International Journal of Innovation Science

An integrated methodology for startup assessment in Quality 4.0 using Thurstone's *Law of Comparative Judgment*

Giovanna Carrera, Domenico A. Maisano, Luca Mastrogiacomo and Fiorenzo Franceschini

DIGEP (Department of Management and Production Engineering), Politecnico di Torino, Turin, Italy

Received 26 February 2025
Revised 16 May 2025
15 July 2025
Accepted 2 August 2025

Abstract

Purpose – This paper addresses the challenge of early-stage evaluation of startups within the Quality 4.0 framework, which is characterized by various principles (e.g. continuous improvement, integration of modern technology, data-driven decision-making, etc.). Startups often lack historical data and operate in uncertain markets, making traditional evaluation methods inadequate. To overcome these limitations, this study introduces an integrated assessment methodology based on Quality 4.0 principles and Thurstone's Law of Comparative Judgment (LCJ). The purpose of this paper is to provide a structured, adaptable framework for early-stage startup evaluation, ensuring a more reliable and systematic assessment of critical success factors.

Design/methodology/approach – This study develops an integrated evaluation methodology for startups by combining expert judgment with data-driven Quality 4.0 techniques. It is inspired by Thurstone's LCJ and focuses on key evaluation criteria, including team characteristics, patents and market potential, strategic alliances, financial health and funding, and innovation and technology. A real-world case study is conducted to validate the approach, demonstrating its effectiveness in providing structured insights for stakeholders involved in the innovation ecosystem. Only aggregate and anonymized data were analyzed, ensuring confidentiality for both experts and startups.

Findings – The proposed methodology enhances the accuracy and reliability of startup evaluations by combining expert judgment with systematic assessment processes, aligned with Quality 4.0 principles. The real-world case study highlights the method's practical benefits, particularly in identifying high-potential startups despite limited historical data.

Research limitations/implications – This study primarily focuses on early-stage startups, which may limit its applicability to more mature companies with established track records. In addition, while the methodology integrates expert judgment with systematic assessment tools, the subjectivity of expert evaluations may introduce biases. Future research could refine the approach by incorporating advanced artificial intelligence-driven analytics to enhance objectivity. Furthermore, broader validation across different industries and international contexts would help assess the model's generalizability and effectiveness in diverse innovation ecosystems.

Originality/value – This paper contributes to Quality 4.0 applications in startup evaluation, by integrating systematic methodologies with expert-driven comparative judgment. It introduces a novel framework inspired by Thurstone's LCJ, offering a structured yet flexible approach to assessing early-stage startups. Unlike

International Journal of Innovation Science
Emerald Publishing Limited
1757-2223
DOI 10.1108/IJIS-02-2025-0089

traditional models that struggle with high uncertainty and limited historical data, this methodology exploits a blend of human expertise and technology-driven insights.

Keywords Startup assessment, Open innovation, Thurstone's law of comparative judgment, Quality 4.0, Continuous improvement

Paper type Research paper

1. Introduction

Traditional evaluation methods for large corporations, such as *financial statement analysis*, *discounted cash flow* (DCF), *market comparables* (Comps), *asset-based evaluation*, and *risk-adjusted return methods*, face significant limitations when applied to startups. These methods often fail to account for the unique potential of startups, given their lack of reliable historical data and their operation in volatile and uncertain markets (Köseoğlu, 2023).

Recent data make the extent of the problem clear: about 90% of startups eventually fail, 10% within the first year and almost half by the fifth year. Even among startups financed by venture-capital funds, about 75% never return investors' capital (Gompers and Kaplan, 2019), and mutual-fund mark-to-market data indicate that early-stage shares are on average overvalued by about 40% (Brown and Gredil, 2022). These figures illustrate both the economic stakes and the shortcomings of conventional valuation metrics, reinforcing the need for tailored assessment methods. Moreover, startups frequently innovate with new technologies and business models, which can escape the confines of traditional metrics (Armstrong, 2006).

In this context, there is a clear need for an assessment method that adapts to the rapid changes and evaluates startups based on both quantitative and qualitative potential. The principles of Quality 4.0 – emphasizing continuous improvement, integration of modern technology, data-driven decision-making, together with the synergy between human and system intelligence – offer a robust framework for this challenge. Quality 4.0 employs advanced data collection and analysis techniques to capture the dynamic and innovative essence of startups during assessment. Specifically, this study addresses the following research questions:

RQ1. Which evaluation criteria, aligned with Quality 4.0 principles, are most suitable for assessing early-stage startups?

RQ2. How can expert judgments related to these evaluation criteria be effectively aggregated to construct a reliable ranking of startup?

To address these research questions and evaluate startups, the methodology presented here utilizes Thurstone's *Law of Comparative Judgment* (LCJ). LCJ is not just a comparative method; it is a dynamic, adaptive process that aligns seamlessly with the continuous-improvement *ethos* of Quality 4.0. LCJ's statistical rigor allows to effectively discern relative perceptions, making it particularly suitable for environments where traditional metrics may be less effective. By enhancing traditional evaluation models with an early-stage flexible framework, this approach ensures the relevance of the startup evaluation process across various contexts and sectors. Figure 1 presents a conceptual scheme that summarizes the methodology's key phases, which are examined in detail later in the article:

- identification of the startups;
- definition of the evaluation criteria;

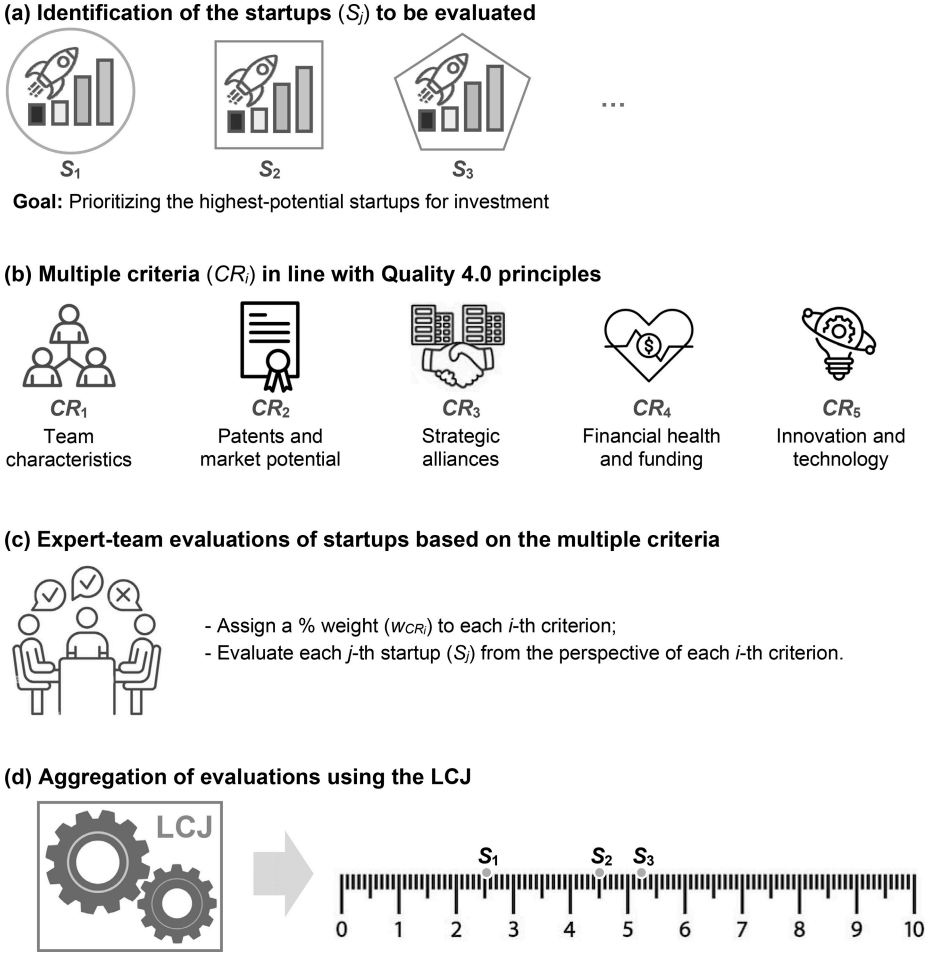

Figure 1. Conceptual scheme of the methodology developed in the paper
Source: Figure created by authors

- evaluation of startups by an expert team; and
- aggregation of the evaluations, so as to find the most promising and incentive-worthy startups.

The remainder of this paper is organized as follows: Section 2 provides an overview of state-of-the-art techniques for assessing startups, as alternatives to the proposed method. Section 3 proposes a general formulation of the startup assessment. Section 4 revisits Thurstone's LCJ, outlining its basic assumptions and practical applications. Section 5, through a case study, shows how LCJ can be successfully applied to assess startups. The concluding section 6 summarizes the key findings, practical implications, limitations, and prospects for future development.

2. Literature review

Assessing early-stage startups presents unique challenges due to limited data and high uncertainty, making traditional evaluations inherently subjective (Yildirim and Maz, 2025). Nevertheless, effective benchmarking of startups is crucial for stakeholders like investors, incubators, and accelerators, as it can yield significant economic returns and stimulate the development of new ideas (El Hanchi and Kerzazi, 2020).

Integrating Quality 4.0 principles into this context enhances the evaluation process with modern technology and data-driven insights, thereby improving decision-making, managing risk and strategically allocating resources. In the dynamic and competitive startup environment, where resources are limited and challenges are complex, accurate assessment can mean the difference between success and failure.

At the risk of oversimplifying, the scientific literature on early-stage startup evaluation can be schematized from three different angles, as illustrated below:

(1) *Traditional vs. modern methods.* Literature extensively addresses methodologies for evaluating and prioritizing startups based on various criteria (Valiris *et al*., 2005). This area of research has evolved through contributions from numerous authors, each offering unique perspectives and innovations. Prominent scholars such as Blank and Dorf (2020), Kotter *et al*. (2015) have explored various aspects of startup evaluation, from validating business ideas to developing competitive strategies. However, many of these earlier approaches now require revisiting in light of today's digital analytics capabilities. There is a growing need to update traditional evaluation models by harnessing big data, machine learning and real-time metrics (i.e. some of the pillars of Quality 4.0), to better predict long-term startup success. Indeed, recent studies have begun to integrate such tools – for example, Wei (2025) proposes a hybrid quantitative framework that uses entropy-based weighting and simulation to evaluate startups on strategic criteria like scalability, adaptability and risk exposure – but overall the literature is still evolving toward fully data-driven startup assessments.

(2) *Context-specific frameworks.* A recurring theme in the literature is the adaptability of assessment methodologies to the diverse business needs and contexts in which evaluations are conducted. The challenge lies in evaluating a variety of startups, each with unique characteristics. Evaluation criteria vary according to business needs and market contexts, making prioritization a dynamic and personalized process (Galende, 2006). For companies conducting these assessments – be they investment firms, corporate innovation departments, or technology incubators – the goal is typically to identify startups that best align with their strategic objectives and values. This has led to the development of structured frameworks targeting context-specific performance indicators. For example, Arshi *et al*. (2021) introduce the SECURE model as a dedicated framework for measuring startup performance in emerging economies, underscoring the need to capture unique dimensions of startup success (such as scalability and resilience) that traditional models might overlook. Overall, literature emphasizes the importance of a multidimensional approach, considering not only financial and market aspects but also innovation, sustainability and social impact (Smith and Cordina, 2014; Manigart *et al*., 1997; Bottani and Rizzi, 2008).

(3) *MCDM and quantitative tools.* To systematically compare startups on such multiple criteria, researchers have increasingly turned to multicriteria decision-making

(MCDM) techniques and other structured analytical tools. Key analytical approaches discussed in the literature include:

- *Balanced scorecard* combined with a multi-attribute rating technique, which has been adapted to startup decision contexts as a way to link strategic objectives with evaluation criteria (Valiris *et al*., 2005).
- *Analytic hierarchy process* (AHP), which has been applied to startup investment decisions to derive weighted rankings of alternatives. For instance, Kyrylych and Povstenko (2023) demonstrate an AHP-based model that organizes evaluation criteria into thematic clusters and calculates a global priority for each startup, allowing investors to see which venture scores highest overall on a combination of quantitative and qualitative factors. Such an approach not only introduces more rigor into the comparison process but also makes the rationale behind decisions more transparent.
- *Fuzzy logic and linguistic models* have gained traction to handle the uncertainty and vagueness in expert judgments. Lin *et al*. (2021) propose a hesitant fuzzy linguistic MCDM model for startup evaluation, enabling assessors to use linguistic terms (like "high" or "low" performance) which are then converted into numerical scores for comparison. This fuzzy approach accommodates the imprecise nature of evaluating novel ventures, where crisp data may be unavailable. In fact, by using tools like fuzzy sets or information envelopment analysis, these models can gauge a startup's efficiency or potential even with scant historical data.
- *Hybrid data-drive frameworks* combine multiple techniques to exploit their complementary strengths. As noted above, Wei (2025) integrates entropy-based weighting (to objectively determine criterion importance from data variability) with a simulation component (to test how robust startup evaluations are under different scenarios). By blending qualitative judgment with quantitative analysis, these hybrid frameworks aim to improve the accuracy of predictions and provide deeper insights into a startup's potential under uncertainty.

Together, the incorporation of these analytical techniques is making startup assessments more evidence-based. An added benefit of quantitative scoring models (like AHP or scorecards) is the ability to compare and even portfolio-manage startup opportunities. Rather than a binary *go*/*no-go* decision on one venture, investors can evaluate a pool of startups and allocate resources proportionally to their scores – thus spreading risk according to each startup's evaluated merit.

In summary, the state-of-the-art in startup assessment is moving toward integrated, multi-criteria frameworks that combine the strengths of expert judgment and advanced analytics. The literature underscores that effective startup evaluation must be flexible enough to account for different contexts and criteria, yet structured enough to allow systematic comparison. Traditional evaluation criteria (market size, financial projections, etc.) are now complemented by deeper analysis of qualitative factors (such as team competence, innovation capacity and adaptability to change) in what is increasingly a data-rich decision environment. However, despite notable progress, few existing methods fully capture the synergy between human expertise and cutting-edge analytical tools. This gap sets the stage for an approach that fuses expert comparative judgment (as exemplified by Thurstone's LCJ) with the data-driven *ethos* of Quality 4.0 – an approach that this study seeks to develop and validate.

3. The problem of startup assessment: criteria definition and expert evaluation

This section aims to define the multicriteria problem of startup evaluation, which is addressed and solved in the following sections. It is divided into two subsections corresponding, respectively, to (i) the definition of the evaluation criteria and (ii) the construction of the assessment matrix by a team of experts.

3.1 Definition of evaluation criteria

Building on insights from recent literature and professional practice – and consistent with the twin paradigms of Quality 4.0 and Industry 4.0 emphasized by Khourshed (2023), Khourshed and Gohar (2023) and Khourshed *et al.* (2023) – five key criteria for evaluation have been identified (cf. *RQ1*):

(1) *Team characteristics* (CR_1). A strong, cohesive team builds investor confidence and positively influences the startup's growth (Blank and Carmeli, 2021; Kotter *et al.*, 2015). The quality of human capital and the ability to establish strong external relationships are crucial factors (Hatch and Dyer, 2004; Fernández-Olmos and Ramírez-Alesón, 2017).

(2) *Patents and market potential* (CR_2). Patents reduce information asymmetry between the startup and external evaluators and the number of patents is positively correlated with the startup's economic value (Jeon, 2019; Conti *et al.*, 2013). Market potential involves the startup's target market size, growth prospects and demand for its product or service.

(3) *Strategic alliances* (CR_3). Strategic alliances provide access to resources and signal the quality of the startup. The ability to form such alliances is often valued by investors, especially those focused on financial returns (Mamédio *et al.*, 2019).

(4) *Financial health and funding* (CR_4). Financial stability, including cash position and investments in research and development, positively influences startup valuation, while high debt levels may have a negative effect (Sievers *et al.*, 2013).

(5) *Innovation and technology* (CR_5). This criterion evaluates the startup's innovation level and technological advancements, including unique selling propositions, intellectual property and market disruption potential (Noh *et al.*, 2018).

The proposed methodology – which will be described in detail in Section 5 – aims to provide a simple, user-friendly tool for startup evaluators while employing an agile model based on minimal assumptions and straightforward comparisons. This approach relies on statistical assumptions about expert preferences (Franceschini *et al.*, 2022) and is easy to implement, even by non-specialists. The Thurstone's LCJ – recalled in Section 4 – constitutes a well-established basic model for achieving the intended goal, offering an indirect measurement framework: expert evaluations are statistically transformed into global-scale values, a procedure generally referred to as *scaling*. Alternative scaling paradigms include the *Rasch model* (Rasch, 1966; Bond and Fox, 2007) and *conjoint analysis* (Luce and Tukey, 1964). However, LCJ is preferred here because it is conceptually straightforward, easier to communicate outside the psychometric field where it originated and is based on comparatively fewer restrictive assumptions (Pollitt, 2012).

3.2 Construction of the assessment matrix by the team of experts

This subsection presents a general formulation of the startup-assessment problem, which can be concisely captured by the *assessment matrix* in Figure 2, whose columns list the individual startups (S_j) and whose rows represent the evaluation criteria (CR_i). Each entry

	S_1	...	S_j	...	S_m	
CR_1	r_{11}	...	r_{1j}	...	r_{1m}	w_{CR_1}
⋮	⋮		⋮		⋮	⋮
CR_i	r_{i1}	...	r_{ij}	...	r_{im}	w_{CR_i}
⋮	⋮		⋮		⋮	⋮
CR_n	r_{n1}	...	r_{nj}	...	r_{nm}	w_{CR_n}

Figure 2. Assessment matrix for startup evaluation. This matrix represents the framework used for assessing startups (S_j) against various criteria (CR_i). Each cell (r_{ij}) of the matrix quantifies the performance of a startup from the perspective of a given criterion, while the weights (w_{CR_i}, expressed as percentages) denote that criterion's relative importance

Source: Figure created by authors

in the assessment matrix (r_{ij}) reflects the performance level of the *j*th startup (S_j) with respect to the *i*th criterion (CR_i). The evaluations are carried out collectively by a team of startup experts, with adequate documentary and/or direct knowledge of the startups under consideration. To guarantee assessments that are sufficiently comprehensive yet not overly cumbersome, the team should combine complementary competences while remaining relatively lean – typically no more than six to eight members (Franceschini and Rossetto, 1995).

The importance of each criterion varies depending on the evaluative context or strategic objectives; the decision regarding the weights can be made collectively by the expert team itself or established *a priori* by decision-makers (Franceschini *et al.*, 2022). The performance of startups for each individual criterion is assessed on a four-level ordinal scale: "high," "medium," "low" or "null." Nevertheless, more "powerful" measurement scales (interval or *ratio*) may be adopted, provided that the resulting values are later converted to an ordinal scale (Stevens, 1946; Franceschini and Maisano, 2019; Franceschini *et al.*, 2022).

The next section shows how LCJ can aggregate these multiple-criteria evaluations into a comprehensive positioning of the startups under study.

4. Basics of Thurstone's LCJ

Thurstone's LCJ posits that it is easier for individual respondents to compare two alternatives side by side, rather than evaluating each one in isolation. This principle forms the foundation for various research methodologies that rely on direct comparisons between pairs of alternatives (Thurstone, 1927; Franceschini and Maisano, 2015; Franceschini *et al.*, 2022). In this application, the classical LCJ is adapted to a real-world context in which a panel of experts (i.e. respondents) performs multiple-criteria evaluations of startups (i.e. alternatives) and these evaluations must be merged into an overall assessment (cf. Section 5).

The general problem addressed by Thurstone's LCJ can be summarized as follows: a group of experts provide their assessments on one or more specific criteria of certain alternatives. Each criterion is defined as a specific attribute (i.e. a property or feature of interest) of the alternatives that elicits a response from the group of respondents. The individual assessments are combined according to the procedure summarized in the following steps:

- *Collection of individual assessments*: Respondents provide their evaluations based on specific criteria for the alternatives in question.

- *Aggregation of assessments*: The assessments related to the individual criteria are then aggregated to obtain a global measure for each alternative. This step often involves creating a proportion matrix that reflects the frequency with which one alternative is preferred over another (Thurstone, 1927; Edwards, 1957).
- *Transformation to z-scores*: The aggregated data is then transformed into *z*-scores, which are standard scores that indicate how many standard deviations an alternative is from the mean. This step normalizes the data, allowing for meaningful comparisons (Franceschini and Maisano, 2015; Franceschini *et al*., 2022).
- *Scaling*: Finally, the *z*-scores are used to determine the average values of the attributes of the alternatives, placing them on a "preference" scale that quantifies the positioning of the alternatives (Edwards, 1957; Kelly *et al*., 2022).

This methodology captures the complexity of respondent evaluations while facilitating the interpretation of the data through advanced statistical models. A more detailed presentation of the Thurstone' LCJ model can be found in Edwards (1957) and in Franceschini *et al.* (2022).

The ability of the LCJ model to convert multiple-criteria evaluations into aggregated, quantifiable data makes it a powerful tool for analyzing and comparing various alternatives. It also offers a structured approach that enables detailed and statistically rigorous analyses, which can be especially valuable in fields like startup assessment and scouting.

5. Practical application of the method in startup evaluation: a case study

This section outlines the application of Thurstone's LCJ to the comparison of startups, focusing on its implementation in a real-world case study within a corporate environment. The assessment concerns eight startups of interest (S_1, […], S_8), which are the alternatives of the problem. These startups were drawn from the Piedmont's aerospace–digital-manufacturing cluster, one of the Europe's major hubs for Quality 4.0 technologies. Focusing on this innovation-intensive region provides a coherent economic context, while offering sectoral variety (e.g., additive manufacturing, artificial intelligence [AI]-enabled inspection, satellite Internet of Things, etc.). The sample therefore balances regional representativeness (all startups belong to the same policy ecosystem targeted by the funding partner) and technological diversity (capturing different early-stage challenges). The expert panel – which is better described in Section 5.1 – was composed to reflect this scope, combining venture-capital analysts, corporate R&D managers and a university technology-transfer officer, each with direct exposure to the cluster. This purposive sampling improves both contextual relevance and external validity.

The specific attribute considered is each startup's performance level – interpreted as its development and success potential – for every criterion. The benchmarking analysis adopts the five criteria described in Section 2 (CR_1, […], CR_5). The weights collectively decided by the expert team correspond to evidence from large samples in entrepreneurship research. *Innovation and technology* (CR_5) received the highest relative importance (33%), echoing longitudinal studies that link patent-based technological advantage to superior growth and valuation (Conti *et al*., 2013; Noh *et al*., 2018). *Team characteristics* (CR_1) ranked second (22%), in line with venture-capital findings that management quality is the strongest predictor of funding success and survival (Manigart *et al*., 1997; Blank and Carmeli, 2021). The moderate weights for *strategic alliances* (CR_3) and *financial health* (CR_4) likewise mirror meta-analyses showing these factors to be supportive, though not dominant, predictors of early performance (Mamédio *et al*., 2019; Sievers *et al*., 2013). Reproducing

these empirical patterns within a small-sample expert setting underscores the validity of the proposed criteria weights.

5.1 Evaluation process

The startup evaluations are conducted by a group of experts whose experience in startup assessment is both deep and diverse. In this case, a panel of six experts with heterogeneous seniority and complementary backgrounds operated: two venture-capital analysts, one serial entrepreneur, one university technology-transfer officer and two industrial R&D managers. Prior to the large-scale data collection, a pilot test was conducted, to increase the reliability and clarity of the assessment criteria prior to their formal application [1].

The evaluation process begins with the construction of the assessment matrix shown in Figure 2. Each *j*th startup (columns of the table in Figure 2) is evaluated against each *i*th criterion (rows of the table in Figure 2) using an ordinal scale represented by the following symbols: "△" for "low," "○" for "medium," "●" for "high," and "∅" for "null" degree of performance; Table 1 illustrates the four levels of this scale. Figure 3 exemplifies the assessment matrix that includes the evaluations assigned by the expert group. As an example, with reference to the *team characteristics* criterion (CR_1, cf. Section 2), the expert group decided to assign the highest value on the scale, "●," to startup S_7. The intermediate value, "○," was given to startups S_2 and S_5, while a null evaluation, "∅," was assigned to the remaining startups.

Content of the assessment matrix can be visualized graphically in the form of performance profiles, as exemplified in Figure 4.

5.2 Methodological approach

The assessment matrix serves as the starting point for initiating the comparison process of startups. The evaluation criteria (CR_i) and their respective weights, along with the specific assessments of each startup by the expert group, form the foundation on which Thurstone's LCJ methodology is built. The technique involves the following structured steps (*a*, *b*, *c* and *d*):

- Transformation of the assessment matrix into rankings

  Focusing on the *i*th row (CR_i) of the assessment matrix, the startups are ordered based on the characterizations assigned by the expert group (Figure 5(a) and (b)). This ranking includes *regular* startups (S_1–S_8) and four *dummy/anchor* startups – $S_{\bullet}$, $S_{\circ}$, $S_{\triangle}$ and $S_{\varnothing}$ – to preserve the levels of information content. The introduction of these dummy alternatives in the rankings is crucial to ensure that no part of the available information is lost (Franceschini and Maisano, 2019). In this regard, it is important to note that the same ranking can be achieved with different quantitative/qualitative evaluations of the alternatives of interest.

Table 1. Ordinal scale of the degree of performance of se *j*th startup (S_j) with respect to a *i*-th criterion (CR_i)

Degree of performance of S_j with respect to CR_i	
Intensity	Symbol
Null	∅
Low	△
Medium	○
High	●

Source(s): Table created by authors

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	w_{CR_i}
CR_1		○			○		●		0.22
CR_2		●	○		○	●	△	△	0.20
CR_3	●			○	○	○	●		0.14
CR_4			△		△			●	0.11
CR_5		○	●		○	●	○		0.33

Figure 3. Assessment matrix containing the expert panel's evaluations. CR_1–CR_5 represent the five evaluation criteria considered in the case study, while S_1–S_8 denote the eight startups assessed
Source: Figure created by authors

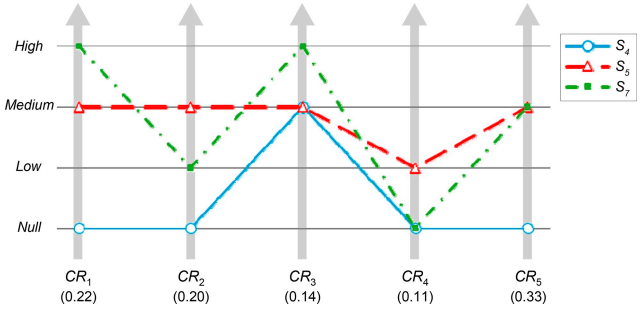

Figure 4. Performance profiles of the startups across the individual criteria (shown on the horizontal axis, with their respective weights in parentheses). To maintain clarity, only three profiles are displayed – S_4, S_5, and S_7. Profiles S_5 and S_7 consistently outperform or at least match S_4 across all five criteria. The head-to-head comparison between S_5 and S_7 is less clear-cut: S_7 leads on CR_1 and CR_3, S_5 on CR_2 and CR_4, while the two are equivalent on CR_5
Source: Figure created by authors

In the LCJ framework, the dummy alternatives are *postulated* to follow a normal distribution with an unknown mean and the same variance as the *regular* alternatives (Franceschini and Maisano, 2015; Franceschini *et al.*, 2022). These dummy alternatives will be used to "anchor" the scaling derived from the LCJ process, as explained later:

- Transformation of rankings into paired-comparison relationships

Although experts do not explicitly state their preferences between pairs of startups, pairwise comparisons can be indirectly derived for each CR_i from the rankings present in the assessment matrix (Franceschini *et al.*, 2022). Each ranking can be systematically converted into paired comparisons. Figure 5(b) and (c), illustrates this conversion process for each CR_i. Since the case study includes twelve total startups – i.e. eight regular (S_1–S_8) and four dummy ones ($S_●$, $S_○$, $S_△$ and $S_∅$) – the total number of paired comparisons is $C_2^{(8+4)} = \binom{8+4}{2} = \frac{12 \cdot 11}{2} = 66$:

- Thurstone's LCJ application

Next, a proportion (p_{jk}) is associated with each jkth paired comparison. In this study, which involves multiple-criteria evaluations, the proportion p_{jk} denotes the fraction of criteria for which startup S_j is judged to outperform startup S_k. This proportion is determined based on the weights (w_{CR_i}) assigned to each evaluation criterion by the expert group, as explained below. The binary

(a) Assessment matrix

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8
CR_1		○			○		●	
CR_2		●	○		○	●	△	△
CR_3	●			○	○	○	●	
CR_4			△		△			●
CR_5		○	●		○	●	○	

(b) Rankings (including *dummy* startups: $S_●$, $S_○$, $S_△$ and $S_∅$)

CR_1: $(S_●\sim S_7) > (S_○\sim S_2\sim S_5) > S_△ > (S_∅\sim S_1 \sim S_3\sim S_4\sim S_6\sim S_8)$

CR_2: $(S_●\sim S_2\sim S_6) > (S_○\sim S_3\sim S_5) > (S_△\sim S_7\sim S_8) > (S_∅\sim S_1 \sim S_4)$

CR_3: $(S_●\sim S_1\sim S_7) > (S_○\sim S_4\sim S_5\sim S_6) > S_△ > (S_∅ \sim S_2\sim S_3\sim S_4\sim S_8)$

CR_4: $(S_●\sim S_8) > S_○ > (S_△\sim S_3)> (S_∅\sim S_1\sim S_2 \sim S_4\sim S_5\sim S_6\sim S_7)$

CR_5: $(S_●\sim S_3\sim S_6) > (S_○\sim S_2\sim S_5\sim S_7) > S_△ > (S_∅\sim S_1\sim S_4\sim S_8)$

(c) Paired-comparison relationships

S Pairs	CR_1	CR_2	CR_3	CR_4	CR_5
S_1, S_2	<	<	>	~	<
S_1, S_3	~	<	~	<	<
S_1, S_4	~	~	<	~	~
S_1, S_5	<	<	<	<	<
S_1, S_6	~	<	<	~	<
S_1, S_7	<	<	<	~	<
S_1, S_8	~	<	~	<	~
S_1, $S_∅$	~	~	>	~	~
S_1, $S_△$	<	<	>	<	<
S_1, $S_○$	<	<	>	<	<
S_1, $S_●$	<	<	~	<	<
S_2, S_3	>	>	~	<	<
S_2, S_4	>	>	>	~	>
S_2, S_5	>	~	<	<	~
S_2, S_6	>	~	<	~	<
S_2, S_7	<	>	<	~	~
S_2, S_8	>	>	~	<	>
S_2, $S_∅$	>	>	~	~	>
S_2, $S_△$	>	>	<	<	>
S_2, $S_○$	~	>	<	<	~
S_2, $S_●$	<	~	<	<	<
S_3, S_4	~	>	<	>	>
S_3, S_5	<	~	<	~	>
S_3, S_6	~	<	<	>	~
S_3, S_7	<	>	<	>	>
S_3, S_8	~	>	~	<	>
S_3, $S_∅$	~	>	~	>	>
S_3, $S_△$	<	>	<	~	>
S_3, $S_○$	<	~	<	<	>
S_3, $S_●$	<	<	<	<	~
S_4, S_5	<	<	~	<	<
S_4, S_6	~	<	~	>	<
S_4, S_7	<	<	<	~	<
S_4, S_8	~	<	~	<	>
S_4, $S_∅$	~	~	>	~	~
S_4, $S_△$	<	<	>	<	<
S_4, $S_○$	<	<	~	<	<
S_4, $S_●$	<	<	<	<	<
S_5, S_6	>	<	~	>	<
S_5, S_7	<	>	<	>	~
S_5, S_8	>	>	>	<	>
S_5, $S_∅$	>	>	>	>	>
S_5, $S_△$	>	>	>	~	>
S_5, $S_○$	~	~	~	<	~
S_5, $S_●$	<	<	<	<	<
S_6, S_7	<	>	<	~	>
S_6, S_8	~	>	>	<	>
S_6, $S_∅$	~	>	>	~	>
S_6, $S_△$	<	>	>	<	>
S_6, $S_○$	<	>	~	<	>
S_6, $S_●$	<	~	<	<	~
S_7, S_8	>	~	>	<	>
S_7, $S_∅$	>	>	>	~	>
S_7, $S_△$	>	~	>	<	>
S_7, $S_○$	>	<	>	<	~
S_7, $S_●$	~	<	~	<	<
S_8, $S_∅$	~	>	~	>	~
S_8, $S_△$	<	~	<	>	<
S_8, $S_○$	<	<	<	>	<
S_8, $S_●$	<	<	<	~	<
$S_∅$, $S_△$	<	<	<	<	<
$S_∅$, $S_○$	<	<	<	<	<
$S_∅$, $S_●$	<	<	<	<	<
$S_△$, $S_○$	<	<	<	<	<
$S_△$, $S_●$	<	<	<	<	<
$S_○$, $S_●$	<	<	<	<	<

Figure 5. (a) Assessment matrix; (b) transformation of the assessment matrix into *m* rankings, being *m* the number of criteria; (c) paired-comparison relationships between startups
Source: Figure created by authors

comparison between pairs of alternatives is weighted by the respective criterion weight (w_{CR_i}), which indirectly represents the proportion of experts supporting that specific comparison. Naturally, the following condition holds: $\sum_{\forall i} w_{CR_i} = 1$.

In summary, for each (*jk*th) comparison between pairs of startups, there will be a number of pairwise comparisons equal to the number of criteria present in the assessment matrix.

Figure 5(c) illustrates the construction of these relationships from the assessment matrix. For clarity, for each CR_i and each jkth paired comparison, it is useful to define three binary coefficients as follows:

$$\begin{aligned} c_{i,jk}^{(>)} &= 1, \text{ if } S_j > S_k (\text{otherwise } 0), \\ c_{i,jk}^{(\sim)} &= 1, \text{ if } S_j \sim S_k (\text{otherwise } 0), \\ c_{i,jk}^{(<)} &= 1, \text{ if } S_j < S_k (\text{otherwise } 0). \end{aligned} \tag{1}$$

It is evident that these coefficients are mutually exclusive, and they maintain a complementary relationship, as shown by the condition: $c_{i,jk}^{(>)} + c_{i,jk}^{(\sim)} + c_{i,jk}^{(<)} = 1$.

A general coefficient that expresses the degree of preference of S_j over S_k from the perspective of the ith criterion can be defined as:

$$c_{i,jk} = 1 \cdot c_{i,jk}^{(>)} + 1/2 \cdot c_{i,jk}^{(\sim)} + 0 \cdot c_{i,jk}^{(<)}, \in \{0, 1/2, 1\}, \tag{2}$$

For symmetry, it follows that: $c_{i,jk} = 1 - c_{i,kj}$. The computation of each single binary coefficient and the corresponding global values ($c_{i,jk}$) for all the paired-comparison relationships is shown in Figure 6. As last step, the coefficients $c_{i,jk}$ are aggregated in the following weighted sum, which determines the p_{jk} values (Maisano *et al.*, 2024).

$$p_{jk} = \sum_{\forall i} (c_{i,jk} \cdot w_{CR_i}), \; p_{jk} \in [0,1]. \tag{3}$$

In adapting the LCJ (cf. Section 4), the p_{jk} proportions are calculated by aggregating prior comparisons through a weighted sum. Conceptually, this value represents the weighted fraction of criteria for which the startup S_j has a greater influence than the startup S_k. For example, for the paired comparison (S_1, S_3) (at the top of Figure 6), the corresponding value for p_{13} would result in: $p_{13} = 0.5 \cdot 22\% + 0 \cdot 20\% + 1 \cdot 14\% + 0 \cdot 11\% + 0 \cdot 33 = 0.250$.

The p_{jk} values can be aggregated into a P matrix of proportions, which – consistently with the LCJ – includes elements that are symmetrical with respect to the main diagonal and complementary to each other with respect to the unit [2]. Table 2(a) contains the P matrix that results from the paired-comparison relationships in Figure 5(c). Starting from now, the traditional LCJ (cf. Section 4) is applied, determining the Z matrix (see Table 2(b)) and, subsequently, the *interval* scaling (x) of the various startup alternatives (see Table 2(c)).

For benchmarking, the same expert data were re-processed with a conventional *Weighted Borda's Count* (WBC) procedure (Franceschini *et al.*, 2022). The LCJ ranking showed a high concordance with the WBC results (Pearson's $R^2 = 0.87$), with both methods identifying startups S_4, S_1 and S_8 as the first, second and third most promising candidates, respectively:

- Scale anchoring

Drawing from the methodology outlined in Franceschini and Maisano (2019), the *interval* scaling derived from the Thurstone's LCJ represents the startups' overall performance level – information that is useful for guiding future investments and support actions – and is calculated through a weighted aggregation of the performances with respect to the five criteria introduced in Section 2. Because its zero point is arbitrary, this interval scale may assume negative values, which can complicate practical interpretation (cf., the x_k values in Table 2(c)). The interval scale can, however, be *anchored* to a *ratio* scale with an *absolute zero*, by exploiting the two dummy alternatives: $S_\varnothing$, representing a fictitious

S pairs	CR_1 (w_{CR_1}=22%)				CR_2 (w_{CR_2}=20%)				CR_3 (w_{CR_3}=14%)				CR_4 (w_{CR_4}=11%)				CR_5 (w_{CR_5}=33%)				p_{jk}
	$c^{(>)}_{1,jk}$	$c^{(\sim)}_{1,jk}$	$c^{(<)}_{1,jk}$	$c_{1,jk}$	$c^{(>)}_{2,jk}$	$c^{(\sim)}_{2,jk}$	$c^{(<)}_{2,jk}$	$c_{2,jk}$	$c^{(>)}_{3,jk}$	$c^{(\sim)}_{3,jk}$	$c^{(<)}_{3,jk}$	$c_{3,jk}$	$c^{(>)}_{4,jk}$	$c^{(\sim)}_{4,jk}$	$c^{(<)}_{4,jk}$	$c_{4,jk}$	$c^{(>)}_{5,jk}$	$c^{(\sim)}_{5,jk}$	$c^{(<)}_{5,jk}$	$c_{5,jk}$	
S_1, S_2	0	0	1	**0.0**	0	0	1	**0.0**	1	0	0	**1.0**	0	1	0	**0.5**	0	0	1	**0.0**	0.195
S_1, S_3	0	1	0	**0.5**	0	0	1	**0.0**	1	0	0	**1.0**	0	0	1	**0.0**	0	0	1	**0.0**	0.250
S_1, S_4	0	1	0	**0.5**	0	1	0	**0.5**	1	0	0	**1.0**	0	1	0	**0.5**	0	1	0	**0.5**	0.570
S_1, S_5	0	0	1	**0.0**	0	0	1	**0.0**	1	0	0	**1.0**	0	0	1	**0.0**	0	0	1	**0.0**	0.140
S_1, S_6	0	1	0	**0.5**	0	0	1	**0.0**	1	0	0	**1.0**	0	1	0	**0.5**	0	0	1	**0.0**	0.305
S_1, S_7	0	0	1	**0.0**	0	0	1	**0.0**	0	1	0	**0.5**	0	1	0	**0.5**	0	0	1	**0.0**	0.125
S_1, S_8	0	1	0	**0.5**	0	0	1	**0.0**	1	0	0	**1.0**	0	0	1	**0.0**	0	1	0	**0.5**	0.415
S_1, $S_\emptyset$	0	1	0	**0.5**	0	1	0	**0.5**	1	0	0	**1.0**	0	1	0	**0.5**	0	1	0	**0.5**	0.570
S_1, $S_\triangle$	0	0	1	**0.0**	0	0	1	**0.0**	1	0	1	**1.0**	0	0	1	**0.0**	0	0	1	**0.0**	0.140
S_1, $S_\circ$	0	0	1	**0.0**	0	0	1	**0.0**	1	0	1	**1.0**	0	0	1	**0.0**	0	0	1	**0.0**	0.140
S_1, $S_\bullet$	0	0	1	**0.0**	0	0	1	**0.0**	0	1	0	**0.5**	0	0	1	**0.0**	0	0	1	**0.0**	0.070
S_2, S_3	1	0	0	**1.0**	1	0	0	**1.0**	0	1	0	**0.5**	0	0	1	**0.0**	0	0	1	**0.0**	0.490
S_2, S_4	1	0	0	**1.0**	1	0	0	**1.0**	0	0	1	**0.0**	0	1	0	**0.5**	1	0	0	**1.0**	0.805
S_2, S_5	0	1	0	**0.5**	1	0	0	**1.0**	0	0	1	**0.0**	0	0	1	**0.0**	0	1	0	**0.5**	0.475
S_2, S_6	1	0	0	**1.0**	0	1	0	**0.5**	0	0	1	**0.0**	0	1	0	**0.5**	0	0	1	**0.0**	0.375
S_2, S_7	0	0	1	**0.0**	0	0	1	**0.0**	0	0	1	**0.0**	0	1	0	**0.5**	0	1	0	**0.5**	0.220
S_2, S_8	1	0	0	**1.0**	0	0	1	**0.0**	0	1	0	**0.5**	0	0	1	**0.0**	1	0	0	**1.0**	0.620
S_2, $S_\emptyset$	1	0	0	**1.0**	0	0	1	**0.0**	0	1	0	**0.5**	0	1	0	**0.5**	1	0	0	**1.0**	0.675
S_2, $S_\triangle$	1	0	0	**1.0**	0	0	1	**0.0**	0	0	1	**0.0**	0	0	1	**0.0**	1	0	0	**1.0**	0.550
S_2, $S_\circ$	0	1	0	**0.5**	0	0	1	**0.0**	0	0	1	**0.0**	0	0	1	**0.0**	0	1	0	**0.5**	0.275
S_2, $S_\bullet$	0	0	1	**0.0**	0	1	0	**0.5**	0	0	1	**0.0**	0	0	1	**0.0**	0	0	1	**0.0**	0.100
S_3, S_4	0	1	0	**0.5**	1	0	0	**1.0**	0	0	1	**0.0**	1	0	0	**1.0**	1	0	0	**1.0**	0.750
S_3, S_5	0	0	1	**0.0**	0	1	0	**0.5**	0	0	1	**0.0**	0	1	0	**0.5**	1	0	0	**1.0**	0.485
S_3, S_6	0	1	0	**0.5**	0	0	1	**0.0**	0	0	1	**0.0**	1	0	0	**1.0**	0	1	0	**0.5**	0.385
S_3, S_7	0	0	1	**0.0**	1	0	0	**1.0**	0	0	1	**0.0**	1	0	0	**1.0**	1	0	0	**1.0**	0.640
S_3, S_8	0	1	0	**0.5**	1	0	0	**1.0**	0	1	0	**0.5**	0	0	1	**0.0**	1	0	0	**1.0**	0.710
S_3, $S_\emptyset$	0	1	0	**0.5**	1	0	0	**1.0**	0	1	0	**0.5**	1	0	0	**1.0**	1	0	0	**1.0**	0.820
S_3, $S_\triangle$	0	0	1	**0.0**	1	0	0	**1.0**	0	0	1	**0.0**	0	1	0	**0.5**	1	0	0	**1.0**	0.585
S_3, $S_\circ$	0	0	1	**0.0**	0	1	0	**1.0**	0	0	1	**0.0**	0	0	1	**0.0**	1	0	0	**1.0**	0.530
S_3, $S_\bullet$	0	0	1	**0.0**	0	0	1	**0.0**	0	0	1	**0.0**	0	0	1	**0.0**	0	1	0	**0.5**	0.165
S_4, S_5	0	0	1	**0.0**	0	0	1	**0.0**	0	1	0	**0.5**	0	0	1	**0.0**	0	0	1	**0.0**	0.070
S_4, S_6	0	1	0	**0.5**	0	0	1	**0.0**	0	1	0	**0.5**	0	1	0	**0.5**	0	0	1	**0.0**	0.235
S_4, S_7	0	0	1	**0.0**	0	0	1	**0.0**	0	0	1	**0.0**	0	1	0	**0.5**	0	0	1	**0.0**	0.055

Figure 6. Calculation of the $c_{i,jk}$ values (bolded) by combining the binary coefficients ($c^{(>)}_{i,jk}$, $c^{(\sim)}_{i,jk}$, $c^{(<)}_{i,jk}$) obtained from the paired-comparison relationships, through equation (2). Next, the relevant p_{jk} values (in the last column) are calculated by applying equation (3)

Source: Figure created by authors

startup with *zero relevance* across all criteria, is fixed at the absolute zero of the ratio scale, whereas $S_\bullet$, representing a fictitious startup with the *maximum conceivable relevance* across all criteria, is placed at the upper bound of that scale (conventionally set at the value 10). This converts the interval scale (x) into a new scale (y) within the range [0, 10] through a linear transformation:

$$\frac{y_k - 0}{10 - 0} = \frac{x_k - x_\emptyset}{x_\bullet - x_\emptyset} \rightarrow y_k = 10 \cdot \frac{x_k - x_\emptyset}{x_\bullet - x_\emptyset}, \tag{4}$$

where:

$x_\emptyset$ and $x_\bullet$ are the scale values of $S_\emptyset$ and $S_\bullet$, respectively, resulting from the LCJ;
x_k is the scale value of a generic S_k, resulting from the LCJ; and
y_k is the relevant transformed scale value in the conventional range [0, 10].

The new scale (y_k) has a conventional unit and an absolute zero-point (indicating the attribute's absence); it is therefore classifiable as a *ratio* scale (Stevens, 1946; Franceschini *et*

Table 2. (a) *P* matrix, (b) *Z* matrix and (c) scaling resulting from the application of the proposed procedure to the case study

	S_1	S_2	S_3	S_4	S_5	S_6	S_7	S_8	$S_\varnothing$	$S_\triangle$	$S_\circ$	$S_\bullet$
(a) P matrix												
S_1	0.500	0.195	0.250	0.535	0.140	0.305	0.125	0.415	0.570	0.140	0.140	0.070
S_2	0.805	0.500	0.490	0.805	0.475	0.375	0.220	0.620	0.675	0.555	0.275	0.050
S_3	0.750	0.490	0.500	0.250	0.514	0.615	0.360	0.290	0.180	0.760	0.810	0.070
S_4	0.430	0.195	0.250	0.500	0.070	0.235	0.055	0.250	0.570	0.140	0.070	0.000
S_5	0.860	0.525	0.514	0.930	0.500	0.400	0.475	0.890	1.000	0.945	0.445	0.000
S_6	0.695	0.625	0.615	0.765	0.600	0.500	0.585	0.780	0.945	0.670	0.600	0.265
S_7	0.875	0.780	0.360	0.945	0.525	0.415	0.500	0.790	0.945	0.790	0.525	0.180
S_8	0.585	0.380	0.290	0.750	0.110	0.220	0.210	0.500	0.655	0.210	0.110	0.055
$S_\varnothing$	0.430	0.325	0.180	0.543	0.000	0.055	0.055	0.345	0.500	0.000	0.000	0.000
$S_\triangle$	0.860	0.455	0.765	0.860	0.055	0.330	0.210	0.790	1.000	0.500	0.000	0.000
$S_\circ$	0.860	0.695	0.810	0.930	0.555	0.400	0.475	0.890	1.000	1.000	0.500	0.000
$S_\bullet$	0.930	0.950	0.930	1.000	1.000	0.735	0.820	0.945	1.000	1.000	1.000	0.500
(b) Z matrix												
S_1	0.000	0.859	0.674	−0.176	1.080	0.510	1.150	0.125	−0.176	1.080	1.080	1.476
S_2	−0.859	0.000	0.025	−0.859	0.062	−0.318	0.772	−0.305	−0.455	−0.125	0.597	1.281
S_3	−0.675	−0.025	0.000	−0.674	0.035	0.292	−0.358	−0.553	−0.915	−0.214	−0.075	−0.974
S_4	0.176	0.859	0.674	0.000	1.475	0.722	1.598	0.675	−0.176	1.080	1.476	3.000
S_5	−1.080	−0.062	−0.035	−1.475	0.000	0.253	0.063	−1.227	−3.000	−1.598	0.138	3.000
S_6	−0.510	−0.318	−0.292	−0.722	−0.253	0.000	−0.215	−0.772	−1.598	−0.439	−0.253	0.628
S_7	−1.150	−0.772	−0.358	−1.598	−0.063	0.215	0.000	−0.806	−1.598	−0.806	−0.063	0.915
S_8	−0.214	0.305	0.553	−0.674	1.227	0.772	0.806	0.000	−0.399	0.806	1.227	1.598
$S_\varnothing$	0.176	0.453	0.915	0.176	−3.000	1.598	1.598	0.399	0.000	3.000	3.000	3.000
$S_\triangle$	−1.080	0.125	0.214	−1.080	1.598	0.439	0.806	−0.806	−3.000	0.000	3.000	3.000
$S_\circ$	−1.080	−0.597	0.074	−1.476	−0.138	0.253	−0.063	−1.226	−3.000	−3.000	0.000	3.000
$S_\bullet$	−1.476	−1.281	−0.874	−3.000	−3.000	−0.628	−0.915	−1.599	−3.000	−3.000	−3.000	0.000
(c) Scaling												
Σ_k	−7.773	−0.453	2.23	−11.56	−0.977	4.108	5.242	−6.095	−17.317	−3.216	7.127	19.924
x_k	−0.647	−0.037	0.185	−0.963	−0.081	0.342	0.437	−0.508	−1.443	−0.268	0.594	1.667
y_k	2.56	4.53	5.25	1.55	4.39	5.75	6.06	3.01	0	3.78	6.56	10

Note(s): Σ_k is the summation of the values reported in the kth column of the Z matrix; the x_k values concern the *interval* scaling resulting from the Thurstone's LCJ; the y_k values concern the *ratio* scaling downstream of the anchoring in equation (4)
Source(s): Table created by authors

al., 2022). Although Thurstone's LCJ computational processes may seem complex, they are actually straightforward and can be fully automated using standard spreadsheet software, such as MS Excel® (Franceschini and Maisano, 2019). Figure 7 provides a graphical summary of the final (ratio) scaling, illustrating the startups' overall performance.

6. Concluding remarks

6.1 Key findings and practical implications

This paper introduced a useful method that enables diverse stakeholders – venture capitalists, private-equity investors, accelerators and incubators, economic-development agencies, public funding bodies, universities, research institutes and crowdfunding platforms – to assess and position start-ups in line with Quality 4.0 principles. Those principles guided the selection of the key evaluation dimensions: *team characteristics*, *patents and market potential*, *strategic alliances*, *financial health and funding*, and *innovation and technology* (see *RQ1*). The methodology distinguishes itself from other support tools through its simplicity and the limited amount of information required from experts. The core of this methodology lies in the assessment matrix, where, for each analysis criterion, the team of experts is asked to determine the intensity of the relationship linking each alternative (startup) to the criterion in question, on a four-level ordinal scale. Based on the startups' evaluations for each criterion, a ranking of the startups is established. From this ranking, pairwise comparisons are derived, leading to a comprehensive positioning of the evaluated startups. The steps following the creation of the assessment matrix can be fully automated on any software platform or spreadsheet (e.g. MS Excel®). The methodology also allows for scenario analysis by varying the configuration of weights assigned to individual criteria, as well as sensitivity analyses by adjusting some of the weights and/or the intensity ratings of specific alternatives (startups).

One of the strengths of the proposed adaptation of Thurstone's LCJ is that it asks experts to express their assessments with simple, practitioner-friendly response modes – chiefly *ordinal* scales or straightforward preference rankings (Franceschini *et al.*, 2022). This is especially valuable when evaluating early-stage startups, for which detailed quantitative data are often unavailable. Once these basic assessments have been gathered, the probabilistic structure of LCJ transforms them into an *interval* scaling and – after scale anchoring – a 0–10 *ratio* scaling that helps to identify the most promising and incentive-worthy startup alternatives (cf. *RQ2*). This scaling may give the whole ecosystem a shared yard-stick for judging young ventures. For professional investors, it may work as a first-pass filter: only the companies that land in the upper band of the scale deserve time-consuming due-diligence, while the same scores, read in reverse, warn against premature enthusiasm. Policy makers can turn the numbers into transparent thresholds when distributing grants or tax breaks and, by tracking average scores across sectors, monitor the "health" of their start-up landscape almost in real time. Managers of accelerators and incubators can overlay each start-up's radar plot on an internal skills audit, design coaching that tackles the most obvious gaps and repeat the assessment mid-program to check whether support is actually moving the needle.

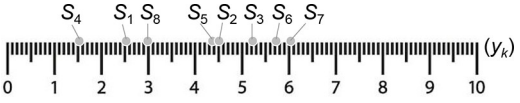

Figure 7. Graphical representation of the startups' final (*ratio*) scaling
Source: Figure created by authors

Universities and research centers gain a fair gatekeeping tool for choosing which spin-outs to back, while the history of scores accumulated across cohorts becomes a natural data set for studying what drives success. Even crowdfunding platforms can translate the underlying ratios into intuitive bronze/silver/gold badges, so that casual investors can gauge quality at a glance without having to wade through technical reports.

Because the assessments are produced collectively, consensually and collaboratively by the expert panel, the approach helps reduce the subjectivity that inevitably arises when decisions rely on sparse or unstructured information. Individual opinions are systematically fused into a group consensus, ensuring that the final evaluation embodies the collective wisdom of the panel, minimizes personal bias, and enhances the robustness of the assessment. This makes it easier to interpret the results and clearly identify which startups stand out.

The method is also flexible, versatile enough to accommodate emerging or less-structured criteria that may be difficult to quantify using traditional methods. Statistically speaking, the method is robust, since it provides a solid framework for analyzing preferences and choices, enabling deeper insights into the factors driving decisions. A relevant aspect is that it helps identify key factors that influence how a startup is perceived. These insights can be invaluable for refining the scouting process and making better decisions in the future. In short, the methodology not only makes the process more systematic but also helps ensure that decisions are well-informed and based on shared data.

6.2 Limitations

While Thurstone's LCJ offers many advantages for evaluating startups, there are some limitations to consider. First, because the empirical test was confined to eight startups from a single regional aerospace–digital-manufacturing cluster, the findings cannot be automatically generalized to other geographies or industries. Replication studies in service-oriented, bio-tech or emerging-market ecosystems are therefore required before broader claims can be made. Second, the five Quality 4.0-aligned criteria were optimized for technology-intensive ventures; sectors with different risk profiles (e.g. med-tech or social enterprises) may call for context-specific adaptations, such as adding regulatory readiness or social-impact indicators. Third, operational scalability remains a challenge: the complexity of the LCJ model grows quadratically with the number of startups, which could be onerous for accelerators screening hundreds of applicants (Franceschini *et al.*, 2022). Potential remedies include hierarchical clustering to reduce the comparison set, algorithmic sampling of representative pairs and crowd-sourced or AI-assisted judgment platforms to distribute the evaluation workload.

6.3 Future development

Regarding the future, there is room to explore how cultural and contextual factors influence expert judgments. Understanding these influences could lead to more tailored and effective evaluation processes, making Thurstone's method more adaptable to different regions or industries. In future research we will evaluate replacing LCJ with more advanced models that can handle *incomplete* or *partial* assessments – for instance, situations in which the expert panel lacks reliable information to score every startup on every criterion. A promising candidate is the ZM_{II} method proposed by Franceschini and Maisano (2019) for analogous multiple-criteria problems. We also plan to integrate AI-based techniques capable of assisting the expert team – for example, by mining historical data sets on comparable startups and surfacing statistically significant patterns (Morande *et al.*, 2023). Such AI support could further mitigate the inherent subjectivity of expert judgment by supplying evidence-based, supplementary insights that enrich and stabilize the collective evaluation.

Research paradigm and epistemological stance: The study is grounded in a pragmatic approach, viewing a startup's "quality" as a latent construct that can be analyzed through observable indicators and expert perceptions. A mixed-methods design is therefore adopted, integrating qualitative expert judgments with quantitative scaling via Thurstone's LCJ. This approach aligns with design-science research principles and with the evidence-based ethos of Quality 4.0, which emphasizes continuous improvement through data-driven insights.

Acknowledgements

This research has been conducted within the framework of the "MUR DM 352/Leonardo-Theoretical and practical methods and tools for the management of a complex open innovation network" collaboration between Politecnico di Torino and Leonardo S.p.A.

Ethics statement

The research protocol was approved under Politecnico di Torino's Research Ethics Guidelines. No personally identifiable data were collected, ensuring minimal risk and full confidentiality for both experts and companies.

Notes

[1.] The six experts completed the evaluation procedure of two early-stage startups that were not included in the final sample, separately; inter-rater agreement was then examined using Kendall's W (0.78) on the four-level ordinal scale, indicating substantial consistency (Franceschini *et al.*, 2022). Minor refinements were made to the criterion descriptors, after which the panel confirmed their validity.

[2.] By combining the relationships $c_{i,jk}^{(>)} + c_{i,jk}^{(\sim)} + c_{i,jk}^{(<)} = 1$, $c_{i,jk} = 1 - c_{i,kj}$, and $\sum_{\forall i} w_{CR_i} = 1$, it can be easily deduced that $p_{jk} = 1 - p_{kj}$.

References

Armstrong, M. (2006), "Competition in two-sided markets", *The RAND Journal of Economics*, Vol. 37 No. 3, pp. 668-691.

Arshi, T.A., Rao, V., Islam, S. and Morande, S. (2021), "SECURE–a new business model framework for measuring start-up performance", *Journal of Entrepreneurship in Emerging Economies*, Vol. 13 No. 3, pp. 459-485.

Blank, S. and Dorf, B. (2020), *The Startup Owner's Manual: The Step-By-Step Guide for Building a Great Company*, John Wiley and Sons, New York, NY.

Blank, T.H. and Carmeli, A. (2021), "Does founding team composition influence external investment? The role of founding team prior experience and founder CEO", *The Journal of Technology Transfer*, Vol. 46 No. 6, pp. 1869-1888.

Bond, T.G. and Fox, C.M. (2007), *Applying the Rasch Model (Second Edition)*, Psychology Press, New York, NY.

Bottani, E. and Rizzi, A. (2008), "An adapted multi-criteria approach to suppliers and products selection -An application oriented to lead-time reduction", *International Journal of Production Economics*, Vol. 111 No. 2, pp. 763-781.

Brown, G.W. and Gredil, O. (2022), "Are mutual funds' startup marks too high? ", *Journal of Financial Economics*, Vol. 146 No. 3, pp. 711-736.

Conti, A., Thursby, J. and Thursby, M. (2013), "Patents as signals for startup financing", *The Journal of Industrial Economics*, Vol. 61 No. 3, pp. 592-622.

Edwards, A.L. (1957), *Techniques of Attitude Scale Construction*, Irvington Publishers, New York, NY.

El Hanchi, S.K.L., (2020), "Startup innovation capability from a dynamic capability-based view: a literature review and conceptual framework", *Journal of Small Business Strategy*, Vol. 30 No. 2, pp. 72-92.

Fernández-Olmos, M. and Ramírez-Alesón, M. (2017), "How internal and external factors influence the dynamics of SME technology collaboration networks over time", *Technovation*, Vols 64-65, pp. 16-27.

Franceschini, F. and Maisano, D.A. (2015), "Prioritization of QFD customer requirements based on the law of comparative judgments", *Quality Engineering*, Vol. 27 No. 4, pp. 437-449.

Franceschini, F. and Maisano, D. (2019), "Fusing incomplete preference rankings in design for manufacturing applications through the ZM_{II}-technique", *The International Journal of Advanced Manufacturing Technology*, Vol. 103 Nos 9-12, pp. 3307-3322.

Franceschini, F. and Rossetto, S. (1995), "QFD: the problem of comparing technical/engineering design requirements", *Research in Engineering Design*, Vol. 7 No. 4, pp. 270-278.

Franceschini, F., Maisano, D.A. and Mastrogiacomo, L. (2022), *Rankings and Decisions in Engineering*, Springer International Publishing, Cham, Switzerland.

Galende, J. (2006), "Analysis of technological innovation from business economics and management", *Technovation*, Vol. 26 No. 3, pp. 300-311.

Gompers, P. and Kaplan, S.N. (2019), "What are we learning about venture capital? ", *Journal of Economic Perspectives*, Vol. 33 No. 3, pp. 173-198.

Hatch, N.W. and Dyer, J.H. (2004), "Human capital and learning as a source of sustainable competitive advantage", *Strategic Management Journal*, Vol. 25 No. 12, pp. 1155-1178.

Jeon, H. (2019), "Patent protection and R&D subsidy under asymmetric information", *International Review of Economics and Finance*, Vol. 62, pp. 332-354.

Kelly, K.T., Richardson, M. and Isaacs, T. (2022), "Critiquing the rationales for using comparative judgement: a call for clarity", *Assessment in Education: Principles, Policy and Practice*, Vol. 29 No. 6, pp. 674-688.

Khourshed, N.F. (2023), "Investigating the critical success factors for integrating lean six sigma and industry 4.0", *Quality Management Journal*, Vol. 30 No. 4, pp. 259-278.

Khourshed, N. and Gohar, N. (2023), "Developing a systematic and practical road map for implementing quality 4.0", *Quality Innovation Prosperity*, Vol. 27 No. 2, pp. 96-121.

Khourshed, N.F., Elbarky, S.S. and Elgamal, S. (2023), "Investigating the readiness factors for industry 4.0 implementation for manufacturing industry in Egypt", *Sustainability*, Vol. 15 No. 12, p. 9641.

Köseoğlu, S.D. (2023), "A practical guide for startup valuation: an analytic approach", *Contributions to Finance and Accounting*, Springer Nature Switzerland, Cham., 1st ed., pp. 1-6.

Kotter, J.P., Porter, M. and Olmsted Teisberg, E. (2015), *Leadership, Strategy, and Innovation: Health Care Collection*, Harvard Business Review Press, Boston.

Kyrylych, T. and Povstenko, Y. (2023), "Multi-criteria analysis of Startup investment alternatives using the hierarchy method", *Entropy (Basel)*, Vol. 25 No. 5, p. 723.

Lin, M., Chen, Z., Chen, R. and Fujita, H. (2021), "Evaluation of startup companies using multicriteria decision making based on hesitant fuzzy linguistic information envelopment analysis models", *International Journal of Intelligent Systems*, Vol. 36 No. 5, pp. 2292-2322.

Luce, R.D. and Tukey, J.W. (1964), "Simultaneous conjoint measurement: a new type of fundamental measurement", *Journal of Mathematical Psychology*, Vol. 1 No. 1, pp. 1-27.

Maisano, D.A., Carrera, G., Mastrogiacomo, L. and Franceschini, F. (2024), "A new method to prioritize the QFDs' engineering characteristics inspired by the law of comparative judgment", *Research in Engineering Design*, Vol. 35 No. 4, pp. 343-353.

Mamédio, D., Rocha, C., Szczepanik, D. and Kato, H. (2019), "Strategic alliances and dynamic capabilities: a systematic review", *Journal of Strategy and Management*, Vol. 12 No. 1, pp. 83-102.

Manigart, S., Wright, M., Robbie, K., Desbrières, P. and De Waele, K. (1997), "Venture capitalists' appraisal of investment projects: an empirical European study", *Entrepreneurship Theory and Practice*, Vol. 21 No. 4, pp. 29-43.

Morande, S., Arshi, T., Gul, K. and Amini, M. (2023), "Harnessing the power of artificial intelligence to forecast startup success: an empirical evaluation of the SECURE AI model", SocArXiv p3gyb, Center for Open Science.

Noh, H., Seo, J.-H., Sun Yoo, H. and Lee, S. (2018), "How to improve a technology evaluation model: a data-driven approach", *Technovation*, Vols 72-73, pp. 1-12.

Pollitt, A. (2012), "Comparative judgement for assessment", *International Journal of Technology and Design Education*, Vol. 22 No. 2, pp. 157-170.

Rasch, G. (1966), "An item analysis which takes individual differences into account", *British Journal of Mathematical and Statistical Psychology*, Vol. 19 No. 1, pp. 49-57.

Sievers, S., Mokwa, C.F. and Keienburg, G. (2013), "The relevance of financial versus nonfinancial information for the valuation of venture capital-backed firms", *European Accounting Review*, Vol. 22 No. 3, pp. 467-511.

Smith, J.A. and Cordina, R. (2014), "The role of accounting in high-technology investments", *The British Accounting Review*, Vol. 46 No. 3, pp. 309-322.

Stevens, S.S. (1946), "On the theory of scales of measurement", *Science*, Vol. 103 No. 2684, pp. 677-680.

Thurstone, L.L. (1927), "A law of comparative judgments", *Psychological Review*, Vol. 34 No. 4, p. 273.

Valiris, G., Chytas, P. and Glykas, M. (2005), "Making decisions using the balanced scorecard and the simple multi-attribute rating technique", *Performance Measurement and Metrics*, Vol. 6 No. 3, pp. 159-171.

Wei, Y.M. (2025), "A hybrid multi-criteria decision-making framework for the strategic evaluation of business development models", *Information*, Vol. 16 No. 6, p. 454.

Yildirim, N. and Maz, Y. (2025), "A thematic analysis of literature on the startup assessment criteria: startup success factors revisited", *Human-Centric, Sustainable, and Resilient Organizations in the Digital Age*, pp. 303-328.

Further reading

Franceschini, F. and Maisano, D. (2018), "A new proposal to improve the customer competitive benchmarking in QFD", *Quality Engineering*, Vol. 30 No. 4, pp. 730-761.

Franceschini, F., Galetto, M., Maisano, D. and Mastrogiacomo, L. (2015), "Prioritisation of engineering characteristics in QFD in the case of customer requirements orderings", *International Journal of Production Research*, Vol. 53 No. 13, pp. 3975-3988.

Galetto, M., Franceschini, F., Maisano, D. and Mastrogiacomo, L. (2018), "Engineering characteristics prioritisation in QFD using ordinal scales: a robustness analysis", *European J. of Industrial Engineering*, Vol. 12 No. 2, pp. 151-174.

Jun, S., Park, S. and Jang, D. (2015), "A technology valuation model using quantitative patent analysis: a case study of technology transfer in big data marketing", *Emerging Markets Finance and Trade*, Vol. 51 No. 5, pp. 963-974.

Maranell, G. (1974), *Scaling: A Sourcebook for Behavioral Scientists*, 1st ed., Routledge.

Corresponding author

Fiorenzo Franceschini can be contacted at: fiorenzo.franceschini@polito.it