

Evaluation Criteria: A Structured Benchmark for Supplier, Offer and Procurement Decisions

Evaluation criteria are defined benchmarks against which buyers systematically compare and assess offers, suppliers or procurement processes. They translate corporate objectives such as cost, quality, delivery capability and sustainability into verifiable characteristics with weighted importance. In industrial procurement they form the backbone of structured decisions – from the request for quotation through supplier selection to ongoing supplier management. Anyone who defines evaluation criteria cleanly reduces subjectivity, creates comparability and documents decisions in an auditable manner.

Key facts

- Evaluation criteria are weighted, verifiable characteristics used to assess offers, suppliers and procurement processes.
- They are frequently divided into categories such as price, quality, logistics, compliance, technology and sustainability.
- A deliberate balance between price and non-price criteria prevents wrong decisions caused by a pure price focus.
- In strategic procurement, evaluation criteria support supplier evaluation; in operational procurement, they support offer comparison.
- Documented criteria and weights increase objectivity, traceability and defensibility of awards.
- Modern criteria catalogues increasingly take account of supplier cyber risk, conflict minerals and the circular economy.

Definition: evaluation criteria

Evaluation criteria are operationalised characteristics with which the suitability of an offer or supplier is measured. Each criterion is given a definition, a measurement method and a weight that expresses its relative importance within the overall evaluation.

Distinction from performance indicators

Evaluation criteria are not identical to KPIs: key figures such as the complaint rate or delivery reliability provide the raw data, while criteria define the extent to which these data feed into a decision. A supplier evaluation, for example, uses the complaint rate as a criterion within the quality block and weights this block with a fixed percentage. Only the combination of measurement, scale and weight produces a complete criterion. This distinction is important because key figures can be measured continuously, whereas criteria are tailored specifically to the procurement project at hand. In ongoing supplier management, the same data lead to an evaluation score, while in a one-off award they may be weighted differently in the offer evaluation.

Hard facts and soft facts

Common practice distinguishes between quantifiable hard facts and qualitatively assessed soft facts. Hard facts include purchase prices, payment terms such as payment deadline and cash discount terms, lead times, credit rating indicators or documented certificates. Soft facts cover cooperation, responsiveness in engineering support, innovative strength or the quality of communication – aspects that can be mapped using structured rating scales. The mistake many organisations make is either to ignore soft facts entirely or to capture them without a measurement method. Both distort the result. A fixed scale with written level descriptions has proven effective, enabling several assessors to reach comparable judgements.

Legal and regulatory framework

In public procurement, evaluation criteria are governed by law and must be disclosed before offers are submitted. In the private sector no comparable obligations apply, but documented criteria are nevertheless recommended for compliance and tax audit reasons and to defend against supplier protests. Criteria have a complementary effect alongside contractual instruments: a quality assurance agreement operationalises quality requirements, a performance bond secures adherence to dates, and quality audits verify retrospectively whether the requirements set out in the criteria catalogue are actually met. General statements on the legal situation do not replace a case-by-case review.

Methods and approach: building criteria catalogues systematically

A robust criteria catalogue does not emerge from a standard form but from the requirements of the specific procurement project. A multi-stage approach has proven effective: define requirements, derive criteria, set weights and methodically safeguard the evaluation.

Deriving and operationalising criteria

The starting point is the functional specification or technical specification, as applicable, prepared by the specialist department. From this, the buyer derives evaluation fields: technical suitability, manufacturability, cost structure, delivery capability, quality system and compliance. Each criterion is operationalised, for example "process capability of the critical manufacturing process with a documented Cpk value".

Weighting and scoring models

Weighting assigns a ranking to competing objectives. Percentage weights at category and criterion level are customary, summing to 100 per cent. Typically, the price share in industrial plant and parts business lies between 30 and 50 per cent; for raw materials it can be significantly higher and for engineering services significantly lower. Scoring scales of 1 to 5 or 1 to 10 have established themselves as scoring methods, supplemented by exclusion criteria that are decided in advance by pass or fail. A supplier without the required level of certification is eliminated regardless of its price – this should be documented as a preceding knockout criterion.

Evaluation in the request for quotation

In a structured request for quotation, criteria and weights are either disclosed to bidders or at least fixed internally before offers are opened. The latter prevents subsequent adjustments in favour of a preferred supplier. The evaluation process usually runs in two stages: first, the buyer checks the formal and technical minimum requirements; then the specialist departments and the procurement team assess the remaining offers using the scoring matrix. For complex services, a proof-of-concept procurement can

serve as an additional criterion: the bidder must deliver a defined sample against which technical criteria are assessed using real results instead of reference statements. Price comparisons should be based on total cost of ownership and include process costs and follow-up costs.

Key KPIs: evaluation in ongoing supplier management

In day-to-day operations, the focus shifts from the one-off award to continuous supplier evaluation. Here, evaluation criteria support ongoing KPIs fed from system data and quality processes.

Quality and delivery performance

The two core dimensions of a supplier evaluation are quality and delivery reliability. On the quality side, the complaint rate – frequently defined as the number of complaints per million delivered units (PPM) – dominates, along with the results of quality audits and the proportion of parts that have passed series approval. Logistics-oriented criteria measure delivery reliability in per cent, adherence to the confirmed date and compliance with agreed lot sizes. All key figures should be captured on a defined basis: without correct master data in master data management and a consistent material numbering system, a data basis that distorts the result quickly emerges. Material classes help to map different requirements depending on the commodity segment.

Cost and commercial criteria

Commercial evaluation fields assess whether a supplier actually delivers the agreed added value. These include savings achieved against the baseline, the transparency of price formation, currency and material price index clauses, and payment behaviour with regard to agreed payment terms. On the buyer side, it is additionally measured how strongly pooling levers in the supplier portfolio are exploited: whoever consolidates identical requirements across multiple sites typically achieves better terms and simultaneously simplifies the evaluation, because fewer suppliers with homogeneous criteria catalogues need to be assessed. The efficiency of collaboration in the ordering process – for example the degree of automation via punch-out catalogues – can also be included as a criterion.

Compliance, risk and sustainability

Modern criteria catalogues increasingly hold suppliers accountable for sustainability and risk topics. These include the origin of conflict minerals, recycling and circular economy aspects in packaging and material selection, occupational safety and environmental management, as well as supplier cyber risk, which can be a significant failure factor in networked manufacturing processes. Such criteria are frequently collected via self-disclosures, certificates and audit reports. For suppliers of critical pre-products, a risk-based deepening is recommended rather than a blanket review of all suppliers, which would disproportionately increase the effort involved.

Risks and controls: typical sources of error with evaluation criteria

Evaluation criteria are only as good as their maintenance and application. Typical errors arise from a lack of currency, methodological errors in weighting and an inadequate data basis.

Weighting errors and false objectivity

A common mistake is overemphasis on price while follow-up costs are neglected: a cheap initial article can become more expensive over its lifetime due to higher process costs, complaints and changeover times. Conversely, excessive weighting of quality and risk can systematically disadvantage efficient

suppliers. The apparent objectivity of mathematically clean scores is also a risk: if subjective impressions lie behind precisely sounding decimal grades, the model suggests an accuracy that does not exist. Controls for this are documented level descriptions, at least two independent assessors per soft fact criterion and a plausibility check of the overall result by the team before a decision is made.

Data quality as a weak point

In practice, evaluations frequently fail not because of the criteria but because of the data. Duplicate item numbers, missing supplier master records or incomplete goods receipt entries distort the complaint rate and delivery reliability. Countermeasures include clean master data management, reference data for uniform units and designations, and clear responsibilities for data maintenance. In addition, the data source should be documented for each criterion: system key figures from ERP and the QM module are more reliable than memory notes from phone calls. Regular residual checks show early on whether a criterion is still measurable or is in fact being filled purely by gut feeling.

Single source and safeguards

When evaluation criteria lead to a narrow supplier selection, a single-source risk arises that should be managed separately. Single-source management requires deeper risk criteria such as second-source planning, safeguards via performance bonds or expressly agreed capacity reserves. In the quality area, defined quality gates between supplier process steps help ensure that evaluation deficits do not only become visible at the end customer. Lot and batch traceability enable precise backtracking in the event of a complaint, making it possible to assess whether an evaluation deficit is isolated or systemic.

Current developments: digitalisation and sustainability are changing the criteria

The procurement criteria catalogue is not a static instrument. Digitalisation, supply chain laws and a growing awareness of supply chain risks are expanding the classic categories with new dimensions.

Supplier enablement and digital evaluation

Digital platforms are changing how criteria are collected and maintained. Supplier enablement describes the process of systematically connecting suppliers to digital processes – from electronic catalogues and punch-out integration to automated request for quotations and self-disclosure in the supplier portal. The higher the supplier's level of digitalisation, the easier it is to evaluate criteria continuously rather than at isolated points. Criteria catalogues thereby become living instruments that are continuously fed with new data, instead of annually updated Excel lists. This also reduces the effort involved in the request for quotation, because master data and certificates, once captured, can be reused multiple times.

Sustainability and supply chain laws

Regulatory requirements are shifting sustainability from voluntary acknowledgement to a mandatory category. Criteria on the origin of conflict minerals, CO2 intensity of manufacturing, circular economy and human rights due diligence are moving to the centre of supplier evaluation. For buyers, this means underpinning sustainability criteria with measurable indicators – for example documented proof of origin, energy intensity values per unit produced or certificates from recognised standards. Interpreting this data requires sensitivity: comparability between industries is limited, and standard values should be understood as guidance, not as an exact benchmark.

AI-supported offer evaluation

AI systems increasingly support the preliminary analysis of large volumes of offers: they extract line items from documents, automatically compare technical specifications and create enquiries with structured criteria catalogues as a template. However, the methodological foundation remains the human-defined evaluation matrix – AI accelerates the analysis but does not replace the decision on criteria and weights. A low-risk entry point is to automate repetitive awards with standardised criteria catalogues and to continue evaluating complex, strategic procurements manually. In both cases, documentation of the evaluation logic remains central.

Typical metrics

Delivery reliability: > 95%

Proportion of items delivered on time; varies by industry.

Complaint rate: < 500 PPM

Rejected units per million units delivered; highly volume-dependent.

Price share in overall evaluation: 30–50%

Common guideline value in industrial procurement.

Scoring compliance: > 90%

Proportion of evaluations completed and maintained on time in the system.

Practical example

A medium-sized mechanical engineering company with annual revenue of around €200 million and approximately 400 active suppliers is revising its supplier evaluation after a maintenance outage revealed a risk under the EU Supply Chain Act. Previously, the criteria catalogue had been predominantly price-oriented; sustainability and risk criteria were entirely absent.

- New category structure with five blocks: price and terms, quality, delivery performance, compliance and sustainability, technology and cooperation – weights between 15 and 30 per cent per block.
- Introduction of measurement methods for soft facts with defined level descriptions and two-stage assessment by procurement and the specialist department.
- Risk-based deepening for A-suppliers: quality audits, proof of origin for conflict minerals, cyber risk checks and documented second-source planning instead of blanket audits of all suppliers.
- Integration of the key figures complaint rate, delivery reliability and savings into the ERP system with quarterly automatic updates of the supplier evaluation.

Conclusion

Evaluation criteria are the central instrument for moving procurement decisions from gut feeling to verifiable foundations. Their value depends less on the mathematical sophistication of the scoring model than on the quality of the criteria themselves: they must be derived in line with requirements, measured in an operationalisable way and given consistent weights. The data basis from master data management, ERP and QM systems deserves the same attention as the catalogue itself. Current topics such as supplier cyber risk, conflict minerals and the circular economy expand the classic categories and demand measurable indicators rather than statements of intent. Those who review criteria catalogues regularly and consistently document the evaluation gain not only better suppliers but also a defensible decision-making basis vis-à-vis management, internal audit and external reviewers.

Frequently asked questions

What are evaluation criteria in procurement?

Evaluation criteria are defined, weighted characteristics used to systematically assess offers, suppliers or procurement processes. Typical categories are price, quality, delivery capability, compliance, technology and sustainability. Each criterion is given a measurement method and a weight that defines its importance in the overall evaluation. They ensure comparability and document decisions in an auditable manner.

How should evaluation criteria be weighted correctly?

Weights are derived from the objectives of the procurement project, typically as a percentage distribution across category and criterion levels totalling 100 per cent. In an industrial setting, the price share is often between 30 and 50 per cent, but it can deviate significantly depending on the item being procured. It is important that weights are fixed and documented before offers are opened. In addition, knockout criteria should be defined that lead to exclusion regardless of the score.

Which criteria belong in a supplier evaluation?

Proven core areas are quality (for example complaint rate, audit results, series approval), delivery performance (delivery reliability, adherence to agreed dates), price and terms (savings, payment behaviour) as well as technology and cooperation. In addition, compliance and sustainability criteria such as conflict minerals, circular economy and supplier cyber risk are coming into focus. The selection of criteria should be tiered according to supplier risk.

What is the difference between a KPI and an evaluation criterion?

A KPI measures a performance metric continuously, such as the complaint rate or delivery reliability in per cent. The evaluation criterion defines the extent to which this measurement feeds into an overall evaluation or decision and how heavily it is weighted. Key figures are therefore the raw data, while criteria are the evaluation logic. In ongoing supplier management, the same KPIs feed into a continuous scoring process.

Must evaluation criteria be disclosed in tenders?

In public procurement, evaluation criteria must in principle be disclosed before offers are submitted; the applicable procurement regulations govern the details on a case-by-case basis. In the private sector there is no comparable general obligation, but disclosing categories and weights is recommended in order to offer bidders fair competitive conditions. Internally, criteria should be fixed and documented no later than before offers are opened. Binding statements require a case-by-case review.

How often should supplier evaluations be updated?

In practice, six-monthly or annual updates have become established for the majority of the supplier base. A-suppliers and critical pre-products are often evaluated quarterly, C-suppliers less frequently. It is crucial that key figures – such as complaint rate and delivery reliability – are captured continuously from systems so that the evaluation is based on current data. Triggers such as complaints or audit results should additionally trigger an extraordinary re-evaluation.