Faculteit Geneeskunde en Levenswetenschappen

master in systeem- en procesinnovatie in de gezondheidszorg

Masterthesis

The Resilience Analysis Grid in Healthcare: Exploring Feasibility

David Dirkx

Scriptie ingediend tot het behalen van de graad van master in systeem- en procesinnovatie in de gezondheidszorg

PROMOTOR :

Prof. dr. Jochen BERGS

2024
2025

Faculteit Geneeskunde en Levenswetenschappen

master in systeem- en procesinnovatie in de gezondheidszorg

Masterthesis

The Resilience Analysis Grid in Healthcare: Exploring Feasibility

David Dirkx
Scriptie ingediend tot het behalen van de graad van master in systeem- en procesinnovatie in de gezondheidszorg

PROMOTOR :
Prof. dr. Jochen BERGS

Contents

List of Figures

List of Tables

Dankwoord

Abstract

Background: Healthcare systems face increasing stress due to rising demand, workforce shortages, and resource constraints. Insights from resilience engineering and Safety-II perspectives highlight the need to strengthen healthcare organisations' capacity for resilient performance. Hollnagel's Resilience Analysis Grid (RAG) is proposed as a tool to assess four core potentials of resilience: respond, monitor, learn, and anticipate.

Objective: This study explored the feasibility and face validity of the proposed items of the RAG as a potential basis for survey instruments in healthcare.

Methods: A cross-sectional internet survey was distributed internationally using convenience sampling. For each original RAG item, participants rated feasibility (*I can answer this question*) and face validity or perceived relevance (*This question seems relevant to me*) on a 5-point Likert scale.

Results: Of 32 initial entries, 18 participants provided responses to at least one RAG item and were included; only 11 completed the full survey. Respondents were mainly employed in Belgium, with diverse roles, including nurses, physicians, and managers. Across all four RAG potentials, median scores converged at 4 (*Somewhat agree*). Mean scores and positive proportions (proportion top-2 answers) were generally above 50%, though confidence intervals were wide. Relevance was often rated higher than feasibility, suggesting that some items were considered important but difficult to answer. Informal feedback from non-participants indicated that a lack of familiarity with resilience concepts and the difficult wording of items contributed to non-completion.

Conclusion: This exploratory study highlights both the potential and the limitations of applying a ready-made RAG in healthcare. The RAG items were perceived mainly as feasible and relevant by respondents, though difficulties in understanding suggest that contextualization and theoretical framing are necessary for optimal use.

Introduction

Healthcare systems are increasingly under pressure. Scarcity of resources, rising costs, workforce shortages, and a rising demand for care create structural challenges that threaten the sustainability of high-quality service delivery. (1) These issues are not only acute, but are expected to intensify in the future, raising concerns about how healthcare organisations can continue to function effectively under such constraints. The COVID-19 pandemic exposed vulnerabilities, revealing inadequate infrastructure, resource allocation, and readiness to respond to crises. In other words, it showed how quickly systems can be overwhelmed when demand surges while staff and resources are limited.

In this context, the concept of resilience has gained importance. Resilience has been defined as "*the intrinsic ability of a system to adjust its functioning prior to, during, or following changes and disturbances so that it can sustain required operations under both expected and unexpected conditions.*" (2) Health system resilience was otherwise defined by Kruk et al. as "*the capacity of health actors, institutions, and populations to prepare for and effectively respond to crises; maintain core functions when a crisis hits; and, informed by lessons learned during the crisis, reorganise if conditions require it.*" (3)

Resilience engineering is described by Woods and Hollnagel as a paradigm for safety management, focusing on helping people cope with complexity under pressure to achieve success. (4) A core insight is that both successes and failures result from everyday adaptations, meaning that understanding normal work is as important as investigating adverse events.

This perspective aligns closely with Safety-II thinking, which shifts attention from preventing errors alone to studying how and why things usually go right. (5,6) Safety-II emphasises learning from successful performance as a basis for improving safety and quality. Together, resilience engineering and Safety-II provide a broader view of safety, one that accounts for adaptability and learning in complex systems.

The definition of resilience was made more concrete by Hollnagel et al., pointing to four abilities that are necessary for a system to be resilient. (2) These four abilities (or potentials) are: the ability to respond to events, monitor ongoing developments, anticipate future threats and opportunities, and learn from past failures and successes. Together, they define resilience as an emergent property of a system.

A system is not resilient, but has the potential for resilient performance. (7) To operationalise the insights of resilience engineering, Hollnagel further suggested the Resilience Analysis Grid (RAG), which offers a structured way to assess an organisation's potential for resilient performance. (7,8)

The RAG is by definition based on resilience engineering principles. These principles advocate for a systems view rather than a Cartesian, reductionist view. Successful performance and safety in large complex systems depend on adopting a proactive approach. (9) Healthcare is an adaptive, complex system, where variability is inevitable. This variability is key to both the success and failure of the system. (10) While resilience is vital for organisations across all sectors, it is even more significant in healthcare: when dealing with people's lives and wellbeing, the stakes are high, and the potential consequences of failure can be disastrous.

Considering this, studying a healthcare system's potential for resilient performance seems necessary, and the RAG appears to be the perfect tool for this assessment.

The present study explores whether the original RAG items can be applied directly as a survey instrument in healthcare, by assessing their feasibility (*can practitioners answer the questions?*) and face validity (*do they perceive them as relevant?*).

Methods

Study design and objective

We conducted a cross-sectional survey to assess the feasibility and face validity of applying Hollnagel's Resilience Analysis Grid (RAG) to healthcare settings. The study had two main objectives: 1. to determine the extent to which participants were capable of answering each RAG question as proposed (feasibility), and 2. to assess the extent to which participants perceived each question as relevant for assessing organisational resilience in their context (face validity).

The reporting of this survey was informed by CHERRIES (Checklist for Reporting Results of Internet E-Surveys) and CROSS (Checklist for Reporting of Survey Studies). (11,12)

Setting and participants

Recruitment used convenience sampling: an open survey link was broadly and internationally distributed through professional networks, with encouragement to reshare. The only inclusion criterion was currently working in a formal healthcare organisation; solo practitioners and employees or owners of small practices were excluded. Participation was voluntary; no incentives were offered.

Instrument

The survey was based on the RAG, which organises items around the four "resilience potentials": anticipate, monitor, respond, and learn. For each RAG item, respondents were shown the original question from Hollnagel's original publication. (7) Participants were asked to assess the feasibility and face validity of these items by evaluating two statements: *I can answer this question* and *This question seems relevant to me*, using a 5-point Likert scale (*Strongly disagree; Somewhat disagree; Neither agree nor disagree; Somewhat agree; Strongly agree*).

Participant characteristics (country of employment, profession, role) were collected to describe the sample and enable analyses of subgroup summaries. No identifiable data were collected.

The anonymous online questionnaire was administered via a web survey platform (Qualtrics[XM®]), provided by Hasselt University. After an information page with a short educational video, participants proceeded to the characteristics set (3 questions) and item set (39 questions of 2 statements each). One open field for reflection and ideas was offered at the end. Participants were informed of a 30 to 40-minute survey time. The survey remained open during a window of 1 month; reminders were disseminated through the same channels midway through the window.

Data management and analysis

Data were collected and are stored on Hasselt University servers, stripped of all possible identifiable (meta-)information.

Data were analysed using R (version 4.5.1) and RStudio (version 2025.5.1.513). (13,14)

Given the feasibility aim and expected small item-level sample sizes, analyses were mainly descriptive. Respondent characteristics were summarized. For each RAG item, descriptive statistics were presented (mean, median, min, max), alongside an inferential element: the positive proportion of responses (answers valued ≥4, '*somewhat agree*' and '*strongly agree*') with 95% Wilson confidence interval (CI).

Incomplete survey responses were excluded only if no feasibility or face validity items were answered. For the item-level analyses, all available data were used; if a response to a specific item was missing, it was excluded from the analysis of that item only.

Results

Convenience sampling during a one-month period resulted in 32 participants opening the survey and completing at least one question on participant characteristics. Of these, 18 provided responses to at least one item on feasibility or face validity and were included in the analyses. All other entries were excluded. Ultimately, 11 participants completed the entire survey. The varying sample size is reported consistently at the item level throughout the text, tables, and figures.

Participant characteristics

Profession	N	Percent
Nurse	5	29.4%
Physician	4	23.5%
(care) project manager	1	5.9%
Administrative staff	1	5.9%
Chief Nursing Officer (Directeur patiëntenzorg)	1	5.9%
Directeur (+ psychotherapeut, ergotherapeut, creatief therapeut)	1	5.9%
Nurse, PhD-researcher	1	5.9%
Nursing student	1	5.9%
Postdoc with nursing background	1	5.9%
Psychologist	1	5.9%

Country of Employment		
Country	N	Percent
Belgium	14	77.8%
Italy	2	11.1%
Netherlands	1	5.6%
NA	1	5.6%

Role	N	Percent
Direct patient care	4	22.2%
Head nurse or team lead	4	22.2%
Administrative or managerial responsibilities	3	16.7%
Educational or training duties,Research	2	11.1%
Direct patient care,Administrative or managerial responsibilities	1	5.6%
Direct patient care,Administrative or managerial responsibilities,Educational or training duties,Research	1	5.6%
Direct patient care,Educational or training duties,Research	1	5.6%
Direct patient care,Head nurse or team lead,Administrative or managerial responsibilities,Educational or training duties	1	5.6%
NA	1	5.6%

Table 1. Participant characteristics. *Participants' professions, roles, and countries of employment.*

Of the 18 participants, 14 were employed in Belgium, 2 in Italy, and 1 in the Netherlands. Five participants worked as nurses and four as physicians, while the remaining participants held other functions, including management and administrative functions. Only four reported being exclusively engaged in direct patient care; all others combined clinical work with additional responsibilities, or focused solely on non-clinical roles.

Item-level responses

Likert items are strictly ordinal, for which medians and interquartile ranges can be calculated. However, given the small sample size and the limited variability of medians (often four across items), item-level means were calculated to better illustrate differences between responses in figures. (15)

Full results for all items across four investigated abilities of the RAG are offered in the tables of the Appendix (Table A1-4, p.12).

The ability to respond

According to E. Hollnagel, a system can only respond appropriately to what occurs if it has prepared responses and resources at the ready or if it has the flexibility to make necessary resources available. (7) The original RAG proposes 10 items to investigate a system's ability to respond to changing circumstances and adversity.

The Ability to Respond – feasibility and relevance

Figure 1. The ability to respond - feasibility and relevance. *Dumbbell plot of Likert-scale item means and varying sample sizes.*

Table A1 (p.12) and Figure 1 show that the questions on the ability to respond have been answered 17 or 18 times (n=17 or 18). Descriptive measures were reported in detail in Table A1 (p.12).

Across the board, feasibility was generally scored positively, with only two questions for which the positive proportion of answers was smaller 50%. This proportion was larger than 70% in 3 cases, of which CIs only excluded the 50% mark in one case, namely the very first question: "*What are the events for which the system has a prepared response?*"

Perceived relevance of questions (face validity) scored similarly: only 1 question scored a positive proportion of answers smaller than 50%. This proportion was larger than 70% in 4 cases, of which

the lowest value of the CIs scored greater than 50% in 2 cases: the same question as feasibility, and "*When is a response activated?"*

The ability to monitor

Resilient performance is dependent on monitoring the internal and external situation of a system. What happens in the outside world and within the boundaries of a system should be monitored to respond adequately and timely. In healthcare, as in other industries, adequate monitoring relies on indicators. (7) The RAG proposes 10 items that explore indicators and their measurement.

The Ability to Monitor – feasibility and relevance
n = 14 started, 14 completed

Figure 2. The ability to monitor - feasibility and relevance. *Dumbbell plot of Likert-scale item means and sample sizes.*

Table A2 (p.15) and Figure 2 show the results of 14 responses (n=14) to the questions on the ability to monitor. Descriptive measures are reported in detail in Table A2 (p.15).

Scores were more diverse than previously. Feasibility was scored negatively in 5 cases, with a positive proportion below 50%. A positive proportion above 70% was only measured in 2 cases, of which only 1 had a CI excluding the 50% mark.

Scores for perceived relevance of questions were not identical: only two questions scored a positive proportion of answers smaller than 50%. This proportion was larger than 70% in 3 cases, of which

the lower end of the CIs scored greater than 50% in 2 cases. Relevance scored relatively higher than feasibility in this section, sometimes with significant discrepancy between the scores for feasibility and perceived relevance, as is made visually obvious in Figure 2.

The ability to Learn

The ability to learn from the past is what informs the ability to anticipate and ultimately respond in the future. It combines the results of the monitoring effort and the knowledge that emerges from investigating past responses in or by the system. The effectiveness of the effort to learn as a system depends on both the method and the selection of events to learn from. (7) Again, 10 items are proposed as essential to learning as a system.

The Ability to Learn – feasibility and relevance
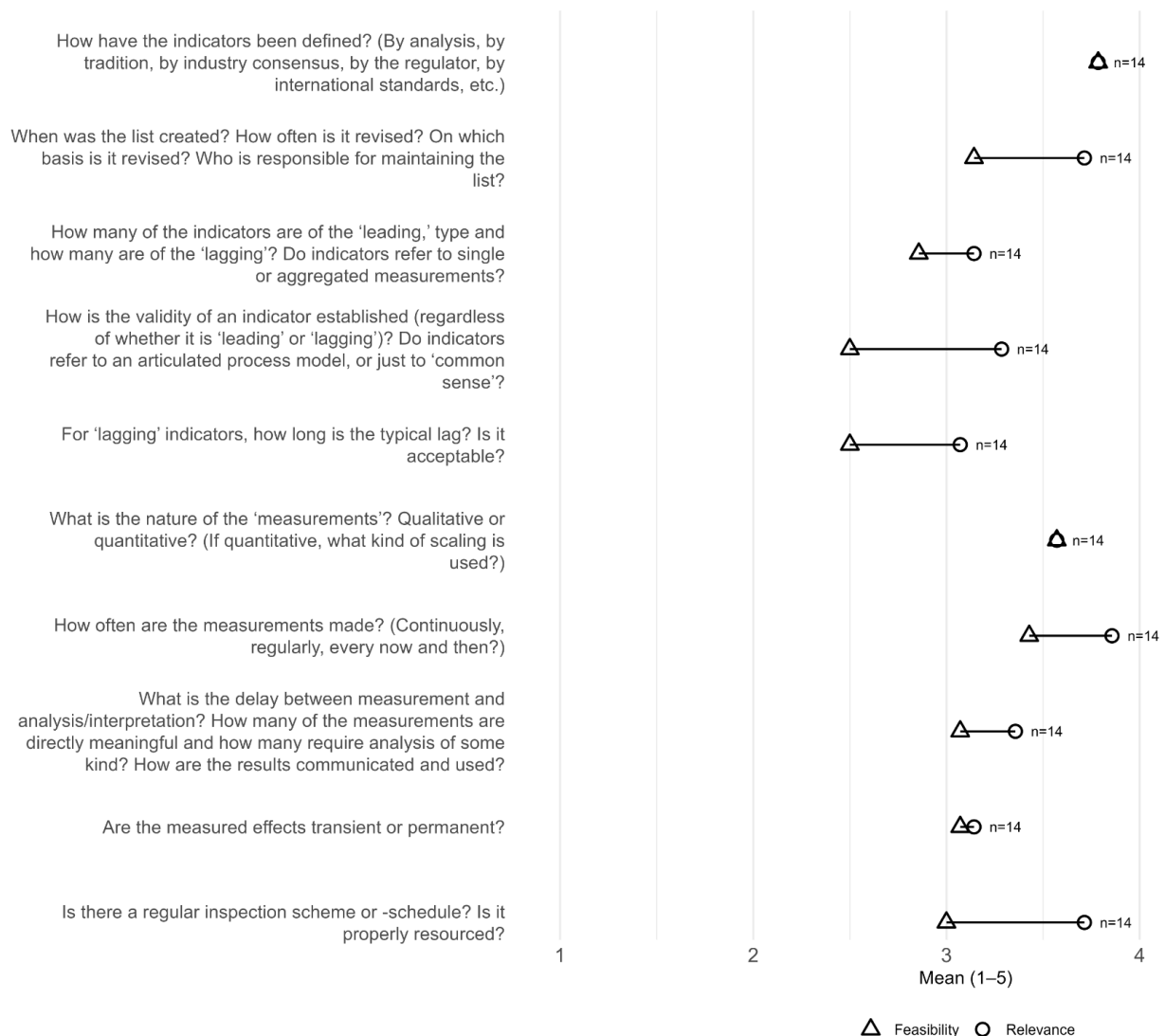n = 11 started, 11 completed

Figure 3. The ability to learn - feasibility and relevance. *Dumbbell plot of Likert-scale item means and sample sizes.*

Table A3 (p.18) and Figure 3 show the results of 11 responses (n=11) to the questions on the ability to learn. Descriptive measures are reported in detail in Table A3 (p.18).

Scores were less diverse. Across the board, scores for both feasibility and perceived relevance were positive. Both means and positive proportions were always greater than 50%. Positive proportions for both feasibility and perceived relevance were greater than 70% in 6 cases, with the lowest score of the CIs greater than 50% in 3 cases each. Nevertheless, these scores are not always linked: in 8 questions, the scores for relevance were higher than those for feasibility. In one case, relevance was scored 4 or 5 in 100% of entries: *Does the system try to learn from successes as well as from failures?*

7

The ability to Anticipate

The potential for resilient performance of a system is greatly increased by its readiness for the future. A system should anticipate future events and conditions, both beneficial and detrimental to its functioning. (7) In this case, nine items are proposed.

The Ability to Anticipate – feasibility and relevance
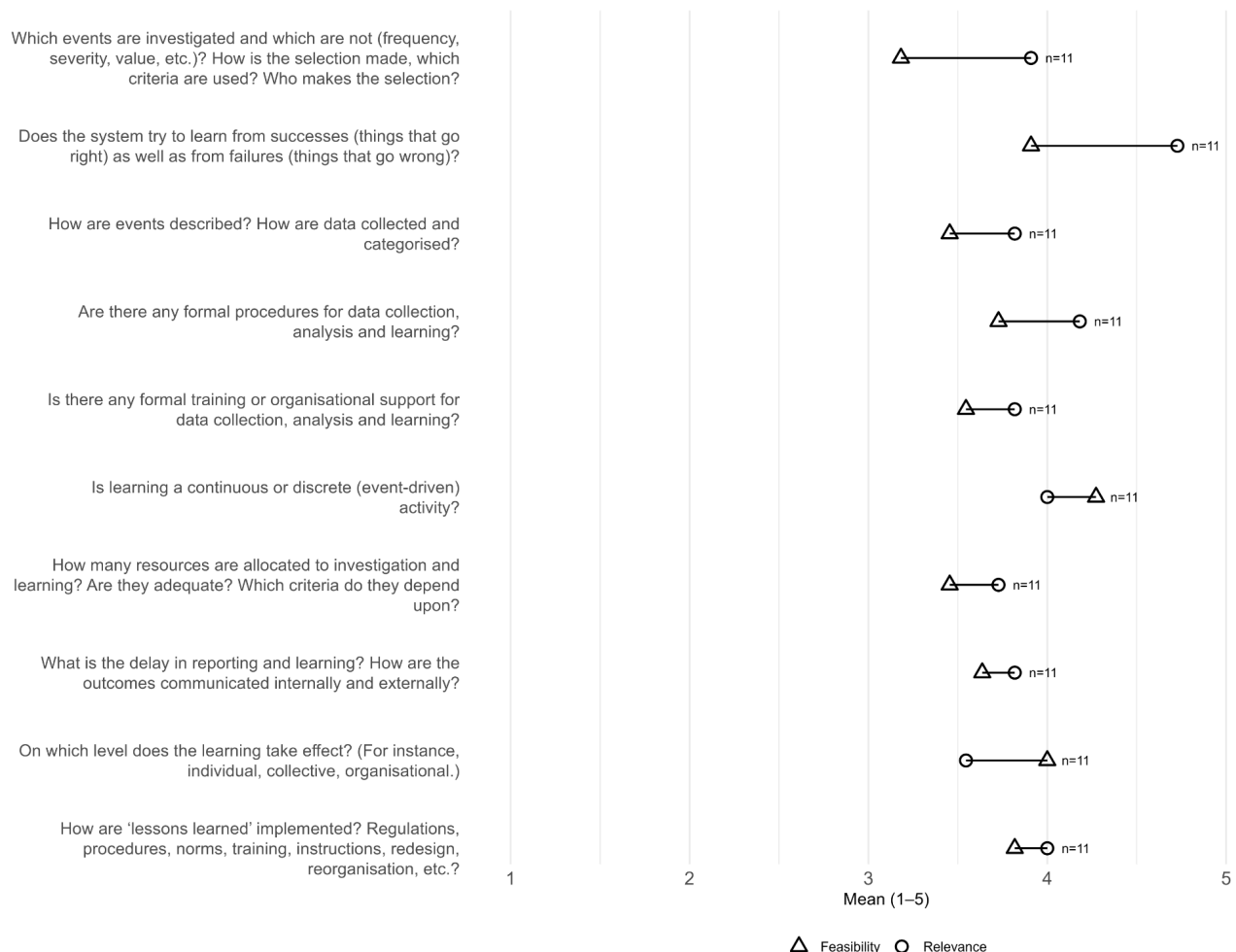n = 11 started, 11 completed
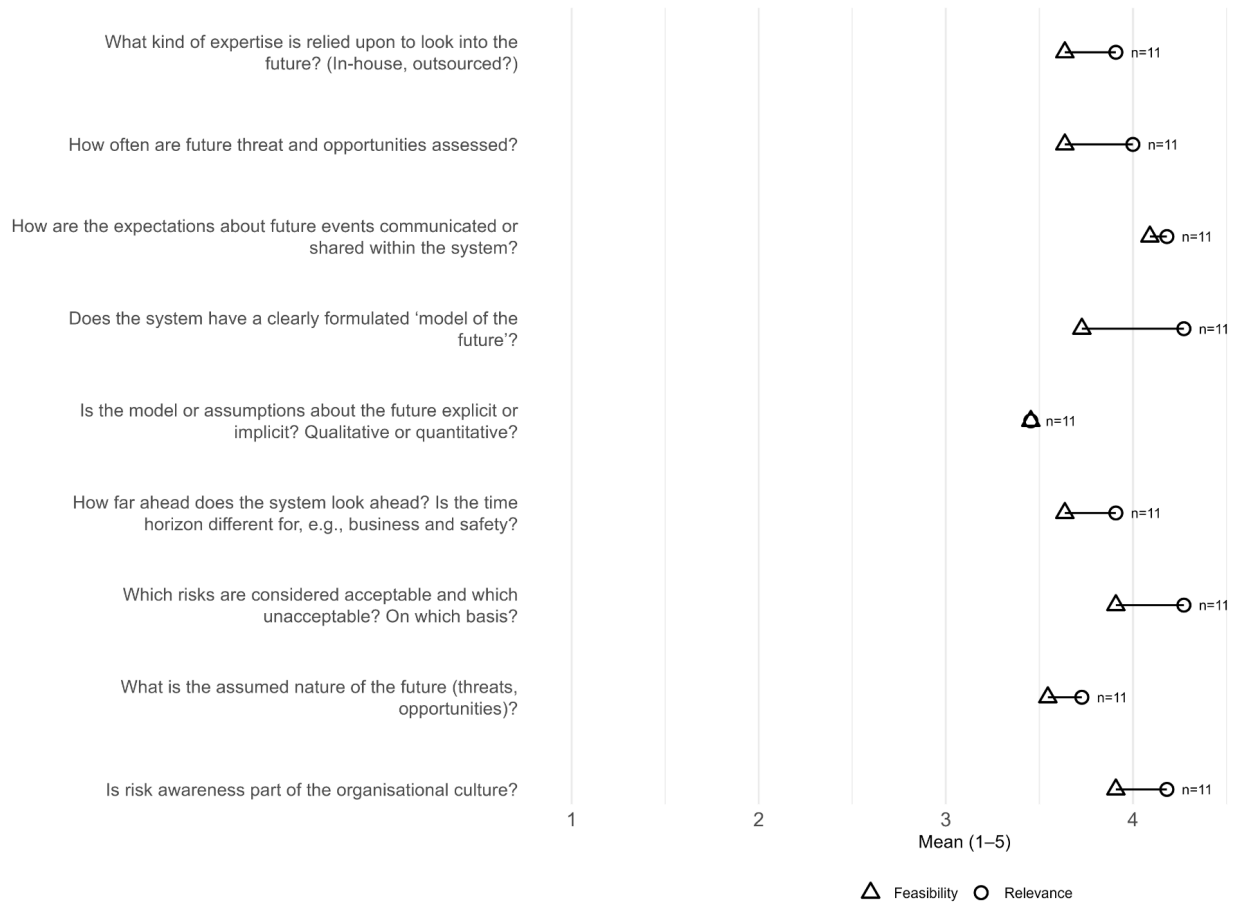
Figure 4. The ability to anticipate - feasibility and relevance. *Dumbbell plot of Likert-scale item means and sample sizes*

Table A4 (p.21) and Figure 4 show the results of 11 responses (n=11) to the questions on the ability to anticipate. Descriptive measures are reported in detail in Table A4 (p.21).

Again, across the board, scores for both feasibility and perceived relevance were positive. Both means and positive proportions were always greater than 50%. Positive proportions for feasibility were greater than 70% in 6 cases, with the lowest score of the CIs greater than 50% once. Perceived relevance was rated with a positive proportion greater than 70% in 7 cases, with lower CI values greater than 50% 5 times.

Overall, as Figure 4 shows, mean scores for relevance were equal to or higher than those for feasibility.

Discussion

Sample size and generalizability

This exploratory study was based on a relatively small convenience sample, which limits the generalizability of the findings and complicates the possibility for meaningful statistical regression. Of 32 initial participants, only 18 answered some questions on feasibility or face validity. Just 11 completed the full survey. All responses to questions on feasibility and face validity (perceived relevance) were included, leading to a varying sample size for all items. The specific sample sizes were reported throughout this paper.

The small sample size resulted in wide CIs for the positive proportion of Likert-scale scores. Item-level medians often converged at 4. Means and positive proportions were presented to illustrate variation.

This study should be interpreted as an exploration of feasibility and face validity rather than as a robust validation of the RAG.

Feedback from non-participants

The possibility of significant response bias must be acknowledged, as more critical perspectives may not have been captured by this survey.

During the timeframe of data collection, several individuals (n=7) contacted us to explain why they chose not to complete the survey. Interestingly, in all cases, these individuals argued that the questions were too difficult to understand. Additionally, 2 participants made similar remarks in the free-text section of the survey responses. Each one of the seven individuals who contacted us expressed the need for more context and a better understanding of the theory that informed the RAG. Although efforts were made to accommodate this need, none of these individuals eventually filled out the survey. As a result, a significant amount of data was not recorded.

The present findings are likely to reflect only the views of those who were more comfortable with the theory and wording. The results of this survey, based on the included sample of 11 to 18 participants, varying across survey items, could have been greatly impacted by the participation of these seven individuals.

Response patterns

Across items, mean scores and the proportion of positive responses were generally above 50%. This indicates that participants who completed the survey tended to regard the items as both feasible and relevant. What stands out is that perceived relevance is often scored higher than feasibility. This suggests that participants considered these questions and topics as important for assessing organisational resilience, even if they found them difficult to answer.

Contextualization of the RAG

The RAG was developed as a flexible tool, intended to be adapted and contextualized in collaboration with the group under study. (7) The difficulties experienced by some non-participants underline the importance of this principle.

In published literature, the RAG is described as a guideline for constructing a resilience profile, advising the development of original questions. (9) A multi-step approach is advised: first, constructing a survey based on the RAG foci and knowledge and experience of the participants, and only then administering the survey to the entire organisation. (7,16) Even when efforts are made to provide a ready-to-use RAG, contextualization is key: Chuang et al. propose the ED-RAG, which is to be used solely for the assessment of the potential of resilient performance of hospital emergency departments. (17)

Direct involvement of participants in contextualizing the items may improve both comprehensibility and engagement in future applications of the RAG. In addition, informal feedback from non-participants suggests that some theoretical knowledge of resilience engineering concepts is appreciated and may be necessary to enable a high-quality assessment.

Limitations and recommendations

A small sample size limits the generalizability of the findings (n=11 to n=18). Significant response bias should be considered, after researchers were made aware of no less than seven participants who did not complete the survey, on the grounds that it was too difficult to understand.

Despite its limitations, this study offers initial insights into how the RAG may be perceived in a healthcare setting. The overall alignment of feasibility and face validity (perceived relevance) ratings suggests that, for those who did respond, the items were both answerable and relevant. The findings of this study affirm the wide recommendation to always contextualize the suggested items of Hollnagel's Resilience Analysis Grid. Future work should involve co-development of the survey items together with the observed healthcare teams.

An interesting yet complex angle for future research would be to study the link between systematic resilient performance measures using the RAG and the impact thereof on outcomes and patient safety.

Conclusion

This cross-sectional study examined the feasibility and face validity of the Resilience Analysis Grid by E. Hollnagel in healthcare through an online survey. While responses indicated that most items were perceived as both answerable and relevant, important limitations were underline by the small sample size, wide confidence intervals, and feedback from non-participants. The findings of this study suggest that applying a standardised RAG without adaptation hinders comprehensibility and participation. Consistent with existing literature, the study supports the need to contextualize RAG items in collaboration with practitioners and to provide adequate theoretical framing when introducing resilience concepts. Future research could focus on linking resilience assessment more directly to patient safety and organisational outcomes.

Declaration on AI tool use

The author confirms that Grammarly (Grammarly Inc.), provided by Hasselt University, was used for language corrections. Generative AI (OpenAI ChatGPT-5) was applied only in statistical analysis, for the correction and automation of R code. The results of this code have been manually checked for their solidity. Generative AI was not used for the writing of this thesis.

References

1. OECD, European Commission. Health at a Glance: Europe 2024: State of Health in the EU Cycle [Internet]. OECD; 2024 [cited 2025 May 16]. (Health at a Glance: Europe). Available from: https://www.oecd.org/en/publications/health-at-a-glance-europe-2024_b3704e14-en.html

2. Hollnagel E, Pariès J, Woods D, Wreathall J, editors. Resilience Engineering in Practice: A Guidebook. Ashgate; 2011. (Ashgate Studies in Resilience Engineering).

3. Kruk ME, Myers M, Varpilah ST, Dahn BT. What is a resilient health system? Lessons from Ebola. The Lancet [Internet]. 2015 May [cited 2025 Jan 11];385(9980):1910–2. Available from: https://doi.org/10.1016/S0140-6736(15)60755-3

4. Hollnagel E, Woods DD, Leveson N, editors. Resilience Engineering: Concepts and Precepts. Ashgate; 2006. 397 p.

5. Hollnagel E, Wears RL, Braithwaite J. From Safety-I to Safety-II: A White Paper. The Resilient Health Care Net; 2015 p. 43.

6. Hollnagel E. Safety-II in practice: developing the resilience potentials. London & New York: Routledge, Taylor & Francis Group; 2018. 130 p.

7. Hollnagel E. RAG - Resilience Analysis Grid. Introduction to the Resilience Analysis Grid (RAG). [Internet]. 2015 [cited 2024 Oct 11]. Available from: https://erikhollnagel.com/onewebmedia/RAG%20Outline%20V2.pdf

8. Alders M, Rafferty AM, Anderson JE. Using the Resilience Assessment Grid to Analyse and Improve Organisational Resilience of a Hospital Ward. In: Nemeth CP, Hollnagel E, editors. Advancing Resilient Performance [Internet]. Switzerland: Springer Nature; 2022 [cited 2024 Oct 11]. 161 p. Available from: https://doi.org/10.1007/978-3-030-74689-6

9. Safi M, Thude BR, Brandt F, Clay-Williams R. The application of resilience assessment grid in healthcare: A scoping review. Tsutsumi Y, editor. PLoS ONE [Internet]. 2022 Nov 4 [cited 2025 Jan 17];17(11):e0277289. Available from: https://doi.org/10.1371/journal.pone.0277289

10. Verhagen MJ, De Vos MS, Sujan M, Hamming JF. The problem with making Safety-II work in healthcare. BMJ Qual Saf [Internet]. 2022 May [cited 2025 Aug 25];31(5):402–8. Available from: https://qualitysafety.bmj.com/lookup/doi/10.1136/bmjqs-2021-014396

11. Eysenbach G. Improving the Quality of Web Surveys: The Checklist for Reporting Results of Internet E-Surveys (CHERRIES). J Med Internet Res [Internet]. 2004 Sept 29 [cited 2025 Aug 24];6(3):e34. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC1550605/

12. Sharma A, Minh Duc NT, Luu Lam Thang T, Nam NH, Ng SJ, Abbas KS, et al. A Consensus-Based Checklist for Reporting of Survey Studies (CROSS). J Gen Intern Med [Internet]. 2021 Oct [cited 2025 Aug 24];36(10):3179–87. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC8481359/

13. R Core Team. R: a language and environment for statistical computing [Internet]. Vienna, Austria: R Foundation for Statistical Computing; 2025. Available from: https://www.R-project.org/

14. Posit team. RStudio: Integrated development environment for R [Internet]. Boston, MA: Posit Software, PBC; 2025. Available from: http://www.posit.co/

15. Sullivan GM, Artino AR. Analyzing and Interpreting Data From Likert-Type Scales. J Grad Med Educ [Internet]. 2013 Dec [cited 2025 Aug 25];5(4):541–2. Available from: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3886444/

16. Alders MDL. A reflective process for analysing organisational resilience to improve the quality of care [Internet]. [London]: King's College; 2019. Available from: https://kclpure.kcl.ac.uk/portal/

17. Chuang S, Ou JC, Hollnagel E, Hou SK. Measurement of resilience potential - development of a resilience assessment grid for emergency departments. PLoS One. 2020;15(9):e0239472.

Appendix - Tabled results and descriptive statistics

All questions in the following tables are directly copied from '*RAG - Resilience Analysis Grid*' by Eric Hollnagel. (7)

The Ability to Respond – results

Question	Type	N	Pos. proportion ≥4 (95% CI)	Mean	Median	Min	Max
What are the events for which the system has a prepared response? - I can answer this question	Feasibility	17	76% (53–90%)	3.88	4	1	5
What are the events for which the system has a prepared response? - This question seems relevant to me	Relevance	17	76% (53–90%)	3.76	4	1	5
How were these events selected? (eg. tradition, regulator requirements, design basis, experience, expertise, risk assessment, industry standard, etc.) - I can answer this question	Feasibility	18	72% (49–88%)	3.83	4	1	5
How were these events selected? (eg. tradition, regulator requirements, design basis, experience, expertise, risk assessment, industry standard, etc.) - This question seems relevant to me	Relevance	18	72% (49–88%)	3.72	4	1	5
When was the list created? How often is it revised? On which basis is it revised? Who is responsible for maintaining and evaluating the list? - I can answer this question	Feasibility	18	56% (34–75%)	3.50	4	1	5
When was the list created? How often is it revised? On which basis is it revised? Who is responsible for maintaining and evaluating the list? - This question seems relevant to me	Relevance	18	61% (39–80%)	3.61	4	1	5
When is a response activated? What is the triggering criterion or threshold? Is the criterion absolute or does it depend on internal / external factors? Is there a trade-off between, e.g., safety and productivity? - I can answer this question	Feasibility	18	72% (49–88%)	3.89	4	1	5

Table A1. The Ability to Respond - results. *Descriptive statistics per question (feasibility and perceived relevance). N; the proportion "positive" results (≥4) of Likert 1-5 (with 95% Wilson CI); mean; median; minimum; maximum. Item-level n varies due to missing data.*

The Ability to Respond – results

Question	Type	N	Pos. proportion ≥4 (95% CI)	Mean	Median	Min	Max
When is a response activated? What is the triggering criterion or threshold? Is the criterion absolute or does it depend on internal / external factors? Is there a trade-off between, e.g., safety and productivity? - This question seems relevant to me	Relevance	18	83% (61–94%)	3.72	4	1	5
How was the specific type of response list decided? How is it ascertained that it is adequate? (Empirically, or based on analyses or models?) - I can answer this question	Feasibility	18	61% (39–80%)	3.22	4	1	5
How was the specific type of response list decided? How is it ascertained that it is adequate? (Empirically, or based on analyses or models?) - This question seems relevant to me	Relevance	18	61% (39–80%)	3.39	4	1	5
How fast is full response ability available? How fast can an effective response be implemented? - I can answer this question	Feasibility	17	47% (26–69%)	3.35	3	2	5
How fast is full response ability available? How fast can an effective response be implemented? - This question seems relevant to me	Relevance	17	53% (31–74%)	3.29	4	1	5
For how long can a 100% effective response be sustained? What is the minimum acceptable response level and how long can it be sustained? - I can answer this question	Feasibility	17	24% (10–47%)	2.71	2	1	5
For how long can a 100% effective response be sustained? What is the minimum acceptable response level and how long can it be sustained? - This question seems relevant to me	Relevance	17	35% (17–59%)	3.00	3	1	5

Table A1. The Ability to Respond - results - continued. *Descriptive statistics per question (feasibility and perceived relevance). N; the proportion "positive" results (≥4) of Likert 1-5 (with 95% Wilson CI); mean; median; minimum; maximum. Item-level n varies due to missing data.*

The Ability to Respond – results

Question	Type	N	Pos. proportion ≥4 (95% CI)	Mean	Median	Min	Max
What is the criterion for ending the response and returning to a "normal" state? - I can answer this question	Feasibility	17	65% (41–83%)	3.65	4	2	5
What is the criterion for ending the response and returning to a "normal" state? - This question seems relevant to me	Relevance	17	71% (47–87%)	3.76	4	1	5
How many resources are allocated to ensure response readiness (people, equipment, materials)? How many are exclusive for the response potential? Who is responsible for maintaining the response ability? - I can answer this question	Feasibility	16	69% (44–86%)	3.50	4	1	5
How many resources are allocated to ensure response readiness (people, equipment, materials)? How many are exclusive for the response potential? Who is responsible for maintaining the response ability? - This question seems relevant to me	Relevance	17	65% (41–83%)	3.59	4	1	5
How is the readiness to respond maintained? How and when is the readiness to respond verified? - I can answer this question	Feasibility	17	53% (31–74%)	3.29	4	1	5
How is the readiness to respond maintained? How and when is the readiness to respond verified? - This question seems relevant to me	Relevance	17	59% (36–78%)	3.41	4	1	5

Table A1. The Ability to Respond - results - continued. *Descriptive statistics per question (feasibility and perceived relevance). N; the proportion "positive" results (≥4) of Likert 1-5 (with 95% Wilson CI); mean; median; minimum; maximum. Item-level n varies due to missing data.*

14

The Ability to Monitor – results

Question	Type	N	Pos. proportion ≥4 (95% CI)	Mean	Median	Min	Max
How have the indicators been defined? (By analysis, by tradition, by industry consensus, by the regulator, by international standards, etc.) - I can answer this question	Feasibility	14	79% (52–92%)	3.79	4.0	1	5
How have the indicators been defined? (By analysis, by tradition, by industry consensus, by the regulator, by international standards, etc.) - This question seems relevant to me	Relevance	14	79% (52–92%)	3.79	4.0	1	5
When was the list created? How often is it revised? On which basis is it revised? Who is responsible for maintaining the list? - I can answer this question	Feasibility	14	50% (27–73%)	3.14	3.5	1	5
When was the list created? How often is it revised? On which basis is it revised? Who is responsible for maintaining the list? - This question seems relevant to me	Relevance	14	79% (52–92%)	3.71	4.0	2	5
How many of the indicators are of the 'leading,' type and how many are of the 'lagging'? Do indicators refer to single or aggregated measurements? - I can answer this question	Feasibility	14	36% (16–61%)	2.86	2.5	1	5
How many of the indicators are of the 'leading,' type and how many are of the 'lagging'? Do indicators refer to single or aggregated measurements? - This question seems relevant to me	Relevance	14	43% (21–67%)	3.14	3.0	2	5
How is the validity of an indicator established (regardless of whether it is 'leading' or 'lagging')? Do indicators refer to an articulated process model, or just to 'common sense'? - I can answer this question	Feasibility	14	21% (8–48%)	2.50	2.0	1	5

Table A2. The Ability to Monitor - results. *Descriptive statistics per question (feasibility and perceived relevance). N; the proportion "positive" results (≥4) of Likert 1-5 (with 95% Wilson CI); mean; median; minimum; maximum. Item-level n varies due to missing data.*

The Ability to Monitor – results

Question	Type	N	Pos. proportion ≥4 (95% CI)	Mean	Median	Min	Max
How is the validity of an indicator established (regardless of whether it is 'leading' or 'lagging')? Do indicators refer to an articulated process model, or just to 'common sense'? - This question seems relevant to me	Relevance	14	50% (27–73%)	3.29	3.5	1	5
For 'lagging' indicators, how long is the typical lag? Is it acceptable? - I can answer this question	Feasibility	14	21% (8–48%)	2.50	2.0	1	5
For 'lagging' indicators, how long is the typical lag? Is it acceptable? - This question seems relevant to me	Relevance	14	50% (27–73%)	3.07	3.5	1	4
What is the nature of the 'measurements'? Qualitative or quantitative? (If quantitative, what kind of scaling is used?) - I can answer this question	Feasibility	14	71% (45–88%)	3.57	4.0	1	5
What is the nature of the 'measurements'? Qualitative or quantitative? (If quantitative, what kind of scaling is used?) - This question seems relevant to me	Relevance	14	64% (39–84%)	3.57	4.0	2	5
How often are the measurements made? (Continuously, regularly, every now and then?) - I can answer this question	Feasibility	14	57% (33–79%)	3.43	4.0	1	5
How often are the measurements made? (Continuously, regularly, every now and then?) - This question seems relevant to me	Relevance	14	71% (45–88%)	3.86	4.0	2	5
What is the delay between measurement and analysis/interpretation? How many of the measurements are directly meaningful and how many require analysis of some kind? How are the results communicated and used? - I can answer this question	Feasibility	14	50% (27–73%)	3.07	3.5	1	5

Table A2. The Ability to Monitor - results - continued. *Descriptive statistics per question (feasibility and perceived relevance). N; the proportion "positive" results (≥4) of Likert 1-5 (with 95% Wilson CI); mean; median; minimum; maximum. Item-level n varies due to missing data.*

The Ability to Monitor – results

Question	Type	N	Pos. proportion ≥4 (95% CI)	Mean	Median	Min	Max
What is the delay between measurement and analysis/interpretation? How many of the measurements are directly meaningful and how many require analysis of some kind? How are the results communicated and used? - This question seems relevant to me	Relevance	14	64% (39–84%)	3.36	4.0	1	5
Are the measured effects transient or permanent? - I can answer this question	Feasibility	14	43% (21–67%)	3.07	3.0	1	5
Are the measured effects transient or permanent? - This question seems relevant to me	Relevance	14	43% (21–67%)	3.14	3.0	1	5
Is there a regular inspection scheme or - schedule? Is it properly resourced? - I can answer this question	Feasibility	14	43% (21–67%)	3.00	3.0	1	5
Is there a regular inspection scheme or - schedule? Is it properly resourced? - This question seems relevant to me	Relevance	14	64% (39–84%)	3.71	4.0	2	5

Table A2. The Ability to Monitor - results - continued. *Descriptive statistics per question (feasibility and perceived relevance). N; the proportion "positive" results (≥4) of Likert 1-5 (with 95% Wilson CI); mean; median; minimum; maximum. Item-level n varies due to missing data.*

The Ability to Learn – results

Question	Type	N	Pos. proportion ≥4 (95% CI)	Mean	Median	Min	Max
Which events are investigated and which are not (frequency, severity, value, etc.)? How is the selection made, which criteria are used? Who makes the selection? - I can answer this question	Feasibility	11	55% (28–79%)	3.18	4	1	5
Which events are investigated and which are not (frequency, severity, value, etc.)? How is the selection made, which criteria are used? Who makes the selection? - This question seems relevant to me	Relevance	11	73% (43–90%)	3.91	4	2	5
Does the system try to learn from successes (things that go right) as well as from failures (things that go wrong)? - I can answer this question	Feasibility	11	82% (52–95%)	3.91	4	2	5
Does the system try to learn from successes (things that go right) as well as from failures (things that go wrong)? - This question seems relevant to me	Relevance	11	100% (74–100%)	4.73	5	4	5
How are events described? How are data collected and categorised? - I can answer this question	Feasibility	11	55% (28–79%)	3.45	4	1	5
How are events described? How are data collected and categorised? - This question seems relevant to me	Relevance	11	73% (43–90%)	3.82	4	1	5
Are there any formal procedures for data collection, analysis and learning? - I can answer this question	Feasibility	11	73% (43–90%)	3.73	4	1	5
Are there any formal procedures for data collection, analysis and learning? - This question seems relevant to me	Relevance	11	82% (52–95%)	4.18	4	2	5
Is there any formal training or organisational support for data collection,	Feasibility	11	64% (35–85%)	3.55	4	1	5

Table A3. The Ability to Learn - results. *Descriptive statistics per question (feasibility and perceived relevance). N; the proportion "positive" results (≥4) of Likert 1-5 (with 95% Wilson CI); mean; median; minimum; maximum. Item-level n varies due to missing data.*

The Ability to Learn – results

Question	Type	N	Pos. proportion ≥4 (95% CI)	Mean	Median	Min	Max
analysis and learning? - I can answer this question							
Is there any formal training or organisational support for data collection, analysis and learning? - This question seems relevant to me	Relevance	11	64% (35–85%)	3.82	4	2	5
Is learning a continuous or discrete (event-driven) activity? - I can answer this question	Feasibility	11	82% (52–95%)	4.27	4	3	5
Is learning a continuous or discrete (event-driven) activity? - This question seems relevant to me	Relevance	11	73% (43–90%)	4.00	4	2	5
How many resources are allocated to investigation and learning? Are they adequate? Which criteria do they depend upon? - I can answer this question	Feasibility	11	64% (35–85%)	3.45	4	1	5
How many resources are allocated to investigation and learning? Are they adequate? Which criteria do they depend upon? - This question seems relevant to me	Relevance	11	64% (35–85%)	3.73	5	1	5
What is the delay in reporting and learning? How are the outcomes communicated internally and externally? - I can answer this question	Feasibility	11	73% (43–90%)	3.64	4	1	5
What is the delay in reporting and learning? How are the outcomes communicated internally and externally? - This question seems relevant to me	Relevance	11	64% (35–85%)	3.82	4	2	5
On which level does the learning take effect? (For instance, individual, collective, organisational.) - I can answer this question	Feasibility	11	82% (52–95%)	4.00	4	2	5

Table A3. The Ability to Learn - results - continued. *Descriptive statistics per question (feasibility and perceived relevance). N; the proportion "positive" results (≥4) of Likert 1-5 (with 95% Wilson CI); mean; median; minimum; maximum. Item-level n varies due to missing data.*

The Ability to Learn – results

Question	Type	N	Pos. proportion ≥4 (95% CI)	Mean	Median	Min	Max
On which level does the learning take effect? (For instance, individual, collective, organisational.) - This question seems relevant to me	Relevance	11	64% (35–85%)	3.55	4	2	5
How are 'lessons learned' implemented? Regulations, procedures, norms, training, instructions, redesign, reorganisation, etc.? - I can answer this question	Feasibility	11	73% (43–90%)	3.82	4	2	5
How are 'lessons learned' implemented? Regulations, procedures, norms, training, instructions, redesign, reorganisation, etc.? - This question seems relevant to me	Relevance	11	82% (52–95%)	4.00	4	2	5

Table A3. The Ability to Learn - results - continued. *Descriptive statistics per question (feasibility and perceived relevance). N; the proportion "positive" results (≥4) of Likert 1-5 (with 95% Wilson CI); mean; median; minimum; maximum. Item-level n varies due to missing data.*

The Ability to Anticipate – results

Question	Type	N	Pos. proportion ≥4 (95% CI)	Mean	Median	Min	Max
What kind of expertise is relied upon to look into the future? (In-house, outsourced?) - I can answer this question	Feasibility	11	64% (35–85%)	3.64	4	2	5
What kind of expertise is relied upon to look into the future? (In-house, outsourced?) - This question seems relevant to me	Relevance	11	82% (52–95%)	3.91	4	2	5
How often are future threat and opportunities assessed? - I can answer this question	Feasibility	11	73% (43–90%)	3.64	4	1	5
How often are future threat and opportunities assessed? - This question seems relevant to me	Relevance	11	73% (43–90%)	4.00	4	3	5
How are the expectations about future events communicated or shared within the system? - I can answer this question	Feasibility	11	82% (52–95%)	4.09	4	2	5
How are the expectations about future events communicated or shared within the system? - This question seems relevant to me	Relevance	11	82% (52–95%)	4.18	4	3	5
Does the system have a clearly formulated 'model of the future'? - I can answer this question	Feasibility	11	73% (43–90%)	3.73	4	1	5
Does the system have a clearly formulated 'model of the future'? - This question seems relevant to me	Relevance	11	91% (62–98%)	4.27	4	3	5
Is the model or assumptions about the future explicit or implicit? Qualitative or quantitative? - I can answer this question	Feasibility	11	73% (43–90%)	3.45	4	1	5

Table A4. The Ability to Anticipate - results. *Descriptive statistics per question (feasibility and perceived relevance). N; the proportion "positive" results (≥4) of Likert 1-5 (with 95% Wilson CI); mean; median; minimum; maximum. Item-level n varies due to missing data.*

The Ability to Anticipate – results

Question	Type	N	Pos. proportion ≥4 (95% CI)	Mean	Median	Min	Max
Is the model or assumptions about the future explicit or implicit? Qualitative or quantitative? - This question seems relevant to me	Relevance	11	64% (35–85%)	3.45	4	1	5
How far ahead does the system look ahead? Is the time horizon different for, e.g., business and safety? - I can answer this question	Feasibility	11	73% (43–90%)	3.64	4	2	5
How far ahead does the system look ahead? Is the time horizon different for, e.g., business and safety? - This question seems relevant to me	Relevance	11	73% (43–90%)	3.91	4	2	5
Which risks are considered acceptable and which unacceptable? On which basis? - I can answer this question	Feasibility	11	73% (43–90%)	3.91	4	2	5
Which risks are considered acceptable and which unacceptable? On which basis? - This question seems relevant to me	Relevance	11	82% (52–95%)	4.27	4	3	5
What is the assumed nature of the future (threats, opportunities)? - I can answer this question	Feasibility	11	64% (35–85%)	3.55	4	2	5
What is the assumed nature of the future (threats, opportunities)? - This question seems relevant to me	Relevance	11	64% (35–85%)	3.73	4	2	5
Is risk awareness part of the organisational culture? - I can answer this question	Feasibility	11	73% (43–90%)	3.91	4	1	5
Is risk awareness part of the organisational culture? - This question seems relevant to me	Relevance	11	91% (62–98%)	4.18	4	3	5

Table A4. The Ability to Anticipate - results - continued. *Descriptive statistics per question (feasibility and perceived relevance). N; the proportion "positive" results (≥4) of Likert 1-5 (with 95% Wilson CI); mean; median; minimum; maximum. Item-level n varies due to missing data.*