publications

MDPI

Article

Integrating Scientific Readiness and Added Scientific Value: A Multiplicative Conceptual Model for Research Assessment

Eldar Knar

Institute of Philosophy, Political Science and Religion Studies, Ministry of Science and Higher Education of the Republic of Kazakhstan, Str. Kurmangazy, 29, Almaty City 050010, Kazakhstan; eldarknar@gmail.com

Abstract

Modern approaches to research assessment often treat two independent dimensions: the degree of readiness (the progress of scientific maturity) and the Added Scientific Value (novelty, rigor, reproducibility, and societal significance). The present study aims to develop an integrative model that links the stepwise logic of Scientific Readiness Levels (SRLs) with the continuous evaluation of Added Scientific Value (ASV). To achieve this, we construct a two-dimensional quantitative framework, SRL × ASV, in which discrete readiness levels are combined with the multiplicative aggregation of normalized qualitative indicators—including novelty, rigor, reproducibility, openness, impact, and collaboration. Advancement to a higher readiness level becomes possible only when the evidential criteria are simultaneously satisfied and the ASV threshold value is reached. The model is formally defined and illustrated via the example of Kazakhstan's national science funding system. The results demonstrate that the integration of SRL and ASV creates a reproducible and transparent decision-making structure in which weak components cannot be compensated by strong ones, and excessive dependence on purely metric-based evaluations is eliminated. By uniting maturity and quality within a single coherent system, the SRL + ASV model bridges the gap between developmental staging and scientific value, transforming research assessment into a transparent, evidence-based, and continuously improvable process applicable to the governance of responsible and reproducible science.

Keywords: scientific readiness; added scientific value; quantitative assessment; reproducibility; responsible evaluation; scientific governance; metrics

check for updates

Academic Editor: Andrew Kirby

Received: 20 August 2025
Revised: 28 October 2025
Accepted: 31 October 2025
Published: 10 November 2025

Citation: Knar, E. (2025). Integrating Scientific Readiness and Added Scientific Value: A Multiplicative Conceptual Model for Research Assessment. *Publications*, *13*(4), 57. https://doi.org/10.3390/publications13040057

Copyright: © 2025 by the author. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (https://creativecommons.org/licenses/by/4.0/).

1. Introduction

Research assessment has long been divided into two largely independent dimensions: maturity or readiness (how far a project has advanced from idea to implementation) and added scientific value (its novelty, rigor, reproducibility, and anticipated impact). In engineering and applied fields, early attempts to formalize readiness can be traced to the classical TRL (technology readiness level) framework, which has nine stages (NASA, 2019; European Commission, 2014). By spanning the trajectory from the observation of basic principles to verified deployment, the TRL has become embedded in project management and technological roadmaps. However, the TRL, which is designed primarily for technological artifacts and product life cycles, is only indirectly suited for evaluating the scientific content of research.

The success of the TRL nonetheless inspired analogous scales for science. The European Space Agency (ESA), for example, has developed and applied SRLs (scientific readiness levels) as a standard procedure for evaluating missions and instruments, with

explicit evidence requirements at each stage (ESA, 2023). These practices highlight the institutional demand for readiness scales but leave a crucial question unresolved: how can research readiness be systematically linked with scientific value without reducing the latter to narrow proxies?

Moreover, the debate over the quality and reliability of scientific findings has intensified. Ioannidis's (2005) seminal analysis revealed that, given typical statistical power, biases, and multiple hypothesis testing, a substantial proportion of published results are likely false positives. The large-scale Open Science Collaboration further demonstrated the limited reproducibility of psychological findings. These results spurred institutional reforms in research evaluation, from the *DORA* (DORA, 2012) and the Leiden Manifesto (Hicks, 2012) to the adoption of the FAIR data principles (Jacobsen et al., 2020). While differing in scope and emphasis, these initiatives converge on the call to move beyond journal-based metrics and strengthen transparency, openness, reproducibility, and multidimensional evaluation. However, even such landmark reforms focus primarily on qualitative criteria and practices, whereas the link between quality and staged project progress remains fragmented across separate instruments.

In everyday research management and R&D practice at the level of organizations, funding agencies, and ministries, journal indicators continue to serve as simplified proxies for value—for example, the impact factor, SJR quartiles, and derivative rankings. Even the originators of the impact factor emphasize its limitations at the level of individual articles or researchers. Moreover, SJR quartiles, while useful for comparing journals, have effectively become a "currency" in decision-making, inevitably producing distortions. This problem exemplifies Goodhart's law (Goodhart, 2013): when a measure becomes a target, it ceases to function as a reliable measure. In the context of research evaluation, this dynamic fosters metric-driven strategies—from editorial and authorial citation optimization to the choice of publication venues, primarily for formal scores rather than substantive scientific contributions. Therefore, a model in which metrics serve as inputs within an explicit decision-making structure rather than as self-sufficient goals is needed.

Our work proposes precisely such a model: a two-axis quantitative system of "*readiness* $\times$ *value*." The maturity axis is defined by the scientific readiness level (SRL) scale (Knar, 2024a), whereas the new axis of added scientific value (ASV) aggregates the key components of quality—novelty, rigor, reproducibility, openness, impact, and collaboration—into a normalized feature space. Unlike existing approaches, we integrate these two axes into a single-stage index: its multiplicative form requires both sufficient maturity and high value, eliminating spurious "leadership" of projects that excel in one dimension while failing in the other. The choice of the geometric mean to calculate the integral value reflects complementarity and the inherent "weakest-link penalty" in systematic scientific work (for example, high citation counts or prestigious venues cannot compensate for irreproducibility). At the institutional level, the model is reinforced by a threshold-based logic for transitions between SRLs and explicit funding prioritization rules, making it compatible with portfolio management and programmatic evaluation procedures (including those that are analogous to SRL protocols).

We define the scientific readiness level (SRL) as a discrete metric of research staging, reflecting progress from a conceptual idea to results that are replicated, validated, and integrated. Each SRL is not achieved through the incremental accumulation of points but through the appearance of a new class of evidence of scientific robustness (e.g., publications, peer reviews, replications, multicenter trials, or societal impact). Thus, in our interpretation, SRL does not measure "quality within a level" but rather the emergence of a new type of evidence that enables advancement to the next stage.

In contrast, we define added scientific value (ASV) as a continuous metric of coherence and rigor within a fixed SRL. It aggregates normalized quality indicators—novelty, methodological rigor, reproducibility, data openness, strength of peer review, normalized citation impact, external influence, and collaboration—into an integral index ranging from 0 to 1. In this sense, the ASV measures the "vertical dimension of quality" within a single SRL, enabling the distinction between weak and strong realizations at the same stage of maturity.

In this framework, the SRL represents the discrete axis of "when a project is sufficiently mature to generate the next class of evidence," whereas ASV represents the continuous axis of "how strong and valuable a project is within its current class." Their integration into the SRL + ASV model bridges the gap between the language of staging (SRL) and the language of quality (ASV).

Accordingly, our contribution is threefold. First, we formalize SRL as a discrete staging metric and ASV as a weighted system of features with transparent normalization and parameters and verifiable checklists. This resolves ambiguities in the operationalization of "registerable value" criteria (novelty, rigor, reproducibility, openness, etc.), aligning them with contemporary principles of open science and responsible metrics. Second, we introduce a quantitative integrator that provides explicit managerial levers for disciplinary adaptation and funding policies without undermining comparability. Third, we demonstrate the application of the model on synthetic and near-real data, while outlining an external validation framework based on real-world cases, drawing on established SRL practices (ESA) and the recommendations of the *DORA* and the Leiden Manifesto for the responsible use and calibration of metrics.

Taken together, these elements form a unified, reproducible, and scalable framework that connects the agendas of quality and openness with those of technological and scientific maturity—precisely where the gap between managerial reporting and scientific methodology is most evident today.

Finally, we deliberately situate our model within the broader context of "*assessment reform.*" The *DORA* explicitly warns against the use of journal-based metrics at the level of individual evaluation; the Leiden Manifesto articulates ten principles for the responsible application of indicators; and the FAIR principles emphasize machine readability, accessibility, and reusability of data. All these norms function as "*external constraints*" in the design of any evaluative model. Our system incorporates these constraints constructively: metrics are treated as data rather than as goals; transition rules and prioritization rely on thresholds and checklists rather than on the fetishization of individual indicators; and calibration mechanisms for parameters (weights, thresholds) are explicitly designed to align with independent expert judgments and retrospective data. In this way, we seek to minimize the risks of Goodhart's law and ensure the verifiability of outcomes, linking the model to real governance processes and the requirements of specific disciplines.

2. Literature Review

The research and management literature on the assessment of scientific projects has historically evolved along two almost independent lines: readiness/maturity scales and systems for evaluating the value or quality of results (bibliometrics, meta-research, open science, and related approaches). In this review of related work, we systematize the key directions and demonstrate how their intersection generates the need for an integrative "*readiness* × *value*" model, which we subsequently formalize as SRL + ASV.

As already noted, the classical technology readiness level (TRL) framework describes nine stages of technological maturity, ranging from the observation of basic principles to verified deployment in the target environment. Official definitions, exit criteria, and

best practices for assessment are continuously updated (including system-wide coverage through TRA guides), ensuring the reproducibility of decisions and alignment with system and mission life cycles. These documents establish a normative language of readiness and specify the evidentiary base that is required at each stage. A similar concept was transferred into the scientific domain by the European Space Agency (ESA) in the form of scientific readiness levels (SRLs). The SRL guidelines (versions 1.1 and 2.0) formalize nine levels of scientific maturity for missions and instruments, specify evidentiary requirements, define key questions for scientific review, and outline procedures for self-assessment by mission teams. A critical innovation of SRLs is the explicit linkage of *scientific progress* to clear, verifiable evidence at each level, thereby making readiness an operational component of the project cycle.

Among contemporary approaches to formalizing scientific readiness levels, Knar's (2024a) study on the *level of scientific readiness with ternary data types*, where the SRL is represented as a three-dimensional vector (FRL—Fundamental Readiness Level; ARL—Applied Readiness Level; and IRL—Innovation Readiness Level), is particularly noteworthy. Each coordinate reflects the degree of fundamental, applied, and innovation maturity of a project, coded in trits from 0 to 9, which produces positional codes such as "7.0.5." This yields a multidimensional representation of readiness, allowing for both visual and quantitative comparisons of projects. Our SRL + ASV model, in contrast, preserves a one-dimensional readiness scale but augments it with an axis of internal value (ASV). The integration of these axes through a multiplicative index, S, ensures complementarity and structural manageability. In the future, Knar's multidimensional approach and our integrative framework could be synthesized into a richer evaluative architecture, where the internal value of each component (FRL, ARL, IRL) is assessed via ASV indicators and subsequently aggregated via threshold logic and improvement mapping. As a related contribution, Knar (2024b) also proposed the aggregated recursive *K-index* as a new scientometric indicator of added value and research outcomes at the level of individual publications.

In parallel, systems engineering has advanced the integration of the TRL with Integration Readiness Levels (IRLs) to form the System Readiness Level (SRL), an index of a system's maturity that accounts for both the readiness of components and the readiness of their integration (Lemos & Chagas Junior, 2016; Sauser et al., 2006). This line of work emphasizes that progress to a higher level is only justified when there is evidence not only at the level of elements but also at their interfaces. For our purposes, this provides an important methodological lesson: readiness refers not only to the "height" of individual features but also to the joint coherence of the knowledge or technology architecture as a whole.

Similar principles of staged assessment have been applied in medicine through knowledge readiness levels (KRLs), which measure the maturity of medical research in terms of its progress toward clinical improvement (Engel et al., 2019). They are also used for evaluating the relative completeness of data for answering specific questions via data readiness levels (DRLs) (Guan et al., 2017).

In recent decades, systems for evaluating scientific value and quality have undergone critical reappraisal. The *DORA* emphasizes the inadmissibility of using journal-based metrics (e.g., JIF) to evaluate individual articles, researchers, or decisions on funding and hiring, shifting the focus instead toward expert review and a diversity of indicators. The Leiden Manifesto articulates ten principles for the responsible use of metrics, including contextualization, alignment with mission, transparency, and awareness of systemic effects (such as behavior being distorted by metric pressure). Similarly, the *Metric Tide* report (HEFCE/UKRI) develops the framework of *responsible metrics*, delineating both

the limitations of indicators in research and management and best practices for their use (Wilsdon, 2015).

The institutional debate also involves warnings relating to Goodhart's law and the closely related Campbell's law (Campbell, 1979). Both highlight the systemic risk that arises when an indicator is transformed into a target KPI, thereby undermining its measurement validity and distorting behavior (Manheim, 2018). We regard this as an inherent vulnerability of any metric-centric policy framework for evaluation.

The reproducibility crisis has radically reshaped the quality agenda. Ioannidis's theoretical analysis demonstrated the high likelihood of false positive findings when low statistical power, multiple hypotheses, and biases converge; the large-scale Open Science Collaboration (2015) empirically confirmed the limited reproducibility of a range of classic effects. These findings triggered institutional reforms targeting data, algorithms, and workflows, directly reinforcing practices of openness and reuse. Thus, the emerging consensus in the literature is that evaluation must reward rigor, replication, and openness rather than proxy measures of popularity.

The shift from individual indicators to an integral measure of "value" naturally raises the problem of correct aggregation. Within the traditions of multiattribute utility theory (Kennedy-Martin et al., 2020) and multicriteria decision analysis (MCDA) (Velasquez & Hester, 2013), additive and multiplicative forms of utility are widely discussed. Several studies provide methodological justifications for the use of the weighted geometric mean, which naturally enforces a "weakest-link penalty" and satisfies key axioms for aggregating ratios and comparative assessments. Recent MCDA surveys underscore the robustness of the geometric mean as a technically convenient and interpretable approximation in many applications, while also stressing the importance of deliberate weight calibration and validation (Krejčí & Stoklasa, 2018).

To summarize, the literature reveals three stable conditions:

(i) Readiness/maturity (TRL/SRL) relies on formal guidelines and evidence-based checklists;
(ii) The evaluation of value shifts toward responsible metrics, article-level indicators, open data, and reproducibility;
(iii) Indicators are susceptible to strategic use, unless they are embedded in verifiable procedures and counterbalanced by expert judgment.

Here, we identify a research gap: the absence of an integrated, parameterized model that unites *staging* and *value* within a framework that is compatible with real-world decision-making. Our proposed SRL + ASV model addresses this gap. The discrete SRL axis specifies "*where we are*" in terms of maturity. The continuous ASV axis, computed via a weighted geometric mean of key components (novelty, rigor, reproducibility, publication prestige, citation impact, external influence, openness, collaboration), specifies "*how valuable it is*." The multiplicative index and threshold logic tie these dimensions together into operational rules for advancement, analogous to the ESA's SRL practices but with an explicit "quality axis" and managed calibration of it. We argue that such a construction aligns with the international trajectory of responsible metrics, minimizing the risks of Goodhart's law and the fetishization of single indicators.

2.1. Theoretical Foundations of the ASV Components

The Added Scientific Value (ASV) framework integrates eight conceptual dimensions—novelty (N), rigor (R), reproducibility (P), openness (O), collaboration (C), impact (I), validation (V), and localization (L)—each rooted in established theories of research evaluation and scientometrics. Together, they capture the epistemic, methodological, and societal aspects of research quality.

Novelty (*N*) reflects the *epistemic originality* of research, echoing Merton's (Merton & Shapere, 1974) concept of "priority and discovery" and recent network-based measures of combinatorial innovation (Uzzi et al., 2013). Novelty is essential for distinguishing incremental progress from paradigm-shifting research.

Rigor (*R*) originates from Popperian's (Popper, 2005) view of scientific testability and falsifiability, reinforced by later meta-research on methodological soundness (Ioannidis, 2005). It represents transparency in research design, statistical validity, and control over biases.

Reproducibility (*P*) responds to the *replication crisis* in multiple disciplines (Nosek et al., 2018). It emphasizes the ability of independent researchers to obtain similar results under equivalent conditions, forming a cornerstone of scientific reliability (Nosek et al., 2018).

Openness (*O*) draws upon the FAIR (Findability, Accessibility, Interoperability, and Reuse) data principles (Wilkinson et al., 2016) and the *Declaration on Research Assessment*, positioning open access and data transparency as measurable quality criteria.

Collaboration (*C*) is derived from network science and team science research (Wuchty et al., 2007), which demonstrate that interdisciplinary and international collaborations enhance innovation, reproducibility, and impact.

Impact (*I*) reflects both scientific and societal influence, in accordance with the multidimensional frameworks proposed by Bornmann (2013) and Aagaard (2015), and shifts the focus from raw citation counts to normalized, field-specific, and policy-relevant impacts.

Validation (*V*) denotes the *strength and credibility of evidence* (Munafò et al., 2017), including peer review transparency and triangulation of results across methods or datasets.

Localization (*L*) incorporates the contextual and cultural relevance of research—its alignment with national priorities and local knowledge systems—ensuring that global standards remain adaptable to diverse epistemic environments.

Collectively, these eight components provide a theoretically grounded, multidimensional conception of research quality that integrates epistemic integrity, methodological soundness, and societal value. Unlike traditional citation-based metrics, the ASV emphasizes *qualitative coherence and cross-domain reproducibility* as the essence of scientific advancement.

2.2. Relation to Multicriteria Decision Models (MCDMs)

The conceptual architecture of the SRL + ASV model partially overlaps with the logic of multicriteria decision-making (MCDM) frameworks, which aim to rank or evaluate alternatives on the basis of multiple, often conflicting, criteria. In recent decades, models such as the AHP (analytic hierarchy process) (R. W. Saaty, 1987), TOPSIS (Technique for Order Preference by Similarity to Ideal Solution) (Hwang & Yoon, 1981), VIKOR (VlseKriterijumska Optimizacija I Kompromisno Resenje) (Opricovic & Tzeng, 2004), and CRITIC (Criteria Importance Through Intercriteria Correlation) (Diakoulaki et al., 1995) have been widely applied to prioritize alternatives in engineering, policy, and management contexts. These methods typically employ *weighted aggregation* or *distance-based ranking* to optimize a predefined objective function.

While the SRL + ASV framework shares MCDM's general intent—to balance multiple qualitative and quantitative indicators—it diverges from classical decision models in several fundamental aspects.

First, SRL + ASV is not a static ranking algorithm but a two-dimensional epistemic model. The SRL (scientific readiness level) dimension represents *processual* maturity—the stage of scientific development and methodological consolidation—whereas the ASV (added scientific value) dimension expresses *qualitative enrichment*, capturing novelty,

rigor, reproducibility, and openness. Together, they form an orthogonal system of readiness + value rather than a unidimensional optimization task.

Second, MCDM frameworks presuppose the commensurability of criteria, assuming that all indicators can be normalized into a common decision space. In contrast, the ASV intentionally preserves partial incommensurability among epistemic components (e.g., novelty vs. validation), reflecting the multidimensional nature of scientific value, which resists full reduction to numerical comparability. In this sense, SRL + ASV functions as a *meta-evaluative structure* rather than a decision rule.

Third, the mathematical aggregation in SRL + ASV is multiplicative and interpretive, not compensatory. Classical MCDMs (TOPSIS, VIKOR) rely on compensatory logic, where a high score on one criterion can offset a low score on another. The SRL + ASV model instead applies a weighted geometric mean, which penalizes weak dimensions and rewards balanced epistemic profiles, thus aligning with the integrity principle of scientific assessment.

Fourth, MCDM methods optimize outcomes under fixed weights, whereas SRL + ASV integrates dynamic, discipline-sensitive weighting derived from expert calibration (Delphi–AHP (Analytic Hierarchy Process) synthesis). The model thus adapts to the epistemic context rather than imposing a universal weighting scheme.

Finally, SRL + ASV incorporates a temporal and developmental logic that is absent from conventional MCDM frameworks. Readiness levels (SRLs 1–10) imply progression over time, enabling longitudinal tracking of research evolution, whereas the ASV captures the dynamic accrual of quality through openness, collaboration, and validation.

In summary, although SRL + ASV can be represented mathematically through MCDM-like operations, it outperforms decision-analytic optimization by embedding an epistemological structure into the evaluative process. It is therefore more accurately described as a hybrid readiness–value architecture, integrating the procedural maturity of science (SRL) with the multidimensional valuation of research quality (ASV).

2.3. Comparative Frameworks for Research Evaluation Systems

In addition to technological and readiness-oriented frameworks, a substantial body of research addresses systemic approaches to research evaluation. These include both ex ante and ex post assessment models, as well as performance-based research funding systems (PRFSs) that explicitly link evaluation outcomes to institutional or financial decisions.

Ex ante evaluation refers to the assessment of research quality and feasibility prior to implementation, typically during proposal or grant selection (Geuna & Martin, 2003). It focuses on peer review, expert panels, and the predictive validity of proposed outcomes. In contrast, ex post evaluation assesses results after project completion, emphasizing outputs, impact, and efficiency (Bozeman & Melkers, 1993). Many national systems employ hybrid models, combining ex ante selection with ex post auditing or impact assessment.

Performance-based research funding systems (PRFSs) represent a distinct and institutionalized category. Their core principle is to distribute public research funds based on measurable outcomes such as publications, citations, patents, or verified societal impact. Examples include the UK Research Excellence Framework (REF), the Australian Excellence in Research for Australia (ERA) initiative, the New Zealand Performance-Based Research Fund (PBRF), and the Norwegian Publication Indicator (Hicks, 2012; Aagaard, 2015). Each system translates evaluations into quantifiable performance signals but differs in terms of governance design, data aggregation, and weight structures.

The REF (UK) is perhaps the most extensively studied model. Conducted periodically (every 6–7 years), it combines expert peer review with quantitative indicators and case studies assessing impact. The strengths of the REF lie in its transparency and multidimen-

sional assessment (outputs, environment, and impact), but it has been criticized for its administrative burden and potential behavioral distortions such as "gaming" and strategic hiring (Sivertsen, 2019). The ERA (Australia), which is coordinated by the Australian Research Council, relies more heavily on bibliometric indicators, journal rankings, and field-based citation benchmarks. The ERA's comparative advantage is its scalability and discipline sensitivity, whereas its main weakness lies in its overreliance on citation-based metrics in certain fields (Jajo & Peiris, 2021).

Beyond these national systems, comparative studies (Hicks et al., 2015; Glänzel & Moed, 2013) highlight a continuum between indicator- and peer-driven approaches. Indicator-heavy systems (ERA, PRFS) excel in efficiency and coverage but risk superficiality and Goodhart-type distortions; peer-heavy systems (REF, DFG's *Forschungsrating*) emphasize contextual validity but are more costly and slower to update. An emerging consensus suggests that effective evaluation architectures combine both logics—quantitative evidence embedded in qualitative peer interpretation—an equilibrium that SRL + ASV explicitly seeks to operationalize.

Furthermore, recent analyses of evaluation timing (Marx & Bornmann, 2016) stress that ex ante and ex post approaches should not be seen as alternatives but as stages in a feedback cycle. Evaluation before funding ensures feasibility and alignment with priorities, whereas evaluation after completion ensures accountability, learning, and improvement. This dynamic feedback aligns closely with our SRL + ASV framework, where readiness levels guide ex ante decision thresholds and added scientific value enables ex post improvement loops.

Overall, the international experience with the PRFS, REF, ERA, and similar systems underscores both the necessity of structured evaluation and the dangers of metric overreliance. These lessons reinforce our central proposition: evaluation systems must integrate maturity (readiness) and quality (value) within reproducible, auditable procedures—precisely the dual-axis logic that is formalized in SRL + ASV.

2.4. Comparative Positioning and Novel Contribution of SRL + ASV

To clarify the conceptual novelty of the SRL + ASV framework, it is important to distinguish it from existing readiness-based checklists, multicriteria decision models, and established assessment programs.

Table 1 summarizes these relationships and highlights how SRL + ASV bridges process- and value-oriented paradigms of research evaluation.

Table 1. Comparative positioning of SRL + ASV in relation to established evaluation frameworks.

Framework	Core Logic	Evaluation Focus	Aggregation Logic	Limitations	Novel Contribution of SRL + ASV
ESA (2023) SRL	Sequential maturity checklist	Process readiness	Discrete, linear stage model	Ignores research quality and epistemic value	Adds continuous value dimension (ASV) to readiness stages
MCDM (AHP, TOPSIS, VIKOR, CRITIC)	Multicriteria optimization	Weighted scoring of alternatives	Additive or compensatory	High-value metrics can offset weak rigor	Introduces noncompensatory dual filter (readiness × value)
Composite Indicators (OECD, JRC)	Normalized indices across domains	Statistical comparability	Additive or mixed	Focus on macrolevel benchmarking	Applies composite logic to microlevel research units

Table 1. *Cont.*

Framework	Core Logic	Evaluation Focus	Aggregation Logic	Limitations	Novel Contribution of SRL + ASV
Responsible Metrics (*DORA*, CoARA, Metric Tide)	Qualitative and policy-driven	Ethics, transparency, diversity	Narrative and contextual	Lacks operational quantification	Translates qualitative principles into reproducible indicators
SRL + ASV (this study)	Dual filter of readiness × value	Process + quality integration	Multiplicative, noncompensatory	Requires empirical calibration	Hybrid model uniting maturity logic and value-based metrics

The SRL + ASV framework thus contributes a hybrid architecture that unifies the structural discipline of readiness assessment with the normative goals of responsible research evaluation. It operationalizes fairness and contextualization within a reproducible quantitative structure while preserving noncompensatory integrity between procedural maturity and epistemic merit.

3. Materials and Methods

3.1. Conceptual Research Design and Methodology

This study follows the conceptual research approach, which focuses on developing and formalizing theoretical constructs rather than collecting or statistically analyzing empirical data. Conceptual research seeks to integrate fragmented theoretical domains into a coherent framework, producing new intellectual syntheses through abstraction, comparison, and logical modeling (Meredith, 1993; Jaakkola, 2020). Accordingly, the present work does not test hypotheses using datasets but derives a reproducible conceptual model (SRL + ASV) that unites readiness-based staging and value-based evaluation within a single decision architecture.

The study adopts a constructive–conceptual approach, combining design science (Hevner et al., 2004) with conceptual modeling traditions in evaluation and systems theory. This hybrid design allows for the development of an operational model while preserving theoretical abstraction. The research is therefore nonempirical but structured, oriented toward model construction, formalization, and validation through internal coherence and external correspondence to established frameworks (e.g., TRL, SRL, *DORA*, FAIR).

The inquiry is deductive and synthetic. It proceeds from the critical synthesis of existing frameworks (technology readiness levels, scientific readiness levels, responsible metrics, and open science principles) toward the formulation of an integrative model. The logic of model construction follows three methodological steps:

Abstraction—identification of universal dimensions of research assessment (maturity and value);

Mapping—formal representation of relationships between components;

Integration—creation of a dual-axis architecture linking readiness levels and quality indicators.

The study relied exclusively on secondary and conceptual materials: theoretical frameworks (NASA TRL, ESA SRL), international declarations (*DORA*, Leiden Manifesto, FAIR principles), and meta-research works on reproducibility, evaluation, and responsible metrics. No new empirical or survey data were collected. Instead, these sources provided normative and comparative reference points for model construction.

Analytical methods used include comparative modeling, system structuring, and deductive integration. The analytical instrument is the SRL + ASV matrix, which links discrete readiness levels to continuous quality dimensions. The model was validated conceptually by assessing its internal consistency, theoretical completeness, and alignment with known evaluation standards (conceptual validity, referential adequacy, and logical coherence).

Interpretation proceeded through conceptual analysis and reflexive validation rather than statistical testing. The adequacy of the model was evaluated based on three aspects:

Construct validity—does the SRL + ASV model faithfully represent the underlying constructs of maturity and value?

Theoretical integration—does it unify the previously separate paradigms of readiness and quality evaluation?

Practical applicability—can it inform institutional decision-making processes (illustrated through the Kazakhstan case)?

Conceptual modeling, by nature, prioritizes internal logical consistency and systematic synthesis over empirical generalization. Its main limitation is the absence of direct statistical validation, compensated for here by a correspondence with existing normative frameworks and the reproducible logic of the construction. The model is thus a conceptual artifact—a structured theoretical system that is designed to be operationalized and empirically tested at a later stage.

3.2. Model Architecture and Operational Logic

We interpret our work as a construct-oriented study. That is, rather than extracting "effects" from an external dataset, we derive and operationalize a coherent evaluative system that explicitly links the stepwise maturity of a scientific project (SRL) with its continuous added scientific value (ASV). The methodological aim is to eliminate the traditional divide between the language of staging ("*how far has it progressed?*") and the language of quality ("*how valuable and reliable is it?*") by proposing a reproducible algorithm. In this framework, the decision to advance a project is derived from the class of presented evidence (SRL) and the coherence of quality practices (ASV). Accordingly, the procedure is designed as a managed architecture with explicit inputs, normalization, aggregation, calibration, threshold conditions, and embedded loops for monitoring and auditing.

The SRL is treated as a discrete ten-level scale ranging from *pre-idea* (Level 0) to *impact* (Level 9). Each level requires a new class of evidence that cannot be substituted by merely strengthening the previous one: for example, the transition from *evidence to replication* requires an independent external replication rather than simply enlarging the initial sample. The *validation*, *application*, *scaling*, and *impact* levels impose progressively higher demands for multicentricity, standards, and robustness of effects. This ensures qualification-based rather than cumulative progression: adding "points" without presenting the necessary class of evidence does not lead to advancement.

The ASV is defined as a vector of quality components $X = (N, R, P, V, C, I, O, L)$, where

N—novelty;

R—methodological rigor;

P—reproducibility/replicability;

V—venue quality and robustness of peer review;

C—field- and age-normalized citation impact;

I—external influence (e.g., *policies*, *standards*, *implementation*, *benchmarks)*;

O—openness (data/code/protocols, *preregistration/registered reports)*;

L—collaboration and multicentricity.

Each component $x_i \in [0, 1]$, with normalization and scale anchoring ensuring comparability. The choice of these dimensions is grounded in the contemporary quality and

reproducibility agenda (rigor, openness, replication) and institutional norms of *responsible metrics*.

The aggregate added value, A, is constructed as a multiplicative aggregation of normalized components with discipline-calibrated weights w_i:

$$A = \prod_{i=1}^{8} x_i^{w_i,}, \sum w_i = 1$$

where x_i are the normalized values, and w_i are the weighting coefficients, which sum to 1.

The geometric form is deliberate: it enforces a "*weakest-link penalty*". No single strong dimension (e.g., citation count or journal prestige) can compensate for deficits in reproducibility, openness, or rigor. Thus, the ASV rewards coherence across practices rather than performance on one or two indicators. In practice, the weights w_i are program- or discipline-specific (see the calibration section), but the principle of multiplicativity remains invariant.

The disciplinary weights (w_i) were calibrated through a two-stage process combining Delphi expert evaluation and analytic hierarchy process (AHP) ranking. In the first stage, experts (n = 12, representing STEM, social sciences, and humanities) independently ranked the relevance of each ASV component within their field. In the second stage, normalized consistency ratios were computed to ensure logical coherence (T. L. Saaty, 2002). The resulting weights were scaled to unity ($\Sigma w_i = 1$) and validated through sensitivity analysis, confirming model stability within $\pm 10\%$ parameter variation.

Once A is calculated, the project is compared against the requirements of the current SRL k. Advancement depends on two coupled checks:

1. The evidence checklist for level k (e.g., presence of an external replication for advancement from *evidence to replication*; multicenter validation for *validation*);
2. The value threshold for level k *is as follows*:

$$A \geq \tau_k$$

where $A \in [0, 1]$ is the aggregated ASV, and $\tau_k \in (0, 1)$ is a calibrated threshold for level k.

Only when both conditions are satisfied is the project considered ready to advance to SRL $k + 1$. If $A < \tau_k$ or the checklist is incomplete, the decision shifts to "revision."

This dual-filter logic prevents metric-centrism (where strong bibliometric signals mask methodological weakness) and blocks advancement without the required new class of evidence.

If the checklist is satisfied but $A < \tau_k$, the project receives an improvement roadmap—targeted actions to raise deficient components, x_i. If $A \geq \tau_k$, advancement to level $k + 1$ is confirmed. If the checklist is not satisfied, the inequality $A \geq \tau_k$ alone is insufficient, since new evidence is mandatory. This preserves the functional symmetry between the SRL (staging) and ASV (quality).

A refusal to advance does not constitute a "final verdict." Instead, the system automatically generates an improvement plan, prioritizing deficits by leveraging (e.g., opening data/code, conducting independent replication, enhancing design rigor, and aligning with benchmarks/KPIs). KPIs and milestones are then established; at interim review points, A is recalculated, SRLs are reassessed, and adjustments are made as needed. Thus, evaluation becomes a managed trajectory of development rather than a one-off procedure.

The methodology incorporates formal appeal windows and an audit trail: the entire evaluation process (artifacts, normalization, computations, decisions) is stored in a verifiable documentation package. All computational steps (tabular data, aggregation, diagrams) are executed in a reproducible computing environment; the results (including hyperlinks to sources) are versioned and can be rerun for verification. This eliminates "black box" effects

and supports trust in the process. The outcomes of appeals and audits are fed back into calibration loops (updating weights, thresholds, and parameter profiles).

All these elements are integrated in the SRL + ASV evaluation architecture: inputs (artifacts and sources) undergo normalization and verification, followed by ASV integration and then consolidation with the SRL to form a combined stage indicator (used for monitoring). Decisions are rendered via the checklist–threshold logic: if "not ready," the improvement loop is triggered; if "ready," advancement is recorded, and requirements for the next SRL are updated. At every stage, computational reproducibility and documentation of decisions are maintained.

In summary, the methodology defines a transparent and reproducible pathway from artifacts to decisions. The SRL specifies the class of required evidence, while the ASV evaluates the coherence of quality practices within the stage. Advancement occurs only when both conditions are jointly satisfied. Multiplicative aggregation eliminates compensatory masking of critical weaknesses, whereas calibrated parameters allow for adaptation to disciplines and program priorities without sacrificing comparability. Finally, monitoring, appeals, and auditing transform evaluations into a learning system, where the outcomes of each cycle refine the parameters of the next.

3.3. Threshold Calibration and Dual-Filter Logic

The dual-filter logic constitutes the operational core of the SRL + ASV framework. It ensures that research evaluation proceeds through two distinct and sequential verification layers: one is structural, related to *readiness*, and one is qualitative, related to *value*. This design prevents premature advancement of projects that are not methodologically or conceptually mature, while simultaneously avoiding rewarding inflation for work that lacks scientific integrity or originality.

First filter: Scientific Readiness Level (SRL)

The first filter functions as a structured checklist assessing whether the project or study meets the essential evidentiary and procedural requirements for its claimed maturity stage. Each SRL is associated with concrete indicators, such as the presence of validated hypotheses, reproducible data, documented protocols, and independent verification. Only after all mandatory readiness criteria are fulfilled does a project become eligible for the second filter.

In practice, these criteria are calibrated through expert consensus and benchmarking against established frameworks. This allows for contextual adaptation to disciplinary norms while maintaining international comparability.

Second filter: Adding scientific value (ASV)

The second filter evaluates the *intrinsic quality* of the research output through the eight ASV components: novelty, rigor, reproducibility, openness, collaboration, impact, validation, and localization. Unlike the binary logic of SRLs, the ASV filter uses continuous normalized indicators that reflect the relative standing of a project within its field.

Calibration of ASV thresholds involves defining discipline-specific reference percentiles (e.g., national or institutional benchmarks) that distinguish between baseline, advanced, and exemplary performance. Thresholds are determined empirically via expert review and iterative comparison with known exemplars in each domain.

Integration of the two filters

The two filters operate hierarchically: SRL governs eligibility, and ASV determines advancement. This ensures that readiness (process maturity) and value (scientific quality) remain analytically distinct yet mutually reinforcing.

A project cannot progress to a higher SRL stage without meeting both conditions: (1) full compliance with SRL criteria and (2) attainment of at least the minimal ASV thresh-

old for its discipline. Conversely, achieving ASV excellence cannot compensate for a lack of methodological maturity.

This *noncompensatory interaction* is a defining feature of the SRL + ASV architecture, ensuring balance between procedural rigor and epistemic merit.

Calibration and validation

The thresholds for both filters are refined through continuous calibration. During pilot testing, the model can be applied to a sample of projects from different SRLs, allowing experts to iteratively adjust cutoff points on the basis of observed consistency and fairness. Sensitivity analyses—varying expert weights, indicator normalizations, and percentile cutoffs—help verify that the thresholds produce stable and transparent classifications.

Ultimately, this dual-filter design establishes an auditable decision trail, making the evaluation process reproducible, resistant to metric manipulation, and adaptable across scientific disciplines.

3.4. Integrity Safeguards and Anti-Manipulation Design

While no aggregation scheme can entirely eliminate strategic behavior, the SRL + ASV architecture reduces its feasibility through structural and procedural safeguards.

The noncompensatory design discourages metric-centrism by penalizing weak performance in any critical component. In addition, integrity checks, such as independent verification of reproducibility claims, dual-role expert reviews, and random audits of evaluation records, ensure that compliance is evidence-based rather than declarative.

All evaluations generate transparent metadata, enabling community scrutiny and reputational accountability. Together, these mechanisms shift incentives from performative compliance toward substantive scientific quality.

Since the SRL + ASV model relies on a multiplicative (geometric) aggregation of normalized indicators, the 'weakest link' effect is structurally embedded: any drop in one parameter (e.g., reproducibility or openness) immediately reduces the overall added scientific value. Moreover, the theoretical form of the model allows for testing local stability under small variations in the weight coefficients w_i. In the next stage of research, we plan to conduct a series of sensitivity tests on cloud infrastructure, introducing perturbations of each w_i by $\pm$5–10% while maintaining the normalized sum of weights ($\Sigma w_i = 1$). These tests enable a quantitative assessment of the model's elasticity, the construction of stability-phase surfaces, and a comparative analysis of geometric versus arithmetic aggregation. The results will be presented in a separate publication devoted to the computational validation and dynamic properties of the SRL + ASV architecture.

4. Results

4.1. The Scientific Readiness Level (SRL) Scale

We propose a ten-level scale of scientific readiness levels (SRLs), presented in Table 2.

Clarification of the SRL 3–4 boundary (Table 2):

SRL 3 ("evidence") corresponds to internally validated studies in which data integrity and methodological reliability have been established within the originating team.

SRL 4 ("independent reproduction") requires verification under conditions of independence—distinct datasets, research teams, or institutional contexts.

Minimal independence may be satisfied by alternative data sources, replication by unaffiliated researchers, or institutional separation between the original and verifying teams.

For borderline cases (e.g., multicenter data within one consortium, replication by close coauthors), the study may be classified as a qualified replication (SRL 3.5) until a third-party verification confirms full external reproducibility.

Table 2. Scale of scientific readiness levels (SRLs).

Level	Attribute	Criterion
0	Pre-Idea	Intuition/sketches without hypothesis formulation
1	Idea	Clear formulation of a hypothesis/research question; minimal preliminary literature review
2	Design	Research protocol/design; preregistration/analysis plan
3	Evidence (Pilot/Preprint/Article)	Initial results obtained; preprint or journal article published
4	Replication	Independent confirmation/replication of results
5	Method/Tool	Stable method/dataset/code, ready for external reproduction
6	Validation	External validation (multisite, multidata); meta-assessment of effect robustness
7	Application	Experimental implementation/prototype with users; KPI measurement
8	Scaling	Standardization; compatibility; benchmarking
9	Impact	Sustained influence (policy/industry/science); recognition and reproducibility by default

Conceptually, our scale is analogous to the classical SRL framework (Knar, 2024a), but it is more formalized, compact, and authentic. Table 2 delineates ten sequential levels of scientific readiness, ranging from the pre-idea state (Level 0) to sustained societal, industrial, and scientific impacts (Level 9). Together, they form a coherent trajectory for a scientific project: from the intuitive articulation of a research idea, through design and initial evidence, to independent confirmation, standardization, implementation, scaling, and, ultimately, durable influence.

The internal logic of the scale is dual. First, each level requires a qualitatively new class of evidence that cannot be substituted by intensifying the previous one: adding another pilot study cannot substitute for external replication, just as increased citations cannot replace multicenter validation. Second, progression is cumulative: evidence from "lower" levels retains its significance but becomes a background against which more stringent checks and operationalized outcomes must be demonstrated.

- At Level 0 (pre-idea), the project exists as unformalized intuitions and associative sketches. The focus is on crystallizing the "seed of a hypothesis," mapping the context, and identifying knowledge gaps.
- Level 1 (idea) marks the transition from intuitive to articulated: a hypothesis is formulated, minimal literature reconnaissance is performed, and operational variables and expected relationships are specified.
- Level 2 (design) raises the bar to a reproducible protocol: preregistration, plan for analysis, inclusion/exclusion criteria, primary and secondary endpoints, assumptions, and strategies for ruling out alternative explanations.
- Level 3 (evidence) represents the first empirical materialization of a hypothesis: pilot results are obtained, and/or a preprint or article is published. The focus is on the validity of a basic empirical link, even in limited samples, with proper reporting and data availability.

- Level 4 (replication) demands independent confirmation: external groups, alternative samples/models/sites, and consistent derivation of outcomes under contextual variation.
- Level 5 (method/tool) captures the stabilization of research tools: methods, datasets, or codes are "packaged" such that external researchers can apply them and reproduce results without undue adjustment.
- Level 6 (validation) shifts attention to multisite, cross-data, and cross-paradigm robustness: meta-analysis, heterogeneity assessment, predictive testing on holdout data, robustness checks, and stress tests of the protocol.
- Level 7 (application) marks initial implementation in real-world settings: experimental deployments, pilot KPIs among users/partners, and justified trade-offs between accuracy, cost, and speed.
- Level 8 (scaling) requires compatibility and standards: benchmarks, performance profiles, cross-platform comparisons, fixed interfaces, and alignment with industrial/regulatory norms.
- Level 9 (impact) implies durable, default reproducibility and influence: incorporation into practices, standards, and policy, with broad recognition and reproducibility being maintained without active intervention.

This description and the explicit criteria listed in Table 1 provide minimal and verifiable foundations for advancement decisions, preventing the substitution of one requirement for another.

Overall, the SRL scale ensures qualification-based rather than cumulative progression: each new step introduces a distinct class of verifiable evidence, preventing the mere "accumulation of points" in one dimension from driving advancement. For example, the transition from Level 3 to Level 4 cannot be achieved simply by enlarging the sample size in a single center or by increasing the statistical power; it requires replication as a methodological operation. This mitigates risks associated with metric-driven management (Goodhart's law): indicators serve as signals of the evidentiary class, not as ends in themselves.

The downward dependency of *hypothesis–design–evidence–replication–validation–application–scaling–impact* captures the evolution from explanatory to predictive power and from local validity to cross-context robustness. At the intermediate levels (4–6), the scale resonates strongly with the agenda of open science and the reproducibility crisis: here, responsible methodology demands external replications, open data/code, robustness-oriented design, and multicenter testing—transforming "best practices" into mandatory criteria for progression.

Table 1 provides an operational "language of agreement" among research teams, management, and funding programs: advancement is not the sum of metrics but the crossing of binary checkpoints ("yes/no") corresponding to classes of evidence. This interpretation aligns with the practices of institutional readiness scales in science and technology, where progression to a higher level is only permitted upon presentation of a new class of evidence (e.g., from laboratory demonstration to testing in relevant environments; from a local effect to multisite robustness). In this context, Table 1 is also convenient for threshold calibration: at each level, a minimum integral *added value* (in our SRL + ASV model) can be set in accordance with the qualitative checklists of that level, thereby linking "staging" to "value" without conflating the two.

In practice, some projects will inevitably "stall" between levels—particularly during transitions from Level 3 to 4 and Level 5 to 6. The reasons are familiar: an external group replicates the effect only partially; repositories exist but are insufficient for seamless reuse; and meta-analyses are contradictory owing to task heterogeneity. Table 1 is useful precisely because it does not "average" such situations into scores such as "3.5." Instead, it

forces action either to complete the missing class of evidence (e.g., achieve full openness and conduct an independent replication) or redesign the study. In combination with the improvement map in the SRL + ASV framework, this structure allows for the diagnosis of precise deficits, identifying the exact "bottleneck" preventing advancement.

Although the overall structure of the levels is cross-disciplinary, their substantive contents differ. In biomedicine, validity often requires multicenter trials and preregistrations; in computational sciences, open repositories, benchmarks, and reproducibility across diverse hardware–software configurations are needed; and in the humanities, finer criteria of interpretive reliability, extended openness of corpora, and transparency of annotation protocols are called for. The tabular scale thus provides a scaffold into which disciplines can "embed" their evidentiary classes without losing cross-disciplinary comparability.

By itself, the stepwise SRL scale answers the question "*How far has the project advanced?*" but does not guarantee that progress is accompanied by scientific value. Integration of Table 1 with the quantitative ASV axis addresses this gap: each stage is associated with a threshold of integral value, and the overall stage index requires joint coherence of maturity and quality. Table 1 remains a stable normative framework for "classes of evidence," while the quantitative component adapts to programmatic horizons and risk priorities.

Because advancement depends on the emergence of a new class of evidence, rather than on the growth of indicator sums, the incentive to "optimize" individual metrics at the expense of methodological rigor is reduced. Citations, venue prestige, and even repository indicators—while important signals—do not substitute for replication, open data/code, multicenter validation, and standardization. In this way, the SRL scale, coupled with the ASV, creates an "anti-Goodhart" framework: metrics function as telemetry, not as the steering mechanism.

Table 1 is "strict" only insofar as it is required for clear project decisions; otherwise, it remains "permeable" to feedback. For example, if a project at Level 5 struggles to reach Level 6 due to data heterogeneity, the scale directly signals which ASV components must be strengthened (e.g., P, O, and R—reproducibility, openness, and rigor) to achieve sustainable validation. Thus, the scheme functions as a roadmap management tool: it not only states "*where we are*" but also suggests "*what to do next*," preserving verifiability and comparability across projects.

Conceptually, Table 1 is fully compatible with traditional SRL methods (notably ESA SRL) and can be mapped onto operational regulations of project–program cycles, where level advancement constitutes a formal event in the lifecycle. The stage names (validation, scaling, impact) and their substantive cores resonate with institutional checklists and documentation.

Accordingly, Table 1 can serve as a shared layer of dialog between disciplines and institutions: sufficiently universal to avoid being a "one-size-fits-all suit" yet specific enough to support formal decisions on progression and prioritization.

4.2. The Added Scientific Value (ASV) Scale

We interpret the ASV paradigm as an integral "*vertical*" measure of value within a stage. Conceptually, it is composed of submetrics that are normalized to the interval of [0, 1].

In this context, we propose eight universal components of the ASV scale, presented in Table 3.

Table 3 describes the "vertical" dimension of assessment: the ASV represents the integral and highlights the scientific value of a project. Unlike the stepwise SRL scale, which fixes the transition across classes of evidence, the ASV quantifies quality and significance *within a given stage* by means of normalized components $x_i \in [0, 1]$, which are then

aggregated into a single index. In this manuscript, we define the ASV precisely as such an *intrascale integral value.* Each component is specified in a component–attribute–criterion format, making them transparent for calibration and auditing.

Table 3. Added scientific value (ASV) scale.

Component	Attribute	Criterion
N	Novelty	Originality and heuristic productivity
R	Rigor	Methodological rigor (design, statistics, testing of assumptions)
P	Reproducibility	Replicability/repositories of data and code
V	Venue Quality	Quality of venue (journal/quartile/strength of peer review)
C	Citability	Field- and age-normalized citation indicators (FNCI/percentile)
I	Impact	External influence (policy, industry, standardization, benchmarks)
O	Openness	Open science practices (data, code, preregistration, reporting)
L	Collaboration	Collaboration and external verification (international, multisite)

We conceptualize the ASV as a component vector $X = (N, R, P, V, C, I, O, L)$, where each coordinate represents a distinct aspect of quality: novelty (N), rigor (R), reproducibility (P), venue quality/peer review strength (V), field- and age-normalized citability (C), external impact (I), openness (O), and collaboration/multicentricity (L).

For each component, Table 3 specifies its attribute and operational criterion, thereby standardizing data collection and verification (e.g., for P: "replicability/repositories of data and code"; for V: "venue quality and peer review strength"). Each component x_i is normalized to [0, 1] according to a predefined scheme (anchor values, transformation scales, data sources). This ensures comparability and allows for aggregation—precisely in line with our methodological formulation that "*ASV is computed from submetrics normalized to [0, 1].*"

The choice of these eight components is not arbitrary. It synthesizes insights from meta-research on quality and replicability with institutional norms of *responsible metrics* (*DORA*, Leiden Manifesto) and *open science* (FAIR). Thus, the ASV is embedded within international standards of responsible evaluation rather than being substituted for it.

- Novelty (N) captures originality and heuristic productivity. Methodologically, expert scales with anchors are preferable (e.g., 0 = incremental/replication with no new claim; 0.5 = locally new combination or method transfer; 1.0 = qualitatively new concept/method/class of phenomena), supplemented by textual justification and debiasing procedures (blind review, independent secondary evaluation).
- Rigor (R) is the determinant of validity: preregistration (where applicable), near-optimal statistical power, correct handling of multiple testing, robustness checks, and sensitivity to modeling choices and hyperparameters. A checklist approach with weighted "critical" practices is recommended.
- Reproducibility (P) has two facets: replicability (independent reproduction) and reproducibility (full openness of data/code/protocols, down to bit-for-bit). The evaluation of P should account for both the number and diversity of replications (different centers, datasets, and computational environments).

- Venue Quality (V) is not "journal-centrism" but rather a proxy for the quality of the review environment. Since the *DORA* and the Leiden Manifesto caution against overreliance on journal metrics, V should be treated as a *subtle signal*: rigor of review, transparency of processes, availability of open reviews, and postponement oversight. This is captured in the criterion "venue quality/peer review strength."
- Citability (C) must be strictly normalized by field and age (e.g., the FNCI percentile) to avoid penalizing disciplines with low baseline citation levels or rewarding hypercited fields disproportionately. The scale emphasizes this field-normalized interpretation, directly addressing the main methodological critique of raw citation counts.
- Impact (I) reflects external effects: citations in policy documents and standards, industrial implementations, practical KPIs, and benchmarks that "shift the field." To ensure objectivity, documented sources (policy reports, standards, patents, industry reports) and transparent counting rules are essential.
- Openness (O) measures the completeness of open science practices: preregistration/registered reports, open data, open code, open reviews, and protocols. As the backbone of reproducibility, O may justifiably be assigned greater weight in disciplines that are data- and computation-intensive.
- Collaboration (L) captures multicentricity, internationalization, and multi-institutional verification of hypotheses. Practical evaluation can use indicators such as "international authorship structure" and "number of independent centers," capped by an upper saturation limit.

In summary, the ASV provides a transparent, auditable, and internationally aligned framework for quantifying scientific value *within stages.* Its structure operationalizes the principles of responsible metrics and open science while minimizing the risks of metric fetishization.

Notably, additive aggregators allow for "compensation" of critical failures; for example, a high venue quality (V) or citability (C) score could mask zero reproducibility, which is methodologically unacceptable. In contrast, the geometric aggregator enforces complementarity: a destructive deficit in any component is penalized exponentially. We believe that this logic precisely reflects the aims of responsible evaluation: the ASV does not allow projects to "succeed" by proxy metrics while having zero openness or replication.

Naturally, the ASV presupposes an evidentiary basis for each component x_i. This includes links to repositories, DOIs, peer review protocols, replication reports, and policy/standard documents. For novelty (N) and, to some extent, impact (I), expert panels are needed, with interrater reliability (ICC) being assessed and debiasing procedures being applied (blind review, independent secondary evaluation).

Importantly, ASVs do not exist in isolation. Growth in the integral indicator A acquires operational meaning only in conjunction with the SRL. This is the rationale for introducing thresholds in our model: a project at level k is deemed ready to advance only if the SRL checklist is satisfied and $A \geq \tau_k$. This coupling prevents metric-centrism: one cannot compensate for the absence of a fundamentally new class of evidence (e.g., external replication) by higher citation counts or journal prestige; conversely, rigor and openness alone cannot justify advancement without the evidentiary class required by the SRL step. This logic is consistently applied throughout this manuscript (see the linkage of the SRL and ASV sections and the subsequent SRL + ASV threshold matrix).

In summary, the ASV operationalizes "quality" as the ensemble of reproducible practices rather than a surrogate for journal rankings. The geometric aggregator, combined with discipline-calibrated weights, safeguards the accountability of metrics: no single coordinate functions as a "joker." Through normalization to $[0, 1]$, all the components become comparable, and the final value of A is transparently interpretable and easily thresholded.

Together with the SRL, the ASV establishes a closed management loop, moving from the diagnosis of deficits (improvement map) to targeted actions (opening data, organizing independent replication, standardizing interfaces, etc.).

Normalization of novelty (N) and impact (I) partly depends on expert judgment; this requires procedures to strengthen interrater agreement and periodic recalibration of anchors. For venue quality (V) and citability (C), strict field and age normalization are necessary to avoid distortions across disciplines. Weights and thresholds require empirical calibration against historical decisions and panel data. The methodological design of Table 2 is adapted to this task: components are specified in a unified "attribute–criterion" format, whereas aggregation and thresholds remain modular and adjustable.

Thus, Table 2 establishes a clear, reproducible, and calibratable morphology of scientific value. It does not replace the SRL, nor does it dissolve into "journal-centric" indicators. In contrast, the ASV makes quality assessments explicit and verifiable, transforming metrics into data rather than goals. The table provides the missing "vertical backbone" that links evidentiary classes to measurable scientific value at every stage—from early demonstrations to scalable impact.

In this configuration, the SRL + ASV model directly fulfills the mandates of the contemporary agenda of responsible evaluation and open science.

4.3. SRL + ASV Evaluation Architecture

In the proposed evaluation architecture, we abandon the conventional separation between the *maturity* and *value* of research, treating them instead as mutually necessary dimensions of a single management task. The process is designed so that every decision—whether to advance a project, pause it, or require revisions—emerges not from the accumulation of abstract points but from verifiable classes of evidence and documented quality practices. The architecture links the discrete levels of scientific readiness to the continuous scale of added scientific value, making the source of each judgment transparent and recording both the provenance of the data and the rationale for each decision.

Feedback loops are deliberately embedded into the system: if a project is not ready for advancement, the outcome is not a verdict of "failure" but a development plan that specifies exactly which elements of research practice need strengthening—openness, reproducibility, independent replications, quality of the review environment, or external impact. Conversely, when a project is ready, progression is not framed as a reward for cumulative scores but as a shift in the required evidentiary class: at the next stage, new forms of validation and a higher level of reporting discipline are expected. In this sense, the architecture is not an external tribunal but rather a grammar of scientific work itself, where advancement is driven not by the rhetoric of metrics but by substantive practices.

Figure 1 illustrates the evaluation architecture, which is based on the integration of two dimensions: the SRL and ASV.

- The left side (SRL) represents a stepwise scale of readiness, tracing the trajectory from initial hypotheses through publications, replications, and international validations to mature results that are recognized by the scientific community and influencing policy or practice. Each SRL is defined by the appearance of a qualitatively new class of evidence.
- The right side (ASV) represents a continuous measure of within-stage quality, incorporating dimensions such as novelty, methodological rigor, reproducibility, openness, independence of peer review, field-normalized citation impact, societal influence, and collaboration. These indicators are aggregated into an overall ASV ranging from 0 to 1.

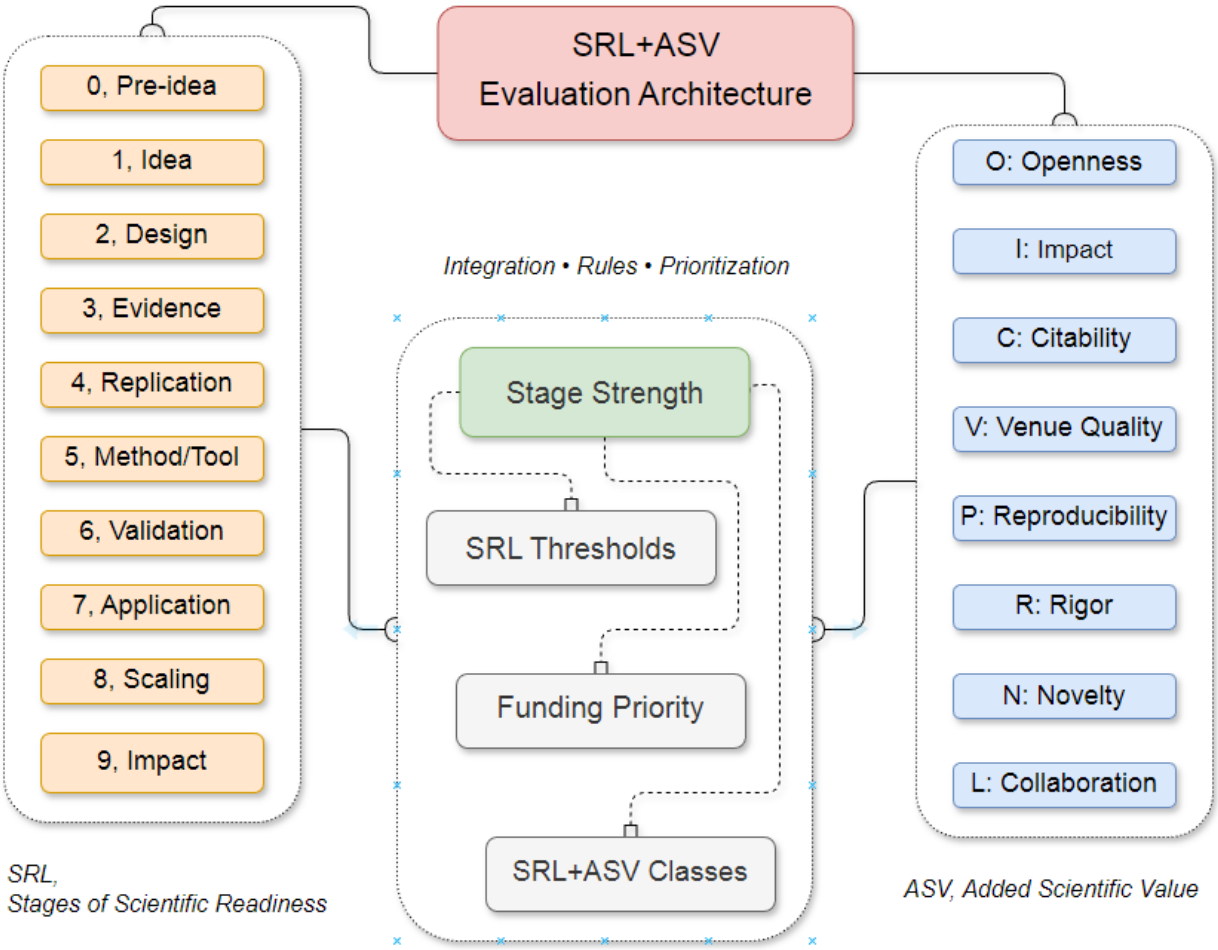


Figure 1. SRL + ASV evaluation architecture.

Integration within the SRL + ASV model is governed by the principle of the "dual filter." First, a project must satisfy the evidentiary checklist for its SRL. Second, it must meet the threshold requirement for quality ($ASV \geq \tau_k$). Only when both conditions are satisfied can the project advance to the next level of readiness. If the ASV threshold is not met, an improvement map is generated, identifying which aspects require targeted enhancement.

This division of functions addresses one of the central limitations of current evaluation systems: the SRL governs horizontal progress (advancement to a new stage of maturity), whereas the ASV governs vertical depth (how well the current stage has been mastered). As a result, a project cannot "leapfrog" to a higher level solely on the basis of formal signals (e.g., a low-quality publication), nor can high citation counts justify progression in the absence of the new evidentiary class required by SRL.

The methodology therefore combines discrete logic (SRL) with continuous logic (ASV). Its practical value lies in enabling the use of SRL + ASV in grant competitions and institutional policies: the SRL guarantees the comparability of projects across stages, whereas the ASV differentiates "strong" projects from "weak" ones within the same stage. Flexibility is ensured through the calibration of ASV weights and thresholds to specific disciplines and national contexts.

To avoid self-referential loops, the recalculation of the ASV after each improvement cycle is performed relative to an external benchmark (disciplinary or institutional median), ensuring stability and convergence. The iterative process thus functions as a bounded optimization with asymptotic convergence to equilibrium rather than an uncontrolled feedback loop.

Figure 1 thus demonstrates that the SRL + ASV architecture is a balanced and reproducible evaluation mechanism, where readiness and value reinforce one another. The simplicity of the logic (SRL step + ASV threshold) makes the model operationally convenient, whereas adjustable parameters allow for adaptation across diverse research and governance settings.

In summary, the SRL + ASV evaluation architecture defines a transparent, reproducible pathway from evidence to decision. It unites two complementary logics—staged readiness and within-stage value—into a framework that is both rigorous and adaptable, simultaneously ensuring comparability, accountability, and constructive guidance for research trajectories.

The block structure presented in Table 4 demonstrates that the SRL + ASV architecture is not conceived as a mechanical algorithm but rather as a management system in which each stage has its own rationale and function. Input data and their validation establish trust in the process; the integration of value captures the quality of research practices; and the synthesis of readiness and value underscores their joint necessity. The *decision gate* translates evaluation into action: progression confirms readiness for a new class of evidence, whereas "nonreadiness" is transformed into a structured improvement plan. Finally, the reproducibility layer keeps the entire architecture within the boundaries of transparency and auditability, ensuring its alignment with contemporary norms of open science and responsible metrics. Thus, each block is not merely a technical operation but a meaningful link in the logic of "readiness + value."

Table 4. Components of the SRL + ASV architecture and their roles.

Block	Meaning	Purpose in the Architecture
Input Data and Sources	The SRL, ASV components, and supporting documents (protocols, open data, publications, replications, policy documents)	Ensures transparency: every claim must have verifiable sources, establishing trust in the system
Normalization and Input Validation	Harmonizing SRL and ASV components to a common scale; validating data accessibility and reference integrity	Eliminates inconsistency: all indicators are comparable, and data availability is guaranteed
Value Integration (ASV)	Aggregation of ASV components into an integral value	Demonstrates coherence of quality practices and prevents "gaming" of isolated indicators
Readiness–Value Integration	Combination of the SRL and ASV into a composite indicator	Highlights the principle that progress is only possible when maturity aligns with value
Decision Gate	Comparison of the ASV with the threshold for the SRL and verification of the evidentiary checklist	Transforms evaluation into a managerial action: advancement or revision
Improvement Loop (Improvement Map)	Identification of deficits, prioritization of components with the greatest growth potential, and creation of a targeted development plan	Makes the system constructive: "not ready" becomes a roadmap for progress
Advancement (Next Level)	Formal recognition of progression to the next SRL, updating of requirements, and restarting of the cycle	Ensures continuity and momentum in project development
Reproducibility Layer	Transparent computational environment with preservation of all data, codes, and documentation	Eliminates the "black box": results are reproducible and subject to audit

The process does not begin with "scores" but with verifiable entities: levels of maturity and markers of scientific value that are anchored in documents and digital artifacts. At the input stage, there are not abstract numbers but rather substantiated claims, protocols and preregistrations, open datasets and codes, publications, reports of independent replications, and evidence of practical uptake. This sets the tone for the entire architecture: rather than comparing "attractive indices," the system verifies evidence. Once normalized to working scales, the architecture does not prematurely conflate maturity with quality: first, a continuous picture of value within the current stage is assembled; only then is it aligned with the stepwise logic of readiness. Importantly, the integrative value is not a mere sum of points: it functions as a coherence check, ensuring that no single indicator can dictate the outcome. At this juncture, the two axes meet: readiness defines the required class of evidence, whereas value demonstrates whether the project upholds the required quality at its current stage.

The branching point that follows is decisive. Advancement signals readiness for the next evidentiary class. A call for revision is not a defeat but a trajectory—identifying which practices can be strengthened with minimal effort to cross the threshold. The entire process is embedded in a reproducible computational environment: calculations, visualizations, and linked sources form a single trace that can be verified and re-executed.

The most important architectural feature is that it removes the longstanding ambivalence between the language of "progress" and the language of "quality." Traditionally, these domains have been siloed: maturity conceived of as formal staging and value as a collection of heterogeneous, often conflicting indicators. Here, the connection is constructive: at each level of maturity, specific classes of evidence are needed; within that level, value is measured as the coherence of key practices, without distortion from one or two highly visible proxies. There is no room for substitution effects: journal prestige cannot stand in for transparency and reproducibility; citation surges cannot replace independent replication; and a striking impact cannot excuse the neglect of methodological rigor. By design, the architecture resists metric-driven strategies that risk Goodhart's law being applicable. If one indicator is artificially inflated without support from the others, the system will still both assign the evidentiary class and require consistency of practices.

Equally crucial is the embedded feedback loop. Denial of progression does not terminate the process but rather structures its continuation: an improvement map, prioritization of deficits, and a return to evidence gathering. In this way, evaluation becomes an instrument of management rather than a scoreboard of accomplishments.

From a practical perspective, the architecture provides a language of transparent, manageable trade-offs. Research programs and funding bodies inevitably vary in priorities: some emphasize external impacts and rapid translation, whereas others stress accuracy, openness, and robustness. The proposed scheme allows these emphases to shift without breaking the methodology. Policymakers can temporarily increase the role of applied impact or, conversely, raise expectations for openness and replication. The underlying logic remains intact: decisions are always made at the intersection of evidentiary class and practice coherence. This adaptability makes the framework suitable across disciplines—from biomedicine to computational sciences to the humanities. Each field may tailor its evidentiary content, but the governing logic of progression and quality remains unified.

An important corollary is the possibility of systematic calibration against empirical data. Because each node of the architecture is tied to explicit sources, adjustments to thresholds or weights can be anchored in evidence: historical panel decisions, the downstream trajectories of funded projects, or the robustness of effects under independent checks. In this way, the architecture not only aligns with the principles of responsible

metrics but also operationalizes them: indicators document the basis of decisions rather than dictating outcomes.

In summary, the SRL + ASV evaluation architecture exemplifies what a mature, responsible yet pragmatic system of research assessment can look like. Its strength lies not in the elegance of the schematic diagram but in the fact that behind each block are verifiable entities, and that behind each arrow, a meaningful transition occurs. Decisions no longer emerge from the accumulation of convenient numbers; they arise from evidentiary classes and disciplined research practices. If a project advances, it is prepared to meet the requirements of the next stage and has demonstrated sufficient quality at the current stage. If it stalls, the architecture transforms "nonreadiness" into a concrete plan for improvement—ensuring that the process itself continues.

From this perspective, high achievement is not a fortuitous sum of disparate signals but rather the point of convergence between maturity and value. Specifically, for this reason, the framework is suitable not only for reporting but also for guiding the daily conduct of scientific work, where each step must be both justified and reproducible.

Field normalization in SRL + ASV relies on publicly available scientometric and institutional datasets such as OpenAlex, Crossref, and Dimensions, applying percentile-based scaling within standardized subject categories and adjusting for publication age cohorts. For the humanities and social sciences, where replication and citation patterns differ, calibration gives greater weight to qualitative ASV components—openness, collaboration, and localization—to maintain fairness and contextual validity across disciplines.

4.4. Scope of Application and Metric Accessibility

The proposed SRL + ASV conceptual model is intentionally designed as a *meta-framework* that can be adapted to multiple levels of research assessment. Its two-dimensional architecture—linking maturity (readiness) and quality (value)—is scalable across different entities, from individual researchers to national research systems. The model does not prescribe fixed indicators but provides a logical template within which appropriate metrics can be mapped and normalized according to the context.

1. *Individual researchers and research groups.*
   At the microlevel, the SRL defines the developmental stage of a researcher's scientific trajectory—from exploratory idea formulation (SRL 1–3) to reproducible, cumulative contributions (SRL 7–10). The ASV component captures added value in terms of novelty, methodological rigor, openness, interdisciplinarity, and societal relevance. Combined, these metrics enable the construction of *SRL + ASV profiles* for career evaluation, mentoring, and capacity-building programs.
2. *Research projects and proposals.*
   For project evaluation, the SRL reflects project readiness (conceptual, experimental, validation, or dissemination stages), whereas the ASV measures expected or demonstrated value (innovation potential, transparency, collaboration, and reproducibility). The model thus supports ex ante funding decisions and ex post performance audits, ensuring continuity between proposal assessment and postcompletion evaluation.
3. *Journals, institutions, and research infrastructures.*
   At the meso- and macrolevels, the SRL can describe the institutional maturity of research environments—e.g., data management, methodological standardization, or openness policies—while the ASV can assess the epistemic value added by journals or institutions (e.g., share of open access content, replication studies, or interdisciplinary connectivity). The framework thereby enables cross-comparison without resorting to citation-centric metrics.

4. *Awards, programs, and policy instruments.*
   The model also applies to strategic decision-making and recognition mechanisms. SRL + ASV can be used to evaluate nominations or grant schemes in terms of their contribution to research culture, identifying whether they reward genuine scientific value or metric performance alone.

The SRL + ASV framework is compatible with both *quantitative* and *qualitative* indicators. All six ASV components—novelty, rigor, reproducibility, openness, impact, and collaboration—can be computed or approximated by existing open access data and tools:

Novelty and *Impact*: via bibliometric or semantic proximity analysis (OpenAlex, Crossref, Dimensions);

Rigor and *reproducibility*: through methodological checklists and replication scores, available in open repositories (e.g., OSF (The Open Science Framework));

Openness: from open access and FAIR data compliance indices;

Collaboration: from coauthorship networks or ORCID-based affiliation graphs.

Where direct computation is not feasible, expert-based scoring or AI-assisted text mining can provide normalized estimates that are consistent with the model's logic. Because SRL thresholds are defined qualitatively and hierarchically, the combination of both dimensions (SRL + ASV) remains robust, even when individual indicators vary in data availability.

Ultimately, the SRL + ASV model serves as a *translational* framework—linking high-level conceptual maturity with operational, data-driven evaluation—making it applicable to researchers, projects, journals, institutions, and policy systems alike.

4.5. The Kazakhstan Case

To further substantiate the need to introduce and integrate the SRL and ASV into assessment practice, we illustrate their relevance through a national case.

In recent years, Kazakhstan, like many other countries, has grappled with the challenge of objectively evaluating research projects in the distribution of public funding. The TRL framework has traditionally served as a universal maturity scale and is particularly suited to applied and engineering tasks. The TRL is effective in describing the trajectory of technology from concept to mass production, which explains its widespread adoption in programs for applied research and innovation.

This logic is also institutionalized in Kazakhstan's new Law on Science and Technology Policy (Law of the Republic of Kazakhstan dated 1 July 2024 No. 103-VIII ZRK), where the level of technological readiness is explicitly defined as a measure of the development and completion of research stages (*Article 16. Technology Readiness Levels*).

However, the direct transposition of the TRL into the domain of fundamental scientific research is, in our view, a methodological error. The developmental logic of scientific inquiry differs profoundly from that of technology. The critical parameters include the novelty of ideas, the rigor of the methodology, the reproducibility of the results, the openness of the data, and the extent of international collaboration. Moreover, scientific projects may remain at early stages of maturity while already generating substantial added value for science and society—an aspect that the TRL, by design, fails to capture.

This creates a serious methodological challenge within Kazakhstan's funding mechanisms—both state research grants (GFs) and Program-Targeted Funding (PTF). When the TRL is applied, projects of high scientific significance but without immediate technological prototypes are systematically undervalued. As a result, fundamental research is disadvantaged relative to applied work, skewing the national research portfolio.

We therefore propose an alternative: adopting the SRL framework, tailored to the dynamics of scientific progress, in combination with the ASV index, which captures the

added scientific value of a project. This two-dimensional system makes it possible to account simultaneously for a project's maturity and its qualitative attributes, thereby enabling more accurate and equitable resource allocation.

As noted, Kazakhstan's two principal funding instruments—GFs (grants for basic and applied research) and PTF (program-targeted funding for national priorities)—face the same structural problem: how to rank proposals when some are still at the ideation stage, whereas others are close to practical implementation. Traditionally, this is addressed through expert councils. While the two-tiered evaluation process (via the National Center for State Scientific and Technical Expertise (NCSTE) and the National Science Council (NSC)) is already relatively robust and high-level, criticisms concerning the transparency and reproducibility of decisions remain.

Here, modernization is both necessary and feasible. Integrating SRL + ASV would improve trust in the process and increase the efficiency of funding allocation. To illustrate, we present a process flowchart for Kazakhstan's grant competitions (GFs and PTF) based on SRL + ASV. The diagram depicts the full pathway of a proposal: from submission and assignment of an SRL, through calculation of the ASV and comparison against thresholds, to the final management decision (e.g., full funding, conditional funding, or revision).

The block diagram in Figure 2 visualizes the full decision-making architecture for Kazakhstan's grant- and program-targeted funding competitions when the SRL and ASV are adopted as the central evaluative instruments. Unlike linear models (e.g., the classical TRL), this scheme operates as a cyclical, recursive system. Proposals and their supporting evidence enter the system, undergo successive layers of verification, normalization, and calibration against disciplinary and policy parameters, and then re-enter the cycle as knowledge for future rounds.

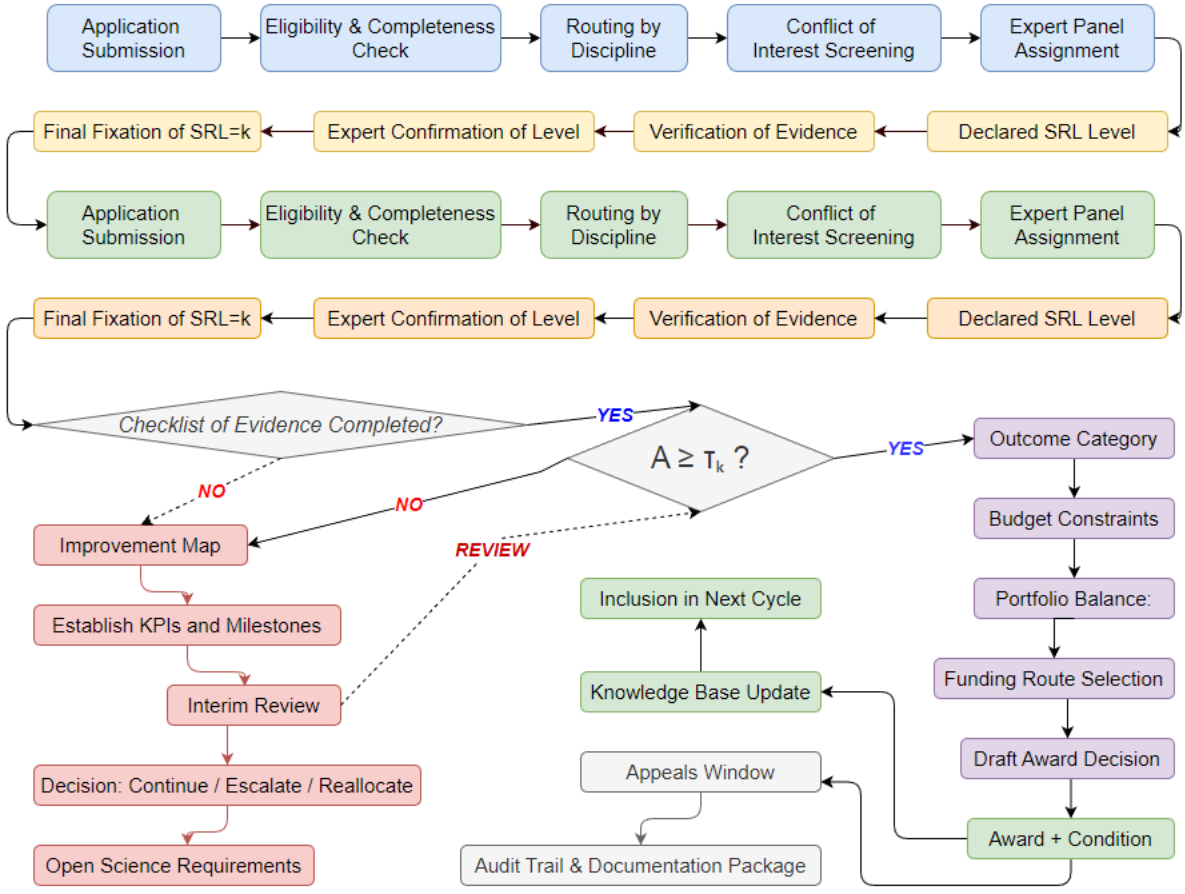


Figure 2. Kazakhstan SRL + ASV grant funding—extended.

The upper tier, "Intake and Governance," oversees proposal submission and formal eligibility screening. Proposals then proceed to expert panels, where the declared SRL is verified against objective evidence (protocols, publications, and pilot data). This ensures that an SRL cannot be assigned merely declaratively—a safeguard that is absent in simplified models.

The next block, "ASV scoring," is central. It captures the eight components of added scientific value—novelty, rigor, practical significance, reproducibility, collaboration, interdisciplinarity, openness, and leadership. Each component is source-verified, normalized with disciplinary sensitivity, and aggregated into an integral ASV score, A. Importantly, the model explicitly accounts for uncertainty and missing data, ensuring that incomplete projects are neither ignored nor artificially advantaged.

The "Calibration and Policy" block introduces disciplinary weights (e.g., biomedical sciences and mathematics may carry different expectations for publication patterns) and policy coefficients, reflecting strategic national priorities (e.g., green energy). Threshold values, τ, specify the minimum ASV required at each SRL for funding eligibility.

Projects then enter the "Decision Engine." Two tests are applied: first, confirmation of the evidence checklist for the claimed SRL (e.g., at least one publication for SRL ≥ 3); second, comparison of the aggregated ASV score against the calibrated threshold τ. Depending on the outcome, a project may receive full funding, conditional funding, a seed grant, or targeted recommendations for improvement.

The process does not end with funding allocation, as the architecture includes the "Improvement and Monitoring" and "Appeals and Audit" modules. The former loops the project back into the system via improvement maps, KPI setting, and interim reassessments of SRL + ASV. The latter ensures transparency, provides an appeal mechanism, and creates an audit trail—a full data package documenting the basis of each decision. Finally, the "outputs" block encompasses funding decisions, knowledge-based updates, and the return of experiential feedback into the system.

Analysis of the scheme highlights three major advantages. First, it integrates quantitative and qualitative methods of evaluation. In traditional grant systems, decisions often rely on expert voting, which may be subjective and poorly reproducible. Here, expert judgment remains important, but it is anchored in an algorithmic logic, where the SRL and ASV function as reference metrics. This reduces variance in evaluation, increases transparency, and strengthens trust among the scientific community.

Second, the architecture introduces multilayered quality control. Each block—SRL, ASV, calibration, decision, monitoring, and audit—acts as a filter to prevent "weak" projects from advancing to large-scale funding. Moreover, unlike rigid cutoffs, the system allows for alternative pathways: seed grants for promising but underdeveloped projects and conditional funding for those requiring additional publications or open data commitments. This flexibility supports a more comprehensive research ecosystem.

Third, the design is inherently cyclical and adaptive. The monitoring and audit modules feed the results back into the Calibration and Policy module. The system therefore "learns" over time: if certain project types are consistently under- or overvalued, parameter thresholds can be recalibrated. This dynamic characteristic aligns the model with the principles of adaptive governance in national science policy.

In summary, the extended SRL + ASV architecture transforms grant allocation from a static assessment into a responsive management system. Combining evidence-based readiness with reproducible measures of value ensures that funding decisions are not only transparent and equitable but also capable of evolving alongside national priorities and the broader scientific landscape.

Several challenges and opportunities accompany the implementation of the proposed architecture. First, deploying SRL + ASV would require significant institutional effort: training experts to work with the new framework, building a digital platform for calculating integrated indicators, and ensuring the openness and accessibility of data. Second, there is a risk of "*metric* pressure"—researchers may focus narrowly on maximizing ASV scores while neglecting the substantive quality of their work. This risk can be mitigated if the *improvement map and monitoring* blocks emphasize not only numerical values but also qualitative changes in the project.

On the other hand, introducing SRL + ASV could markedly increase trust in Kazakhstan's grant funding system. Many researchers currently criticize the procedures for being opaque and overly subjective. A transparent block diagram, where each step is logical and traceable, could significantly alter these perceptions. Furthermore, the introduction of ASV components would likely stimulate more responsible engagement with open data, repositories, and international collaborations—practices that are fully aligned with global trends in science policy.

In this respect, the "Kazakhstan SRL + ASV Grant Funding—Extended" framework is not only a visualization tool but also a fundamentally new approach to structuring research assessment. It combines the rigor of quantitative metrics with the flexibility of managerial decision-making and the cyclicality of feedback. Such an architecture can simultaneously support breakthrough research and ensure baseline quality in national science output.

Its central value lies in its potential to become a unifying framework for Kazakhstan's national science policy—either replacing or complementing existing TRL-based mechanisms. While the TRL captures technological readiness, SRL + ASV enables the assessment of scientific maturity and value specifically. When the state is investing heavily in science and expecting returns not only in the form of technologies but also in the form of robust, high-quality knowledge, such a system seems particularly timely.

In the context of Kazakhstan, the adoption of SRL + ASV would yield multiple benefits:

- For researchers, clear guidance on what steps are needed to strengthen future grant applications (e.g., improving openness, reproducibility, and normalized citation performance, rather than tailoring proposals merely to please expert reviewers).
- For funding actors (Ministry of Science and Higher Education, NCGTE, National Science Council, Science Fund), a tool for systematizing and transparently monitoring projects, which, in the long term, can support the design of national research roadmaps.
- For society and the state, a guarantee that public resources are distributed not only on the basis of subjective expert opinion but also through a reproducible model that balances maturity with scientific value.

In summary, the SRL + ASV model has the potential to establish a new standard for evaluating both grant- and program-targeted funding in Kazakhstan, combining international transparency with local relevance. In contrast, relying on TRLs alone for research assessment introduces methodological distortions and reduces the efficiency of the national funding system. The SRL captures the trajectory of scientific maturity from early ideas to durable outcomes, whereas the ASV measures the project's intrinsic quality in terms of novelty, rigor, reproducibility, and openness. Together, these two dimensions provide a transparent and fair instrument for evaluating research proposals. Crucially, SRL + ASV can be introduced without dismantling existing procedures, instead complementing them and enhancing trust within the research community.

This national application example is presented as a prospective illustration of how the SRL + ASV framework could be implemented in future national funding programs (e.g., GF/PTF, Global Flow/Process Transfer Function). Pilot testing on anonymized proposals,

comparisons with conventional peer review processes, and subsequent monitoring are planned as part of the next phase of empirical validation.

5. Discussion

The proposed "*readiness* × *value*" framework reconfigures the conventional logic of expert evaluation in research. The discrete SRL axis captures shifts between distinct classes of evidence—from first publications and independent replications to multicenter validation and sustainable impact—while the continuous ASV axis measures the internal coherence of quality practices *within* each level (novelty, rigor, reproducibility, review quality, field-normalized citations, external impact, openness, and collaboration). Together, these two axes eliminate the traditional disconnect between the "language of stages" and the "language of quality."

Progression across SRLs becomes not a matter of rhetorical persuasion aimed at reviewers but rather a formal event that is contingent on presenting a new class of evidence while also demonstrating a minimum vertical value threshold ($ASV \geq \tau$) at the current stage. This makes SRL advancement reproducible and explainable for research teams, who receive not only a decision map but also an improvement map—a clear indication of which components must be strengthened to cross the threshold.

The SRL scale proposed in this paper resonates with established operational frameworks (including the naming of levels such as *validation*, *application*, *scaling*, and *impact*) and supports interdisciplinary dialog with a shared foundation. It is sufficiently universal to apply across diverse fields of knowledge yet specific enough to guide formal decisions on progression and prioritization. Crucially, progression is treated as the introduction of a new class of evidence, not the accumulation of points. Reinforcing existing practices cannot substitute for meeting new requirements (e.g., external replication cannot be "compensated" for merely by increasing the sample size in the original cohort). This addresses the persistent inequity in competitions where early-stage fundamental projects are judged alongside mature applied ones: within the SRL framework, projects are comparable *within their respective levels*, not in a single undifferentiated pool.

The ASV index is constructed as a component vector $X = (N, R, P, V, C, I, O, L)$, normalized to $[0, 1]$, and subsequently aggregated. The key methodological decision is the use of a multiplicative aggregator, which "penalizes" weak links: no single standout component (such as a prestigious publication venue or high citation counts) can offset deficiencies in rigor, reproducibility, or openness. In this way, the model incentivizes the coherence of quality practices and reduces the risk of metric-centrism, where one indicator distorts the meaning of an integrated assessment.

We explicitly align this logic with the international agenda of responsible metrics and the risks articulated in Goodhart's Law: indicators should function as compasses, not as ends in themselves. Practically, the integration of the ASV with SRL checklists closes the most frequent vulnerabilities of expert assessment: the "loud venue" effect, the penalization of disciplines with different citation dynamics, and the substitution of genuine evidence with appearances.

Equally important, the ASV rests on normalizations that mitigate cross-disciplinary asymmetries. For example, C (citability) is assessed as field- and age-normalized citation performance (FNCI/percentile), V (venue quality) is the strength and transparency of peer review rather than a surrogate for the impact factor, O (openness) is the completeness and licensing of open data/code/protocols, and P (reproducibility) is the feasibility of actual independent replication. All the components are linked to verifiable sources and repositories.

This construction is therefore both compatible with the FAIR principles (findable, accessible, interoperable, reusable) and declarations on responsible research assessment (*DORA*, Leiden Manifesto), while also being adaptable to the managerial requirements of specific programs and disciplines through weights, thresholds, and policy coefficients. Such managed calibration enables shifts in emphasis (e.g., temporarily prioritizing openness and multicenter collaboration for laboratories joining international consortia) without compromising the comparability of the core framework.

At the level of science policy, the proposed model integrates naturally with existing "readiness scales." If the TRL has historically described the maturity of technologies within the engineering life cycle—supported by the extensive guidance from NASA and the U.S. DoD—the European Space Agency's SRL formalizes the scientific readiness of missions and instruments by specifying which questions and classes of evidence are required at each stage. Our proposed SRL axis is fully compatible with this logic, but it adds the missing "vertical quality dimension" (ASV), thereby addressing a well-recognized gap in the literature: the absence of a parameterized model that jointly integrates *stage* and *value* into decision-making. The result is a framework that simultaneously aligns with international standards (SRLs), incorporates the lessons of responsible research assessment (*DORA*, Leiden, FAIR), and remains operational for real-world selection procedures.

Notably, in national contexts where the TRL is already institutionalized in legislation and funding regulations, its uncritical extension into the domain of fundamental research produces methodological distortions. A scientific project may remain at an early stage of maturity while nonetheless delivering high added value to science and society—something that the TRL cannot capture. The Kazakhstan case illustrates how SRL + ASV disentangles these dimensions and enables more equitable rankings: applications are compared within their SRL, whereas funding decisions (seed grants, conditional funding, transition to programmatic instruments) are tied to the ASV profile and the associated improvement map rather than rhetorical force in the proposal. This marks a shift from a "one-off verdict" trajectory to a managed trajectory with explicit *KPI*s and interim reassessments, effectively transforming the system into a learning trajectory.

Methodologically, the distinctive feature of our approach lies in its dual filter:

(i) a checklist of evidence classes for the current SRL and (ii) a threshold for the aggregated ASV.

This two-layered logic protects against substitution effects (e.g., strong publication or citation indicators masking a weak methodology) while also preventing formal level-upgrading without the required new class of evidence. In combination with normalizations and discipline-specific weights, the model provides a controllable balance of "fairness" across fields with very different citation dynamics, verification practices, and traditions of openness. Importantly, the model does not eliminate the role of expert judgment. Rather, it structures the space in which such judgment operates, requiring experts to argue for weights and thresholds explicitly, rather than replacing debate with intuition.

In discussing the risks and boundaries of applicability, two points merit emphasis. The first is sensitivity to calibration: the choice of weights and thresholds must be grounded in disciplinary consensus and periodically revised in light of monitoring, with the system "tightening the screws" prematurely in contexts where infrastructures for openness and replication are still emerging. The second is resilience to metric gaming: the multiplicative aggregation combined with the mandatory checklist makes the model relatively resistant to inflating single indicators, but this only holds if strict source verification is enforced (repositories, licenses, open reviews) and if the FAIR principles are supported in the governance infrastructure (metadata, persistent identifiers, accessibility).

In other words, SRL + ASV reduces but does not eliminate strategic behavior—hence, the importance of embedding audit trails, appeal windows, and feedback loops into parameter calibration.

Taken together, our discussion converges on three key claims:

1. SRL + ASV fills a critical gap in both the literature and practice: the lack of a single, parameterized model that integrates maturity and value into operational decisions.
2. The model is interoperable with international readiness standards (ESA SRL) and resonates with the contemporary agenda of responsible research assessment (*DORA*, Leiden, FAIR), enhancing its applicability across institutions and national systems.
3. Its architecture is adaptive: through monitoring cycles, improvement maps, audits, and recalibration, the system becomes self-adjusting and -learning—an essential property for national research ecosystems where infrastructures of openness and reproducibility are still under construction.

In this sense, "readiness meets value" not in abstraction but in a procedure through which managerial decisions can be made, explained, and reproduced.

6. Conclusions

In this study, we derived and operationalized an integrated framework for research assessment in which stepwise scientific readiness levels (SRLs) are combined with the continuous measure of added scientific value (ASV). Within this framework, managerial decisions about advancing a project are not only the product of rhetorical persuasion but also the result of verifiable classes of evidence and the coherence of quality practices. Conceptually, the model closes the longstanding gap between the "language of stages" and the "language of quality." The discrete SRL axis marks transitions that require fundamentally new types of evidence (from initial publications and independent replications to multicenter validation and durable impacts), whereas the continuous ASV axis quantifies the "vertical dimension of quality" within each level—novelty, rigor, reproducibility, peer review strength, field- and age-normalized citation performance, openness, collaboration, and societal impact. Together, the SRL and ASV form not only an index but also a governable decision-making procedure with transparent inputs, auditable transformations, and reproducible outcomes.

The central methodological decision is the use of multiplicative aggregation of ASV components with discipline-specific weights and SRL-specific thresholds. This structure intentionally penalizes weak links and rewards the coherence of practices: neither prestigious venues nor rapid citation growth can compensate for deficits in openness, reproducibility, or methodological rigor. In this way, the model minimizes metric-centrism and resists "inflating" individual indicators. The threshold condition ($A \geq \tau_k$) complements the evidence checklist required for level k. Advancement is thus only permitted when both criteria are met. As a result, decisions become explainable (the reasons for acceptance or rejection are explicit), and trajectories become manageable (improvement maps highlight priority deficits and milestones).

Our work focuses on management architecture. We designed, implemented, and visualized block schemes of varying complexity—from compact to extended versions, including large-format and English-language adaptations. These schemes illustrate not only the logic of SRL + ASV but also the entire cycle of a real grant competition: application intake, verification, and routing; assessment based on the SRL and ASV; calibration of policy parameters; decision-making under portfolio constraints; monitoring cycles with interim reassessment; appeal windows and audit trails; and feedback of lessons into the recalibration of weights and thresholds. In this way, the model shifts from an "academic"

construct to a usable system, and assessment itself becomes a learning process in which each subsequent cycle improves upon the previous cycle.

From a science–policy perspective, the framework is flexible and portable: weights, thresholds, and policy coefficients can be tuned to local priorities without compromising its core logic. This enables an emphasis on openness and reproducibility where infrastructures are still emerging, stricter requirements for multicenter validation in fields nearing implementation, or the incentivization of international collaboration and standardization where these are strategic goals. Embedded appeals and audit mechanisms, along with the requirement of computational reproducibility, increase community trust and provide a verifiable trail of decisions—from input artifacts to final outcomes.

Of course, no model is universal. SRL + ASV remains sensitive to calibration: weights and thresholds must be set through disciplinary consensus and periodically revised through monitoring. Its resilience to strategic behavior depends not only on the multiplicative index but also on the triad of aggregation, a mandatory evidence checklists and strict verification of sources (repositories, licenses, open reviews). Precisely because of this triad, the model is already usable as a standard of reproducible peer review that makes funding policies both predictable and adaptive, whereas for researchers, it is transparent and navigable—clarifying what to improve, why, and in what sequence.

Ultimately, what we have proposed is not "just another metric" but rather a decision-making construct in which readiness genuinely meets value. Theoretically, it bridges stage and quality; methodologically, it provides explicit rules of aggregation and thresholds; computationally, it offers reproducible procedures and visualizations; institutionally, it delivers a governance architecture that explains why and how each decision was made while leaving a roadmap for the next step.

We believe that this framework can serve as the basis for a new compact between researchers and institutions—one in which the *quality of knowledge* and the *maturity of research products* are assessed together, and progress is secured not by eloquence but by verifiable, auditable evidence.

To demonstrate operational feasibility, a pilot validation protocol has been designed. The SRL + ASV model will be applied to 20 research projects from three domains (STEM, social sciences, and the humanities). ASV indicators are extracted via OpenAlex, OSF (The Open Science Framework), and institutional reports, whereas SRLs are assigned through expert consensus. Correlation and rank order validation assess internal coherence (Kendall's τ, Spearman's ρ). Although the pilot is not yet executed, the procedure ensures reproducibility for future implementation.

Funding: This research was funded by the Science Committee of the Ministry of Science and Higher Education of the Republic of Kazakhstan (grant No. AP26199077).

Data Availability Statement: Data derived from public domain resources.

Conflicts of Interest: The author declares no conflicts of interest.

References

Aagaard, K. (2015). How incentives trickle down: Local use of a national bibliometric indicator system. *Science and Public Policy*, *42*(5), 725–737. [CrossRef]

Bornmann, L. (2013). What is societal impact of research and how can it be assessed? A literature survey. *Journal of the American Society for Information Science and Technology*, *64*, 217–233. [CrossRef]

Bozeman, B., & Melkers, J. (1993). *Evaluating R&D impacts: Methods and practice*. Springer. [CrossRef]

Campbell, D. T. (1979). Assessing the impact of planned social change. *Evaluation and Program Planning*, *2*(1), 67–90. [CrossRef]

Diakoulaki, D., Mavrotas, G., & Papayannakis, L. (1995). Determining objective weights in multiple criteria problems: The CRITIC method. *Computers and Operations Research*, *22*(7), 763–770. [CrossRef]

DORA. (2012). *San Francisco Declaration on Research Assessment (DORA)*. Available online: https://sfdora.org (accessed on 1 September 2025).

Engel, C. C., Silberglitt, R., Chow, B. G., Jones, M. M., & Grant, J. (2019). *Development of a knowledge readiness level framework for medical research*. RAND Corporation. [CrossRef]

ESA. (2023). *Scientific Readiness Levels (SRL) handbook*. ESA-EOPSM-SRL-MA-4267. Available online: https://missionadvice.esa.int/wp-content/uploads/2024/01/Scientific-Readiness-Levels-Handbook-document-v2.0-Web.pdf (accessed on 1 September 2025).

European Commission. (2014). *Horizon 2020—Work programme 2014–2015: General annexes, part 19—Annex G: Technology Readiness Levels (TRL)*. European Commission. Available online: https://ec.europa.eu/research/participants/data/ref/h2020/wp/2014_2015/annexes/h2020-wp1415-annex-g-trl_en.pdf (accessed on 1 September 2025).

Geuna, A., & Martin, B. R. (2003). University research evaluation and funding: An international comparison. *Minerva*, *41*, 277–304. [CrossRef]

Glänzel, W., & Moed, H. F. (2013). Opinion paper: Thoughts and facts on bibliometric indicators. *Scientometrics*, *96*(1), 381–394. [CrossRef]

Goodhart, C. A. E. (2013). Goodhart's law. *Libellio*, *9*(4), 29–33.

Guan, H., Gentimis, T., Krim, H., & Keiser, J. (2017). First study on data readiness level. *arXiv*, arXiv:1702.02107v1. Available online: https://arxiv.org/pdf/1702.02107.pdf (accessed on 1 September 2025).

Hevner, A. R., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS Quarterly*, *28*(1), 75–105. [CrossRef]

Hicks, D. (2012). Performance-based university research funding systems. *Research Policy*, *41*(2), 251–261. [CrossRef]

Hicks, D., Wouters, P., Waltman, L., de Rijcke, S., & Rafols, I. (2015). The Leiden Manifesto for research metrics. *Nature*, *520*(7548), 429–431. [CrossRef] [PubMed]

Hwang, C.-L., & Yoon, K. (1981). *Multiple attributes decision making methods and applications*. Springer.

Ioannidis, J. P. A. (2005). Why most published research findings are false. *PLoS Medicine*, *2*(8), e124. [CrossRef] [PubMed]

Jaakkola, E. (2020). Designing conceptual articles: Four approaches. *AMS Review*, *10*(1–2), 18–26. [CrossRef]

Jacobsen, A., Azevedo, R. d. M., Juty, N., Batista, D., Coles, S., Cornet, R., Courtot, M., Crosas, M., Dumontier, M., Evelo, C. T., Goble, C., Guizzardi, G., Hansen, K. K., Hasnain, A., Hettne, K., Heringa, J., Hooft, R. W. W., Imming, M., Jeffery, K. G., . . . Schultes, E. (2020). FAIR principles: Interpretations and implementation considerations. *Data Intelligence*, *2*, 10–29. [CrossRef]

Jajo, N. K., & Peiris, S. (2021). Statistical analysis of ERA and the quality of research in Australian universities. *Journal of Applied Research in Higher Education*, *13*(2), 420–429. [CrossRef]

Kennedy-Martin, M., Slaap, B., Herdman, M., van Reenen, M., Kennedy-Martin, T., Greiner, W., & Boye, K. S. (2020). Which multiattribute utility instruments are recommended for use in cost–utility analysis? A review of national health technology assessment (HTA) guidelines. *European Journal of Health Economics*, *21*(8), 1245–1257. [CrossRef] [PubMed]

Knar, E. (2024a). Level of scientific readiness with ternary data types. *arXiv*, arXiv:2410.09073.

Knar, E. (2024b). Recursive index for assessing value added of individual scientific publications. *arXiv*, arXiv:2404.04276. [CrossRef]

Krejčí, J., & Stoklasa, J. (2018). Aggregation in the analytic hierarchy process: Why weighted geometric mean should be used instead of weighted arithmetic mean. *Expert Systems with Applications*, *114*, 97–106. [CrossRef]

Lemos, J. C., & Chagas Junior, M. d. F. (2016). Application of maturity assessment tools in the innovation process: Converting system's emergent properties into technological knowledge. *RAI: Revista de Administração e Inovação*, *13*(2), 145–153. [CrossRef]

Manheim, D. (2018). *Building less flawed metrics: Dodging Goodhart and Campbell's laws*. Munich Personal RePEc Archive.

Marx, W., & Bornmann, L. (2016). Change of perspective: Bibliometrics from the point of view of cited references. *Scientometrics*, *109*, 1397–1415. [CrossRef]

Meredith, J. (1993). Theory building through conceptual methods. *International Journal of Operations & Production Management*, *13*(5), 3–11. [CrossRef]

Merton, R. K., & Shapere, D. (1974). The sociology of science: Theoretical and empirical investigation. *Physics Today*, *27*(8), 52–53. [CrossRef]

Munafò, M. R., Nosek, B. A., Bishop, D. V. M., Button, K. S., Chambers, C. D., Percie Du Sert, N., Simonsohn, U., Wagenmakers, E.-J., Ware, J. J., & Ioannidis, J. P. A. (2017). A manifesto for reproducible science. *Nature Human Behavior*, *1*(1), 0021. [CrossRef]

NASA. (2019). *Technology readiness level definitions*. NASA.gov.

Nosek, B. A., Ebersole, C. R., DeHaven, A. C., & Mellor, D. T. (2018). The preregistration revolution. *Proceedings of the National Academy of Sciences of the United States of America*, *115*(11), 2600–2606. [CrossRef]

Open Science Collaboration. (2015). Estimating the reproducibility of psychological science. *Science*, *349*(6251), aac4716. [CrossRef] [PubMed]

Opricovic, S., & Tzeng, G. H. (2004). Compromise solution by MCDM methods: A comparative analysis of VIKOR and TOPSIS. *European Journal of Operational Research*, *156*(2), 445–455. [CrossRef]

Popper, K. (2005). *The logic of scientific discovery*. Taylor and Francis. [CrossRef]

Saaty, R. W. (1987). The analytic hierarchy process—What it is and how it is used. *Mathematical Modeling*, *9*(3–5), 161–176. [CrossRef]

Saaty, T. L. (2002). Decision making with the analytic hierarchy process. *Scientia Iranica*, *9*(3), 215–229. [CrossRef]

Sauser, B., Verma, D., Ramirez-Marquez, J., & Gove, R. (2006, April 7–8). *From TRL to SRL: The concept of systems readiness levels*. Conference on Systems Engineering Research (pp. 1–10), Los Angeles, CA, USA.

Sivertsen, G. (2019). Understanding and evaluating research and scholarly publishing in the social sciences and humanities (SSH). *Data and Information Management*, *3*(2), 61–71. [CrossRef]

Uzzi, B., Mukherjee, S., Stringer, M., & Jones, B. (2013). Atypical combinations and scientific impact. *Science*, *342*(6157), 468–472. [CrossRef]

Velasquez, M., & Hester, P. (2013). An analysis of multicriteria decision-making methods. *International Journal of Operations Research*, *10*(2), 56–66.

Wilkinson, M. D., Dumontier, M., Aalbersberg, I. J., Appleton, G., Axton, M., Baak, A., Blomberg, N., Boiten, J.-W., da Silva Santos, L. B., Bourne, P. E., Bouwman, J., Brookes, A. J., Clark, T., Crosas, M., Dillo, I., Dumon, O., Edmunds, S., Evelo, C. T., Finkers, R., . . . Mons, B. (2016). The FAIR guiding principles for scientific data management and stewardship. *Scientific Data*, *3*, 160018. [CrossRef] [PubMed]

Wilsdon, J. (2015). *The metric tide: Report of the independent review of the role of metrics in research assessment and management*. SAGE Publications Ltd. [CrossRef]

Wuchty, S., Jones, B. F., & Uzzi, B. (2007). The increasing dominance of teams in the production of knowledge. *Science*, *316*(5827), 1036–1039. [CrossRef] [PubMed]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.