

Mehr als die Summe der Einzelteile: Teaminteraktion in IPA-Auswahlverfahren fundiert bewerten

Wie Auftraggeber die Kooperationsfähigkeit von Bieterteams auf Teamebene erfassen – acht wissenschaftliche Ansätze, sechs Teamkriterien und eine Kriterien-Übungs-Matrix für einen AC-Tag

Dr. Walter Lieberei

26.10.2026

Ein Bieterteam aus fünf hervorragenden Fachleuten ist noch kein hervorragendes Team. Diese Einsicht steht im Zentrum der Integrierten Projektabwicklung (IPA): Planende, Bauausführende und Auftraggeberin schließen sich in einem Mehrparteienvertrag zusammen, teilen Chancen und Risiken und entscheiden gemeinsam „Best for Project“. Ob eine solche Allianz trägt, hängt deshalb weniger davon ab, wie gut jede einzelne Person ist, als davon, wie gut die Beteiligten ihr Wissen, ihre Interessen und ihre Ressourcen miteinander verbinden. Folgerichtig wird die Kooperationsfähigkeit in IPA-Vergaben als Zuschlagskriterium gewertet – meist mithilfe eines Assessment-Centers.

Doch wie misst man die Kooperationsfähigkeit eines Teams fair und möglichst objektiv? Dieser Beitrag geht der Frage aus eignungsdiagnostischer Sicht nach. Er knüpft an unsere Fachbeiträge Nr. 33 („Partner auf dem Prüfstand“) und Nr. 34 („Die Blaupause für Erfolgspartnerschaften“) an und führt sie weiter: Wir stellen acht wissenschaftliche Ansätze vor, mit denen sich Interaktion und Gruppenergebnis erfassen lassen, leiten daraus sechs Teamkriterien und sechs baunahe Übungen ab und verdichten beides zu einer prototypischen Kriterien-Übungs-Matrix für einen AC-Tag mit vier bis sechs Personen.

1. Ausgangslage: Ein fundierter Leitfaden mit einer eignungsdiagnostischen Leerstelle

1.1 Warum Kooperationsfähigkeit zum Zuschlagskriterium wird

Beginnen wir mit einer Situation, wie sie derzeit in vielen Vergabestellen entsteht. Der folgende Fall ist fiktiv, aber typisch; er begleitet uns durch den Artikel. Die Nordwasser AöR, ein kommunaler Wasserver- und Abwasserentsorger, will ihre Kläranlage Hafenmühle im laufenden Betrieb erweitern: neue Belebungsbecken, ein zusätzlicher Faulbehälter, eine vierte Reinigungsstufe. Das Vorhaben ist technisch anspruchsvoll, terminlich eng getaktet und voller Schnittstellen zwischen Tiefbau, Verfahrenstechnik, Elektro- und Leittechnik sowie Betrieb. Die Projektleitung hat sich deshalb für ein IPA-Modell entschieden. Drei Bieterteams haben den Teilnahmewettbewerb überstanden; jetzt soll ein Assessment-Center zeigen, welches Team am besten zur künftigen Allianz passt.

Vergaberechtlich ist dieser Schritt gut begründbar. Kooperationsfähigkeit ist kein Eignungskriterium im Sinne von § 122 GWB, weil dieses auf Befähigung sowie wirtschaftliche, technische und berufliche Leistungsfähigkeit beschränkt ist. Sie kann aber als qualitatives Zuschlagskriterium nach § 127 GWB gewertet werden, weil sie unmittelbar mit dem Auftragsgegenstand – einem auf Kollaboration angelegten Mehrparteienvertrag – zusammenhängt. Zuschlagskriterien, Unterkriterien und ihre Gewichtung sind dabei vorab transparent bekanntzugeben (Rodde & Kersten, 2024, S. 22–24). Die Bedeutung ist erheblich: Der Endbericht schlägt vor, das Assessment-Center je nach Verfahrensvariante mit 30 bis 40 Prozent zu gewichten, und empfiehlt bei nur einer Verhandlungsrunde mindestens 30 Prozent (Rodde & Kersten, 2024, S. 24). Ein Instrument, das über ein Drittel eines Zuschlags im Millionenbereich entscheidet, muss methodisch belastbar sein.

1.2 Was der BBSR-Endbericht leistet

Die maßgebliche deutschsprachige Grundlage für die Gestaltung solcher Verfahren ist der Endbericht „Partnerauswahl und Teambuilding in der Projektallianz / Integrierten Projektabwicklung“, den die Lumico GmbH im For-

schungsprogramm Zukunft Bau für das Bundesinstitut für Bau-, Stadt- und Raumforschung erarbeitet hat (Rodde & Kersten, 2024). Der Bericht leistet viel. Er ordnet die IPA-Grundprinzipien, beschreibt das Verhandlungsverfahren mit vorgeschaltetem Teilnahmewettbewerb, klärt die Abgrenzung zwischen Eignungs- und Zuschlagskriterien, diskutiert Einzel- und Teambewerbungen, schlägt Gewichtungsvarianten vor und entwirft eine detaillierte Roadmap für das Teambuilding nach dem Zuschlag. Die baubetrieblichen, organisatorischen und vergaberechtlichen Fragen sind fundiert und praxisnah dargestellt.

Auch für das Auswahlverfahren selbst enthält der Bericht wertvolle Hinweise: Alle Bieterteams erhalten dieselben Aufgaben, jedes Unterkriterium soll in mindestens zwei Elementen beobachtet werden, die Beobachter werden geschult, bewerten zunächst unabhängig voneinander und dokumentieren ihre Eindrücke, und eine Beobachterkonferenz führt die Ergebnisse zusammen (Rodde & Kersten, 2024, S. 26–33). Das sind bewährte Bausteine guter Assessment-Center-Praxis.

1.3 Wo Klärungsbedarf besteht: Kapitel III.2 aus eignungsdiagnostischer Sicht

Deutlich weniger fundiert ist dagegen das Kapitel III.2 „Ablauf des Auswahlverfahrens zur Bewertung der Kooperationsfähigkeit“ (Rodde & Kersten, 2024, S. 25–33) – und zwar genau dort, wo es um die eignungsdiagnostische Substanz geht: Was genau wird gemessen, auf welcher Ebene, mit welchen Anker, und wie entsteht aus vielen Einzelbeobachtungen ein belastbares Urteil? Das ist kein Vorwurf an einen Bericht, dessen Schwerpunkt erkennbar auf Bau und Recht liegt. Es ist aber ein Klärungsbedarf, den jede Auftraggeberin kennen sollte, bevor sie ein Verfahren mit 30 Prozent Zuschlagsgewicht aufsetzt. Die folgende Tabelle fasst die wichtigsten Punkte zusammen.

Aussage im Endbericht	Eignungsdiagnostische Einordnung	Klärungsbedarf
Unterkriterien wie Kommunikationsfähigkeit, Führungskompetenz, Empathie oder „energetische Präsenz“ (Tabellen 4 und 5, S. 25)	Es handelt sich überwiegend um Merkmale einzelner Personen; die Bewertung erfolgt laut Bericht für „die Teilnehmer“ (S. 33).	Wie wird aus Personenurteilen ein Teamurteil? Eine Regel für die Aggregation fehlt.
„Teamdynamik“ mit 15 von 30 Prozent (Tabelle 4)	Das einzige explizit teambezogene Kriterium trägt die Hälfte des Gewichts, ist aber nicht definiert und nicht verankert.	Operationalisierung als beobachtbare kollektive Prozesse und Ergebnisse
Haltung als „zeit- und situationsunabhängige“ Grundeinstellung (S. 13), die im AC bewertet wird	Ein AC beobachtet Verhalten in Situationen. Der Rückschluss auf eine stabile innere Haltung ist eine weitreichende Interpretation.	Bewertung auf beobachtbares Verhalten und Teamleistung beschränken
Beobachtungshorizont „positive Einstellung“ mit sechs Stufen (Tabelle 6, S. 26)	Die Anker beschreiben Eindrücke der Beobachter („Sie nehmen eine Abwehr wahr“) und bündeln je Stufe drei Aspekte.	Verhaltensverankerte, eindimensionale Anker auf Teamebene
Realitätsnahe Aufgaben stellen sicher, dass „gemessen wird, was gemessen werden soll“ (S. 29–30)	Realitätsnähe stützt die inhaltliche Validität, belegt aber nicht die Vorhersage späteren Projekterfolgs.	Pilotierung und spätere Prognoseprüfung einplanen
Teilnehmende könnten erwünschtes Verhalten „nur sehr kurz“ vorspielen (S. 31)	Plausible Annahme, aber im Bericht nicht belegt	Vorspielen durch Aufgabenkomplexität und mehrere Situationen erschweren, nicht ausschließen
Beobachterkonferenz erzeugt ein „gemeinsames Bewertungsergebnis“ (S. 33)	Konsensdiskussionen sind anfällig für Status- und Reihenfolgeeffekte.	Regelgeleitete Integration der unabhängigen Ratings, Diskussion nur bei definierten Abweichungen

Tabelle 1: Eignungsdiagnostischer Klärungsbedarf in Kapitel III.2 des BBSR-Endberichts (eigene Darstellung auf Basis von Rodde & Kersten, 2024).

Hinzu kommen zwei grundsätzliche Punkte. Erstens räumt der Bericht selbst ein, dass es „nicht möglich“ sei, Haltung oder Verhalten objektiv zu beobachten, und spricht von „subjektiver Objektivität“ (S. 28). Die Eignungsdiagnostik verfügt jedoch über erprobte Mittel, um Objektivität in Durchführung, Auswertung und Interpretation systematisch zu erhöhen – die DIN 33430 und die internationalen AC-Leitlinien beschreiben sie ausführlich (Diagnostik- und Testkura-

torium, 2018; International Taskforce on Assessment Center Guidelines, 2015). Zweitens bleibt offen, wie die Unterkriterien aus den Anforderungen eines konkreten Projekts abgeleitet werden; der Haltungskatalog (Tabelle 5) stammt aus einem Beratungsmodell zur Führung („MAIN 7“), der Kompetenzkatalog (Tabelle 4) wird ohne Herleitung aus einer Anforderungsanalyse von IPA-Teams eingeführt (Rodde & Kersten, 2024, S. 13–14).

Der Kern des Klärungsbedarfs lässt sich in einem Satz fassen: **Das Verfahren will die Kooperationsfähigkeit eines Teams bewerten, misst aber überwiegend Eigenschaften von Personen.** Damit sind wir bei der zentralen Frage dieses Beitrags.

2. Vom Ich zum Wir: Das Team als Bewertungseinheit

2.1 Das Ganze ist mehr – oder weniger – als die Summe seiner Teile

Der Satz „Das Ganze ist mehr als die Summe seiner Teile“ ist eine Kernidee der Gestaltpsychologie und beschreibt treffend, worauf es in einer Projektallianz ankommt. Ein Team erzeugt Ergebnisse, die keine einzelne Person allein hervorbringen könnte: einen Ablaufplan, der Tiefbau, Verfahrenstechnik und Betrieb widerspruchsfrei verbindet; eine Kostenentscheidung, in die verteilte Informationen aus allen Gewerken eingeflossen sind; eine Störungsbewältigung, bei der jemand den entscheidenden Hinweis gibt, ein anderer ihn aufgreift und ein dritter die Konsequenzen für seine Schnittstelle prüft.

Wichtig ist allerdings die Umkehrung: Ein Team kann auch **weniger** sein als die Summe seiner Teile. Koordinationsverluste, Dominanz einzelner Personen, ungeprüfte Annahmen oder das Zurückhalten von Informationen können dazu führen, dass eine Gruppe schlechter abschneidet als ihr bestes Mitglied. Ein Experiment von Almaatouq et al. (2021) zeigt, dass interagierende Gruppen bei komplexen Aufgaben effizienter sein können als das effizienteste Einzelmitglied, bei einfachen Aufgaben aber nicht – Synergiegewinne und Prozessverluste halten sich je nach Aufgabenkomplexität unterschiedlich die Waage. Der Mehrwert eines Teams ist also keine Selbstverständlichkeit, sondern eine **prüfbare Hypothese**. Genau das macht ihn eignungsdiagnostisch interessant: Er lässt sich beobachten.

Für die Nordwasser AöR heißt das: Wenn sie fünf Personen eines Bieterteams einzeln bewertet und die Punktwerte addiert, erfährt sie wenig darüber, ob diese fünf Personen gemeinsam einen tragfähigen Plan für die Erweiterung der Hafenumühle entwickeln würden. Fünf gute Einzelwerte können ein schwaches Team verdecken – und umgekehrt kann ein Team mit unauffälligen Einzelpersonen außergewöhnlich gut zusammenspielen.

2.2 Input, Prozess, Output: drei Ebenen auseinanderhalten

Um die Teamebene sauber zu erfassen, hat sich in der Teamforschung die Unterscheidung von Inputs, Prozessen und Outputs bewährt. Marks, Mathieu und Zaccaro (2001) haben sie präzisiert, indem sie Teamprozesse – also das, was Teammitglieder miteinander tun – von emergenten Zuständen unterschieden, die aus der Zusammenarbeit entstehen, etwa ein gemeinsames Lagebild oder Vertrauen. Für die Eignungsdiagnostik ergeben sich daraus vier Begriffe, die nicht vermischt werden dürfen.

Ebene	Bedeutung	Beispiel im IPA-Kontext
Inputs	Ressourcen und Bedingungen, mit denen ein Team eine Aufgabe beginnt	Fachexpertise, Rollen, Erfahrung, Entscheidungsbefugnisse, Informationszugang
Teamprozesse	Aufeinander bezogene Handlungen, durch die das Team seine Aufgabe bearbeitet	Informationen prüfen, Abhängigkeiten abstimmen, Entscheidungen revidieren
Emergente Zustände	Aus Zusammenarbeit entstehende kognitive, motivationale oder affektive Eigenschaften des Teams	Gemeinsames Lagebild, psychologische Sicherheit, kollektive Zuversicht
Outputs	Gemeinsam erzeugte, bewertbare Ergebnisse	Tragfähige Zielkostenvariante, ausführbarer Ablaufplan, bewältigte Störung

Tabelle 2: Vier Ebenen der Teamdiagnostik (in Anlehnung an Marks et al., 2001).

Die Unterscheidung hat praktische Konsequenzen. Inputs wie Fachkompetenz, Rollenbesetzung und Verfügbarkeit bleiben notwendige Voraussetzungen; sie werden aber in den technischen Konzepten und im Personaleinsatzkonzept geprüft. Das Assessment-Center sollte sich auf das konzentrieren, was nur dort sichtbar wird: die **Prozesse** des Zusammenwirkens und die **Outputs**, die daraus entstehen. Umgekehrt darf eine überzeugende Zusammenarbeit gravierende fachliche Lücken nicht verdecken – deshalb gehört die Qualität des gemeinsamen Ergebnisses ausdrücklich zur Bewertung.

2.3 Konsequenzen für die Bewertungslogik

Wer das Team als Bewertungseinheit ernst nimmt, muss drei Dinge ändern. Erstens werden die Kriterien als **kollektive Vorgänge** formuliert: nicht „Herr M. kommuniziert wertschätzend“, sondern „Widersprechende Informationen aus mehreren Rollen werden geprüft und im gemeinsamen Plan berücksichtigt“. Zweitens müssen die Übungen echte **wechselseitige Abhängigkeit** erzeugen: Eine Reihe paralleler Kurzpräsentationen einzelner Teammitglieder liefert keine Teamdaten. Drittens braucht das Verfahren eine **Messlogik**, die festlegt, wie aus den Beobachtungen mehrerer Beobachter über mehrere Übungen ein Teamprofil entsteht – ohne den Umweg über die Summe von Personenurteilen. Welche wissenschaftlichen Ansätze dafür zur Verfügung stehen, zeigt das folgende Kapitel.

3. Acht Ansätze zur Erfassung von Interaktion und Gruppenergebnis

Die folgenden acht Ansätze sind eine analytische Ordnung für die Verfahrensgestaltung, keine acht voneinander unabhängigen Schulen. Manche beschreiben Konstrukte, andere Messlogiken oder zeitliche Perspektiven. Gerade ihre Kombination ist für ein IPA-Assessment-Center sinnvoll. Für jeden Ansatz fragen wir: Was wird erfasst, wie wird gemessen, und welche Rolle kann er im Auswahlverfahren spielen?

3.1 Ansatz A: Kompositions- und Konfigurationsdiagnostik

Der erste Ansatz fragt, welche personellen und organisatorischen Voraussetzungen ein Team mitbringt und ob deren Verteilung zur Aufgabe passt. Entscheidend ist dabei nicht der Durchschnitt, sondern die Konfiguration: Durchschnittlich hohe Fachkompetenz nützt wenig, wenn an einer kritischen Schnittstelle – etwa zwischen Verfahrenstechnik und Leittechnik – niemand zuständig ist. Gemessen wird über Rollen- und Kompetenzabdeckung, Erfahrung an kritischen Schnittstellen, Kapazitäten, Befugnisse und Vertretungsregeln. Die Forschung zur kollektiven Intelligenz zeigt zugleich, dass individuelle Fähigkeiten je nach Aufgabentyp unterschiedlich stark zur Gruppenleistung beitragen (Riedl et al., 2021). Für IPA ist der Ansatz hoch relevant zur Sicherung notwendiger Voraussetzungen, aber ungeeignet als Ersatz für die Beobachtung des Zusammenwirkens: Ein Input-Profil ist noch kein Teamleistungswert.

3.2 Ansatz B: Mehrdimensionale Teamprozessdiagnostik

Der zweite Ansatz untersucht, wie ein Team seine aufeinander angewiesenen Beiträge organisiert. Grundlage ist die Taxonomie von Marks et al. (2001), die zehn Teamprozesse in drei Bereiche gliedert: Transitionsprozesse in Planungsphasen, Aktionsprozesse in Ausführungsphasen und interpersonelle Prozesse, die durchgängig wirken. Für ein AC lassen sich diese Prozesse in Verhaltensbeobachtung übersetzen.

Prozessbereich	Zugeordnete Prozesse	Übersetzung in IPA-Verhalten
Transition	Auftragsanalyse, Zielspezifikation, Strategieformulierung und Planung	Zielkonflikte offenlegen, Erfolgsmaßstäbe klären, Entscheidungsweg vereinbaren
Aktion	Fortschrittsmonitoring, Systemmonitoring, gegenseitige Unterstützung, Koordination	Risiken und Änderungen beobachten, Engpässe erkennen, Beiträge abstimmen
Interpersonell	Konfliktmanagement, Motivations- und Vertrauensaufbau, Affektmanagement	Gegenargumente sachlich bearbeiten, unter Druck arbeitsfähig bleiben

Tabelle 3: Teamprozesse nach Marks et al. (2001) und ihre Übersetzung in IPA-Verhalten.

Gemessen wird in mehreren Episoden mit standardisierten Anlässen und verhaltensverankerten Teamratings. Für IPA ist der Ansatz der Kern des Beobachtungsmodells. Allerdings sollten die zehn Forschungsdimensionen nicht ungeprüft als zehn gleich zuverlässige Auswahlskalen verwendet werden; für einen AC-Tag sind wenige, klar unterscheidbare Bewertungsbereiche mit wiederholten Beobachtungsebenen zweckmäßiger.

3.3 Ansatz C: Emergente Teamzustände

Der dritte Ansatz richtet den Blick auf Zustände, die aus der Zusammenarbeit entstehen und weitere Leistung ermöglichen: gemeinsame mentale Modelle über Ziele und Abhängigkeiten, ein organisiertes Wissen darüber, wer welche Expertise besitzt (transaktives Gedächtnis), psychologische Sicherheit und kollektive Zuversicht. Im AC lassen sich vor allem kognitive Zustände prüfen, etwa durch eine kurze, getrennte Lagebildabfrage nach einer Arbeitsphase: Hat jedes Teammitglied dasselbe – und vor allem ein **zutreffendes** – Bild von Zielen, kritischen Abhängigkeiten und Zuständigkeiten? Richtigkeit und Übereinstimmung sind dabei getrennt zu bewerten, denn ein gemeinsam geteilter Irrtum ist kein gutes mentales Modell. Für dauerhafte Aussagen über Vertrauen oder Teamkultur ist ein eintägiges Auswahlsetting dagegen nur eingeschränkt geeignet. Formulierungen wie „In dieser Situation wurden Einwände aufgegriffen“ sind belastbar, „Dieses Team besitzt psychologische Sicherheit“ ist es nicht.

3.4 Ansatz D: Interaktions-, Sequenz- und Netzwerkdiagnostik

Der vierte Ansatz fragt, welche Muster zwischen den Beteiligten entstehen und welche Folgen einzelne Beiträge für den weiteren Verlauf haben. Oostrom, Lehmann-Willenbrock und Klehe (2019) haben vorgeschlagen, die in der Teamforschung etablierte feinkörnige Interaktionsanalyse auf Assessment-Center zu übertragen: Verhaltensbeiträge werden zunächst beschrieben und kodiert, nicht sofort bewertet, was die Subjektivität der Urteile reduziert. Die Autoren demonstrieren das Verfahren an 30 Dreiergruppen; die höhere Prognosegüte formulieren sie ausdrücklich als zu prüfende Annahme. Gestützt wird der Ansatz durch die Metaanalyse von Marlow et al. (2018) über 150 Studien mit 9.702 Teams: Die Qualität der Kommunikation hängt deutlich stärker mit Teamleistung zusammen als ihre Häufigkeit, und die Informationselaboration – das gemeinsame Verarbeiten und Weiterentwickeln von Informationen – zeigt den stärksten Zusammenhang. Für die Praxis heißt das: Nicht Redeanteile zählen, sondern Sequenzen. Die Folge „Risiko wird benannt → Rückfrage → Prüfung → Planänderung → Zuständigkeit bestätigt“ enthält mehr diagnostische Information als die Feststellung, dass fünf Personen jeweils mehrfach gesprochen haben. Weil eine Vollkodierung aufwendig ist, empfiehlt sich im AC die Erfassung weniger, vorab definierter kritischer Sequenzen.

3.5 Ansatz E: Direkte kollektive Leistungs- und Synergiediagnostik

Der fünfte Ansatz bewertet, was das Team gemeinsam leistet. Das Team erhält einen Ergebniswert für seinen Plan, seine Entscheidung oder seine Vorlage – gemessen an fachlicher Richtigkeit, erfüllten Randbedingungen, gelösten Zielkonflikten, Zeitbedarf und Robustheit gegenüber veränderten Annahmen. Dafür muss nicht erst jeder Person ein Anteil zugerechnet werden. Darüber hinaus lässt sich der kollektive Mehrwert konzeptionell bestimmen:

Kollektiver Mehrwert = beobachtete Teamleistung – Leistung einer begründeten Vergleichsbedingung

Die Vergleichsbedingung muss zur Aufgabe passen, etwa die beste unabhängig erarbeitete Einzellösung oder die regelgeleitete Zusammenführung unabhängig erstellter Beiträge (Almaatouq et al., 2021). Ein Vorteil gegenüber dem Durchschnitt ist ein schwächerer Nachweis als ein Vorteil gegenüber der besten Einzelleistung. Für IPA ist die direkte Outputbewertung unverzichtbar; ein eigenständiger Synergieindex eignet sich dagegen zunächst als ergänzende Information für die Pilotierung, nicht als gewichtetes Zuschlagskriterium.

3.6 Ansatz F: Kollektive Intelligenz

Der sechste Ansatz fragt, ob es einen gemeinsamen Faktor gibt, der erklärt, weshalb dasselbe Team über verschiedene Aufgaben hinweg gut abschneidet. Woolley et al. (2010) fanden in zwei Studien mit 699 Personen in Gruppen von zwei bis fünf Personen einen solchen „c-Faktor“, der über 43 Prozent der Varianz der Gruppenleistungen erklärte. Er hing nicht stark mit der durchschnittlichen oder maximalen Intelligenz der Mitglieder zusammen, wohl aber mit

ihrer sozialen Sensibilität und mit einer gleichmäßigeren Verteilung der Gesprächsbeiträge. Riedl et al. (2021) bestätigten in einer gemeinsamen Analyse von 22 Studien mit 5.279 Personen in 1.356 Gruppen einen robusten Faktor, der auch Leistungen in weiteren Aufgaben vorhersagte, und fanden, dass der Kooperationsprozess insgesamt bedeutsamer war als die Fähigkeiten der Einzelnen. Kritische Reanalysen bezweifeln allerdings die Stärke des Faktors: Credé und Howardson (2017) kommen nach Prüfung von sechs Stichproben zu dem Schluss, dass der allgemeine Faktor bei vielen Aufgaben nur wenig Varianz erklärt.

Für IPA ist der Ansatz deshalb als Denkraum wertvoll – er begründet, warum Interaktionsqualität zählt –, aber als alleinige Zuschlagsgrundlage ungeeignet. Wenige Bieterteams reichen nicht für eine stabile Faktorenanalyse, und ein frei gewichteter AC-Gesamtwert darf nicht allein deshalb „c-Faktor“ heißen. Ausdrücklich gilt zudem: Demografische Merkmale wie der Frauenanteil, die in diesen Studien korrelativ mit kollektiver Intelligenz zusammenhängen, sind keine zulässigen Auswahlregeln.

3.7 Ansatz G: Teamlernen und Anpassung

Der siebte Ansatz untersucht, wie ein Team sein Vorgehen bei neuen Informationen, Fehlern oder Feedback verändert. Die Metaanalyse von Leblanc, Harvey und Rousseau (2024) verbindet Teamreflexivität – das gemeinsame Hinterfragen von Zielen, Strategien und Prozessen – positiv mit Teamleistung und zeigt zugleich, dass Rahmenbedingungen wie Teamgröße und gemeinsame Bestandsdauer den Zusammenhang beeinflussen. Diagnostisch eignet sich eine Abfolge aus Ausgangsaufgabe, standardisierter Störung, festgelegtem Reflexionsfenster und vergleichbarer Anschlussaufgabe. Beobachtet wird, ob das Team den Anpassungsbedarf erkennt, eine Änderung auswählt, sie gemeinsam umsetzt und in der nächsten Episode ein besseres Ergebnis erzielt. Die bloße Absichtserklärung, künftig besser zu kommunizieren, ist kein Lernnachweis. Für IPA ist der Ansatz sehr relevant, weil Änderungen, Zielkonflikte und neue Beteiligte zum Projektalltag gehören. Alle Teams erhalten dabei dieselbe Hilfestellung, und Lerngewinne werden neben dem absoluten Ergebnisniveau berichtet, nicht an seiner Stelle.

3.8 Ansatz H: Multiteam- und Schnittstellendiagnostik

Der achte Ansatz weitet den Blick über das einzelne Team hinaus. Eine Projektallianz ist ein Multiteam-System: ein Netzwerk voneinander abhängiger Teams mit eigenen Teilzielen und gemeinsamen übergeordneten Zielen. Zaccaro et al. (2020) ordnen die Forschung um drei Kernbereiche – Koordination innerhalb und zwischen Teams, Führungsstrukturen sowie kognitive und motivationale Zustände – und zeigen, dass Teamgrenzen, Distanz zwischen den Teambenutzern und Art der übergeordneten Ziele die Zusammenarbeit prägen. Gemessen werden Informationsübergaben zwischen Fachgruppen, die Abstimmung von Teilplänen, der Umgang mit widersprechenden Prioritäten und die Eskalation über Entscheidungsebenen hinweg. Das Kriterium ist die Funktionsfähigkeit des Gesamtsystems: Ein Teambenutzer kann seine eigenen Ziele hervorragend erfüllen und dabei das Projekt beeinträchtigen. Da die reale Gesamtallianz im Auswahlverfahren meist noch nicht existiert, wird die Anschlussfähigkeit des Bieterteams über standardisierte Gegenrollen geprüft – etwa die Projektleitung der Auftraggeberin, den Betrieb oder einen Fachplaner.

3.9 Die acht Ansätze im Vergleich

Ansatz	Primärer Gegenstand	Eignung für IPA	Empfohlene Rolle im AC
A – Komposition	Inputs und ihre Verteilung	Hoch für Voraussetzungen	Vorab im Personaleinsatzkonzept prüfen
B – Teamprozesse	Koordination und gemeinsame Bearbeitung	Sehr hoch	Kern der Verhaltensbeobachtung
C – Emergente Zustände	Lagebild, Expertiseorganisation	Hoch, Kulturaussagen eingeschränkt	Lagebildabfrage als Ergänzung
D – Interaktion und Sequenzen	Wechselseitige, zeitliche Muster	Sehr hoch, aber aufwendig	Kritische Sequenzen protokollieren
E – Output und Synergie	Gemeinsame Ergebnisse, Mehrwert	Sehr hoch für Outputs	Unverzichtbare Ergebnisbewertung
F – Kollektive Intelligenz	Generalisierte Leistungsfähigkeit	Begrenzt als Auswahlwert	Denkrahmen, kein „Team-IQ“
G – Lernen und Anpassung	Veränderung über Episoden	Sehr hoch	Mindestens eine Störungs- und Transfersequenz
H – Multiteam-System	Integration über Teamgrenzen	Sehr hoch bei Allianzen	Standardisierte Gegenrollen

Tabelle 4: Die acht Ansätze im Vergleich (Eignungseinstufungen als fachliche Transferurteile, keine metaanalytisch geschätzte Rangfolge).

Für das Auswahlverfahren ergibt sich daraus eine klare Architektur: Die Ansätze B, D und H liefern das Beobachtungsmodell für die Prozesse, Ansatz E die Ergebnisbewertung, Ansatz G die dynamische Perspektive und Ansatz C eine kognitive Ergänzung. Ansatz A gehört in die Prüfung der Voraussetzungen, Ansatz F bleibt ein Forschungsrahmen.

4. Von den IPA-Anforderungen zu den Kriterien

4.1 Was eine Projektallianz von ihren Teams verlangt

Kriterien dürfen nicht aus einem allgemeinen Katalog übernommen werden; sie müssen aus den Anforderungen der Tätigkeit abgeleitet werden. Das ist der Kern einer Anforderungsanalyse nach DIN 33430 (Diagnostik- und Testkuratorium, 2018) und eine Grundforderung der internationalen AC-Leitlinien (International Taskforce on Assessment Center Guidelines, 2015). Für IPA lassen sich sechs typische Anforderungen benennen, aus denen sich je ein beobachtbarer kollektiver Mechanismus und ein bewertbares Ergebnis ableiten lassen.

IPA-Anforderung	Relevanter Input	Beobachtbarer kollektiver Mechanismus	Bewertbarer Output
Best for Project / Zielkosten	Fachwissen, Kostenbasis, Mandate	Ziele und Zielkonflikte explizit abgleichen, Alternativen gemeinsam prüfen	Begründete Variante mit konsistenten Kosten-, Termin- und Qualitätsfolgen
Open Book / Risikomanagement	Verteilte Kosten- und Risikoinformationen	Daten offenlegen, Unsicherheit markieren, Gegenprüfung zulassen	Vollständiges Risikoregister mit Maßnahmen und Verantwortlichen
Lean / verlässliche Zusagen	Ressourcen- und Ablaufwissen	Voraussetzungen prüfen, Übergaben abstimmen, Engpässe gemeinsam lösen	Ausführbarer Ablaufplan mit belastbaren Zusagen
Fachübergreifende Planung / BIM	Komplementäre Expertise	Fachbeiträge verknüpfen, betroffene Schnittstellen einbeziehen	Konsistente Entscheidung ohne ungelöste kritische Kollision
Konfliktlösung / kollektive Führung	Entscheidungsregeln, Zuständigkeiten	Einwände prüfen, Expertise situativ führen lassen, Eskalation begründen	Entscheidung oder geregelte Eskalation mit klarer Umsetzung
Anpassung / Lernen	Ausgangsplan, neue Information	Fehlerursache gemeinsam analysieren, Vorgehen ändern und anwenden	Wiederhergestellte Qualität und wirksamer Transfer

Tabelle 5: Ableitung beobachtbarer Teammechanismen aus IPA-Anforderungen (Entwicklungsvorschlag, keine validierte Skala).

4.2 Was ein gutes Teamkriterium auszeichnet

Aus der Tabelle lassen sich Kriterien gewinnen, die fünf Anforderungen erfüllen sollten. Sie sind **anforderungsbezogen**, also aus der Projektaufgabe abgeleitet. Sie sind **kollektiv formuliert** und beschreiben die Verbindung von Beiträgen, nicht Sympathie, Selbstsicherheit oder Redeanteile einzelner Personen. Sie sind **beobachtbar** und lassen sich mit Verhaltensankern abstufen. Sie sind **voneinander unterscheidbar**, damit nicht derselbe Gesamteindruck mehrfach benannt wird. Und sie sind **fair**, weil sie keinen Stil bevorzugen, der mit der Aufgabe nichts zu tun hat – ein ruhiges Team, das Informationen sorgfältig integriert, darf gegenüber einem lebhaften Team, das viel redet, keinen Nachteil haben.

4.3 Sechs Kriterien der kollektiven Kooperationsfähigkeit

Auf dieser Basis schlagen wir sechs Kriterien vor: fünf Prozesskriterien und ein Ergebniskriterium. Die folgende Tabelle zeigt die Verhaltensanker für die Stufen 1, 3 und 5 einer fünfstufigen Skala; die Zwischenstufen 2 und 4 werden vergeben, wenn das beobachtete Verhalten zwischen zwei Anker liegt.

Kriterium	Stufe 1 – schwach	Stufe 3 – mittel	Stufe 5 – stark
K1 Informationsintegration	Relevante Einzelinformationen bleiben ungenutzt; Widersprüche werden übergangen.	Informationen werden gesammelt, aber nur teilweise gegeneinander geprüft.	Das Team verbindet relevante Beiträge, klärt Widersprüche und ändert nachvollziehbar seine Lösung.
K2 Koordination und gegenseitige Absicherung	Parallelarbeit ohne geklärte Übergaben; Zusagen widersprechen sich.	Rollen sind verteilt, Abhängigkeiten werden teilweise überwacht.	Voraussetzungen und Übergaben sind abgestimmt; Engpässe werden erkannt, Unterstützung wird organisiert.
K3 Gemeinsame Entscheidung und Konfliktbearbeitung	Dominanz, Vermeidung oder vorschneller Konsens; Einwände bleiben unbearbeitet.	Alternativen werden diskutiert, Entscheidungsmaßstäbe bleiben teilweise unklar.	Einwände werden sachlich geprüft; die Entscheidung folgt gemeinsamen Kriterien und wird verbindlich festgehalten.
K4 Kollektive Anpassung und Lernen	Die Störung wird ignoriert oder einer Person zugeschoben.	Eine Änderung wird beschlossen, aber nicht vollständig umgesetzt.	Das Team aktualisiert sein Vorgehen gemeinsam und zeigt die Veränderung in der Folgeaufgabe.
K5 Integration über Teamgrenzen	Eigene Teilziele werden zulasten anderer Schnittstellen optimiert.	Nachbarbereiche werden informiert, ihre Anforderungen teilweise berücksichtigt.	Abhängigkeiten und Prioritäten werden mit den Gegenrollen abgestimmt; die Lösung ist anschlussfähig.
K6 Gemeinsame Ergebnisqualität	Wesentliche Randbedingungen verletzt, kritische Abhängigkeiten ungelöst	Randbedingungen überwiegend erfüllt, einzelne Lücken	Fachlich richtig, vollständig, robust und umsetzbar

Tabelle 6: Sechs Teamkriterien mit Verhaltensankern (K1–K5 Prozesse, K6 Ergebnis; Anker vor dem Einsatz anhand von Aufzeichnungen zu kalibrieren).

Die Kriterien lassen sich mit dem Endbericht gut verbinden. Kommunikationsfähigkeit geht in der Informationsintegration (K1) auf, Problem- und Konfliktlösung in K3, Veränderungsbereitschaft und Lernfähigkeit in K4, Führungskompetenz als situativ geteilte Führung in K3 und K5, Motivation und Zielbewusstsein in der Orientierung am gemeinsamen Ergebnis (K6). Die bislang undefinierte „Teamdynamik“ erhält durch K1 bis K5 erstmals einen beobachtbaren Inhalt. Für jedes Kriterium ist zusätzlich die Kategorie „**nicht beobachtbar**“ vorzusehen. Ein Team, in dem kein Konflikt auftrat, erhält nicht automatisch den besten Konfliktwert – deshalb werden Einwände in den Übungen standardisiert ausgelöst.

5. Praxisnahe Übungen im Baukontext

5.1 Konstruktionsprinzipien

Die Kriterien entfalten ihren Wert nur, wenn die Übungen die beabsichtigten kollektiven Anforderungen tatsächlich auslösen. Fünf Prinzipien sind dafür maßgeblich. Erstens **echte Interdependenz**: Die Teammitglieder erhalten unterschiedliche Rolleninformationen zu Kosten, Ausführung, Termin, Betrieb und Risiken, sodass die Aufgabe nur gemeinsam lösbar ist. Zweitens **standardisierte Auslöser**: Störungen, Einwände und neue Informationen werden zu festgelegten Zeitpunkten und mit identischen Skripten eingespielt. Drittens **prüfbare Ergebnisse**: Jede Übung endet mit einem Arbeitsprodukt, das sich anhand einer vorab erstellten Prüfliste bewerten lässt. Viertens **Alltagsnähe statt Extremsituation**: Der Endbericht betont zu Recht, dass übliches Verhalten beobachtet werden soll; künstliche Überforderung erzeugt Stressreaktionen, keine Kooperationsdaten. Fünftens **gleiche Bedingungen**: Material, Zeit, Moderation und Rollenspiel sind für alle Bieterteams identisch, und die Moderation lenkt die Zusammenarbeit nicht.

5.2 Sechs Übungen für den AC-Tag

Für die Nordwasser AÖR lässt sich daraus ein durchgehender Fall „Erweiterung der Kläranlage Hafenmühle“ mit sechs Übungen entwickeln. Ein fortlaufender Fall erhöht den Realismus und spart Einarbeitungszeit; damit die Übungen dennoch unterschiedliche Situationen abbilden, wechseln Aufgabentyp, Auslöser und Gegenüber.

Ü1 – Gemeinsame Auftragsklärung und Lagebild (45 Minuten). Jedes Teammitglied erhält ein eigenes Rollendossier, etwa als künftige Projektleitung, Bauleitung, Polier, Verfahrenstechnik oder Lean-/BIM-Koordination. Nach einer kurzen unabhängigen Vorbereitung erarbeitet das Team ein gemeinsames Lagebild mit priorisierten Projektzielen. Im Anschluss beantworten alle getrennt fünf Fragen zu Zielen, kritischen Abhängigkeiten und Zuständigkeiten (Ansatz C).

Ü2 – Zielkosten-Variantenentscheidung im Open-Book-Modus (60 Minuten). Das Team entscheidet zwischen drei Varianten für den neuen Faulbehälter – Sanierung im Bestand, Ersatzneubau oder modulare Lösung. Die Kosten-, Termin- und Risikodaten sind über die Rollen verteilt; einzelne Informationen widersprechen sich bewusst. Vor der Teamphase notiert jede Person unabhängig eine Empfehlung, sodass sich das Teamergebnis mit der besten Einzelpfehlung vergleichen lässt (Ansatz E). Output: begründete Variante mit Risikoregister.

Ü3 – Lean-Ablaufplanung mit Gewerkeschnittstellen (60 Minuten). Das Team erstellt im Stil einer Pull-Planung einen Ablaufplan für die Bauphase im laufenden Klärbetrieb. Zwei Rollenspieler vertreten die Elektro- und Leittechnik sowie den Anlagenbetrieb und bringen zu definierten Zeitpunkten Einschränkungen ein. Beobachtet werden Übergaben, Zusagen und die Abstimmung mit den Gegenrollen. Output: ausführbarer Ablaufplan mit Zusagen und Hindernisliste.

Ü4 – Störung im Bauablauf (45 Minuten). Eine neue Information entwertet einen Teil des Plans: Bei Erdarbeiten wird eine nicht dokumentierte Altleitung gefunden, zugleich kündigt der Wetterdienst Starkregen an, der die Baugrube und die Zulaufmengen der Anlage betrifft. Das Team muss den Plan revidieren oder begründet eskalieren. Output: revidierter Plan beziehungsweise Eskalationsvorlage.

Ü5 – Schnittstellen- und Eskalationsgespräch (30 Minuten). Das Team vertritt seine Lösung aus Ü4 gegenüber zwei standardisierten Gegenrollen: der Projektleitung der Auftraggeberin und der Betriebsleitung. Beide bringen nach Skript einen Zielkonflikt ein – Betriebssicherheit gegen Terminziel. Hier zeigt sich die Integration über Teamgrenzen und die Qualität gemeinsamer Entscheidungen unter Gegendruck (Ansatz H).

Ü6 – Teamreflexion und Transferaufgabe (50 Minuten). Nach einem standardisierten Reflexionsformat von 15 Minuten, das für alle Teams identisch ist, bearbeitet das Team eine strukturgleiche Anschlussaufgabe – etwa die Ablaufplanung für einen zweiten Bauabschnitt. Bewertet wird, ob die erkannten Verbesserungen tatsächlich in veränderter Bearbeitung und besserem Ergebnis sichtbar werden (Ansatz G).

Übung	Bezeichnung und IPA-Bezug	Dauer	Gemeinsames Arbeitsergebnis	Ansätze
Ü1	Auftragsklärung und Lagebild (Big-Room-Start)	45 Min.	Gemeinsames Lagebild, priorisierte Ziele	B, C, D
Ü2	Zielkosten-Variantenentscheidung (Open Book, Target Value Design)	60 Min.	Begründete Variante, Risikoregister	B, D, E
Ü3	Lean-Ablaufplanung mit Gewerkeschnittstellen	60 Min.	Ablaufplan mit Zusagen und Hindernisliste	B, D, E, H
Ü4	Störung im Bauablauf	45 Min.	Revidierter Plan oder Eskalationsvorlage	B, D, E, G
Ü5	Schnittstellen- und Eskalationsgespräch	30 Min.	Abgestimmte Vereinbarung mit Gegenrollen	B, D, H
Ü6	Teamreflexion und Transferaufgabe	50 Min.	Nachweisbar veränderte Bearbeitung, neuer Output	E, G

Tabelle 7: Übungsarchitektur für einen AC-Tag im Baukontext.

Die haptischen Einstiegsübungen wie die „Marshmallow-Challenge“, die der Endbericht nennt, sind als Eisbrecher denkbar, sollten aber nicht in die Wertung eingehen: Sie haben, wie der Bericht selbst festhält, keinen IPA-Bezug (Rodde & Kersten, 2024, S. 32). Ergänzend kann am Ende ein kurzes Reflexionsgespräch stehen. Es dient der Einord-

nung und dem Feedback, nicht der Bewertung, weil Selbstbeschreibungen andere Informationen liefern als beobachtetes Verhalten.

6. Die Kriterien-Übungs-Matrix für einen AC-Tag

6.1 Rahmenbedingungen

Die Matrix ist für eine Gruppe von vier bis sechs Personen eines Bieterteams und einen AC-Tag ausgelegt. Teilnehmen sollten die Personen, die tatsächlich im Projekt eingesetzt werden, insbesondere die für das Projektleitungsteam vorgesehene Person sowie Projekt- und Bauleitung, ein Polier und – soweit vorgesehen – Lean- oder BIM-Verantwortliche (Rodde & Kersten, 2024, S. 28–29). Weil die Teamgröße die Interaktion beeinflusst, sollte die Auftraggeberin für alle Bieter dieselbe Teilnehmerzahl vorgeben, etwa fünf Personen. Das Beobachterteam besteht aus vier geschulten Beobachtenden der Auftraggeberin, einer eignungsdiagnostischen Leitung und zwei Rollenspielenden; es bleibt über alle Bieterteams konstant. Die fachliche Ergebnisqualität (K6) bewertet ein zweiköpfiges Fachgremium anhand von Prüflisten und ohne Kenntnis der Prozessratings.

Uhrzeit	Element	Hinweis
08:30–09:00	Begrüßung, Ablauf, Bewertungskriterien	Transparenz über Kriterien und Ablauf für alle Teams identisch
09:00–09:45	Ü1 Auftragsklärung und Lagebild	inklusive getrennter Lagebildabfrage
09:45–10:00	Pause, Mini-Beobachterkonferenz	Ratings werden zuerst unabhängig erfasst
10:00–11:00	Ü2 Zielkosten-Variantenentscheidung	unabhängige Einzelempfehlung vorab
11:15–12:15	Ü3 Lean-Ablaufplanung	zwei Rollenspielende als Gewerke und Betrieb
12:15–13:00	Mittagspause	Fachgremium bewertet Outputs aus Ü2 und Ü3
13:00–13:45	Ü4 Störung im Bauablauf	Auslöser nach Skript zu festen Zeitpunkten
14:00–14:30	Ü5 Schnittstellen- und Eskalationsgespräch	standardisierte Gegenrollen
14:45–15:35	Ü6 Teamreflexion und Transferaufgabe	Reflexionsformat für alle Teams identisch
15:35–16:00	Abschluss mit dem Bieterteam	Einordnung und Dank, keine Bewertung
16:00–17:30	Beobachterkonferenz und Dokumentation	regelgeleitete Integration, Dokumentation nach § 8 VgV

Tabelle 8: Exemplarischer Zeitplan für einen AC-Tag (reine Übungszeit 4 Stunden 50 Minuten).

6.2 Die Matrix

Die Matrix ordnet jedem Kriterium die Übungen zu, in denen es bewertet wird. Drei Regeln leiten die Zuordnung: Jedes Kriterium wird in mindestens zwei Übungen beobachtet, damit kein Urteil auf einer einzigen Situation beruht; jede Übung bewertet höchstens drei Prozesskriterien, damit die Beobachtenden nicht überlastet werden; und die Ergebnisqualität wird in allen Übungen mit prüfbarem Arbeitsprodukt bewertet.

Kriterium	Ü1 Lage- bild	Ü2 Ziel- kosten	Ü3 Lean- Planung	Ü4 Stö- rung	Ü5 Schnitt- stelle	Ü6 Trans- fer	Anzahl	Gewicht
K1 Informationsintegration	•	•	•			•	4	15 %
K2 Koordination und Absiche- rung	•		•	•			3	15 %
K3 Entscheidung und Konfliktbe- arbeitung		•		•	•		3	15 %
K4 Anpassung und Lernen				•		•	2	15 %
K5 Integration über Teamgren- zen			•		•		2	10 %
K6 Ergebnisqualität (Fachgremi- um)		•	•	•		•	4	30 %
Prozesskriterien je Übung	2	2	3	3	2	2	14	100 %

Tabelle 9: Kriterien-Übungs-Matrix für einen AC-Tag (• = Kriterium wird in der Übung bewertet; Gewichte als Beispiel, vor Verfahrensbeginn projektbezogen festzulegen und bekanntzugeben).

Die Gewichtung im Beispiel gibt den fünf Prozesskriterien zusammen 70 Prozent und der gemeinsamen Ergebnisqualität 30 Prozent. Diese Verteilung ist eine begründete Setzung, kein empirisch optimierter Wert: Aus den vorliegenden Metaanalysen lassen sich keine Auswahlgewichte herausrechnen, weil sie mit unterschiedlichen Prädiktoren und Kriterien arbeiten. Entscheidend ist, dass die Gewichte aus den Projektrisiken abgeleitet, vor dem Verfahren festgelegt und mit den Unterkriterien bekanntgegeben werden, wie es § 127 Abs. 5 GWB verlangt.

6.3 So wird die Matrix mit Zahlenwerten befüllt

Damit die Matrix fair und möglichst objektiv befüllt wird, folgt die Auswertung sechs festen Schritten.

- **Schritt 1 – Beobachten und protokollieren.** Während jeder Übung notieren die Beobachtenden beschreibend, was geschieht, und halten die vorab definierten kritischen Sequenzen fest, etwa: Wurde der widersprüchliche Kostenhinweis in Ü2 aufgegriffen, geprüft und im Ergebnis berücksichtigt?
- **Schritt 2 – Unabhängig bewerten.** Nach jeder Übung bewertet jede beobachtende Person allein die zugeordneten Kriterien auf der Skala von 1 bis 5 anhand der Verhaltensanker; „nicht beobachtbar“ ist zulässig.
- **Schritt 3 – Zellwerte bilden.** Der Zellwert ist der Mittelwert der vier unabhängigen Ratings. Weichen zwei Ratings um mehr als einen Punkt voneinander ab, wird die Abweichung anhand der Protokolle geklärt; geändert werden darf nur mit Verweis auf dokumentierte Beobachtungen, nicht durch Mehrheitsverhandlung.
- **Schritt 4 – Kriteriumswerte bilden.** Der Kriteriumswert ist der Mittelwert der Zellwerte über die zugeordneten Übungen. K6 wird vom Fachgremium anhand der Prüflisten vergeben.
- **Schritt 5 – Mindestanforderungen prüfen.** Vorab festgelegte, nicht kompensierbare Schwellen werden geprüft, im Beispiel: K6 mindestens 2,5 und kein Kriterium unter 2,0. Hohe Kommunikationswerte dürfen ein fachlich unbrauchbares Ergebnis nicht ausgleichen.
- **Schritt 6 – Teamprofil und Gesamtwert berichten.** Berichtet werden das Profil der sechs Kriteriumswerte, der gewichtete Gesamtwert und Angaben zur Belastbarkeit, etwa die Zahl der Beobachtungen und die Beobachterübereinstimmung.

Wie das konkret aussieht, zeigt das fiktive Ergebnis eines Bieterteams B im Verfahren der Nordwasser AöR. Die Zellen enthalten die Mittelwerte der vier Beobachtenden beziehungsweise die Wertung des Fachgremiums.

Kriterium	Ü1	Ü2	Ü3	Ü4	Ü5	Ü6	Kriteriumswert	Gewicht	Gewichteter Wert
K1 Informationsintegration	3,50	4,00	3,25			4,25	3,75	15 %	0,56
K2 Koordination und Absicherung	3,00		3,50	2,75			3,08	15 %	0,46
K3 Entscheidung und Konfliktbearbeitung		3,75		3,00	3,50		3,42	15 %	0,51
K4 Anpassung und Lernen				2,50		4,00	3,25	15 %	0,49
K5 Integration über Teamgrenzen			3,25		2,75		3,00	10 %	0,30
K6 Ergebnisqualität		4,00	3,50	3,00		4,00	3,63	30 %	1,09
Teamwert (Skala 1–5)								100 %	3,41

Tabelle 10: Befüllte Kriterien-Übungs-Matrix für das fiktive Bieterteam B (Skala 1–5).

Das Profil erzählt mehr als die Gesamtzahl. Team B integriert Informationen gut und lernt erkennbar: Die Anpassung fällt in der Störungsübung noch schwach aus (2,50), in der Transferaufgabe aber deutlich besser (4,00). Die Schwächen liegen in der Koordination unter Druck und an der Schnittstelle zur Auftraggeberin und zum Betrieb. Alle Mindestanforderungen sind erfüllt. Fließt das AC mit 30 Prozent in den Zuschlag ein, lässt sich der Teamwert linear umrechnen: Punkte = (Teamwert – 1) / 4 × 30. Für Team B ergeben sich (3,41 – 1) / 4 × 30 = 18,1 von 30 möglichen Punkten. Das Profil ist zugleich eine wertvolle Grundlage für das Teambuilding nach dem Zuschlag, denn es benennt die Entwicklungsfelder der künftigen Allianz.

Ergänzend – aber ohne Gewicht im Zuschlag – sollten in der Pilotphase einige Kennzahlen mitgeführt werden: der Anteil der vorab definierten entscheidungsrelevanten Hinweise, die im gemeinsamen Ergebnis korrekt berücksichtigt wurden; die Richtigkeit und Übereinstimmung der Lagebilder aus Ü1; der Anteil standardisiert ausgelöster Einwände, die geprüft und begründet übernommen oder verworfen wurden; und der Vergleich des Teamergebnisses aus Ü2 mit der besten unabhängigen Einzelempfehlung. Diese Kennzahlen machen die kollektive Ebene sichtbar und liefern die Daten, mit denen das Verfahren später weiterentwickelt werden kann.

6.4 Fairness und Objektivität sichern

Ein Verfahren mit 30 Prozent Zuschlagsgewicht muss sich an den Qualitätsstandards der Eignungsdiagnostik messen lassen. Die DIN 33430 verlangt unter anderem, dass Verfahren aus einer Anforderungsanalyse abgeleitet, Beobachtungs- und Bewertungsregeln vorab festgelegt, die Beteiligten qualifiziert und die Ergebnisse nachvollziehbar dokumentiert werden (Diagnostik- und Testkuratorium, 2018). Die internationalen AC-Leitlinien ergänzen mehrere verhaltensbezogene Simulationen, geschulte Beobachtende, die Trennung von Beobachtung und Bewertung sowie eine systematische Integration der Daten (International Taskforce on Assessment Center Guidelines, 2015). Für die Kriterien-Übungs-Matrix bedeutet das konkret:

- **Durchführungsobjektivität:** identische Unterlagen, Zeiten und Moderationsregeln; Rollenspieler arbeiten mit Skripten und bringen Einwände zu festgelegten Zeitpunkten ein.
- **Auswertungsobjektivität:** verhaltensverankerte Skalen, unabhängige Ratings vor jeder Diskussion, regelgeleitete Integration statt Konsensverhandlung und eine Bewertung der Ergebnisse ohne Kenntnis der Prozessratings.
- **Interpretationsobjektivität:** Gewichte, Mindestschwellen und Umrechnungsregeln stehen vor dem ersten Bieterteam fest und werden bekanntgegeben.
- **Fairness:** gleiche Teamgröße, gleiche Vorabinformation über Kriterien und Ablauf, keine Bewertung von Stil, Sprachgewandtheit oder demografischen Merkmalen, ein klarer Umgang mit Aufzeichnungen und personenbezogenen Daten.

- **Beobachterqualifikation:** Eine Schulung von etwa vier Stunden, wie sie der Endbericht vorsieht, sollte um eine Kalibrierung an aufgezeichneten Beispielsequenzen ergänzt werden, damit alle Beobachtenden die Anker gleich anwenden.

6.5 Grenzen und Validierung

Auch ein sorgfältig gestaltetes Verfahren liefert eine Momentaufnahme. Das Urteil gilt für die beobachtete Teamkonfiguration unter den Bedingungen des AC-Tags; wird eine Schlüsselperson ausgetauscht, gilt es nicht automatisch fort. Realitätsnahe Übungen stützen die inhaltliche Validität, belegen aber noch nicht, dass das Verfahren späteren Projekterfolg vorhersagt. Dafür braucht es eine Pilotierung mit bestehenden Teams, eine Prüfung der Beobachterübereinstimmung und – über mehrere Vergaben hinweg – den Abgleich mit unabhängig erhobenen Projektkriterien wie der Qualität gemeinsamer Entscheidungen, vermeidbarer Nacharbeit oder der Verlässlichkeit anspruchsvoller Zusagen. Bei nur drei bis fünf Bieterteams je Verfahren ist die Datenbasis klein; projektübergreifende Datensammlungen öffentlicher Auftraggeber wären deshalb ein wichtiger Schritt.

7. Fazit

Der BBSR-Endbericht hat der Integrierten Projektabwicklung in Deutschland eine solide bauliche, organisatorische und vergaberechtliche Grundlage gegeben. Sein Kapitel III.2 lässt jedoch die eignungsdiagnostische Kernfrage offen: Es will die Kooperationsfähigkeit eines Teams bewerten, beschreibt aber überwiegend Haltungen und Kompetenzen einzelner Personen, lässt das einzige teambezogene Kriterium undefiniert und lässt offen, wie aus Einzelbeobachtungen ein Teamurteil entsteht.

Die Teamforschung bietet für diese Lücke tragfähige Antworten. Die acht vorgestellten Ansätze zeigen, dass sich Interaktion und Gruppenergebnis auf unterschiedlichen Wegen erfassen lassen – über Teamprozesse, emergente Zustände, Interaktionssequenzen, direkte Ergebnisbewertung, kollektive Intelligenz, Lernen und Anpassung sowie Schnittstellen zwischen Teams. Für ein IPA-Auswahlverfahren ergibt sich daraus eine klare Konsequenz: Bewertet wird, **wie ein konkret zusammengesetztes Team unter projekttypischen Bedingungen einen gemeinsamen Leistungsbeitrag hervorbringt**. Sechs Teamkriterien, sechs baunahe Übungen mit echter Interdependenz und eine Kriterien-Übungs-Matrix mit festen Auswertungsregeln machen dieses Urteil für einen AC-Tag handhabbar, nachvollziehbar und fair.

Das leitende Eignungsurteil lautet damit nicht mehr „Diese fünf Personen sind kooperativ“, sondern: „Dieses Team nutzt seine Ressourcen in relevanten Situationen verlässlich für gemeinsame Ergebnisse und kann seine Zusammenarbeit anpassen.“ Ob das Ganze tatsächlich mehr ist als die Summe seiner Teile, wird damit vom Glaubenssatz zur beobachtbaren Größe.

8. Ausblick

Drei Entwicklungen werden die Bewertung von Teaminteraktion in IPA-Verfahren in den kommenden Jahren prägen. Erstens die **Validierung**: Je mehr öffentliche Auftraggeber IPA-Assessment-Center einsetzen, desto dringlicher wird eine projektübergreifende Datenbasis, die AC-Befunde mit späteren Projektverläufen verbindet. Erst sie kann zeigen, welche Kriterien und Gewichte tatsächlich vorhersagen, ob eine Allianz trägt. Ein gemeinsamer Kriterienrahmen, wie er hier vorgeschlagen wird, wäre dafür eine gute Ausgangsbasis.

Zweitens die **technische Unterstützung der Interaktionsanalyse**: Digitale Protokollwerkzeuge erleichtern schon heute die Erfassung kritischer Sequenzen. Automatisierte Auswertungen von Sprache oder Gesprächsverläufen könnten den Aufwand der Interaktionsanalyse künftig weiter senken. Bevor sie in Vergabeentscheidungen einfließen, müssen jedoch Kodierqualität, Fehlerrisiken, Fairness und Datenschutz geprüft sein; eine automatische Ableitung von Persönlichkeitsmerkmalen aus Video- oder Sprachdaten ist abzulehnen.

Drittens die **Verbindung von Auswahl und Entwicklung**: Das Teamprofil aus dem Auswahlverfahren ist der natürliche Ausgangspunkt für das Teambuilding nach dem Zuschlag. Ein Entwicklungsassessment mit der realen Gesamtal-

lianz einschließlich der Auftraggeberseite kann die im Auswahlverfahren nur simulierten Schnittstellen sichtbar machen und Entwicklungsfelder früh adressieren. So wird aus einem Auswahlinstrument ein Lernsystem für die Allianz.

Für die Nordwasser AöR heißt das ganz praktisch: Sie gewinnt mit dem AC nicht nur einen Partner, sondern ein erstes gemeinsames Lagebild über die Stärken und Entwicklungsfelder ihrer künftigen Allianz.

Literaturverzeichnis

- Almaatouq, A., Alsobay, M., Yin, M., & Watts, D. J. (2021). Task complexity moderates group synergy. *Proceedings of the National Academy of Sciences*, 118(36), Article e2101062118. <https://doi.org/10.1073/pnas.2101062118>
- Credé, M., & Howardson, G. (2017). The structure of group task performance—A second look at “collective intelligence”: Comment on Woolley et al. (2010). *Journal of Applied Psychology*, 102(10), 1483–1492. <https://doi.org/10.1037/apl0000176>
- Diagnostik- und Testkuratorium (Hrsg.). (2018). *Personalauswahl kompetent gestalten: Grundlagen und Praxis der Eignungsdiagnostik nach DIN 33430*. Springer. <https://doi.org/10.1007/978-3-662-53772-5>
- International Taskforce on Assessment Center Guidelines. (2015). Guidelines and ethical considerations for assessment center operations. *Journal of Management*, 41(4), 1244–1273. <https://doi.org/10.1177/0149206314567780>
- 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), Article 101042. <https://doi.org/10.1016/j.hrmr.2024.101042>
- Marks, M. A., Mathieu, J. E., & Zaccaro, S. J. (2001). A temporally based framework and taxonomy of team processes. *Academy of Management Review*, 26(3), 356–376. <https://doi.org/10.2307/259182>
- Marlow, S. L., Lacerenza, C. N., Paoletti, J., Burke, C. S., & Salas, E. (2018). Does team communication represent a one-size-fits-all approach? A meta-analysis of team communication and performance. *Organizational Behavior and Human Decision Processes*, 144, 145–170. <https://doi.org/10.1016/j.obhdp.2017.08.001>
- Oostrom, J. K., Lehmann-Willenbrock, N., & Klehe, U.-C. (2019). A new scoring procedure in assessment centers: Insights from interaction analysis. *Personnel Assessment and Decisions*, 5(1), 73–82. <https://doi.org/10.25035/pad.2019.01.005>
- Riedl, C., Kim, Y. J., Gupta, P., Malone, T. W., & Woolley, A. W. (2021). Quantifying collective intelligence in human groups. *Proceedings of the National Academy of Sciences*, 118(21), Article e2005737118. <https://doi.org/10.1073/pnas.2005737118>
- Rodde, N., & Kersten, L. (2024). *Partnerauswahl und Teambuilding in der Projektallianz/Integrierten Projektabwicklung: Endbericht* (Forschungsprogramm Zukunft Bau, Az. 10.08.17.7-23.79). Lumico GmbH im Auftrag des Bundesinstituts für Bau-, Stadt- und Raumforschung. <https://ipa-zentrum.de/wp-content/uploads/2024/10/endbericht.pdf>
- Woolley, A. W., Chabris, C. F., Pentland, A., Hashmi, N., & Malone, T. W. (2010). Evidence for a collective intelligence factor in the performance of human groups. *Science*, 330(6004), 686–688. <https://doi.org/10.1126/science.1193147>
- Zaccaro, S. J., Dubrow, S., Torres, E. M., & Campbell, L. N. P. (2020). Multiteam systems: An integrated review and comparison of different forms. *Annual Review of Organizational Psychology and Organizational Behavior*, 7, 479–503. <https://doi.org/10.1146/annurev-orgpsych-012119-045418>