- **Keine Kausalaussage zu Geschäftszahlen.** Keine der zitierten Meta-Analysen zeigt, dass Teamentwicklung Umsatz, Marge oder Entwicklungsgeschwindigkeit steigert. Belegt sind Effekte auf Zusammenarbeit, Prozesse und Klima. Dass diese mittelbar auf Geschäftszahlen wirken, ist plausibel, aber Plausibilität ist kein Beleg. Belegt ist damit genau der Teil, den ihr bei mir beauftragt. Den zweiten Teil der Kette könnt ihr selbst prüfen, denn eure Zahlen habt ihr und ich nicht.
- **Keine Prozentversprechen für deinen Einzelfall.** Die 25 % aus der Debrief-Meta-Analyse sind ein Durchschnitt über 46 Studien. Ein Durchschnitt sagt nichts darüber, wo dein Team landen wird. Deshalb messe ich bei euch, statt zu versprechen.

- **Meta-Analysen mitteln über Kontexte, die selten Startups sind.** Das ist kein Einwand von mir, sondern eine Feststellung der Autoren selbst: Der Großteil der ausgewerteten Studien arbeitet mit Studierenden in simulierten Umgebungen, geprägt durch militärische Forschungsförderung. Die Mechanismen sind übertragbar, die Zahlen nur eingeschränkt.
- **Kleine Studienzahlen erzeugen große Zahlen.** Je enger man einen Teilbereich schneidet, desto spektakulärer wirken die Effekte und desto weniger bedeuten sie. Effektstärken aus weniger als fünf Studien sind nicht interpretierbar, auch dann nicht, wenn sie mir gefallen. Die Befunde, auf die ich mich hier stütze, beruhen auf Dutzenden bis über hundert Einzelstudien.
- **Googles „Projekt Aristotle“ ist keine Studie.** Die interne Untersuchung von 180 Teams, die psychologische Sicherheit als wichtigsten Faktor identifizierte, ist nicht begutachtet und nicht veröffentlicht. Ich nenne sie, weil sie bekannt ist und die Richtung stützt. Tragend sind Edmondson und Frazier et al., nicht Google.
- **Publikationsverzerrung.** Studien mit Wirkung werden häufiger veröffentlicht als Studien ohne. Die berichteten Effekte sind daher eher Ober- als Untergrenzen. Das gilt für die gesamte Literatur, auch für die hier zitierte. Was das verschiebt, ist die Höhe der Effekte, nicht ihre Richtung. Deshalb argumentiere ich mit der Richtung und nicht mit der Nachkommastelle.
- **Ich verspreche kein Ergebnis.** Ich verspreche ein Vorgehen, das dem entspricht, was sich in der Forschung als wirksam gezeigt hat, und eine ehrliche Einordnung, wenn etwas nicht funktioniert.

Ob es bei euch gewirkt hat, messe ich

Auf der vorigen Seite steht, dass es für ein 40-köpfiges Software-Unternehmen in Aachen keine Studie gibt. Das stimmt. Deshalb erzeuge ich in eurem Fall eigene Zahlen.

Was gemessen wird. Psychologische Sicherheit, das am besten belegte Einzelmerkmal funktionierender Teams und genau der Faktor, an dem der Sprint arbeitet. Es ist kein Ersatz für ein Ergebnis, aber der belastbarste verfügbare Indikator dafür, ob sich in der Zusammenarbeit etwas bewegt hat.

Womit. Dem **PsySafety-Check** von Fischer und Hüttermann, einem im deutschsprachigen Raum validierten Kurzinstrument, veröffentlicht in der Zusammenstellung sozialwissenschaftlicher Items und Skalen des GESIS-Leibniz-Instituts. Sieben Fragen, drei Minuten, anonym. Für englischsprachige Teams das Originalinstrument von Amy Edmondson, aus dem der deutsche Fragebogen abgeleitet ist.

Wann. Einmal vor der ersten Retrospektive, ein zweites Mal beim Follow-up-Check nach drei Monaten.
Bewusst nicht direkt nach dem Sprint: Ein Wert unmittelbar nach einem Workshop ist regelmäßig erhöht: frischer Eindruck, gute Stimmung, der Wunsch, dass es etwas gebracht hat. Interessant ist der Wert, wenn der Alltag drei Monate lang dagegen gearbeitet hat.

Was ich damit sagen werde – und was nicht

Ich werde sagen: „Der Wert ist von 4,1 auf 5,3 gestiegen.“ Das ist eine beschreibbare Veränderung, und ihr könnt sie nachvollziehen.

Ich werde **nicht** sagen: „signifikant verbessert“. Bei acht bis fünfzehn Personen gibt es keine belastbare Inferenzstatistik. Wer bei dieser Gruppengröße mit Signifikanz argumentiert, hofft, dass niemand nachfragt.

Das ist keine Schwäche des Verfahrens, sondern eine Frage des passenden Werkzeugs. Inferenzstatistik beantwortet, ob ein Befund über die untersuchte Gruppe hinaus gilt, ob er also auch für andere Teams zuträfe. Diese Frage stellt sich hier nicht: Es geht um euer Team, nicht um Teams im Allgemeinen. Dafür ist die Beschreibung kein schwächerer Ersatz, sondern das richtige Mittel.

Und wenn sich nichts bewegt hat, sage ich das auch. Dann reden wir darüber, woran es lag.

Was ihr davon habt, unabhängig von mir. Ihr bekommt einen Ausgangswert für euer Team, den ihr später erneut erheben könnt, mit einem frei verfügbaren, dokumentierten Instrument. Das bleibt euch, auch wenn wir nie wieder zusammenarbeiten.

Die Erhebung ist anonym; Auswertungen für Gruppen unter fünf Personen weise ich nicht getrennt aus. Dass ich die Werte anonymisiert und ohne Rückschluss auf euer Unternehmen für eigene Auswertungen verwenden darf, halten wir vorab schriftlich fest – oder eben nicht, dann bleiben sie bei euch.

Quellenverzeichnis

Alle Angaben sind so notiert, dass du sie direkt nachschlagen kannst.

- De Dreu, C. K. W., & Weingart, L. R. (2003). Task versus relationship conflict, team performance, and team member satisfaction: A meta-analysis. **Journal of Applied Psychology**, 88(4), 741–749.
- Detert, J. R., & Edmondson, A. C. (2011). Implicit voice theories: Taken-for-granted rules of self-censorship at work. **Academy of Management Journal**, 54(3), 461–488.
- Donovan, J. J., & Radosevich, D. J. (1999). A meta-analytic review of the distribution of practice effect: Now you see it, now you don't. **Journal of Applied Psychology**, 84(5), 795–805.
- Edmondson, A. C. (1999). Psychological safety and learning behavior in work teams. **Administrative Science Quarterly**, 44(2), 350–383.
- Fischer, N., & Hüttermann, H. (2020). PsySafety-Check (PS-C): Fragebogen zur Messung psychologischer Sicherheit in Teams. In **Zusammenstellung sozialwissenschaftlicher Items und Skalen (ZIS)**. GESIS – Leibniz-Institut für Sozialwissenschaften.
- Frazier, M. L., Fainshmidt, S., Klinger, R. L., Pezeshkan, A., & Vracheva, V. (2017). Psychological safety: A meta-analytic review and extension. **Personnel Psychology**, 70(1), 113–165.
- Gersick, C. J. G. (1988). Time and transition in work teams: Toward a new model of group development. **Academy of Management Journal**, 31(1), 9–41.
- Glasl, F. (2013). **Konfliktmanagement. Ein Handbuch für Führung, Beratung und Mediation** (11. Aufl.). Bern: Haupt. (Erstauflage 1980)
- Gollwitzer, P. M., & Sheeran, P. (2006). Implementation intentions and goal achievement: A meta-analysis of effects and processes. **Advances in Experimental Social Psychology**, 38, 69–119.
- Karau, S. J., & Williams, K. D. (1993). Social loafing: A meta-analytic review and theoretical integration. **Journal of Personality and Social Psychology**, 65(4), 681–706.
- Klein, C., DiazGranados, D., Salas, E., Le, H., Burke, C. S., Lyons, R., & Goodwin, G. F. (2009). Does team building work? **Small Group Research**, 40(2), 181–222.
- Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. **Psychological Bulletin**, 119(2), 254–284.
- Lu, L., Yuan, Y. C., & McLeod, P. L. (2012). Twenty-five years of hidden profiles in group decision making: A meta-analysis. **Personality and Social Psychology Review**, 16(1), 54–75.
- Milliken, F. J., Morrison, E. W., & Hewlin, P. F. (2003). An exploratory study of employee silence: Issues that employees don't communicate upward and why. **Journal of Management Studies**, 40(6), 1453–1476.
- Pettigrew, T. F., & Tropp, L. R. (2006). A meta-analytic test of intergroup contact theory. **Journal of Personality and Social Psychology**, 90(5), 751–783.
- Salas, E., DiazGranados, D., Klein, C., Burke, C. S., Stagl, K. C., Goodwin, G. F., & Halpin, S. M. (2008). Does team training improve team performance? A meta-analysis. **Human Factors**, 50(6), 903–933.
- Tannenbaum, S. I., & Cerasoli, C. P. (2013). Do team and individual debriefs enhance performance? A meta-analysis. **Human Factors**, 55(1), 231–245.
- Leblanc, P.-M., Harvey, J.-F., & Rousseau, V. (2024). A meta-analysis of team reflexivity: Antecedents, outcomes, and boundary conditions. **Human Resource Management Review**, 34(4), 101042.

Ein Hinweis in eigener Sache

Eine gern zitierte Zahl fehlt in diesem Dokument bewusst: dass 23 % der gescheiterten Startups „das falsche Team“ als Ursache nennen. Sie stammt aus Selbstauskünften von Gründern in Rückschauen, nicht aus einer kontrollierten Erhebung, und Menschen erklären das eigene Scheitern selten mit sich selbst. Als Aufhänger tauglich, als Beleg nicht.

Fragen zu einzelnen Befunden?

Dann schreib mir. Ich diskutiere lieber über eine Studie als über eine Broschüre.

Levi Neumann · M.Sc. Psychologie · IHK-zertifizierter Trainer

anfrage@levineumann.de · levineumann.de · Aachen