

How to conduct effective risk culture assessments

Niclas Dürst, Jennifer Kunz

Angaben zur Veröffentlichung / Publication details:

Dürst, Niclas, and Jennifer Kunz. 2025. "How to conduct effective risk culture assessments." *Journal of Management Control* 36 (2): 269–314.
<https://doi.org/10.1007/s00187-025-00402-y>.



How to conduct effective risk culture assessments

Niclas Dürst¹ · Jennifer Kunz¹

Received: 4 November 2024 / Accepted: 3 November 2025 / Published online: 8 December 2025
© The Author(s) 2025

Abstract

Assessing risk culture remains a major challenge for financial institutions, as existing tools often fail to capture its multidimensional and evolving nature. While prior research has focused on developing measurement instruments, less attention has been paid to the practical and methodological challenges involved in applying such tools effectively. This study provides practitioner-oriented insights into how risk culture assessments can be designed and implemented more effectively. Drawing on non-item-specific feedback from expert interviews conducted in the context of a broader risk culture measurement initiative aligned with the Financial Stability Board (FSB) framework, it identifies key considerations for improving survey-based assessments, focusing on general implementation challenges rather than scale-specific content. Using the Gioia method, 21 qualitative interviews with risk management experts were analyzed. First-order concepts were extracted and aggregated into eight second-order themes, which were synthesized into a practice-oriented framework for reflecting on the design and application of risk culture assessments. The eight resulting themes highlight critical design considerations: Clarity of Terminology, Optimization of Questionnaire Design, Reduction of Redundancies, Adaptation of Item Order, Limitations of Questionnaires, Scale Design, Time-Series Analysis, and Illuminating Risk Culture from Different Perspectives. These findings reflect both methodological insights and implementation challenges drawn from organizational practice. Rather than proposing a new measurement tool, this study offers a grounded framework for improving the application of existing instruments. By bridging regulatory guidance, expert experience, and qualitative theory-building, it supports the development of more context-sensitive and actionable approaches to risk culture assessment.

Keywords Risk culture assessment · Financial stability board · Gioia method · Expert feedback · Financial institutions

JEL classification M14

Extended author information available on the last page of the article

1 Introduction

In today's complex financial environment, risk management is essential for ensuring organizational resilience and stability (Baghersad & Zobel, 2022; Marquez-Tejon et al., 2022, Lisdiono et al., 2022; Durst et al., 2019). Central to an effective risk management is the development of a robust risk culture, which enhances an organization's ability to manage uncertainties and maintain continuous operations during disruptions (Osman & Lew, 2021; Aven & Ylönen, 2021; Sheedy et al., 2019), ensures mitigating potential threats and aligns risk-taking behaviors with an organization's strategic goals (Agnese & Capuano, 2021). Thus, a strong organizational risk culture can create substantial value for firms (Gatzert & Bockius, 2023).

Assessing and managing risk culture remains a significant challenge for financial institutions, not only due to ambiguous terminology but also because of the inherently multidimensional and embedded nature of the concept. Risk culture encompasses interrelated elements such as leadership behavior (“tone from the top”), employee risk awareness, incentive structures, communication openness, and norms for risk-based decision-making (McConnell, 2013; Financial Stability Board FSB, 2014). Each of these elements poses distinct challenges for measurement: for instance, leadership behavior is often signaled implicitly rather than formally documented, incentive structures may create unintended effects that are difficult to detect, and employee perceptions of communication openness or decision norms are shaped by informal dynamics rather than formal policies. Taken together, these interdependencies operate across both formal structures and informal behavioral patterns, making culture difficult to observe directly or to quantify as a single variable. Consequently, risk culture assessments must focus on specific dimensions rather than attempting to measure the phenomenon in aggregate, which highlights the need for validated instruments capable of capturing such dimensions in a systematic manner (Fernández Muñiz et al., 2020; Sheedy & Jepsen, 2018).

Well-conducted risk culture assessments are essential for practice as they help uncover cultural gaps, such as misalignments between governance frameworks and actual employee behavior or discrepancies between leadership tone and frontline decision-making (Sheedy & Griffin, 2018), and more broadly capture the attitudinal and behavioral aspects emphasized in organization design and “soft tools” perspectives (Previati, 2017). At the same time, they provide an evidence base for evaluating whether cultural change initiatives—such as incentive realignment or leadership interventions—achieve their intended outcomes (Financial Stability Board FSB, 2014; Basel Committee on Banking Supervision BCBS, 2015; Butler & Brooks, 2024).

The global financial crisis of 2007–2009 underscored the importance of establishing a comprehensive risk culture that extends beyond traditional risk management departments. This period revealed significant deficiencies in how financial institutions managed risk, leading to increased emphasis by managers, regulators, and researchers on fostering a proactive risk culture (McConnell, 2013; Power et al., 2013; Financial Stability Board FSB, 2014). In response, regulatory bodies have increasingly emphasized the need for robust risk cultures in financial institutions to mitigate future crises (Basel Committee on Banking Supervision BCBS, 2015; Aus-

tralian Prudential Regulation Authority APRA, 2016; European Banking Authority EBA, 2017).

Despite growing recognition of its importance, the literature on risk culture assessment highlights several gaps: the lack of a standard definition, diverse measurement approaches, and the need for comprehensive frameworks that capture its multidimensional nature (Cimini, 2021; Wiedemann et al., 2020). Traditional quantitative tools often fail to account for the complexity and dynamism inherent in risk culture (Agarwal et al., 2019). Moreover, existing theoretical models do not adequately address the intricate interdependencies that characterize risk culture within organizations (Palermo et al., 2017). The current methodologies used to assess risk culture, such as surveys, interviews, and text analysis, each have their advantages and limitations depending on the context (Bockius and Gatzert, 2023; Agarwal et al., 2019). While various scales for assessing risk culture, such as those developed by Sheedy et al. (2017), Fernández Muñoz et al. (2020), and Ghafoori et al. (2023), provide valuable frameworks, they often face challenges in capturing the context-specific manifestations of risk culture within financial institutions.

This paper addresses this observation. Instead of attempting to refine extant measurement items or scales, it focuses on the practical challenges of applying survey-based instruments for risk culture assessment. Drawing on expert interviews conducted as part of a broader risk culture research initiative aligned with the Financial Stability Board (FSB) framework (Dürst and Kunz, 2025), we identify common obstacles and propose a structured conceptual framework to support more effective application of such tools in practice. To achieve this goal, we focus specifically on non-item-specific feedback by the experts, i.e. observations and insights voiced by experts that were not tied to particular questionnaire items and captured broader reflections on practical, institutional, and methodological challenges related to survey-based risk culture assessments and allowed us to identify cross-cutting design issues beyond the scope of a specific measurement instrument offering transferable insights into general implementation issues. The feedback was analyzed using the Gioia methodology (Gioia et al., 2013). This method involves a structured, inductive coding process that distinguishes between first-order concepts—terms and phrases used by informants themselves—and second-order themes, which are higher-level categories developed by researchers to interpret and organize the data. These themes are then integrated into aggregate dimensions that form the basis for theory building.

In detail, the analysis is based on qualitative interviews with risk management experts from various financial institutions. Applying the Gioia Method, we systematically coded these interviews to extract first-order concepts, which are informant-centric terms and phrases derived directly from the language used by the interviewees. These first-order concepts capture the raw data and reflect the informants' experiences and perspectives.

We then synthesized these first-order concepts into second-order themes and aggregate dimensions, which are more abstract, researcher-centric constructs. These themes represent a higher level of conceptualization, integrating multiple first-order concepts into coherent theoretical categories. This process forms the basis of our relational diagram, which presents a grounded theory model for effective risk culture

assessment. Following Gioia and Chittipeddi (1991) and Corley and Gioia (1991), we adhered to the methodological steps of developing a comprehensive compendium of first-order terms, organizing them into second-order themes, and distilling these into overarching aggregate dimensions.

The analysis revealed eight critical second-order concepts: Clarity of Terminology, Optimization of Questionnaire Design, Reduction of Redundancies in Risk Culture Assessments, Adaptation of Item Order, Limitations of Questionnaires, Scale for Questionnaires, Time-Series Analysis of Risk Culture, and Illumination of Risk Culture from Different Perspectives. These concepts form a hierarchical framework for assessing the effectiveness of operationalizing risk culture.

The present study contributes to the literature by offering a novel theoretical model grounded in qualitative analysis that addresses the limitations of traditional risk culture assessment tools in the financial sector. Traditional methods often fail to capture the complexity and deep uncertainty inherent in financial risk culture. Deep uncertainty refers to situations where neither the probabilities of risks nor their potential outcomes are well understood, making it difficult for conventional assessment methods to be effective (Shortridge et al., 2017). Our findings provide a structured framework that enhances the understanding of the interdependencies among key elements of risk culture. The model not only enriches the theoretical discourse on risk culture but also offers practical insights for developing more effective assessment tools tailored to the complex needs of financial institutions (Bromiley et al., 2015). It contributes to a more nuanced understanding of risk culture from both a practitioner and a theoretical perspective.

The remainder of the paper is organized as follows: ‘Understanding the Bank’s Risk Culture’ reviews the relevant literature. The ‘Methodology’ section describes the research approach, highlighting the use of the Gioia Method for qualitative analysis. ‘Procedure and Experts’ provides an overview of the survey process and introduces the experts involved in the interviews. In ‘Study Results of the Gioia Method’, the key findings are presented, including the development of the grounded model. Lastly, the section ‘Conclusion, Further Research, and Limitations’ discusses the implications of the findings for both theory and practice, and concludes with recommendations for future research.

2 Understanding the Bank’s Risk Culture

Risk culture is widely understood as a subset of an organization’s overall culture, particularly in how it deals with risk-related behavior and decision-making (Australian Prudential Regulation Authority, 2016; Ring et al., 2016).; To situate our study, we briefly review how the literature has conceptualized risk culture in financial institutions. Table 1 provides a selection of key definitions, paraphrased from the original sources for clarity and comparability. A more comprehensive compilation of definitions can be found in Bockius and Gatzert (2023), but only selected examples are included here to maintain focus and consistency.

While definitions vary in wording, most share a focus on how collective values, norms, and behaviors shape risk-related decisions. Our study does not seek to develop

Table 1 Selected definitions of risk culture (Paraphrased)

Source	Definition/Emphasis
Institute of Risk Management (IRM), 2012)	Values, beliefs, knowledge, and understanding about risk shared by a group of people with a common purpose.
FSB (2014)	Norms, attitudes, and behaviors related to risk awareness, risk-taking, and risk management, influencing decision-making.
Kunz and Heitz (2021)	Shared experiences of individuals and norms, values, traditions, and attitudes that lead to particular activities related to the handling of risk and its consideration in decision processes.
Bockius and Gatzert (2023)	Based on a review of 83 publications, the authors identify five core dimensions of risk culture: tone from the top, risk awareness and comprehension, accountability, communication and escalation, and its manifestation in decisions and actions.

a new definition of risk culture but rather focuses on improving how such cultural elements can be assessed through survey design. We therefore treat these definitions as the conceptual backdrop for our methodological investigation.

While our study does not apply broader models of organizational culture in a systematic way, we acknowledge their conceptual relevance—particularly Schein’s (1990) layered view of culture—and reference them at selected points to contextualize the cultural dynamics that influence risk-related behavior.

3 Methodology

The data used in this study were collected during a broader research initiative on risk culture measurement. While a separate publication addresses the statistical validation of a risk culture scale, this paper exclusively focuses on non-item-specific qualitative feedback, which emerged independently of the scale content. The FSB framework summarizes accepted risk culture best practices (McConnell, 2013; Power et al., 2013) and incorporates different regulatory approaches (e.g., Chartered Institute of Internal Auditors, 2014; Senior Supervisors Group, 2009; IRM, 2012; Cass Business School and New City Agenda, 2014).

This paper builds on that previous work and uses the Gioia method to analyze non-item-specific expert feedback collected during the validation process. The Gioia method was used to derive first-order concepts and second-order themes as well as aggregated dimensions from the transcribed interviews.

Unlike classical content analysis, which typically focuses on frequency counts and inter-coder reliability across predefined categories, the Gioia method follows a structured inductive approach aimed at developing new theoretical insights. It emphasizes the progression from first-order concepts—grounded in informants’ own language—to second-order themes and aggregate dimensions constructed by the researchers. Importantly, the Gioia approach seeks not only to categorize data, but to build a grounded model that explains how key concepts are interrelated. This makes it particularly suitable for exploring complex, multi-level phenomena such as the design and implementation of risk culture assessments.

Following the identification of aggregate dimensions, we structured these into a conceptual framework broadly analogous to established phases of scale and survey

instrument development processes (cf. Groves et al., 2009; Boateng et al., 2018). Typically, these processes encompass distinct phases such as conceptualization and item development, instrument design and construction, and application and evaluation. Similarly, we structured our model into foundational concepts, intermediate concepts, and target perspectives. However, we explicitly adapted this structure to reflect the specific conceptual and practical requirements involved in effectively assessing risk culture in financial institutions.

The application of the Gioia method to non-item-specific feedback (NIR) from expert interviews represents an innovative approach to understand the issues relevant to a questionnaire-based risk culture assessment tool in practice. It systematically translates qualitative data into a grounded theory model, capturing the complexity and multifaceted nature of risk culture assessment tools. Gioia et al. (2013), alongside Langley (1999), underscore the Gioia method's efficacy in inductive theory building through precise data analysis and organization. This approach is particularly beneficial for risk culture research, fostering theory generation directly from data, rather than fitting it into pre-existing theories. By focusing on NIR, this study adopts a broader perspective on risk culture, encapsulating elements beyond the scale's initial items. This methodological choice enables the identification of emergent themes and deeper insights into how risk culture is shaped and perceived or can be evaluated. Notably, sources such as Gehman et al. (2018) and Reay et al. (2019) have highlighted the effectiveness of the Gioia method in generating theoretically relevant concepts through systematic qualitative data analysis. Analyzing NIR offers a comprehensive view of risk culture by illuminating the values, attitudes, and beliefs potentially underrepresented in experts' specific responses. As Alvesson and Kärreman (2007) suggest, interpreting qualitative data necessitates a critical reflection on the context and potential biases stemming from the subjective viewpoints of both researchers and respondents.

With this in mind, the primary research questions guiding this study are: (1) What are the key dimensions that shape the design and implementation of survey-based risk culture assessments in financial institutions? (2) How can the key dimensions identified be structured to guide the effective development and refinement of risk culture survey instruments?

Importantly, the resulting model is not intended to provide a new measurement instrument. Rather, it offers a conceptual structure and practice-oriented guidance that supports the more effective application of existing survey tools. Specifically, it highlights typical implementation challenges and organizes them into a structured framework. In doing so, the model helps researchers and practitioners to adapt existing instruments to institutional contexts, improve data quality, and strengthen the practical relevance of risk culture assessments.

4 Procedure and experts

The data analyzed for the present study, was part of an overarching scale development and validation process designed to measure risk culture in line with the FSB (Dürst and Kunz, 2025). This process involved two distinct phases: (1) qualitative

expert interviews and (2) statistical validation using a quantitative dataset. In Phase 1, expert interviews were conducted to ensure content validity as part of the scale validation. The selected experts represented a wide range of professional experience, spanning risk management, risk culture, the FSB framework, and related areas.

The interviews generated two types of qualitative data: (1) feedback directly related to the scale's items (item-relevant or IR feedback) and (2) broader, more general feedback on risk culture that was not tied to individual items (non-item-relevant or NIR feedback). While the IR feedback was utilized in the scale validation study, this paper focuses exclusively on the NIR feedback, offering deeper insights into the broader aspects of risk culture assessment. This type of feedback provides valuable insights into the underlying factors and context of risk culture, extending beyond the specific questionnaire items.

Details of the survey instrument used in the broader scale validation study (documented separately) are not part of this analysis. This paper does not rely on any item-level data or feedback. The current analysis focuses exclusively on non-item-specific insights that emerged independently of specific questionnaire content and reflect broader methodological concerns. We therefore consider this study analytically and conceptually distinct from the measurement instrument itself.

While the data originated from the validation context of a specific scale, the feedback analyzed in this study refers to broader methodological and contextual challenges in conducting survey-based risk culture assessments. The second-order concepts identified are not idiosyncratic to the particular scale under development but reflect recurring implementation issues observed across institutions. To strengthen the validity of these findings, we triangulated them with existing literature and included perspectives from a diverse panel of experts with different institutional backgrounds and roles.

In total, 21 expert interviews were conducted remotely, including one interview with two experts participating together. As two individuals declined to be recorded, and one interview involved two participants, a total of 18 transcripts were produced and analyzed, supplemented by written notes from the unrecorded interviews. The experts were provided with the questionnaire items, the FSB framework, and background information prior to the interviews. These interviews, which lasted between 30 minutes and 3 hours and 44 minutes, resulted in more than 33.5 hours of content, of which 28 hours were recorded and transcribed.

To increase transparency and allow readers to link statements to the interviewees' backgrounds, Table A1 (Appendix) provides an overview of all interviews, including anonymized information on institutional type, role, interview duration, and transcript availability. In aggregated terms, 14 interviewees represented banks, three came from building societies, and four had advisory roles; the sample included four board members, seven senior risk and compliance managers, and several risk culture specialists and consultants. This diversity of perspectives helped ensure that the feedback reflected both institutional breadth and practitioner depth. All interviews were conducted remotely via video call or telephone. Participants received an information sheet and provided explicit consent prior to recording. The procedures for data

collection, storage, and participant consent were approved by the university's data protection officer.

Each expert interview was analyzed individually to ensure that specific insights and nuances were accurately captured. During this initial analysis, the interviews were carefully annotated to highlight the most important findings in relation to risk culture and the improvement of risk culture measurement tools. These annotations serve to identify significant statements or themes directly from the participants, laying the groundwork for generating first-order concepts.

After individual analysis, 270 annotations were relevant for further aggregation into concepts. These were consolidated and reviewed to identify patterns and common themes across different interviews. This step involved aggregating similar annotations to form 18 distinct first-order concepts. These first-order concepts represent direct, observable data points and are derived from the specific terms and descriptions used by the experts during their interviews.

The first-order concepts were then examined and grouped into broader themes, resulting in 8 second-order concepts. This process involves abstracting the first-order concepts into a higher level of thematic aggregation.

We provide representative examples of the hierarchical data structure derived from expert quotations, illustrating the progression from raw data to first-order concepts, second-order themes, and aggregate dimensions in the context of risk culture assessment. To improve readability, the Table is presented in Appendix A (Table A2).

In order to facilitate interpretation and enhance the comprehensibility of our findings, a representative data structure matrix was created. This structure matrix visually maps the relationships between the second-order concepts, providing a clear and structured representation of the conceptual structure underlying risk culture as revealed by the expert interviews. These findings are discussed in the section Relational Diagram of Second-Order Concepts.

5 Study results of the Gioia method

The following sections present the eight aggregate dimensions identified through our analysis for assessing risk culture and the Relational Diagram of Second-Order Concepts.

The expert statements are paraphrased in the text to provide a clear and concise presentation of the key findings. To maintain the readability of the text, only a few quotations are included in the text. Additional relevant quotes are exemplified in Appendix A (Table A2) under each respective concept, allowing for a detailed understanding of the expert insights.

5.1 Clarity in Terminology as a foundation for effective risk culture measurement

The findings from expert feedback underscore the importance of clear definitions for key terms within risk culture assessments. Ambiguity in concepts such as risk appetite, leadership, and management was identified as a major barrier to consistent interpretation. As one expert noted:

The question would be how to define management again. Because it's definitely not just the board of directors, or maybe it is. I just don't know exactly. Whether it's management or the level below that. (Interview 18)

Similar concerns were raised about the meaning of “remuneration” and “risk appetite,” especially in settings where no explanatory guidance is provided to respondents. Experts emphasized that such ambiguity persists even in mature organizations and can significantly undermine both the interpretability and comparability of survey responses.

This observation aligns with longstanding insights from the literature. For example, the term “risk” itself is understood in various ways across organizations—either as downside exposure or as symmetrical variance from performance expectations (March & Shapira, 1987; Holton, 2004). Moreover, assessments often fail to specify whether they refer to strategic, operational, or compliance risks (Power, 2009; Mikes & Kaplan, 2013), contributing further to conceptual confusion.

Clarity in terminology should therefore not only concern item phrasing, but also include a shared understanding of the broader constructs being assessed. This is essential for aligning the mindset of all involved parties and for improving the reliability and practical relevance of the assessment process. As Trimpop (1994) and Holton (2004) emphasize, risk-related concepts are inherently complex and subjective, necessitating precise definitions. MacKenzie et al. (2011) and Boateng et al. (2018) likewise underline that construct clarity is foundational to validity and reliability in measurement development—an imperative that holds especially true in the sensitive context of risk culture.

In the context of risk culture assessments, terminological clarity carries particular weight because key terms are not only survey constructs but also regulatory reference points (e.g., “risk appetite” in FSB or EBA guidance). Ambiguity in these terms can therefore undermine both internal alignment and external credibility. To address this, practitioners may complement surveys with preparatory steps such as the development of shared glossaries, i.e. explanatory notes for ambiguous terms. Such measures help ensure that assessment results are not distorted by inconsistent definitions, thereby improving both reliability and regulatory credibility.

5.2 Reduction of redundancies in risk culture assessments

Several experts pointed out that redundancies in data collection often stem from overlaps with existing regulatory processes. When risk culture surveys replicate questions or topics already addressed in audits or compliance reviews, the added value becomes questionable. One expert illustrated this point succinctly:

If I ascertain whether something is prepared systemically and at the same time this is specified for the institutions from the MaRisk guidelines, this is already somewhat redundant at this point. The question would then be whether this information, which is now necessary here and which is also needed here, is already available elsewhere when you go into operational practice, or whether it has to be available. (Interview 10)

Such duplication not only leads to inefficiencies and survey fatigue but may also obscure more meaningful aspects of risk culture by overemphasizing formal structures. To address this challenge, structured mapping techniques can help identify overlaps. Analogous to ‘assurance mapping’ in risk management (Chen et al., 2016), an ‘assessment mapping’ exercise could systematically compare survey items with existing regulatory reviews, audit procedures, or internal monitoring in order to highlight redundancies. Such mapping makes it transparent where survey items merely duplicate regulatory checks and where they provide genuinely new insights into cultural perceptions and behaviors. This approach directly addresses the concerns raised by experts by providing a structured method to avoid unnecessary duplication, while at the same time preserving the unique value that culture-focused assessments add beyond regulatory checks.

Experts also emphasized the value of methodological plurality. Complementing surveys with qualitative approaches such as interviews or participatory workshops can help capture behavioral nuances and group dynamics that questionnaires may miss. This call for methodological plurality is echoed in the literature: Renn et al. (2022) emphasize that systemic and cultural risks require multidimensional assessment frameworks, combining structured metrics with open-ended exploration.

Ultimately, the choice of assessment methods should be guided by contextual relevance. A modular, case-sensitive approach can help reduce redundancies, increase practitioner buy-in, and enhance the strategic utility of the findings. This not only conserves resources but also sharpens the focus on what truly matters in a given institutional environment. While minimizing operational redundancies is essential for efficiency and relevance, we acknowledge that some conceptual overlap in survey items may be warranted to support measurement reliability—a point further discussed in Sect. 7.

5.3 Optimization of Questionnaire Design in risk culture assessments

Experts emphasized that lengthy or poorly structured questionnaires can negatively affect response quality. Fatigue, cognitive overload, and perceived irrelevance were frequently mentioned. As one expert noted:

I read through the questions, I think I got through to question 80 quite well and then there is this point, then of course there is an effect, a fatigue effect. (Interview 14)

To mitigate such effects, experts recommended segmenting longer instruments, distributing items across different roles, or conducting surveys in multiple stages. This reduces response burden and increases relevance. One expert suggested:

[...] I would rather suggest trying to possibly segment the questionnaire, or only give certain parts of the questionnaire to specific people, so that you don't overwhelm them and to achieve as high a response rate as possible. (Interview 16)

Beyond length, the clarity and accessibility of wording were central concerns. Experts advocated for simpler language and more intuitive phrasing—especially when abstract constructs such as “risk culture” are involved. While examples can help clarify meaning, they may introduce bias or inapplicability across departments. As one expert warned:

[...] The problem with examples is always that you immediately focus on them and then someone says, but that’s not the case here. That’s difficult. [...] (Interview 18)

To further optimize the design, experts suggested tailoring the questionnaire to the specific needs of the organization. Pre-survey interviews can assess the maturity of an organization’s risk culture, allowing for the customization of the questionnaire. This can be done by implementing pre-filtering questions to eliminate irrelevant sections.

And then, if you agree with a company to have them go through the questionnaire and they say, ‘But we don’t have any performance evaluation,’ well, of course, you have to exclude those questions. That’s technically very simple nowadays. (Interview 18)

Experts also recommended the use of dynamic question presentation techniques, where certain questions are hidden or revealed based on prior answers. This approach minimizes unnecessary questions while maintaining the focus on relevant areas.

Finally, the survey should account for differences in organizational size and structure. A ‘one-size-fits-all’ approach may not work for all companies, as noted by an expert:

For example, if you ask the colleagues in the retail banking about financial risks, they won’t know, simply because they don’t understand the financial metrics that express financial risks. They are not familiar with them at all. So, it’s a bit, let me put it this way, I’m not sure if this one-size-fits-all approach will work systemically, in the sense of producing reliable results. (Interview 6)

The expert feedback on optimizing questionnaire design is well-supported by existing literature. For example, research on questionnaire length suggests that splitting long questionnaires can prevent respondent fatigue without compromising data depth (Peytchev & Peytcheva, 2017). Studies have shown that lengthy questionnaires increase the risk of straight-line responses, where participants give identical answers to consecutive items, thereby compromising data quality (Herzog & Bachman, 1981; Burchell & Marsh, 1992). Reducing the number of items, as demonstrated by Moran et al. (2001), can enhance efficiency without sacrificing reliability. Staffini et al. (2022) also emphasize that item reduction techniques improve respondent engagement by eliminating redundancies.

In web surveys, Toepoel et al. (2009) found that increasing the number of items per screen leads to higher nonresponse rates and negative evaluations of the survey layout. Simplifying language helps to reduce cognitive load and improve comprehen-

sion (Warnecke et al., 1997). Smyth et al. (2009) found that clear, motivating introductions to questions improve response quality in surveys, while Ganassali (2008) highlighted that thoughtful design, including short questionnaires and direct wording, significantly affects response quality.

Tailoring questionnaires to the specific needs and maturity of an organization is also important. Dolnicar et al. (2013) demonstrated that aligning survey questions with respondents' profiles improves the relevance and effectiveness of the survey. Similarly, Crispim et al. (2019) show that in the context of project risk management, different maturity levels and project complexity influence the choice of practices. Mature organizations use more targeted and comprehensive practices, leading to better project outcomes. These findings, in relation to our context, imply that it is important to tailor the number and complexity of questions based on the organization's maturity in handling risk culture assessments. The development of tailored questionnaires that consider specific working conditions and organizational contexts has been shown to improve the relevance and applicability of the collected data (Nielsen et al., 2014).

In summary, optimizing risk culture questionnaires requires balancing comprehensiveness with simplicity, ensuring the questionnaire is manageable and aligned with the organization's specific needs. By segmenting the questionnaire, simplifying language, tailoring content, and using dynamic question techniques, organizations can improve response rates and collect more accurate data.

These recommendations, however, must be interpreted in light of broader methodological tensions. While several experts advocated for item reduction and streamlined formats, literature in scale development cautions that eliminating seemingly redundant items can reduce internal consistency and obscure latent constructs (MacKenzie et al., 2011; Boateng et al., 2018). Clifton (2020) further notes that pressure to shorten instruments often shifts emphasis from construct validity to superficial reliability. Our findings suggest that rather than eliminating redundancy categorically, practitioners should evaluate the informational function of each item within the specific risk culture context.

A similar trade-off emerges between tailoring items to organizational subgroups and ensuring cross-role comparability. While segmentation and filtering increase relevance and reduce response burden, they may undermine standardization and comparability across departments—particularly in regulated environments that require consistent reporting. As noted by Dillman et al. (2014), tailoring surveys to specific respondent groups can improve engagement and perceived relevance. Research by Manski and Molinari (2008) further demonstrates that item sequencing affects not only dropout rates and data completeness but also how respondents interpret subsequent questions. However, such adaptations may reduce standardization and complicate the aggregation of results across organizational units. The optimization of questionnaire design should therefore be approached as a context-dependent calibration of scope, depth, flexibility, and sequence—rather than as a mere reduction in item volume.

These findings are particularly relevant for risk culture assessments, because unlike generic surveys, they deal with constructs that are both formalized in regulation and experienced in daily behavior. If questionnaires are too long, too formal, or

misaligned with roles, they risk capturing only compliance-oriented answers while missing the cultural dimension. Optimizing design therefore goes beyond generic survey practice: it is a way of ensuring that cultural phenomena such as tone from the top, openness in communication, or incentive effects are captured as lived experiences across different levels of the organization.

In practice, this can be implemented in different ways. For example, instead of asking all employees a compliance-oriented question such as “*Our institution has a documented risk appetite statement*”—which tends to elicit uniform agreement—practitioners can reframe it into a perception-based item: “*I understand how the risk appetite statement affects my daily decisions.*” This turns a formal requirement into a culture-relevant measure. Another approach is to tailor items where specialized knowledge is required. For instance, questions about complex financial risk metrics may be directed at treasury or risk management staff, while employees in retail banking receive items focused on customer-facing risk behavior. At the same time, culture-relevant constructs such as “tone from the top” remain included across all levels, since differences in perception between leaders and staff are themselves diagnostic of risk culture.

5.4 Adapting item order

Experts provided feedback on the sequence of elements in the risk culture assessment. Some suggested that simple topics should be addressed first, followed by more complex elements. This gradual approach could make it easier for participants to engage with the assessment.

Maybe it’s again an issue with the sequence, so that somehow—well, we had the specific questions later on. Perhaps it really makes sense to bring those to the beginning and then ask about the topic afterward. Then you simply don’t have to do much here, just a bit, for example, as you just said, give examples, but you’ve already talked about it beforehand, and maybe people are already more engaged and can answer more easily. (Interview 12)

In contrast, others suggested starting with the most important elements and working toward less critical topics. These approaches reflect different strategies that can influence the effectiveness and focus of the assessment process. Both methods have their merits and can be selected according to the specific objectives and constraints of the risk assessment.

[...] So, regarding the sequence, I think, in my opinion, that tone from the top is very important, and various elements like compensation, whistleblowing, etc., should actually reflect, or should indeed mirror, the Tone from the Top. [...] So, I would also say, especially when it comes to the top-down process, I find tone from the top actually quite good. If we assume that there is a defined or desired risk culture, then asking to what extent the executive board or the top of the organization or company actually embodies or talks about it is not illogical, I think. (Interview 16)

A structured progression from simple to complex allows for a step-by-step approach, where participants gain confidence and understanding of the topic. This method ensures that respondents are better prepared to tackle more difficult or abstract questions later in the assessment. Prioritizing the discussion of the most important elements first ensures that key aspects of risk culture are addressed when participants are most attentive and focused.

Research shows that the order of questions can influence respondents' answers. For example, the concept of anchoring suggests that initial information can heavily influence subsequent judgments and decisions, creating a bias toward the initial information presented (Gehlbach & Barge, 2012). Additionally, there are contextual effects, where the placement of questions influences subsequent responses by creating contrast and consistency effects and changing perceptions of the importance of topics (Moore, 2002). Prioritizing discussions by importance ensures that the most critical elements of risk culture are addressed when all participants are most attentive and engaged (VonderHaar et al., 2019).

In the context of risk culture assessments aligned with the FSB Framework, 'Tone from the Top' is typically regarded as a key component. Accordingly, starting with this element may be appropriate. For other instruments, however, adjusting the order of elements may be more suitable depending on the specific goals and context of the assessment.

In this sense, adapting item order should not be seen as a stand-alone dimension but as a subdimension of questionnaire optimization. Its relevance for risk culture assessments lies in ensuring that key cultural constructs—such as tone from the top—are sequenced in a way that elicits meaningful perceptions across hierarchical levels, rather than being diluted by generic survey order effects. For instance, some institutions may choose to begin with items on leadership behavior to capture unfiltered perceptions of tone from the top, before moving to technical or procedural aspects of risk culture.

5.5 Limitations of Questionnaires in risk culture assessment

Experts emphasized that questionnaires often lack the depth to uncover the full range of cultural dynamics in financial institutions. As one expert noted:

Have you ever looked into the approach of the Dutch Central Bank in the context of your research? [...] Their approach goes much further. For example, they examine the work within internal committees: How does the executive board interact with each other? Are things also discussed controversially? They also examine the collaboration between the executive board and the supervisory board. [...] They discovered that if decisions are made and everyone agrees with everything, then the issue of risk culture may not be well-addressed. (Interview 7)

Additionally, experts noted that framework orientation and abstraction could be a problem. The use of frameworks such as the Financial Stability Board (FSB) guidelines could be too abstract, lacking concrete standards and not considering specific

banking risks or national characteristics. Some expert feedback indicates dissatisfaction with the FSBs guidance, which is seen as too procedural and not reflective of the deeper, cultural layers of financial institutions. This reliance on a broad framework could benefit from a more tailored approach that considers national implementations and specific bank risks. There may also be a significant gap in how risk culture definitions derived from such frameworks apply in practical scenarios, often being too abstract to determine actual cultural manifestations.

Respondent variability was also mentioned as a limitation. According to the experts, responses may be highly dependent on the personal perspective of the respondent, which may not accurately reflect the true risk culture.

Yes, it's all good. The questions are good, no. So I don't want to criticize the quality of the questions, I think they're essential and right. It's just that because they're essential and right, I have justified doubts about getting honest and realistic answers. Because, of course, they will already show massive blatancy in many companies. I am convinced of that. (Interview 2)

These expert observations are consistent with findings from existing literature. Guldenmund (2007), for instance, points out that questionnaires often capture only superficial compliance aspects, neglecting the more profound cultural elements essential for a thorough understanding of risk culture. Similarly, the importance of adapting questionnaires to local conditions has been emphasized in previous studies. Bolton and Tang (2002) for example developed and validated a new method for cross-functional and gender-specific assessment that avoids the main problems of existing methods. They found that contextually adapted questionnaires based on locally relevant tasks allow for better cultural fit and provide reliable results. This shows that abstract frameworks are often insufficient to capture actual cultural manifestations.

Existing literature suggests that responses can be significantly influenced by individual perceptions and experiences, which limits the validity of the results. Guillemette et al. (2012) show in a study the importance of individual judgements in determining risk tolerance and the possible deviations from the actual risk culture. Self-report questionnaires may lead to misclassification of risk tolerance, potentially resulting in unsuitable investment decisions (Lucarelli et al., 2015). Brenot et al. (1998) discuss the variability in risk perceptions across different cultural backgrounds, highlighting how questionnaire designs must account for cultural biases and personal interpretations in their structure to enhance validity and reliability.

Our point is not framework-specific. While several interviewees referenced the FSB, the underlying limitation is not unique to that guidance. Principle-based frameworks—such as those of the FSB (2014), De Nederlandsche Bank (DNB), (2015), or Committee of Sponsoring Organizations of the Treadway Commission (COSO), (2017)—are designed to be tailored to institutional contexts. Our findings are consistent with that view: the limitation we highlight pertains to the use of questionnaires as a single method. Translating such high-level principles into observable, context-specific behaviors typically requires a multi-method design (e.g., observation of decision forums, targeted interviews, incident narratives) alongside surveys. Accordingly, we interpret Limitations of Questionnaires not as a critique of any particular framework,

but as a methodological reminder that principle-based guidance must always be operationalized through such multi-method approaches.

5.6 Scale for Questionnaire

The design and structure of response scales in questionnaires were another focus of expert feedback. Four recurring themes emerged: the need for neutral or differentiated options, the use of bipolar scales, the challenge of regulatory framing, and the importance of clear, content-rich response labels.

First, several experts noted that traditional Likert scales (e.g., ‘strongly agree’ to ‘strongly disagree’) may oversimplify respondents’ perceptions. Neutral categories and more nuanced gradations (e.g., ‘partially true’) can help reflect the complexity of organizational roles.

I couldn’t answer all the questions. Sometimes I also miss a neutral category because, of course, the market area is different from risk management. (Interview 19)

Second, some experts recommended using bipolar scales with a conceptual optimum in the middle—particularly for assessing whether processes are too weak, too strong, or balanced.

There are now also response categories that are a bit more evaluative. If you potentially have the optimum in the middle, and the further you go out, the less optimal it becomes. [...] (Interview 16)

Third, Likert-scale items tied to regulatory requirements may elicit socially desirable responses. Experts noted that these items often generate uniform agreement, regardless of actual perceptions.

This is all a regulatory requirement. So, but, in this respect, you can formulate it like this. But you won’t find anyone who answers the question with “no”. (Interview 3)

Finally, the clarity of response labels was seen as crucial. Some experts recommended using answer texts that explicitly describe what each level (e.g., good, average, poor) means in practice.

I would indeed make the answers more concrete [...] to at least remove the ambiguity that people might have different understandings of what their criteria mean. (Interview 5)

The expert feedback on the design and structure of response scales is supported by existing research, which has explored the effectiveness of different scale types in enhancing data granularity and improving respondent satisfaction. Preston and Colman (2000), for example, argue that expanding Likert scales can lead to more

nuanced data, although Cheung and Mooi (1994) and Simms et al. (2019) question the overall effectiveness of such scales. Notably, Kulas and Stachowski (2009) found that middle-category endorsements in Likert scales often reflect nuanced positions, offering deeper insights into respondents' attitudes.

Bipolar scales, as suggested by the experts, are also effective for capturing balanced evaluations. These scales allow respondents to indicate where their assessment falls between two opposing concepts, providing a clear picture of whether conditions meet optimal standards. Malhotra et al. (2009) highlight that bipolar Likert scales are particularly useful for evaluating concepts that involve balance and can be combined with branching questions for greater precision. The choice between unipolar and bipolar scales should be determined by the specific research context. Vautier and Pohl (2009) argue that unipolar scales are better suited for measuring the intensity or frequency of single dimensions, whereas bipolar scales are more effective for assessing opposing concepts.

The inclusion of a neutral category, as discussed by the experts, has also been examined in the literature. Dubois and Burns (1975) suggest that while neutral responses can indicate ambivalence or indifference, they are not necessarily more ambiguous than other response categories.

Finally, the literature provides various recommendations on the number of response options in scales. Simms et al. (2019) found that psychometric precision decreases with fewer response options (2 to 5), though criterion validity does not necessarily follow this trend. Preston and Colman (2000) found that scales with 7 to 10 response categories provided the most reliable scores. Krosnick and Presser (2010) similarly observed that scales with two or three points have lower reliability than Likert-type scales with five to seven points, a conclusion echoed by Boateng et al. (2018) in their review of response scale design.

In the context of our framework, the design of response scales is not an end in itself but a methodological means to improve the quality and interpretability of survey data. Scales shape how abstract cultural constructs—such as tone from the top, openness, or accountability—are translated into measurable perceptions, but they do not constitute substantive outcomes of risk culture assessment. We therefore treat “Scale for Questionnaire” as a methodological element situated in the design phase of assessments, and thus as a subdimension of questionnaire optimization.

For risk culture assessments, these insights imply that scale design cannot be treated as a purely technical detail but must be aligned with the cultural and regulatory context of financial institutions. In practice, organizations may choose to introduce neutral or differentiated categories where roles and functions differ substantially, employ bipolar scales to capture under- and over-emphasis of cultural dimensions, and provide clear, content-rich response labels. Finally, providing short explanatory notes for response labels (e.g., clarifying what “rarely” means by specifying “once or twice per year,” or what “frequently” means by specifying “weekly”) helps respondents align their interpretations, reducing ambiguity and improving data quality. In this way, scale design supports both methodological robustness and the practical relevance of risk culture assessments.

5.7 Time-Series Analysis of Risk Culture: a dynamic perspective

The experts also emphasized the need to look at risk culture over time. A snapshot cannot capture the dynamics of risk culture within an institution. They argued that understanding risk culture through a longitudinal analysis provides a more robust and meaningful assessment.

What is difficult and also not helpful in my view is an evaluation at a single point in time. Because there is no absolute good or bad, we have agreed on that, and it is only a matter of whether it is developing in the direction that you ultimately want. (Interview 8)

This approach is consistent with recent academic work, which suggests that analyzing risk culture over time can reveal patterns of adaptation and resistance within an organization. For example, Chen et al. (2023) demonstrate the efficacy of time-series methods in financial risk prediction, highlighting how periodic changes in financial data can reveal underlying patterns in risk culture that static snapshots may miss. Similarly, Caporin and Storti (2020) emphasize the importance of financial time-series analysis in understanding data over time, aligning with the argument that temporal analysis can uncover organizational adaptation patterns. While these studies focus more on risk prediction, the importance of observation over time is also transferable to the assessment of risk culture within financial institutions. Power et al. (2013) explore the challenges in managing and understanding risk cultures within financial organizations, highlighting the dynamic nature of these cultures.

To strengthen the practical relevance, repeated assessments of risk culture can serve as a monitoring tool to evaluate the effectiveness of measures aimed at cultural change. Follow-up surveys can indicate whether communication initiatives by senior leaders have permeated to middle management and front-line staff, or whether incentive structures are being perceived differently after reform. In this sense, time-series analysis is not only about methodological rigor but also provides actionable insights: it allows institutions to track whether interventions are effective, at which organizational levels cultural shifts occur, and where resistance persists. In practice, repeated surveys can be scheduled annually or biannually and complemented by pulse checks on specific themes (e.g., openness in communication). This enables institutions to identify whether changes, such as leadership training or incentive adjustments, are perceived differently at the board, managerial, and staff levels.

5.8 Illuminating Risk Culture from Different Perspectives

To obtain a comprehensive perspective on risk culture, our questionnaire distinguished between three key components: 1) employee behavior, which reflects how employees act within the organization; 2) leadership behavior, which encompasses how leaders conduct themselves; and 3) the systemic component, which pertains to the processes embedded in organizational procedures. This tripartite division of the components in the risk culture assessment has been confirmed by expert feedback. Beyond the division into components, the experts' feedback suggests that it is impor-

tant to examine the risk culture from different perspectives. For example, one expert said that differences in responses between these components to the same question may provide more insightful information than the responses themselves.

They looked at what the leaders say about certain things and what the employees say. And thus, the absolute level of the responses was not important, but rather the discrepancy between the different target groups. For example, if the leaders say, “We practice risk culture daily, and we are convinced that we are training and developing employees effectively towards risk culture”, but the employees say, “This doesn’t really reach us. We don’t quite understand what is expected of us.” Then you already have such a discrepancy, and this discrepancy alone indicates a need for action. And it’s not about right or wrong. Do you understand what I mean? (Interview 11)

His suggestion is rooted in the idea that discrepancies in attitudes and behaviors across different target groups are more indicative of the underlying risk culture dynamics than homogeneous responses. To effectively extract these discrepancies, the expert proposed that as many questions as possible should explore all three dimensions from a perception-based angle. This approach aims to highlight contrasts that might exist between how these groups perceive risk culture elements, potentially uncovering areas of misalignment or differing priorities.

However, this recommendation to include as many questions as possible may seem contradictory to the earlier feedback regarding the risk of overloading respondents and inducing survey fatigue. As discussed in a previous section, experts emphasized the need to strike a balance between depth and clarity in questionnaires to avoid overwhelming participants with too many items, which could compromise the quality of the responses (see section on Optimization of Questionnaire Design). Ensuring that the questionnaire remains manageable while capturing the necessary depth of information is key to its effectiveness.

The importance of examining organizational culture from multiple perspectives is supported in the literature. Aarons et al. (2017) found that discrepancies between supervisor and provider ratings of transformational leadership in mental health clinics were associated with a more negative organizational culture. Similarly, Hall et al. (2020) explored how dependencies between functional entities within large, complex organizations can propagate risks throughout the system. This idea aligns with the notion that differing perceptions within an organization’s components can impact the overall risk culture. Weyman and Clarke (2003) further examined how organizational roles influence risk perceptions in coal mines, finding significant differences in perceived risk among employees, particularly those closer to hazards. These studies support the concept that discrepancies in risk perceptions across different groups within an organization can provide valuable insights into its risk culture.

In the specific context of risk culture, prior research has already emphasized the diagnostic value of comparing perspectives. Quail (2012) stresses that differences between the target appetite, the exhibited appetite, and staff perceptions of appetite are critical signals of cultural (mis)alignment rather than mere measurement noise. Similarly, Sheedy and Griffin (2018) show empirically that divergence between staff

and management perceptions of risk culture reveals areas where cultural messages fail to cascade. These studies resonate with our expert feedback and underscore that perception gaps are substantive data points, not artifacts.

From a practical perspective, discrepancies can be used proactively as a basis for dialogue and intervention. Institutions can, for example, present survey findings disaggregated by hierarchy or function to leadership teams, highlighting where misalignments between leaders and staff perceptions are most pronounced. Complementary workshops or focus groups can then be used to explore why these gaps exist and what organizational measures may close them. In this sense, discrepancies become not only a diagnostic tool but also a starting point for targeted cultural change initiatives, such as leadership communication programs or team-level discussions to align expectations and practices.

5.9 Methodological trade-offs and design decisions in risk culture surveys

Our findings reveal three overarching methodological considerations that are particularly salient in survey-based risk culture assessments.

The first balance concerns comprehensiveness and manageability. Long instruments with overlapping items can improve reliability and capture the complexity of cultural phenomena, but they also risk fatigue, disengagement, and survey burden. Our interviewees repeatedly highlighted redundancies—both psychometric (overlapping items for reliability) and procedural (duplication of audit or compliance data). The challenge is to distinguish between justified and unjustified overlaps. Literature on scale development stresses that some redundancy is often necessary to ensure construct validity and internal consistency (Boateng et al., 2018; MacKenzie et al., 2011). In line with this, Clifton (2020) notes that efforts to streamline questionnaires frequently undermine conceptual depth and measurement quality.

A second methodological tension involves terminological clarity and interpretive flexibility. Precise and consistent terminology is essential for shared understanding and comparability across groups. Yet overly rigid definitions may constrain respondents and obscure differences in how concepts such as “risk” or “management” are understood in practice. Prior research highlights that even foundational terms like “risk” are conceptualized differently across organizations: sometimes narrowly as downside exposure, other times as any deviation from expected performance (March & Shapira, 1987; Holton, 2004). Furthermore, ambiguity often surrounds which types of risk are being referred to in different contexts (Power, 2009; Mikes & Kaplan, 2013). Too much definitional rigidity risks reducing the diagnostic value of discrepancies in perception, while too little clarity leads to ambiguity and noise. The methodological task is to ensure that terms are clear enough for aggregation but open enough to capture diverse lived experiences.

The third trade-off concerns standardization versus contextual tailoring. Our interviews showed that sequencing and response formats influence how respondents interpret and prioritize questions. While standardization improves comparability across units and time, tailoring allows surveys to resonate with specific organizational contexts, maturity levels, and roles. This reflects findings by Manski and Molinari

(2008), who show that item order and skip logic can reduce cognitive burden but may also introduce sequencing bias.

Taken together, these methodological balances illustrate that there are no fixed design rules for risk culture surveys. Instead, effective assessment requires context-sensitive calibration across these three tensions: comprehensiveness versus manageability, clarity versus flexibility, and standardization versus tailoring. The value of our framework lies not in prescribing a single “optimal” solution, but in providing orientation for practitioners and scholars to navigate these methodological choices in ways that best fit their institutional environment and strategic objectives.

5.10 Methodological positioning of the eight second-order concepts in relation to risk culture frameworks and survey design literature

Building on the overarching methodological balances outlined in Sect. 5.9, this section positions the eight second-order concepts derived from our study within the broader landscape of risk culture frameworks and survey design literature. While some concepts (e.g., terminology, redundancy, questionnaire design) directly extend the methodological issues discussed above, others (e.g., time-series analysis, multi-perspective triangulation) introduce novel methodological considerations for risk culture assessment.

Existing frameworks for understanding and shaping risk culture—such as the FSB (2014) principles, Sheedy and Griffin (2018) behavioral dimensions, or Palermo et al. (2017) work on cultural ambiguity—provide conceptual guidance on what constitutes a sound risk culture. However, they offer limited orientation on *how* survey-based assessments should be designed to validly and reliably capture such culture within financial institutions. As a result, critical decisions about instrument structure, item phrasing, response formats, or sequencing are often made without a clear methodological rationale—despite their potential influence on data quality. Our framework aims to address this gap by structuring practitioner-derived insights and aligning them with methodological literature on survey design, thus offering practical guidance for balancing trade-offs (e.g., redundancy vs. fatigue, clarity vs. openness) that are particularly salient in regulated, complex organizational settings.

The following discussion situates the eight second-order concepts derived in our study within this broader methodological and conceptual landscape, highlighting where they complement, refine, or challenge existing models.

For example, *Clarity in Terminology* echoes the FSBs high-level emphasis on fostering a shared understanding of core risk concepts across the institution (Financial Stability Board FSB, 2014). While Sect. 5.9 has already highlighted the balance between clarity and interpretive openness, our findings emphasize its methodological relevance by showing how ambiguous terms—such as “management” or “risk appetite”—can undermine construct validity in survey-based assessments. This observation is consistent with prior research highlighting that even foundational concepts like “risk” are interpreted inconsistently across organizations (March & Shapira, 1987; Holton, 2004), and that terminological clarity is essential for valid and reliable measurement (MacKenzie et al., 2011).

Beyond terminological issues, our framework also addresses a second set of challenges that pertain to the practical design and institutional embedding of survey instruments.

Similarly, *Reduction of Redundancies* extends the discussion in Sect. 5.9 by distinguishing between psychometric redundancies (deliberate item repetition to ensure consistency Boateng et al., 2018) and procedural redundancies (duplications due to overlapping regulatory and internal data collection). This dual perspective is largely absent in existing frameworks but is highly salient in financial institutions operating under tight compliance regimes. By recognizing both sources of redundancy, our framework supports a more strategic review of which overlaps are necessary and which can be streamlined to enhance efficiency without compromising measurement quality.

Optimization of Questionnaire Design and Adapting Item Order reflects debates known from the survey literature (e.g., Dillman et al., 2014; Manski & Molinari, 2008). While Sect. 5.9 has already emphasized the trade-off between standardization and tailoring, our findings underscore the practical implications for risk culture surveys: sequencing, contextualization, and respondent load decisively shape data quality. This perspective is largely absent from prior risk culture models, which often treat surveys as neutral data collection tools rather than embedded organizational interventions.

Third, our framework includes concepts that are novel in the context of risk culture assessment. The idea of Time-Series Analysis of Risk Culture, for instance, responds to existing critiques that risk culture is often assessed statically, despite its dynamic nature (Power et al., 2013; Chen et al., 2023). While temporal measurement is common in behavioral finance and organizational learning, it has rarely been explicitly integrated into risk culture frameworks. This addition reflects a growing awareness among practitioners that trajectory matters more than snapshot in assessing cultural alignment.

Illuminating Risk Culture from Different Perspectives is similarly distinctive. While prior frameworks acknowledge the multi-level nature of culture (e.g., Schein, 1990), our findings show that systematic triangulation—especially between leadership, employee, and systemic perceptions—can reveal misalignments that homogeneous scores would obscure. This approach is consistent with findings in organizational psychology (Aarons et al., 2017) but is rarely applied systematically in financial risk culture diagnostics.

Fourth, several of the eight concepts directly reflect tensions specific to the risk culture domain. Unlike general organizational culture, risk culture is often embedded in highly regulated, politically sensitive, and reputationally exposed environments. This creates unique challenges in measurement: survey fatigue is amplified by compliance overlaps, and surveys often struggle to capture the deeper cultural dynamics underlying formal responses (cf. Guldenmund, 2007, ; Brenot et al., 1998). Our concept Limitations of Questionnaires captures these tensions and advocates for multi-method triangulation—something the FSB framework acknowledges but does not methodologically specify.

Finally, Scale for Questionnaire addresses the psychometric microstructure of risk culture surveys. While scale construction is well-covered in psychometric literature

(e.g., Preston & Colman, 2000; Simms et al., 2019), its application to risk culture remains underexplored. Our findings suggest that response format choices affect not only measurement quality but also the psychological safety of respondents, especially when dealing with normative or regulatory questions.

In sum, the eight concepts identified in our framework do not contradict existing models but enrich them through a design- and practitioner-focused lens. They emphasize that assessing risk culture is not a neutral or purely technical endeavor but a process that is embedded in, and shaped by, organizational, psychological, and regulatory realities. This perspective expands the current discourse and offers a more adaptable, diagnostically useful foundation for future risk culture assessments.

While our eight second-order concepts aim to improve the validity, depth, and relevance of risk culture assessments, their implementation is not cost-neutral. Several recommendations—such as tailoring questionnaires, segmenting target groups, or adopting time-series data collection—require not only conceptual effort but also substantial internal resources. These include staff time for survey participation, coordination with stakeholders (e.g., works councils, compliance, data protection), communication and promotion efforts, and follow-up processes for analyzing and reflecting on the results. In regulated environments, such investments may be justified by compliance expectations or reputational considerations. More generally, robust risk culture has been shown to enhance organizational resilience and performance (Fritz-Morgenthal et al., 2015; Bianchi et al., 2021; Sheedy & Griffin, 2018). To actively develop and sustain such a culture, however, systematic assessment is essential—serving as a diagnostic foundation for targeted interventions and continuous improvement. We therefore encourage organizations to treat our framework as a flexible guide rather than a prescriptive model, adapting its elements to their strategic priorities and operational capacities.

While our study focuses on the methodological design of risk culture assessments, the utility of any such assessment ultimately depends on the clarity and robustness of the underlying content dimensions being measured. Our second-order concepts do not replace substantive components of risk culture—such as tone from the top, accountability, communication, or incentive alignment—but aim to improve how such elements are operationalized, interpreted, and triangulated through survey design. In this sense, our findings complement existing taxonomies and frameworks rather than duplicating or redefining them.

5.11 Functions and contexts of risk culture assessments

Beyond methodological design, the effectiveness of risk culture assessments also depends on who conducts them and under which circumstances survey approaches are most suitable.

The Three Lines Model (Institute of Internal Auditors IIA, 2020) distinguishes governance roles. Operational management (first line) is responsible for objectives and risk in daily business. Risk and compliance functions (second line) provide expertise, support, monitoring, and challenge. Internal audit (third line) offers independent assurance to the governing body. Effective coordination across all lines is essential, with internal audit regarded as the most independent. Arnaboldi and Vasciaveo

(2017) suggest that internal audit can effectively assess risk culture, especially in organizations with mature governance systems. Sinha and Arena (2020) further show that stakeholder perspectives on internal audit's role diverge significantly. While regulators and implementers tend to view it as a control mechanism, consultants and professional bodies emphasize its enabling potential, including behavioral assessments and staff engagement. This heterogeneity underlines that the responsibility for conducting risk culture assessments is not fixed but shaped by institutional context and stakeholder expectations. External actors, such as consultants or supervisors, offer neutrality and comparative benchmarks. However, they may lack contextual knowledge and face acceptance issues if assessments are not embedded in internal dialogue. In practice, a combination of internal and external perspectives may help balance methodological rigor and organizational relevance.

While the choice of function determines ownership and credibility, the situational context shapes the suitability of survey approaches. Surveys are particularly suitable when organizations aim to capture broad perceptions across hierarchical levels and functions in a standardized way, including the identification of subcultures. However, their limited scalability, along with the influence of national and institutional contexts on risk perception and communication (Bockius & Gatzert, 2023; Carretta et al., 2017), constrains their usefulness for benchmarking and cross-organizational comparison. Surveys should therefore not be regarded as a stand-alone instrument, but as one element in a broader, multi-method assessment strategy that triangulates quantitative findings with interviews, workshops, or document analyses. For practitioners, this implies that assigning responsibility for risk culture assessments is less about selecting a single "correct" function and more about ensuring coordination, credibility, and contextual fit. Equally, the individual or function leading such assessments must have sufficient authority and organizational standing to ensure that results are taken seriously and translated into meaningful follow-up actions. Without such mandate and credibility, even methodologically sound assessments risk remaining symbolic exercises with limited impact.

6 Relational Diagram of Second-Order Concepts

Our qualitative analysis, grounded in the Gioia Method, revealed eight relevant second-order concepts that collectively shape the assessment and management of risk culture within financial institutions.

Following the identification of these aggregate dimensions, we structured them into a conceptual framework distinguishing foundational concepts, intermediate concepts, and target perspectives. This structuring reflects their conceptual dependencies and functional roles in shaping how risk culture assessments can be designed and interpreted.

The arrangement broadly aligns with established phases in survey and scale development processes (cf. Groves et al., 2009; Boateng et al., 2018), typically encompassing phases such as conceptualization and item development, instrument design and construction, and application and evaluation. However, we explicitly adapted

this structure to reflect the specific conceptual and practical requirements inherent in assessing risk culture in financial institutions.

The rationale behind this categorization is derived from the nature and function of each concept within the risk culture assessment process. Foundational concepts represent the core elements that underpin all other aspects of the assessment. Intermediate concepts build upon the foundational elements and directly contribute to the operational and methodological aspects of the assessment. Target perspectives represent the desired outcomes of a well-conducted risk culture assessment, focusing on providing a comprehensive understanding of how risk culture evolves and is perceived within an organization. While Sect. 5 presented each theme separately to reflect the structure of expert feedback, the framework developed here differentiates between stand-alone concepts and methodological subdimensions. Accordingly, *Adapting Item Order and Scale for Questionnaire*—discussed as individual sections in the findings—are treated here as subdimensions of *Optimization of Questionnaire Design*. They represent methodological levers that enhance design quality rather than independent second-order concepts in their own right.

This section provides an integrated understanding of how these concepts relate in a structured, start-to-goal sequence, without implying strict causality or dynamic change over time. To guide the interpretation of the relational diagram, it is important to note that the arrows between the concepts represent logical and conceptual dependencies rather than empirically validated causal relationships. The directionality reflects a structured conceptual progression derived from the synthesis of expert feedback and practical considerations regarding the effective assessment of risk culture. No claims regarding temporal dynamics or mutual causality among the concepts are made; instead, the model aims to provide a structured conceptual guide to organizing key factors relevant for survey-based risk culture assessments (See Fig. 1).

6.1 Foundation concept

Clarity in Terminology represents the foundational pillar for effective risk culture assessment. Clear and consistent terminology enhances the precision and reliability of all subsequent methodological steps. It ensures that questionnaire items are interpreted uniformly, facilitates alignment between regulatory and internal frameworks, and clarifies the focus of discussions when structuring or sequencing assessment elements. By establishing a common language, it reduces ambiguity, fosters synergy, and provides the conceptual grounding on which all further methodological refinements build.

6.2 Intermediate concepts

These concepts translate the foundational clarity into methodological and operational design choices. At their core lies *Optimization of Questionnaire Design*, which defines how abstract cultural constructs are transformed into measurable perceptions. This concept encompasses two methodological subdimensions—*Adapting Item Order* and *Scale for Questionnaire*. The former concerns the logical sequencing of items or thematic sections (for example, from general to specific or from tone from

Risk Culture Assessment

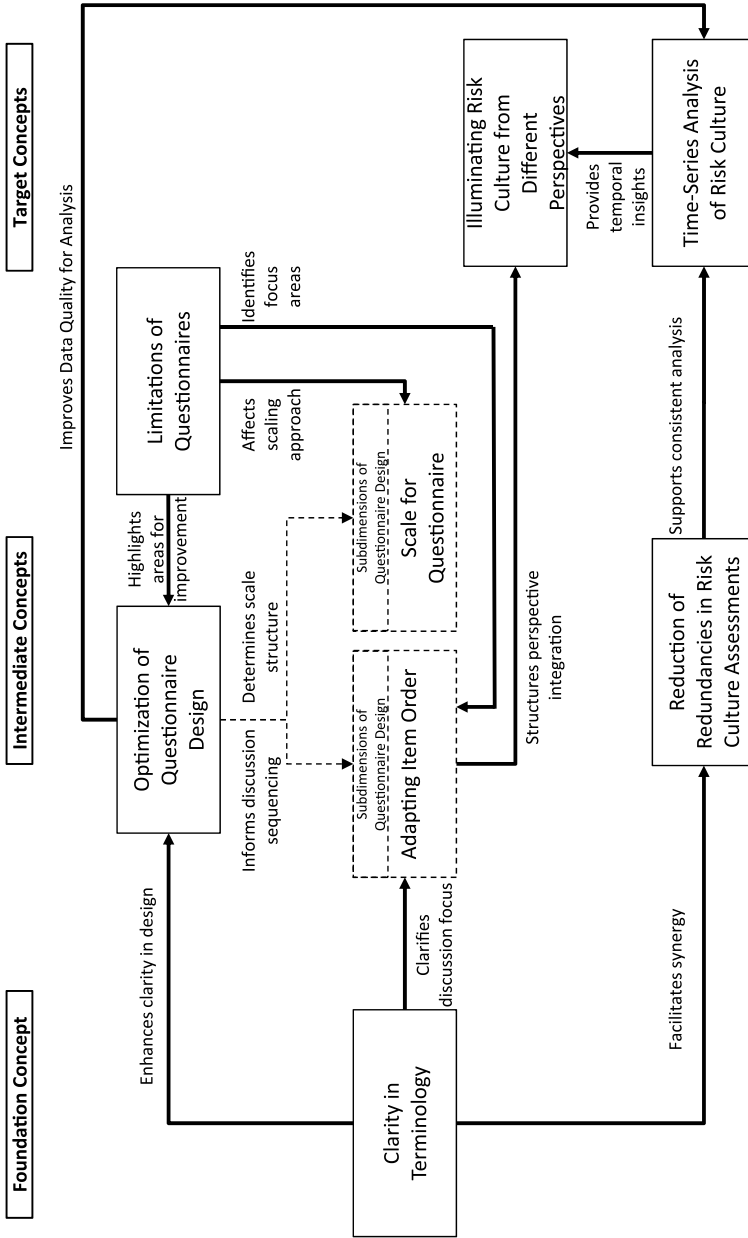


Fig. 1 Relational Diagram of Second-Order Concepts. The directional arrows indicate conceptual dependencies and logical progressions between the identified second-order concepts, based on the synthesis of expert feedback and practical considerations. The framework does not imply dynamic mutual influences or empirically validated causal relationships but serves as a structured guide to organizing relevant factors for effective survey-based risk culture assessment

the top before technical aspects), shaping respondent engagement and the salience of cultural constructs. The latter addresses the construction of response formats, such as the inclusion of neutral categories, the use of bipolar scales, or the development of content-rich response labels. Together, these subdimensions enhance data quality and contextual relevance, ensuring that assessments capture the nuances of institutional settings.

Another intermediate concept, *Reduction of Redundancies in Risk Culture Assessments*, promotes the alignment of internal and regulatory instruments. This alignment supports consistency across assessment cycles, minimizes duplication of effort, and strengthens the comparability of data over time. Such coherence directly enhances the quality of *Time-Series Analysis of Risk Culture*, discussed below as one of the target concepts.

Limitations of Questionnaires constitute a further intermediate concept, drawing attention to the inherent constraints of standardized tools in capturing complex cultural dimensions. Recognizing these limitations encourages continuous refinement of questionnaire design and scale construction, ensuring ongoing methodological rigor. It also informs *Adapting Item Order* by identifying topics that require more nuanced exploration or alternative sequencing. Moreover, by highlighting areas where standardized instruments may fall short, *Limitations of Questionnaires* help define the focus areas later emphasized in *Illuminating Risk Culture from Different Perspectives*.

6.3 Target concepts

The target concepts reflect the intended outcomes of a well-designed risk culture assessment. *Time-Series Analysis of Risk Culture* enables institutions to trace developments and shifts in cultural perceptions over time. It provides a longitudinal perspective that depends on the methodological consistency and alignment established through the intermediate concepts.

Finally, *Illuminating Risk Culture from Different Perspectives* represents the integrative and interpretive dimension of the framework. It combines multiple viewpoints—across hierarchical levels, business functions, and methodological approaches—into a comprehensive understanding of how risk culture manifests within an organization. This perspective benefits from the structured design established through *Optimization of Questionnaire Design* and *Adapting Item Order*, as well as from the temporal insights generated through *Time-Series Analysis*. Together, these connections ensure that risk culture is not examined in isolation but understood as a multidimensional and evolving organizational construct.

In summary, the framework connects conceptual clarity (Foundation Concept), methodological design (Intermediate Concepts), and interpretive integration (Target Concepts). Each level builds upon the previous one, creating a coherent structure that guides both the design and interpretation of risk culture assessments. While not implying causal or temporal processes, the framework provides a structured conceptual map that supports both scholars and practitioners in organizing the key considerations for comprehensive, context-sensitive measurement.

While existing scales, such as those by Sheedy et al. (2017), Fernández Muñoz et al. (2020), and Ghafoori et al. (2023), provide essential insights, they often focus on

broad overviews without addressing the deeper, interrelated aspects of risk culture dynamics. Recent studies, such as those by Kunz and Heitz (2021) on risk culture in banks, and research on risk management and resilience by Monazzam and Crawford (2024) and Eichholz et al. (2024), underline the importance of robust control frameworks. However, these works do not provide a structured model aimed at developing specialized risk culture measurement tools. This framework complements and extends these approaches by introducing a structured, qualitative foundation that captures nuanced elements of risk culture. It provides a comprehensive model that not only aligns with but also advances existing methodologies, offering a more adaptable and thorough tool for the assessment of risk culture in financial institutions.

Thus, the framework provides a coherent, practice-informed structure for organizing the key conceptual considerations relevant to risk culture assessment.

7 Conclusion, limitations, and future research

This study employed the Gioia methodology to analyze expert feedback obtained during the validation of a risk culture scale developed to align with the Financial Stability Board's framework. Our findings highlight eight key concepts that capture recurring design challenges in risk culture assessments: *Clarity in Terminology* as the foundation concept, which influences *Optimization of Questionnaire Design*, *Reduction of Redundancies in Risk Culture Assessments*, and *Adapting Item Order*; these intermediate concepts, in turn, support the target concepts of *Time-Series Analysis of Risk Culture* and *Illuminating Risk Culture from Different Perspectives*. Additionally, recognizing the *Limitations of Questionnaires* ensures a comprehensive and effective assessment process, while the appropriate *Scale for Questionnaire* captures nuanced responses. Together, these structured elements provide a reflective framework that can inform and improve the design and application of risk culture assessments.

While the eight second-order concepts derived in this study offer structured guidance for improving risk culture assessments, they should not be interpreted as universally prescriptive or methodologically exhaustive. Several of the concepts reflect inherent trade-offs in the design and implementation of survey-based instruments. For example, while reducing redundancies may improve clarity and minimize respondent fatigue, some degree of item overlap is methodologically desirable to enhance construct validity and scale reliability (e.g., Boateng et al., 2018; MacKenzie et al., 2011). Similarly, efforts to improve terminological clarity may conflict with the need to preserve interpretive openness across different stakeholder groups. This shows that the proposed framework is not a prescriptive model but a reflective tool to support critical examination, organizational tailoring, and context-sensitive adaptation of existing instruments. Its practical value depends on how effectively it is applied within specific institutional contexts.

Despite the insights generated by this study, several limitations must be acknowledged. First, the specific group of 21 experts may limit the transferability of the findings to other institutions or broader contexts. While the study is grounded in the financial sector and framed by the FSB guidelines, many of the methodological insights—such as trade-offs in questionnaire design, the need for terminological clarity, and the balance between redundancy and reliability—are not unique to banks. They are relevant to any organization where culture influences risk-related behavior, including industries such as healthcare, energy, or the public sector. Nevertheless, contextual factors such as regulatory regimes and institutional maturity may shape how these insights can be applied in practice. While this provided rich and in-depth data, future research could benefit from expanding the pool of experts to include more diverse perspectives from additional institutions or geographical regions to further strengthen the conclusions. Second, the study relies on subjective expert opinions, which may introduce bias due to individual or organizational experiences, internal policies, and personal interpretations of risk culture. While expert insights are crucial, this subjectivity could influence the identification of second-order concepts. Third, the use of the Gioia Method for qualitative analysis, though widely respected, has inherent limitations. The method involves interpretation during the coding and synthesis process, which can lead to variability in how concepts are categorized. Different analytical approaches might yield alternative conceptual structures. Fourth, the study is framed around the Financial Stability Board (FSB) guidelines, which, while comprehensive, may not fully capture the nuances of institutions operating under different regulatory regimes. This focus could limit the applicability of the findings to financial institutions outside of the FSB framework. Finally, the research did not include a quantitative validation of the developed scale. While qualitative methods were appropriate for the exploratory phase, further quantitative analysis is necessary to empirically test the reliability and validity of the proposed risk culture assessment model across a broader set of institutions.

Future research should explore the integration of these findings into the practical application of risk culture scales and extend the analysis to other sectors where risk culture plays a critical role in governance and compliance. Refining assessment tools based on our findings could improve their predictive power and practical utility—particularly in preventing cultural blind spots that may contribute to future financial crises.

The second-order concepts developed through this study offer a framework for a more nuanced understanding of the design challenges that influence the effectiveness of risk culture assessments. These insights are grounded in organizational behavior theories that highlight the role of structures, language, and perception in shaping cultural norms (Schein, 1990; Weick, 1995). They also connect with strategic management perspectives that emphasize the alignment between culture and strategic goals (Kaplan & Norton, 2001), as well as with behavioral corporate finance research on

how organizational culture shapes risk perception and decision-making (Baker et al., 2007; Malmendier & Tate, 2005).

In addition to these theoretical implications, our findings offer direct practical value for financial institutions seeking to enhance their risk culture frameworks in line with global standards such as those set by the Financial Stability Board. The practitioner-informed structure developed here can help institutions tailor and adjust existing tools to better reflect organizational maturity, regulatory expectations, and strategic priorities. This supports both compliance and the development of a more risk-aware culture.

While this study is centered on the survey-based assessment of risk culture, we recognize the growing importance of complementary methodological approaches. Recent contributions have proposed behavioral text analysis (Corritore et al., 2020), diverse empirical and conceptual frameworks (Carretta et al., 2017), and systematic typologies of assessment methods (Bockius & Gatzert, 2023) to capture deeper cultural dynamics. For instance, Carretta et al. (2017) compile conceptual, regulatory, and case-based insights; Sinha and Arena (2020) explore risk culture from the perspective of internal auditors using interview data; and Streicher et al. (2023) offer a theoretically integrative model to map cultural influences across organizational levels. Moreover, an increasing number of studies have developed structured measurement instruments to quantify specific aspects of risk culture (e.g., Ghafoori et al., 2023).

In contrast to these contributions, our study does not aim to introduce a new instrument or theory. Instead, it distills recurring design challenges and reflective practitioner insights that emerge during the real-world application of risk culture surveys. This practitioner-side perspective sheds light on practical tensions, contextual limitations, and usability issues that are often overlooked in conceptually driven measurement research. Future research could build on this contribution by embedding such practitioner-derived frameworks into multi-method diagnostic strategies or by empirically linking our identified challenges to organizational behavior and performance indicators.

Appendix A

Table A1 Sample of expert interviews: institutional type, role, and duration

Interview ID	Type of Institution	Role/Position (anonymized)	Duration (minutes)	Transcript
Interview 1	Advisory	Risk management expert	51	Yes
Interview 2	Advisory	Compliance expert	157	Yes
Interview 3	Building society	Head of Risk Controlling	224	Yes
Interview 4	Building society	Risk Controlling, Board Office	45	Yes
Interview 5	Bank	Head of Compliance – Risk Assessment & Controls	62	Yes
Interview 6	Bank	Head of Compliance	64	Yes
Interview 7	Bank	Board member, Risk Culture expertise	52	Yes
Interview 8	Bank	Risk culture specialist, Group Internal Audit & Regulation	97	Yes
Interview 9	Bank	Department Head, Internal Audit & Regulation	97	Yes
Interview 10	Bank	Controlling, Risk Culture expertise	153	Yes
Interview 11	Bank	Risk culture specialist	90	Yes
Interview 12	Bank	Risk Governance & ICAAP specialist	58	Yes
Interview 13	Bank	Board member, Risk Culture expertise	32	Yes
Interview 14	Bank	Risk culture specialist	56	Yes
Interview 15	Building society	Board member	30	Yes
Interview 16	Advisory	Risk culture expert	90	Yes
Interview 17	Bank	Board member	76	Yes
Interview 18	Bank	Director, Head of Risk Controlling	172	Yes
Interview 19	Bank	Risk Controlling	84	Yes
Interview 20	Bank	Risk Controlling	150	No (notes only)
Interview 21	Advisory	Compliance expert	180	No (notes only)

Table A2 Representative data structure matrix for risk culture assessment

Representative quotations from informants	1st Order Concepts	2nd Order Themes	Aggregate Dimensions
Yes, so the topic of risk appetite, risk propensity, etc. Risk culture, you have to explain to people what that means before you give them the questionnaire (Interview 2)	Explain risk appetite, risk propensity, risk culture etc.	Define terms	1. Clarity in Terminology as a Foundation Risk Culture Measurement
The question would be how to define management again. Because it's definitely not just the board of directors, or maybe it is. I just don't know exactly. Whether it's management or the level below that. (Interview 18)	Define management		
What do you mean by remuneration? Is this the variable remuneration? Is this the total remuneration? (Interview 17)	Define whether fixed or variable remuneration		
You have extracted an incredible amount from the FSB paper, and I used it as well when I gave my presentation at an event. I found it to be the central document that I relied on because I thought it was very helpful for actually understanding what the expectations of a bank are and what needs to be done. However, one of the critical points in answering all these questions is: First, you have to define the risk culture. So, when I ask any question related to risk culture, I must first say what I actually understand by risk culture and what I understand by it in my specific institution. So, at least for you, if you want to achieve comparable results, it is important to define your understanding of risk culture. Now, from the FSB paper 'Corporate Governance Principles for Banks,' I have a definition here, then I used something from MaRisk, then from the EBA Guidelines, then there are a few individual aspects, i.e., the EBA Guidelines and internal governance from 2017, which you can then use, but there is really no one single definition of what risk culture is. This also aligns with what is stated in the FSB paper, but the other goal definitions are not really clear-cut either. And I think that's also the problem you face when you have questionnaires or the challenge you have when you send out such a questionnaire: Depending on who fills it out, they have a completely different understanding of it. (Interview 5)	Definition of the company-specific risk culture		
What I would definitely recommend anyway is to further define the various components. I think processes, deficits, role-modeling values, etc. need to be further defined because otherwise the answers will be too divergent. I would explain a bit further by paragraph or by part what exactly is meant by it. For example, role-modeling. (Interview 16)	Define more precisely		

Table A2 (continued)

Representative quotations from informants	1st Order Concepts	2nd Order Themes	Aggregate Dimensions
<p>If I ascertain whether something is prepared systemically and at the same time this is specified for the institutions from the MaRisk guidelines, this is already somewhat redundant at this point. The question would then be whether this information, which is now necessary here and which is also needed here, is already available elsewhere when you go into operational practice, or whether it has to be available. (Interview 10)</p> <p>Survey item 'As soon as deficits arise in risk management, they are rectified.' Yes, that's true, which makes me smile, because it's so strongly supported by tools and processes. Every bank will have them, that all audit findings are tracked hard. So, there is regular monitoring with reports that define clear target dates by which the findings must be completed. There are review processes to check whether the measures implemented are also suitable for eliminating the weaknesses. This is all a regulatory requirement. (Interview 3)</p> <p>Now, as a non-employment law expert, I don't know to what extent this can be covered by the implementation of the Remuneration Ordinance for Institutions. Let me give you an example: remuneration is aimed at implementing corporate values. I would not know whether compliance with the Remuneration Ordinance for Institutions also covers this issue. But it's worth taking a look. (Interview 10)</p> <p>In the next question, I translate internal audits as audit reports. In the case of risk-relevant decisions, you would then have to involve special functions in the institution in accordance with MaRisk (Standard 3), the Head of Risk Controlling, the Head of Internal Audit and the Compliance Officer. That is standard. In other words, if I understand correctly, this is actually always guaranteed as long as they do their job. And this is also checked regularly. (Interview 10)</p> <p>If I ascertain whether something is prepared systemically and at the same time this is specified for the institutions from the MaRisk guidelines, this is already somewhat redundant at this point. The question would then be whether this information, which is now necessary here and which is also needed here, is already available elsewhere when you go into operational practice, or whether it has to be available. (Interview 10)</p>	<p>Systemic component equals high degree of coverage with other audit schemes/questionnaire redundant-</p> <p>Question in order. Regulatory requirement. Everyone confirms the question. (Added value?)</p> <p>Incentive structures generally covered by the Remuneration Ordinance for Institutions (Institutsvergütungsverordnung)</p> <p>Covered by MaRisk, expert recommendation: delete</p> <p>High degree of coverage with other test schemes/redundant questionnaires</p>	<p>Already covered by other regulatory requirements</p>	<p>2. Reduction of Redundancies in Risk Culture Assessments</p>

Table A2 (continued)

Representative quotations from informants	1st Order Concepts	2nd Order Themes	Aggregate Dimensions
<p>[...] so we have guidelines on money laundering, so to speak, yes, hooks on it, then there is somehow the report on the following money laundering violations and we put it all together first, from that you can generate a picture without having to talk to anyone, let's say, at length. You might just have to request the report somewhere and you have to put it on somehow, I don't need to conduct any big interviews, I can get that together halfway in the situation, so to speak, and then I can draw a line under it and could now go to the board and say, so we have a money laundering regulation, we had three violations last year, we had four the year before and that seems to be somehow stable, because, I don't know, with three million transactions a year, that doesn't really sound dramatic, everything seems to be in the green. (Interview 12)</p> <p>Well, I think, because if you approach others, you will likely encounter similar people to me, I assume. They are reasonably competent but not deeply involved. They will read through it and, when they get to question 40 or 50, just quickly tick off the answers—ok, tick, tick, tick—that's my assumption so they can finish the questionnaire. The shorter it is, the higher the response rate will be. (Interview 14)</p> <p>Yes, so the questionnaire is certainly quite appropriate for addressing the topic. I dare to say that if someone critically works through this questionnaire, they will be busy with it for a good half hour. The question is, to what extent will the person then stop. saying, 'Come on, I've had enough, I don't want to continue,' or to what extent will the respondent not be overwhelmed by the input to the point where they can no longer answer the questions with sufficient precision. (Interview 7)</p> <p>I read through the questions, I think I got through to question 80 quite well and then there is this point, then of course there is an effect, a fatigue effect. (Interview 14)</p> <p>[...] for example, if you ask the colleagues in the retail banking about financial risks, they won't know, simply because they don't understand the financial metrics that express financial risks. They are not familiar with them at all. So, it's a bit, let me put it this way, I'm not sure if this one-size-fits-all approach will work systemically, in the sense of producing reliable results. (Interview 6)</p>	<p>Analyzing Additional (Regulatorily Mandated) Data</p>	<p>Number of questions too high+ do not ask all questions if necessary</p>	<p>3. Optimization of Questionnaire Design in Risk Culture Assessments</p>

Table A2 (continued)

Representative quotations from informants	1st Order Concepts	2nd Order Themes	Aggregate Dimensions
<p>And now the question is, how do you make this measurable? And that was, to put it cautiously, my point of criticism: I believe the sheer number of questions, yes, if you start from the top and go to the bottom, it becomes somewhat difficult, and I also don't believe you will get consistent results. You will get heterogeneous results because people tend to forget, after question 10, how they already answered a similar question 1, and then they have to look back. (Interview 6)</p>	<p>Inhomogeneous results due to number of questions, evaluation difficult</p>		
<p>Yes, well, of course, it seems overwhelming at first—a 16-page comprehensive questionnaire. On the other hand, it's obviously a topic that is also reflected in MaRisk, a topic, an important topic, which certainly requires a broad or comprehensive analysis. In this respect, the extent it has is overwhelming at first glance, but on second glance, I would say it's completely okay. If you can cover the topic of risk culture with 16 pages and see how much importance the regulator now attaches to this topic, then I think it's well-invested time. (Interview 15)</p>	<p>Number of questions okay, as topic is important for supervision</p>		
<p>Yes, well, let's say if you have a web dialogue, like Mentimeter, you probably have something at the top indicating the cultural and dimensional context you are dealing with, right? Which sub-dimension you are in, which aspect you are focusing on. I could imagine these as three headings along with the component, and then comes the question. And for the answers, you have radio buttons. (Interview 1)</p>	<p>Survey via web dialogue with radio buttons</p>	<p>Questionnaire procedure</p>	
<p>[...] and then, if you agree with a company to have them go through the questionnaire and they say, 'But we don't have any performance evaluation,' well, of course, you have to exclude that questions. That's technically very simple nowadays. Just to say, it won't necessarily fit every organization. (Interview 18)</p>	<p>Question technique: partially hide questions if not relevant based on preselection</p>		
<p>[...] of course, with a hundred questions, it's all a bit difficult, and the last twenty or thirty questions tend to get more attention. I would rather suggest trying to possibly segment the questionnaire, or only give certain parts of the questionnaire to specific people, so that you don't overwhelm them and to achieve as high a response rate as possible. (Interview 16)</p>	<p>Division of the questionnaire into several parts</p>		

Table A2 (continued)

Representative quotations from informants	1st Order Concepts	2nd Order Themes	Aggregate Dimensions
<p>You have now taken the FSB paper one-to-one as a basis and formulated questions from it. I have to be honest and say that this has its difficulties when you look through it. Why? Because this FSB, this FSB paper, contains a whole series of redundancies in the explanations and if you look at the questions, eh, there are often things that are unrelated to the topic clustered together. [...] Ehm, so I would recommend that you think again about clustering the topics differently. Of course with the FSBs main categories, I fully support that, but that you take out these redundancies a bit. (Interview 11)</p> <p>But again, don't you think one could, because the questions are so similar, evaluate the results from multiple perspectives? So, couldn't one ask a question that essentially addresses three aspects or topics at once in the background, instead of asking the question again in a different formulation? (Interview 6)</p> <p>So let's take an example in relation to the questionnaire. [...] 'Managers promote an active discussion of the understanding of risk culture in the company, e.g. through personal discussions and exchanges with employees.' 'Do you have the feeling that your boss regularly talks to you about risk culture?' (Interview 1)</p> <p>Code of conduct is also a good keyword. It is such a nice topic, it is formally required anyway, it is a formal component of risk culture. Do I have a code of conduct, yes/no. Has it been communicated in an appropriate manner, yes/no? These are purely formal questions from an auditor's perspective. And then the questionnaire can be based on that, if it is appropriately formulated, if the formal components are there. Then I know that the target should be known. Then I can start by looking at the individual principles, for example, whatever the word is that is in the code of conduct, I can then use that as the target, and adapt the questions from that, yes, adapt them accordingly. From the formal to the informal or from the formal to the practical. (Interview 9)</p> <p>[...] perhaps work more with examples in some places or push some questions further back, so to speak, to make them a little more practicable. (Interview 12)</p> <p>I would be careful with examples. [...] The problem with examples is always that you immediately focus on them and then someone says, but that's not the case here. That's difficult. And uh, unless of course you really do multiple choice here. But then it would have to be broadly structured, for example, you would have to offer a selection of 10, 12, 14 categories of what constitutes a risk culture. (Interview 18)</p>	<p>FSB has redundancies within itself or things unrelated to the topic are clustered together. Suggestion, adjust clustering. Beam questions into different areas</p> <p>Beam questions into different areas</p> <p>simplifying the language of the items</p> <p>formulating company-specific questions</p> <p>Break down the level of abstraction using examples</p> <p>be careful with the use examples</p>	<p>Change formulation of the items</p> <p>Using Examples</p>	

Table A2 (continued)

Representative quotations from informants	1st Order Concepts	2nd Order Themes	Aggregate Dimensions
<p>Maybe it's again an issue with the sequence, so that somehow—well, we had the specific questions later on. Perhaps it really makes sense to bring those to the beginning and then ask about the topic afterward. Then you simply don't have to do much here, just a bit, for example, as you just said, give examples, but you've already talked about it beforehand, and maybe people are already more engaged and can answer more easily. (Interview 12)</p> <p>[...] So, regarding the sequence, I think, in my opinion, that tone from the top is very important, and various elements like compensation, whistleblowing, etc., should actually reflect, or should indeed mirror, the Tone from the Top. [...] So, I would also say, especially when it comes to the top-down process, I find tone from the top actually quite good. If we assume that there is a defined or desired risk culture, then asking to what extent the executive board or the top of the organization or company actually embodies or talks about it is not illogical, I think. (Interview 16)</p> <p>Have you ever looked into the approach of the Dutch Central Bank in the context of your research? [...] There is also a very good book by them that explains their supervisory approach. Their approach goes much further. For example, they examine the work within internal committees: How does the executive board interact with each other? Are things also discussed controversially? They also examine the collaboration between the executive board and the supervisory board. They discovered that if decisions are made and everyone agrees with everything, then the issue of risk culture may not be well-addressed. It doesn't necessarily have to be that way. It could be that people simply think alike and therefore have the same opinion. But it could also be that there is someone with a dominant personality and the others don't want to start conflicts with them. There isn't just black and white, but also gray. So, they go quite far. (Interview 7)</p> <p>Yes, it's all good. The questions are good, no. So I don't want to criticize the quality of the questions, I think they're essential and right. It's just that because they're essential and right, I have justified doubts about getting honest and realistic answers. Because, of course, they will already show massive blatancy in many companies. I am convinced of that. (Interview 2)</p> <p>And I think, regarding the German banking supervision, I have received feedback from many, both internally at BaFin and the Bundesbank, along the lines of the FSB wrote a paper on this, and the insights came out of the financial crisis. They realized that this is a huge issue. We had to address it. But we don't really have a plan on how to deal with it. That's where the Dutch approach is good because they go a step further and look into it for themselves. (Interview 7)</p>	<p>Consideration, change the order of questions, from concrete to abstract</p> <p>(Do not) adjust sequence</p> <p>Approach of the Dutch central bank-> Investigation of committee work</p>	<p>Discussion</p> <p>Adjust sequence across all aspects (not within one aspect)</p> <p>Boundaries of the questionnaire</p>	<p>4. Adapting Item Order</p> <p>5. Limitations of Questionnaires in Risk Culture Assessment</p>
	<p>doubts about getting honest and realistic answers</p> <p>Overarching manner of FSB (how to deal with)</p>	<p>Sense of Frameworks (FSB)</p>	

Table A2 (continued)

Representative quotations from informants	1st Order Concepts	2nd Order Themes	Aggregate Dimensions
<p>The challenge is that this FSB paper, apart from discussing incentivization and a few minor details, does not directly address these risk drivers but rather views it all in a very broad, overarching manner. [...] (Interview 8 and Interview 9)</p> <p>As you know, the whole thing came to the fore as a result of the financial market crisis in 2008, 2007/2008, when people started to look at what had caused it to happen in the first place. The cultural issue was one of the key points that led to the FSB, the Financial Stability Board, addressing the issue and developing the guidelines to which you refer. I must honestly say that I was also very, very impressed by the guidelines at the time. When I first looked into the topic, I also came across this guide relatively quickly. And I think it is still as relevant today as it was back then, because it is also the only comprehensive, systematic approach to dealing with the topic from a supervisory perspective. (Interview 7)</p> <p>That's why I also said that this was the FSBs first exercise in dealing with this and that you have to be realistic about supervision. Supervision is, of course, to some extent an internal audit of the company. And supervision typically looks at such technical issues. And, um, I mean, I've also given presentations on how to examine risk culture in a company. And of course this dimension that you are highlighting is also included. So an audit will also look at whether a strategy has been defined, how it is communicated, whether there is training on it and so on. So there is also this technical consideration. But this is not a cultural issue. Yes? That is a necessary prerequisite for culture to be promoted, but it is not the promotion of culture per se. Yes? Because if you don't have a risk strategy, then it will be extremely difficult to introduce and develop a standardised risk perspective among employees. (Interview 11)</p> <p>Yes, in my view, culture is something that needs to develop. It's not enough to conduct a one-time measurement if you're serious about it. You can always say, 'measure it or forget it.' With the measures you implement, you should demonstrate how you are evolving. It's important to see where the development is heading. Culture is a process of development. (Interview 7)</p> <p>And if we take this further, a time series is particularly interesting. We develop this, and we can even say that starting from a low level, and then you can see precisely, because we have been doing this here, and this concept has been pursued by the company for two and a half years, we are improving there. (Interview 4)</p>	<p>Overarching manner of FSB</p> <p>FSB guide still up to date and convincing</p> <p>FSB equals technical consideration, but that is not yet the promotion of culture</p>	<p>View in time series</p> <p>Developing a risk culture is important and should be measured.</p> <p>Time series important</p>	<p>6. Time-Series Analysis of Risk Culture: A Dynamic Perspective</p>

Table A2 (continued)

Representative quotations from informants	1st Order Concepts	2nd Order Themes	Aggregate Dimensions
<p>What is difficult and also not helpful in my view is an evaluation at a single point in time. Because there is no absolute good or bad, we have agreed on that, and it is only a matter of whether it is developing in the direction that you ultimately want. (Interview 8)</p> <p>Because there is no absolute good or bad, we have agreed on that, and it is only a matter of whether it is developing in the direction that you ultimately want. (Interview 8)</p> <p>Yes, of course, the system and the shaping of the system, and the influence of the leaders on the system, are completely relevant. Yes, those at the top have the greatest influence, of course. I would even say that the leaders have the greatest influence in the end. It's the role modeling that is relevant. What good is a great system in the company if the leaders consistently model that it doesn't have to be adhered to. (Interview 2)</p> <p>12:40 (Interviewer Leader): Maybe we can revisit, when we talk about the structure, these components derived from the dimensions of the FSB paper. Leaders, employees, and systemic aspects. Do you find something like this useful, or do you feel something is missing?</p> <p>13:10 (Interviewee): I found this immediately understandable: one is the perception—what I believe I am doing, and the other is what actually reaches the employee level, and the systemic aspect as the organizational framework that should encompass everything. I don't think anything is missing. (Interview 5)</p> <p>The division into three components is considered useful from the expert's perspective. The systemic component corresponds to the approach used by auditors, such as financial auditors or BaFin. They would assess through peer groups. (Protocol note on Interview 20)</p> <p>They looked at what the leaders say about certain things and what the employees say. And thus, the absolute level of the responses was not important, but rather the discrepancy between the different target groups. For example, if the leaders say, 'We practice risk culture daily, and we are convinced that we are training and developing employees effectively towards risk culture,' but the employees say, 'This doesn't really reach us. We don't quite understand what is expected of us.' Then you already have such a discrepancy, and this discrepancy alone indicates a need for action. And it's not about right or wrong. Do you understand what I mean? (Interview 11)</p>	<p>Time series important</p> <p>Time series analysis important. Example of error culture</p> <p>Discussion Balance of components. Perhaps managers have more influence than systemic components</p> <p>Components are complete and understandable</p>	<p>Feedback on components</p>	<p>7. Illuminating Risk Culture from Different Perspectives</p>

Table A2 (continued)

Representative quotations from informants	1st Order Concepts	2nd Order Themes	Aggregate Dimensions
I couldn't answer all the questions. Sometimes I also miss a neutral category because, of course, the market area is different from risk management. I found this aspect generally very difficult, because, for example, goal achievement is certainly measured by different criteria in the market than in risk management. (Interview 19)	Does it make sense to include a neutral category?	Topic Scale for Questionnaire	8. Scale for Questionnaire
[...] But if you assess this on a scale, for example, on a scale from 1 to 10, you can do it that way. Then I understand. [...] (Interview 3)	number of scale points		
There are now also response categories that are a bit more evaluative. If you potentially have the optimum in the middle, and the further you go out, the less optimal it becomes. The advantage is that you actually ask for an assessment of how sufficient it is: are the structures perhaps adequate, or are they overstructured, or are they almost non-existent. That would be another possibility. If, naturally, as you said, it goes from 'not at all,' 'does not apply at all,' to 'fully applies,' then the question is, where is the optimum. The optimum is probably on one side. Only the other side, where it might be too much or too regulated, is not so visible. (Interview 16)	Likert scale vs. optimum in the centre		
[...] This is all a regulatory requirement. So, but, in this respect, you can formulate it like this. But you won't find anyone who answers the question with 'no'. (Interview 3)	yes and no answers (e.g. for regulatory requirements) problematic with Likert scale		
The expert points out a problem with the Likert scale: There are some questions that must be met according to regulatory requirements. There are then different intensities of findings. (Notes from Interview 20)			
I would indeed make the answers more concrete, i.e., not just have 'bad,' 'average,' 'good,' but include answer texts that reflect these categories with content. This is to at least remove the ambiguity that people might have different understandings of what their criteria mean. (Interview 5)	specific options for the answers		

Funding Open Access funding enabled and organized by Projekt DEAL. This research did not receive any specific grant from funding agencies in the public, commercial, or not-for-profit sectors.

Data availability The data gathered and analysed during the current study are not publicly available because of confidentiality reasons.

Declarations

Generative AI and AI-assisted technologies in the writing process During the preparation of this manuscript, the authors utilized OpenAI's ChatGPT and DeepL to assist with improving the readability and language quality. OpenAI's ChatGPT was employed to refine language, suggest alternative phrasings, and provide perspectives that helped enhance the clarity and coherence of the manuscript. DeepL facilitated translation and further language enhancement. The authors maintained full control over the use of these tools, carefully reviewing, critically assessing, and significantly editing all AI-generated content. The core scientific ideas, research, and data presented in this manuscript are the sole contributions of the authors. The authors take full responsibility for the content of this publication, ensuring that it accurately represents their original research and findings. The use of AI technologies was strictly limited to aiding the writing process and did not involve the creation or modification of the scientific content or data.

Competing interests The authors declare that they have no conflict of interest.

Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons licence, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons licence, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons licence and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this licence, visit <http://creativecommons.org/licenses/by/4.0/>.

References

- Aarons, G. A., Ehrhart, M. G., Farahnak, L. R., Sklar, M., & Horowitz, J. (2017). Discrepancies in leader and follower ratings of transformational leadership: Relationship with organizational culture in mental health. *Administration and Policy in Mental Health, 44*(4), 480–491. <https://doi.org/10.1007/s10488-015-0672-7>
- Agarwal, A., Gupta, A., Kumar, A., & Tamilselvam, S. G. (2019). Learning risk culture of banks using news analytics. *European Journal of Operational Research, 277*(2), 770–783. <https://doi.org/10.1016/j.ejor.2019.02.045>
- Agnese, P., & Capuano, P. (2021). The culture of risk governance in financial institutions. *Journal of Risk Management in Financial Institutions, 14*(3), 301–313.
- Alvesson, M., & Kärreman, D. (2007). Constructing mystery: Empirical matters in theory development. *Academy of Management Review, 32*(4), 1265–1281. <https://doi.org/10.5465/AMR.2007.26586822>
- Arnaboldi, F., & Vasciaveo, C. (2017). Role of internal audit on risk culture. In A. Carretta, F. Fiordelisi, & P. Schwizer (Eds.), *Risk culture in banking* (pp. 403–430). Springer International Publishing.
- Australian Prudential Regulation Authority (APRA). (2016). Risk culture. Information paper.
- Aven, T., & Ylönen, M. (2021). How the risk science can help us establish a good safety culture. *Journal of Risk Research, 24*(11), 1349–1367. <https://doi.org/10.1080/13669877.2020.1871056>
- Baghersad, M., & Zobel, C. W. (2022). Organizational resilience to disruption risks: Developing metrics and testing effectiveness of operational strategies. *Risk Analysis, 42*(3), 561–579. <https://doi.org/10.1111/risa.13769>
- Baker, M., Ruback, R., & Wurgler, J. (2007). Behavioral corporate finance. In *Handbook of corporate finance: Empirical corporate finance* (1st ed., Vol. 1, pp. 145–186, Handbooks in finance). Elsevier North-Holland.

- Basel Committee on Banking Supervision (BCBS). (2015). Guidelines - corporate governance principles for banks.
- Bianchi, N., Carretta, A., Farina, V., & Fiordelisi, F. (2021). Does espoused risk culture pay? Evidence from European banks. *Journal of Banking and Finance*, 122, 105767. <https://doi.org/10.1016/j.jbanfin.2020.105767>
- Boateng, G. O., Neilands, T. B., Frongillo, E. A., Melgar-Quiñonez, H. R., & Young, S. L. (2018). Best practices for developing and validating scales for health, social, and behavioral research: A primer. *Frontiers in Public Health*, 6, 149. <https://doi.org/10.3389/fpubh.2018.00149>
- Bockius, H., & Gatzert, N. (2023). Organizational risk culture: A literature review on dimensions, assessment, value relevance, and improvement levers. *European Management Journal*. <https://doi.org/10.1016/j.emj.2023.02.002>
- Bolton, P., & Tang, A. M. (2002). An alternative approach to cross-cultural function assessment. *Social Psychiatry & Psychiatric Epidemiology*, 37(11), 537–543. <https://doi.org/10.1007/s00127-002-0580-5>
- Brenot, J., Bonnefous, S., & Marris, C. (1998). Testing the cultural theory of risk in France. *Risk Analysis*, 18(6), 729–739. <https://doi.org/10.1111/j.1539-6924.1998.tb01116.x>
- Bromiley, P., McShane, M., Nair, A., & Rustambekov, E. (2015). Enterprise risk management: Review, critique, and research directions. *Long Range Planning*, 48(4), 265–276. <https://doi.org/10.1016/j.lrp.2014.07.005>
- Burchell, B., & Marsh, C. (1992). The effect of questionnaire length on survey response. *Quality & Quantity*. <https://doi.org/10.1007/BF00172427>
- Butler, T., & Brooks, R. (2024). Time for a paradigm change: Problems with the financial industry's approach to operational risk. *Risk Analysis*, 44(6), 1285–1304. <https://doi.org/10.1111/risa.14240>
- Caporin, M., & Storti, G. (2020). Financial time series: Methods and models. *Journal of Risk and Financial Management*, 13(5), 86. <https://doi.org/10.3390/jrfm13050086>
- Carretta, A., Fiordelisi, F., & Schwizer, P. (2017). *Risk culture in banking*. Cham: Springer International Publishing.
- Cass Business School and New City Agenda. (2014). A report on the culture of British retail banking.
- Chartered Institute of Internal Auditors. 2014. Governance and risk report 2014: Internal audit's perspective on the management of risk.
- Chen, L., Lin, M., & Chen, L. (2023). Financial risk prediction based on time series. *Academic Journal of Management and Social Sciences*, 5(1), 169–174. <https://doi.org/10.54097/ajmss.v5i1.14073>
- Chen, Y. S. A., Decaux, L., & Showalter, S. (2016). Mapping assurance: Internal auditors can facilitate efforts to document the organization's combined assurance activities. *Internal Auditor*, 73(6), 53–56
- Cheung, K. C., & Mooi, L. C. (1994). A comparison between the rating scale model and dual scaling for likert scales. *Applied Psychological Measurement*, 18(1), 1–13. <https://doi.org/10.1177/014662169401800101>
- Cimini, R. (2021). A systematic and bibliometric review on risk culture: A novel theoretical framework. *Journal of Risk Finance*, 22(2), 153–168. <https://doi.org/10.1108/JRF-06-2020-0123>
- Clifton, J. D. W. (2020). Managing validity versus reliability trade-offs in scale-building decisions. *Psychological Methods*, 25(3), 259–270. <https://doi.org/10.1037/met0000236>
- Committee of Sponsoring Organizations of the Treadway Commission (COSO). (2017). *Enterprise risk management: Aligning risk with strategy and performance*. Durham, NC.
- Corritore, M., Goldberg, A., & Srivastava, S. B. (2020). The new analytics of culture. *Harvard Business Review*, 98(1), 76–83.
- Crispim, J., Silva, L. H., & Rego, N. (2019). Project risk management practices: The organizational maturity influence. *International Journal of Managing Projects in Business*, 12(1), 187–210. <https://doi.org/10.1108/IJMPB-10-2017-0122>
- De Nederlandsche Bank (DNB). (2015). *Supervision of behaviour and culture: Foundations, practice & future developments*. Amsterdam. https://www.dnb.nl/media/1gmkp1vk/supervision-of-behaviour-and-culture_tcm46-380398-1.pdf?utm_source=chatgpt.com
- Dillman, D. A., Smyth, J. D., & Christian, L. M. (2014). *Internet, phone, mail, and mixed-mode surveys: The tailored design method*. Wiley.
- Dolnicar, S., Grün, B., & Yanamandram, V. (2013). Dynamic, interactive survey questions can increase survey data quality. *Journal of Travel and Tourism Marketing*, 30(7), 690–699. <https://doi.org/10.1080/10548408.2013.827546>
- Dubois, B., & Burns, J. A. (1975). An analysis of the meaning of the question mark response category in attitude scales. *Educational and Psychological Measurement*, 35(4), 869–884. <https://doi.org/10.1177/001316447503500414>

- Durst, S., Hinteregger, C., & Zieba, M. (2019). The linkage between knowledge risk management and organizational performance. *Journal of Business Research*, *105*, 1–10. <https://doi.org/10.1016/j.jbusres.2019.08.002>
- Dürst, N., & Kunz, J. (2025). The risk culture scale: a measurement tool to comprehensively assess banks' risk culture. *Abacus*. <https://doi.org/10.1111/abac.12370>
- Eichholz, J., Hoffmann, N., & Schwering, A. (2024). The role of risk management orientation and the planning function of budgeting in enhancing organizational resilience and its effect on competitive advantages during times of crises. *Journal of Management Control*, *35*(1), 17–58. <https://doi.org/10.1007/s00187-024-00371-8>
- European Banking Authority (EBA). (2017). Guidelines on internal governance under directive 2013/36/EU.
- Fernández Muñiz, B., Montes Peón, J. M., & Vázquez Ordás, C. J. (2020). Misconduct and risk climate in banking: Development of a multidimensional measurement scale. *Global Policy*, *11*(S1), 73–83. <https://doi.org/10.1111/1758-5899.12752>
- Financial Stability Board (FSB). (2014). *Guidance on supervisory interaction with financial institutions on risk culture*. A Framework for Assessing Risk Culture., July 15, 2015.
- Fritz-Morgenthal, S. G., Hellmuth, J., & Packham, N. (2015). Does risk culture matter? The relationship between risk culture indicators and stress test results. *Journal of Risk Management in Financial Institutions*, *9*(1), 71–84.
- Ganassali, S. (2008). The influence of the design of web survey questionnaires on the quality of responses. *Survey Research Methods*, *2*(1), 21–32. <https://doi.org/10.18148/srm/2008.v2i1.598>
- Gehlbach, H., & Barge, S. (2012). Anchoring and adjusting in questionnaire responses. *Basic and Applied Social Psychology*, *34*(5), 417–433. <https://doi.org/10.1080/01973533.2012.711691>
- Gehman, J., Glaser, V. L., Eisenhardt, K. M., Gioia, D., Langley, A., & Corley, K. G. (2018). Finding theory–method fit: A comparison of three qualitative approaches to theory building. *Journal of Management Inquiry*, *27*(3), 284–300. <https://doi.org/10.1177/1056492617706029>
- Ghafoori, E., Mata, F., Lauren, N., Faulkner, N., & Tear, M. J. (2023). Measuring risk culture in finance: Development of a comprehensive measure. *Journal of Banking and Finance*, *148*, 106720. <https://doi.org/10.1016/j.jbankfin.2022.106720>
- Gioia, D. A., & Chittipeddi, K. (1991). Sensemaking and sensegiving in strategic change initiation. *Strategic Management Journal*, *12*(6), 433–448.
- Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking qualitative rigor in inductive research. *Organizational Research Methods*, *16*(1), 15–31. <https://doi.org/10.1177/1094428112452151>
- Gioia, D. A., Corley, K. G., & Hamilton, A. L. (2013). Seeking Qualitative Rigor in Inductive Research. *Organizational Research Methods*, *16*(1), 15–31. <https://doi.org/10.1177/1094428112452151>
- Groves, R. M., Fowler, F. J., Couper, M., Lepkowski, J. M., Singer, E., & Tourangeau, R. (2009). *Survey methodology* (Wiley series in survey methodology). Wiley.
- Guillemette, M., Finke, M., & Gilliam, J. (2012). Risk tolerance questions to best determine client portfolio allocation preferences. 25.
- Goldendmund, F. W. (2007). The use of questionnaires in safety culture research – an evaluation. *Safety Science*, *45*(6), 723–743. <https://doi.org/10.1016/j.ssci.2007.04.006>
- Hall, S. N., Gallagher, M. A., & Fenn, D. S. (2020). Risk framework for an organizational system with major components. *Risk Analysis*, *40*(12), 2509–2523. <https://doi.org/10.1111/risa.13572>
- Herzog, A. R., & Bachman, J. G. (1981). Effects of questionnaire length on response quality. *Public Opinion Quarterly*, *45*(4), 549. <https://doi.org/10.1086/268687>
- Holton, G. A. (2004). Defining risk. *Financial Analysts Journal*, *60*(6), 19–25. <https://doi.org/10.2469/faj.v60.n6.2669>
- Institute of Internal Auditors (IIA). (2020). The IIA's three lines model: An update of the three lines of defense.
- Institute of Risk Management (IRM). (2012). Risk culture under the microscope: Guidance for boards.
- Kaplan, R. S., & Norton, D. P. (2001). *The strategy-focused organization : How balanced scorecard companies thrive in the new business environment (Strategy focused organization)*. Harvard Business School Press.
- Krosnick, J. A., & Presser, S. (2010). Question and questionnaire design. In P. V. Marsden & J. D. Wright (Eds.), *Handbook of survey research* (pp. 263–314). Elsevier.
- Kulas, J. T., & Stachowski, A. A. (2009). Middle category endorsement in odd-numbered likert response scales: Associated item characteristics, cognitive demands, and preferred meanings. *Journal of Research in Personality*, *43*(3), 489–493. <https://doi.org/10.1016/j.jrp.2008.12.005>

- Kunz, J., & Heitz, M. (2021). Banks' risk culture and management control systems: A systematic literature review. *Journal of Management Control*, 32(4), 439–493. <https://doi.org/10.1007/s00187-021-00325-4>
- Langley, A. (1999). Strategies for theorizing from process data. *Academy of Management Review*, 24(4), 691. <https://doi.org/10.2307/259349>
- Lisdiono, P., Said, J., Yusoff, H., & Hermawan, A. A. (2022). Examining Leadership Capabilities, Risk Management Practices, and Organizational Resilience: The Case of State-Owned Enterprises in Indonesia. *Sustainability*, 14(10), 6268. <https://doi.org/10.3390/su14106268>
- Lucarelli, C., Uberti, P., & Brighetti, G. (2015). Misclassifications in financial risk tolerance. *Journal of Risk Research*, 18(4), 467–482. <https://doi.org/10.1080/13669877.2014.910678>
- MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and Behavioral research: Integrating new and existing techniques. *MIS Quarterly*, 35(2), 293–334. <https://doi.org/10.2307/23044045>
- Malhotra, N., Krosnick, J. A., & Thomas, R. K. (2009). Optimal design of branching questions to measure bipolar constructs. *Public Opinion Quarterly*, 73(2), 304–324. <https://doi.org/10.1093/poq/nfp023>
- Malmendier, U., & Tate, G. (2005). Ceo overconfidence and corporate investment. *Journal of Finance*, 60(6), 2661–2700.
- Manski, C. F., & Molinari, F. (2008). Skip sequencing: A decision problem in questionnaire design. *The Annals of Applied Statistics*, 2(1), 264–285. <https://doi.org/10.1214/07-AOAS134>
- March, J. G., & Shapira, Z. (1987). Managerial perspectives on risk and risk taking. *Management Science*, 33(11), 1404–1418. <https://doi.org/10.1287/mnsc.33.11.1404>
- Marquez-Tejon, J., Jimenez-Partearroyo, M., & Benito-Osorio, D. (2022). Security as a key contributor to organisational resilience: A bibliometric analysis of enterprise security risk management. *Security Journal*, 35(2), 600–627. <https://doi.org/10.1057/s41284-021-00292-4>
- McConnell, P. J. (2013). A risk culture framework for systemically important banks. *Journal of Risk and Governance*, 3, 24–68.
- Mikes, A., & Kaplan, R. S. (2013). Managing risks: Towards a contingency theory of enterprise risk management. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.2311293>
- Monazzam, A., & Crawford, J. (2024). The role of enterprise risk management in enabling organisational resilience: A case study of the Swedish mining industry. *Journal of Management Control*, 35(1), 59–108. <https://doi.org/10.1007/s00187-024-00370-9>
- Moore, D. W. (2002). Measuring new types of question-order effects: Additive and subtractive. *Public Opinion Quarterly*, 66(1), 80–91.
- Moran, L. A., Guyatt, G. H., & Norman, G. R. (2001). Establishing the minimal number of items for a responsive, valid, health-related quality of life instrument. *Journal of Clinical Epidemiology*, 54(6), 571–579. [https://doi.org/10.1016/S0895-4356\(00\)00342-5](https://doi.org/10.1016/S0895-4356(00)00342-5)
- Nielsen, K., Abildgaard, J. S., & Daniels, K. (2014). Putting context into organizational intervention design: Using tailored questionnaires to measure initiatives for worker well-being. *Human Relations*, 67(12), 1537–1560. <https://doi.org/10.1177/0018726714525974>
- Osman, A., & Lew, C. C. (2021). Developing a framework of institutional risk culture for strategic decision-making. *Journal of Risk Research*, 24(9), 1072–1085. <https://doi.org/10.1080/13669877.2020.1801806>
- Palermo, T., Power, M., & Ashby, S. (2017). Navigating institutional complexity: The production of risk culture in the financial sector. *Journal of Management Studies*, 54(2), 154–181. <https://doi.org/10.1111/joms.12241>
- Peytchev, A., & Peytcheva, E. (2017). Reduction of measurement error due to survey length: Evaluation of the split questionnaire design approach. *Survey Research Methods*, 11(4), 361–368. <https://doi.org/10.18148/srm/2017.v11i4.7145>
- Power, M. (2009). The risk management of nothing. *Accounting, Organizations & Society*, 34(6–7), 849–855. <https://doi.org/10.1016/j.aos.2009.06.001>
- Power, M., Ashby, S., & Palermo, T. (2013). Risk culture in financial organizations: A research report: A research report. London School of Economics.
- Preston, C. C., & Colman, A. M. (2000). Optimal number of response categories in rating scales: Reliability, validity, discriminating power, and respondent preferences. *Acta Psychologica*, 104(1), 1–15. [https://doi.org/10.1016/S0001-6918\(99\)00050-5](https://doi.org/10.1016/S0001-6918(99)00050-5)
- Previati, D. A. (2017). People first: Risk culture swings into action. In A. Carretta, F. Fiordelisi, & P. Schwizer (Eds.), *Risk culture in banking* (pp. 125–154). Springer International Publishing.
- Quail, R. (2012). Defining your taste for risk. *Corporate Risk Canada* (Spring 2012), 25–30

- Reay, T., Zafar, A., Monteiro, P., & Glaser, V. (2019). Chapter 10 presenting findings from qualitative research: One size does not fit All! In T. B. Zilber, J. M. Amis, & J. Mair (Eds.), *The production of managerial knowledge and organizational theory: New approaches to writing, producing and consuming theory* (Vol. 59, pp. 201–216, Research in the Sociology of Organizations): Emerald Publishing Limited.
- Renn, O., Laubichler, M., Lucas, K., Kröger, W., Schanze, J., & Scholz, R. W., et al. (2022). Systemic risks from different perspectives. *Risk Analysis*, 42(9), 1902–1920. <https://doi.org/10.1111/risa.13657>
- Ring, P. J., Bryce, C., McKinney, R., & Webb, R. (2016). Taking notice of risk culture – the regulator’s approach. *Journal of Risk Research*, 19(3), 364–387. <https://doi.org/10.1080/13669877.2014.983944>
- Schein, E. H. (1990). Organizational culture. *The American Psychologist*, 45(2), 109–119. <https://doi.org/10.1037/0003-066X.45.2.109>
- Senior Supervisors Group. (2009). Risk management lessons from the global banking crisis of 2008.
- Sheedy, E. A., Griffin, B., & Barbour, J. P. (2017). A framework and measure for examining risk climate in financial institutions. *Journal of Business & Psychology*, 32(1), 101–116. <https://doi.org/10.1007/s10869-015-9424-7>
- Sheedy, E. A., & Jepsen, D. (2018). How to measure risk culture in Australian superannuation funds. *SSRN Electronic Journal*. <https://doi.org/10.2139/ssrn.3128693>
- Sheedy, E., & Griffin, B. (2018). Risk governance, structures, culture, and behavior: A view from the inside. *Corporate Governance an International Review*, 26(1), 4–22. <https://doi.org/10.1111/corg.12200>
- Sheedy, E., Zhang, L., & Tam, K. C. H. (2019). Incentives and culture in risk compliance. *Journal of Banking and Finance*, 107, 105611. <https://doi.org/10.1016/j.jbankfin.2019.105611>
- Shortridge, J., Aven, T., & Guikema, S. (2017). Risk assessment under deep uncertainty: A methodological comparison. *Reliability Engineering and System Safety*, 159, 12–23. <https://doi.org/10.1016/j.res.2016.10.017>
- Simms, L. J., Zelazny, K., Williams, T. F., & Bernstein, L. (2019). Does the number of response options matter? Psychometric perspectives using personality questionnaire data. *Psychological Assessment*, 31(4), 557–566. <https://doi.org/10.1037/pas0000648>
- Sinha, V. K., & Arena, M. (2020). Manifold conceptions of the internal auditing of risk culture in the financial sector. *Journal of Business Ethics*, 162(1), 81–102. <https://doi.org/10.1007/s10551-018-3969-0>
- Smyth, J. D., Dillman, D. A., Christian, L. M., & McBride, M. (2009). Open-ended questions in web surveys. *Public Opinion Quarterly*, 73(2), 325–337. <https://doi.org/10.1093/poq/nfp029>
- Staffini, A., Fujita, K., Svensson, A. K., Chung, U.-I., & Svensson, T. (2022). Statistical methods for item reduction in a representative lifestyle questionnaire: Pilot questionnaire study. *Interactive Journal of Medical Research*, 11(1), e28692. <https://doi.org/10.2196/28692>
- Streicher, B., Bielefeld, M., & Eller, E. (2023). The risk culture framework: Introducing an integrative framework for holistic risk analysis. *Sage Open*. <https://doi.org/10.1177/21582440231191789>
- Toepoel, V., Das, M., & van Soest, A. (2009). Design of web questionnaires: The effects of the number of items per screen. *Field Methods*, 21(2), 200–213. <https://doi.org/10.1177/1525822X08330261>
- Trimpop, R. (1994). *The psychology of risk taking behavior (Advances in psychology, vol. 107)*. Amsterdam: North-Holland.
- Vautier, S., & Pohl, S. (2009). Do balanced scales assess bipolar constructs? The case of the stai scales. *Psychological Assessment*, 21(2), 187–193. <https://doi.org/10.1037/a0015312>
- VonderHaar, R. L., McBride, D. M., & Rosenbaum, D. A. (2019). Task order choices in cognitive and perceptual-motor tasks: The cognitive-load-reduction (CLEAR) hypothesis. *Attention, Perception & Psychophysics*, 81(7), 2517–2525. <https://doi.org/10.3758/s13414-019-01754-z>
- Warnecke, R. B., Johnson, T. P., Chávez, N., Sudman, S., O’Rourke, D. P., & Lacey, L., et al. (1997). Improving question wording in surveys of culturally diverse populations. *Annals of Epidemiology*, 7(5), 334–342. [https://doi.org/10.1016/S1047-2797\(97\)00030-6](https://doi.org/10.1016/S1047-2797(97)00030-6)
- Weick, K. E. (1995). *Sense making in organizations*. Sage
- Weyman, A. K., & Clarke, D. D. (2003). Investigating the influence of organizational role on perceptions of risk in deep coal mines. *Journal of Applied Psychology*, 88(3), 404–412. <https://doi.org/10.1037/0021-9010.88.3.404>
- Wiedemann, A., Bouten, C., & Hertrampf, P. (2020). Risikokultur in Banken – eine systematische Literaturanalyse. *Zeitschrift für Bankrecht und Bankwirtschaft*, 32(5), 286–304. <https://doi.org/10.15375/zbb-2020-0505>

Authors and Affiliations

Niclas Dürst¹  · Jennifer Kunz¹

✉ Niclas Dürst
niclas.duerst@gmail.com

¹ Chair for Management Accounting and Control, Augsburg University, Universitätsstraße
16, D-86159 Augsburg, Germany