Chapter 6
Inter-rater Agreement Based Risk Assessment Scheme for ICT Corporates

Roberto Cassata, Gabriele Gianini, Marco Anisetti, Valerio Bellandi, Ernesto Damiani, and Alessandro Cavaciuti

Abstract An ISO 9001 audit can be seen as an independent risk assessment on the business, where each 'Nonconformity" or "Opportunity For Improvement" is considered as a potential risk. Nevertheless, their actual impact on the business remains difficult to determine; as a consequence, the urgency of a mitigation plan at corporate level can sometimes be underestimated. This paper proposes a semi-quantitative risk assessment methodology on the ISO 9001 findings relying on a selected panel of experts. The experts' responses are analyzed and validated using a specific statistics test for inter-rater reliability. The proposed methodology has been applied on real findings coming from ISO 9001 internal audits, involving 10 subject matter experts of 7 different countries.

6.1 Introduction

Risk-assessment is a widely discussed research topic, and number of solutions and standards were proposed. Similar to the Delphi Technique [8] our risk assessment methodology estimates the likelihood of an event by asking a panel of experts; but

R. Cassata · A. Cavaciuti
Cisco Systems, San Jose, USA
e-mail: rcassata@cisco.com

A. Cavaciuti
e-mail: acavaciu@cisco.com

G. Gianini · M. Anisetti · V. Bellandi (✉) · E. Damiani
Universitá degli Studi di Milano, Milan, Italy
e-mail: valerio.bellandi@unimi.it

G. Gianini
e-mail: gianini.gabriele@unimi.it

M. Anisetti
e-mail: anisetti.marco@unimi.it

E. Damiani
e-mail: damiani.ernesto@unimi.it

I. Czarnowski et al. (eds.), *Intelligent Decision Technologies*,
Smart Innovation, Systems and Technologies 309,
https://doi.org/10.1007/978-981-19-3444-5_6

instead of running the assessment until participants reach consensus, we measure concordance among raters with a statistics named Kappa Inter-rater agreement, where Kappa is a score expressing the consensus level.

In this paper we specifically target large ICT corporations where the adoption of a risk-based approach translates into: (i) identification of risks and opportunities, (ii) plan of actions to address them, (iii) implementation in a quality management system and (iv) evaluation of effectiveness.

An ISO 9001 audit can be seen as the 1st step of the risk assessment, also called "Risk Identification" in the ISO 31000:2018 Risk Management standard nomenclature [4]. The outcome of the audit is an identified set on nonconformities and opportunities for improvement that have associated risks with potential impacts on the business. For each finding, the actions taken to prevent or mitigate the associated risk and to evaluate the effectiveness of the mitigation are generally local to the audited organization and not propagated to other organizations belonging to the same corporation. This paper proposes a methodology to perform a Risk Assessment using a systematic and structured approach involving a panel of experts who provide their judgments on the findings identified during an audit. The judgments of the experts are evaluated using a specific approach for inter-rater reliability assessment, computing the level of inter-rater agreement as measured by a set of metrics inspired by the Cohen Kappa [2]: each one of those metrics—whose exact formulation depends on the nature of the response categories used in the assessment (nominal, ordinal, numeric)—measures how much the raters agree with each other, and discounts the effect of agreements occurring by chance, computed via probabilistic methods. For eliciting the expert judgment, we designed a survey focused on identifying (i) risk categories, (ii) risk probability, (iii) risk impact and (iv) risk profile cost.

6.2 The Methodology

Figure 6.1 shows the risk Assessment approach as defined by ISO 31000 (left portion) and the relation with our approach (right portion). In our methodology, the Risk Identification is executed during the audit itself. Risk Analysis is done by the panel of selected experts using our structured survey and the Risk Evaluation is summarized in our Business Risk Scorecard.

Findings Selection Our starting point are the finding discovery procedures, according to Fig. 6.1, the auditor is in charge of reviewing all the ISO 9001 findings opened in a given time frame, selecting the relevant ones. During this phase, findings details are collected in order to provide sufficient information to the experts. We consider the following ISO 9001 finding details as most relevant: (i) the finding title, (ii) the finding description, (iii) the requirement and (iv) the potential business impact.

Panel of Experts Selection Criteria The auditor is responsible of the selection of a panel of experts that will analyze the risks associated with the identified findings. The selected experts must be independent and in a position that allows them to

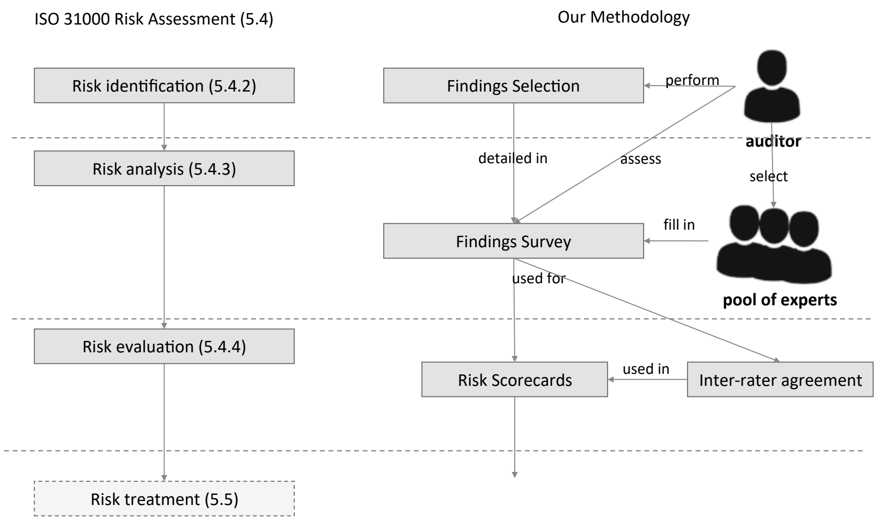

Fig. 6.1 ISO 31000 risk assessment process compared to the proposed methodology

make unbiased judgment. They must have agreed to treat all the ISO 9001 finding information as confidential since they contain sensitive information. The auditors should select the experts based on (i) their experience, (ii) their technical expertise, (iii) their business background, (iv) their knowledge of company processes, (v) their roles in the corporation.

Survey Structure The panel of experts evaluates the findings using survey we specifically designed for this purpose. The survey is focused on typical ICT risks. The survey is articulated in four main assessments: (i) Risk Categories, (ii) Risk Cost Profile, (iii) Risk Probability and (vi) Risk Impact. Experts are asked to select the applicable Risk Categories (multiple choices), to select the Risk Profile Cost (single choice), Risk Probability (single choice), Risk Impact (single choice). In the following we detail each of the above assessment.

Risk Categories assessment The identification of a list of risks associated to an ISO 9001 finding depends to the nature of the business and on the target technology. Table 6.1 shows our risk catalogue made of a set of five main risk categories and a subset of impacted areas. The expert is requested to select for each of the five categories the relevant impacted areas.

Profile Cost assessment Risk treatment is a decision-making process whereby risks are treated by selecting and implementing measures to address the specific risks identified in the risk assessment and subsequent risk analysis. Nowadays projects stakeholders involved in development activities typically do their risk analysis considering two aspects: risk probability and risk impact. Table 6.2 shows a typical risk matrix considering probability and impact.

Table 6.1 The proposed risk catalogue

Risk categories	Impacted areas
Strategic risk Are risks associated to the decisions taken by directors that could have an impact on organization business objectives	Approval of product/service delivery even if the quality goals are not met Decision impacting credit, financial aspects Decision impacting infrastructures availability Decision impacting resources and headcount Other strategic risks
Operational Risk—Legal The risk that a counter-party to a transaction will not be liable to meet its obligations under law	Intellectual property (IP) Brand protection and reputation Legal Lawsuit Other Legal risks
Operational risk Risks related to inadequate processes, resource that could results in ineffective or inefficient product/service delivery and reduction of the organization margins	Inefficient process (process generates waste of resource/time/money/etc.) Ineffective process (process is unable to produce desired results) KPIs are not capturing relevant indicators and/or can't be used for business improvement Other operational risks
Compliance risk Risk arising from failure to comply with process, laws and regulations	Ability to sell product/service in specific countriesRisk arising from failure to comply with process, laws and regulations Health and safety compliance ESD compliance Product/service limitations (e.g: blind Color GUI) Compliance with international standard Other product/service compliance risk
Technical risk Risks related to product/service technical aspects that could result in customer dissatisfaction	Testability Performance Reliability Availability Scalability Security Maintainability Other technical risk

This catalogue is used during the Risk Categories assessment. It is defined based on our expertise in risk assessment evaluation

To have a more accurate risk analysis of corporate business risks it is important to consider another factor: the *cost profile*. For example, a risk could have a negligible initial cost but with an exponential cost profile over time that could results in unsustainable cost.

The cost profile represents the economic impact of a delay in implementing a risk control, where risk control represents a "measure that maintains and/or modifies risk" (ISO 31000:2018 3.8) [4]. In other words the finding risk's cost profile is the impact on the corporate projects costs if no control or mitigation are implemented.

Table 6.2 Probability/impact matrix

	Risk probability		
Risk impact	Rare	Possible	Probable
LOW	1	2	3
MEDIUM	2	4	6
HIGH	3	6	9

Risk Probability/Impact assessment Risk is also analyzed in relation to its potential impact using the risk-matrix in Table 6.2. More formally, the risk score is defined as $score = f_r(Probability, Impact)$ where f is the function that maps probability and Impact to the score using the risk matrix. Any risk generates a cost for the corporate so, we decided to ask to the expert the impact in terms of corporate monetary cost, defined in our case as LOW ($LOW \leq$ 10.000\$), MEDIUM (10.000\$ $< MEDIUM <$ 100.000\$), HIGH ($HIGH \geq$ 100.000\$). We decided to describe the Risk probability in the next 2 years in qualitative terms, such as Rare: $Probability \leq 10\%$, Possible: $10\% < Probability < 50\%$ Probable: $Probability \geq 50\%$.

Table 6.3 shows the survey structure as it is presented to the panel of experts in relation to a given finding.

6.2.1 *Determination of the Inter-rater Agreement*

In order to quantify the level of inter-rater agreement for a given setting it is customary the design, a metric inspired by the Cohen Kappa [2]: i.e. one computes the *observed value* of the agreement π—according to some chosen metric—then compares the outcome to its *expected value* for the case of a random choice of the options by the experts. The *rate of improvement* with respect to the performance of the random choice process is then adopted as the value of the Kappa metric, which takes the following form

$$\kappa \equiv \frac{\pi - \langle \pi \rangle}{1 - \langle \pi \rangle} \tag{6.1}$$

where π is the *observed agreement rate*, and $\langle \pi \rangle$ the *chance-expected agreement rate*. The maximum value achievable by the Kappa metric is 1. Most often, the Kappa issued by a study is used for *benchmarking* with respect to an ordinal scale of qualitative agreement expressions, such as the scale devised by [7]: $\leq 0 \rightarrow$ "Poor agreement" $0.00 - 0.20 \rightarrow$ "Slight agreement", $0.21 - 0.40 \rightarrow$ "Fair agreement", $0.41 - 0.60 \rightarrow$ "Moderate agreement", $0.61 - 0.80 \rightarrow$ "Substantial agreement", $0.81 - 0.99 \rightarrow$ "Almost perfect agreement", $1.00 \rightarrow$ "Perfect agreement".

Table 6.3 Survey structure example described in Sect. 6.3

Engineering quality assessment finding

DESCRIPTION: Confidentially level of the design documents is inconsistent and inappropriate, Design documents information can be shared only when it's necessary and with people who need to know

EVIDENCE: Configuration Management System: J-AX All documents are stored in Joker Directory

REQUIREMENTS: ISO 7.5.3.1 Document Information required by the quality management system shall be controlled to ensure: b) it adequately protected /e.g. from loss of confidentially, improper use, or loss of the integrity

IMPACT: Inappropriate breaches of confidentially can lead to legal action and/or competitors' advantages

Question	Answer options
1. **Strategic Risk**	Approval of product/service delivery even if the quality goals are not met Decision impacting credit, financial aspects Decision impacting infrastructures availability Decision impacting resources and headcount Other: …
2. **Operational Risk—LEGAL**	Intellectual property (IP) Brand Protection and Reputation Legal Lawsuit Other: …
3. **Operational Risk**	Inefficient process (Process generates waste of resources/time/money/etc.) Ineffective process (process) KPI are not capturing relevant indicators and/or can't be used for business improvement Other: …
4. **Compliance Risks**	Ability to sell product/service in specific countries Health and safety compliance ESD Compliance Product/Service limitations (e.g. blind color GUI) Compliance with international standard Other: …
5. **Technical Risk**	Testability Performance Reliability Availability Scalability Security Maintainability Other: …
6. **Risk Assessment—Probability**	RARE POSSIBLE PROBABILE
7. **Risk Assessment—Impact**	LOW MEDIUM HIGH
8. **Risk Assessment—Profile Cost**	Constant Logarithmic Linear Exponential

6.2.2 Kappa Coefficient Formulation

The original definition of κ by Cohen has spawned a number of variants that fit diverse settings. A comprehensive review can be found in [3]. Here we consider the assessment of each object (each Finding) by a number n of raters. The setups most relevant to the present work are the following:

1. Ordinal *mutually exclusive* response categories, with multiple level scale ($k > 2$), assessed by $n > 2$ raters which applies to Probability, Impact and Cost Profile assessment
2. Nominal, multiple-level scale, *non-mutually-exclusive* response categories, assessed by $n > 2$ raters, which applies to Strategic Risk, Operational Risk—Legal, Operational Risk, and Compliance Risk assessment.

Rating with nominal response categories, with single choice over multiple level scale ($k > 2$), **assessed by** $n > 2$ **raters** Let the index $i = 1, 2, \ldots, N$ represent the objects, let the index $j = 1, 2, \ldots, k$ represent the categories, and let the index $h = 1, 2, \ldots, n$ represent the raters. Let r^i_j be the number of raters that have assigned object i to category j. Then to quantify the agreement over the fact that the category j is assigned to object i one can count the number of pairs $r^i_j(r^i_j - 1)/2$. Since we are assuming here that each rater assigns the object to exactly one of the k categories, a natural way for quantifying the agreement in a *purely nominal* setting consists of counting how many rater pairs agree over a category and to compare it to the maximum agreement achievable, i.e. to compute the ratio $\pi^i_j \equiv r^i_j(r^i_j - 1)/n(n-1)$. The agreement over object i is the sum over the categories, and the overall agreement is the average over the number N' of the objects that were rated by at least one pair

$$\overline{\pi} = \frac{1}{N'}\sum_{i=1}^{N'}\pi^i \quad with\ \pi^i = \sum_{j=1}^{k}\frac{r^i_j(r^i_j - 1)}{n(n-1)}$$

Most agreement coefficients share the same definition of the observed agreement: they differ in the expression of the the chance expected agreement. The simplest choice for the chance expected rate is the one by Brennan and Prediger [1], i.e. $\langle\pi\rangle_B = 1/k$: when the rating of an object is a random process, the object is be assigned to any of the k categories with equal probability $1/k$. Plugging $\overline{\pi}$ and $\langle\pi\rangle_B$ in equation (6.1) one obtains the Brennan and Prediger's Kappa.

Rating with ordinal response categories, with single choice over multiple level scale ($k > 2$), **assessed by** $n > 2$ **raters**. One can treat the ordinal setting as a nominal setting enriched with extra structure, which weights differently the disagreements of categories which are located near and far in the ranking: one can for instance stipulate that ratings where an object is assigned to categories closer in the ordering represent less serious a disagreement than ratings where the object is assigned to categories that are located farther in the ordering. This can be formalized introducing a weight matrix $w_{j\ell}$ such that the matrix element $w_{j\ell} = 1$ when $j = \ell$ and such that the the matrix element $w_{j\ell}$ is non-zero, when the categories j and ℓ are meant to be considered a partial agreement. This leads to the definition of a weighted count $\tilde{r}$, which accounts also for the "cross-talk" among categories, and to a corresponding weighted count of the pairwise agreements $r^i_j(\tilde{r}^i_j - 1)$. Overall the average over objects is

$$\overline{\pi} \equiv \frac{1}{N'} \sum_{i=1}^{N'} \left(\sum_{j=1}^{k} \frac{r^i_j(\tilde{r}^i_j - 1)}{n(n-1)} \right) \quad with\ \tilde{r}^i_j \equiv \sum_{\ell=1}^{k} w_{j\ell}\, r^i_\ell \tag{6.2}$$

Possible choices for the weights are linear $w_{j\ell} = 1 - |j - \ell|/(k-1)$, quadratic $w_{j\ell} = 1 - (j - \ell/(k-1))^2$, square root $w_{j\ell} = 1 - (|j - \ell|/(k-1))^{1/2}$ or power of a fixed number, such as $w_{j\ell} = 1/3^{|j-\ell|}$.

Brennan-Prediger agreement coefficient The Brennan-Prediger agreement coefficient [1] is defined by Eq. (6.2) for the observed agreement rate π and by the following definition of the chance expected agreement (index B denotes the Brennan-Prediger definition)

$$\langle \pi \rangle_B \equiv \frac{1}{k^2} \sum_{j,\ell} w_{j,\ell}.$$

The Kappa index for multi-choice nominal categories in the case of n raters. Consider now the *multi-choice* case, i.e. assume that each expert can assign to an object up to k distinct and *non-mutually exclusive properties* (each property being expressed by a response category). Consider a pair of experts, indexed by g and h: with reference to an object i, we denote by $\mathcal{A}^i_g$ the set of options for which rater g has expressed positive opinion, we denote this set cardinality by $a^i_g \equiv card(\mathcal{A}^i_g)$, and call it *response cardinality* for rater g; similarly we denote the response cardinality of rater h by $a^i_h \equiv card(\mathcal{A}^i_h)$. For quantifying the agreement over object i we count only the number x^i_{gh} of response categories in which both experts say $True$, i.e. the Positive Agreements

$$x^i_{gh} \equiv card(\mathcal{A}^i_g \cap \mathcal{A}^i_h)$$

In the case of single-choice constraint on e has $x^i_{gh} \in \{0, 1\}$, whereas in the case of multi-choice $x^i_{gh} \in \{0, 1, \ldots, \min(a^i_g, a^i_h)\}$. The quantity x^i_{gh} has to be compared to a reference value. We use $\min(a^i_g, a^i_h)$ (indeed, given a^i_h and a^i_g, the maximum achievable number of agreements is the minimum of the two numbers) and define the rate of agreement as follows.

$$\pi^i_{gh} \equiv \frac{x^i_{gh}}{\min(a^i_g, a^i_h)} \tag{6.3}$$

Notice that when $\min(a^i_h, a^i_g) = 0$—i.e. when at least one of the raters in the pair does not make any assessment for the object—the quantity π is undefined.

The most straightforward way to generalize the rate of agreement from 2 to n raters (i.e. $\nu = n(n-1)/2$ unordered rater pairs) consists of taking the average of the pairwise agreements over the ν pairs

$$\pi^i_{r.a} \equiv \frac{1}{\nu} \sum_{g=1}^{n-1} \sum_{h=g+1}^{n} \pi^i_{gh} = \frac{1}{\nu} \sum_{g=1}^{n-1} \sum_{h=g+1}^{n} \frac{x^i_{gh}}{\min(a^i_g, a^i_h)} \tag{6.4}$$

we call this *rates-average* definition (it is denoted by the index *r.a.*).

Now we consider that the variable x^i_{gh} is described by an Hypergeometric distribution ([6]) so that

$$\langle x^i_{gh} \rangle = \frac{a^i_g a^i_h}{k} = \frac{\max(a^i_g, a^i_h)\min(a^i_g, a^i_h)}{k}$$

It follows that for the expected value we have

$$\langle \pi^i_{r.a.} \rangle = \frac{1}{\nu} \sum_{g=1}^{n-1} \sum_{h=g+1}^{n} \frac{\langle x^i_{gh} \rangle}{\min(a^i_g, a^i_h)} = \frac{1}{\nu k} \sum_{g=1}^{n-1} \sum_{h=g+1}^{n} \max(a^i_g, a^i_h) = \frac{1}{\nu k} \sum_{h=1}^{n} (h-1)\max(a^i_g, a^i_h)$$

Plugging the last expression and Eq. (6.4) in expression (6.1) defines $\kappa_{r.a.}$.

Notice that this expression refers to a single object i: if needed one can average over the objects which have received at least a pair of ratings. However, in this work we focus on the ratings of individual objects. This is the formulation of $\kappa_{r.a.}$ that we used in this work for the *non-mutually exclusive nominal response* options, which corresponded to the Risk Categories.

6.2.3 Business Risk Scorecard

The *business risk scorecard* was conceived to have a structured and concise report focused on providing the *risk evaluation* ensuring a consistent view with the relevant information necessary to top management to understand easily: (i) the risks categories exposure, (ii) the results of impact and probability assessment and (iii) the results of the cost profile assessment. We extended the traditional risk analysis score defined as a function of probability (likelihood of an event) and the impact (its consequences) to include the cost profile. We call this new score function *Business Risk Severity* and is more formally defined as: $BusinessRiskSeverity = f_s(f_r(Probability, Impact), CostProfile)$, where f_r is the risk score function as defined in Sect. 6.2 and f_s is the severity mapping function that takes the risk score and the cost profiles and map them to the level of severity using the Business Risk Severity matrix defined in Table 6.4.

The inter-rater agreement is computed and reported in the business risks scorecard as an index of the level of reliability agreement of the experts.

Table 6.4 Business risk severity matrix

Risk score f_r	Cost profile			
	Const.	Log.	Linear	Exp.
9				
6				
4				
3				
2				
1				

The Business Risk Severity is depicted in red (High) yellow (Moderate) and green (minor)

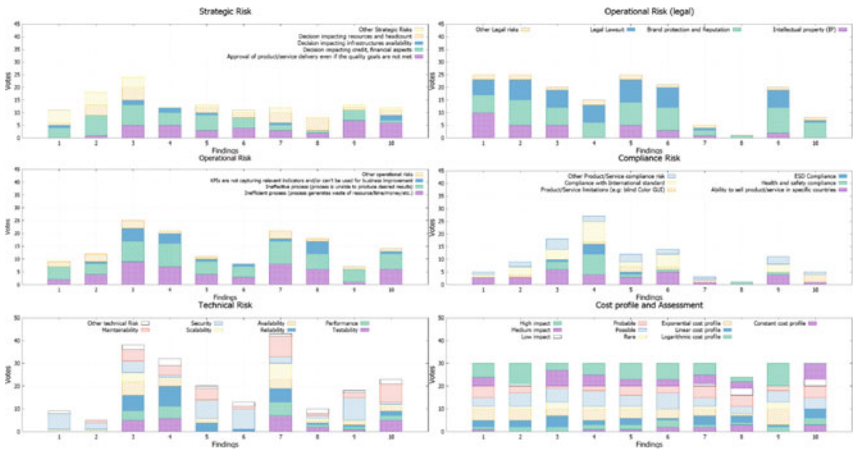

Fig. 6.2 Experts' votes aggregated by findings and risk categories

6.3 Case Study

Having selected the findings the auditor identifies the panel of experts that shows the right competences to evaluate the selected findings according to the criteria in Sect. 6.2. Before starting asking the panel of experts to fulfill our survey, we first presented the ISO 9001 findings (i.e., via brainstorming) we responded to any clarification request. We then submitted the survey to the panel of experts (Survey structure is presented in Sect. 6.2).

Figure 6.2 shows survey results in terms of experts' votes relative to all the findings of our case study aggregated by categories. Note that, the plots of Strategic, Operational, Compliance and Technical categories are affected by multiple choice votes in terms of the absolute numbers presented.

Inter-rarer Agreement. The collected survey data are used to evaluate our inter-rater agreement following the approach in Sect. 6.2.1. Tables 6.5, and 6.6 show the results of the computation of the different *kappas*.

Business Risk scorecard. For each ISO 9001 finding a Business Risk Scorecard is generated as described in Sect. 6.2.3.These risk scorecards summarizes all the impor-

Table 6.5 The value of the ratio-average κ_{ra} for the nominal variables

Categories	01 κ_{ra}	02 κ_{ra}	03 κ_{ra}	04 κ_{ra}	05 κ_{ra}	06 κ_{ra}	07 κ_{ra}	08 κ_{ra}	09 κ_{ra}	10 κ_{ra}
Strategic risk	0.44	0.52	0.14	0.39	0.44	0.15	–0.04	0.56	0.60	0.38
Op erational risk legal	1.00	1.00	0.51	0.68	0.65	0.87	–0.67		0.82	0.58
Operational risk	0.23	–0.06	0.13	0.60	0.33	0.58	0.69	0.44	0.91	0.43
Compliance risk	0.33	0.19	0.32	0.73	0.00	0.46	–0.50		0.30	0.33
Technical risk	0.97	1.00	–0.29	0.40	0.79	0.74	0.52	–0.33	0.93	0.63

The columns refer to the findings, the rows to the different risk categories. The six shades of color give the range in which the number falls, according to Landis and Koch [7]. The cell is white when there are less than two raters expressing a judgment on that case

Table 6.6 The value of the Brennan-Prediger κ_B for the ordinal variables with weight given by $w_{j\ell} = 1/3^{|j-\ell|}$

Categories	01 κ_B	02 κ_B	03 κ_B	04 κ_B	05 κ_B	06 κ_B	07 κ_B	08 κ_B	09 κ_B	10 κ_B
Cost profile	0.08	0.15	0.18	0.02	–0.01	–0.08	–0.03	–0.07	0.31	0.08
Risk probability	0.10	0.19	0.19	0.31	0.11	0.40	0.10	0.31	0.03	0.29
Risk impact	0.31	0.61	0.40	0.29	0.40	0.40	0.10	–0.10	1.00	0.40

See caption of Table 6.5 for the color code conventions tant elements considered by our approach and helps in prioritizing risk mitigation actions.

6.4 Discussion

The ISO 9001 finding owner, who is responsible to implement the correction and the preventive action is not always able to provide a reliable evaluation of the risk impact at corporate level. The reliability and the investment in consistent preventive actions (Risk Treatment) depends on knowledge of the different corporate entities and the ability to understand the complexity of the risk. Moreover, each risk can have multiple ramifications impacting several aspects of the business like: financial, infrastructure, brand reputation, security, health, safety, etc. In order to improve Risk Assessment accuracy, we propose an innovative approach that involves a pool of experts coming from different areas of the business and implements a methodology based on four new key ideas:

1. a defined *Risk Catalogue* (see Table 6.1) with the list of the typical ICT risks,
2. the new concept of *Profile Cost* (see Sect. 6.2),
3. the introduction of the *Inter-rater Kappa* (see Sect. 6.2.1) to measure the agreement among the panel of experts and
4. the adoption of the new *Business Risk Scorecard* (see Sect. 6.2.3) to have a structured and concise report on the *risk evaluation* to support top management in the decision making process.

An advantage of the proposed methodology is the low interaction among the experts and this could be an important factor that would naturally reduces bias and further investigations will be conducted on this aspect.

6.5 Conclusions

Uncertainty is a key concept in risk conceptualization and risk assessment; several methodologies to conduct risk assessment are described in ISO 31010 [5]. We think that the approach presented in this research is a valid methodology that potentially could be generalized to become a valid risk assessment technique. The proposed methodology has been applied on real findings coming from several different ISO 9001 internal audits and 10 subject matter experts from 7 different European countries have been involved.

References

1. Brennan, R.L., Prediger, D.J.: Coefficient kappa: some uses, misuses, and alternatives. Educ. Psychol. Measure. **41**(3), 687–699 (1981)
2. Cohen, J.: A coefficient of agreement for nominal scales. Educ. Psychol. Measure. **20**(1), 37–46 (1960)
3. Gwet, K.L.: Handbook of Inter-rater Reliability: The Definitive Guide to Measuring the Extent of Agreement Among Raters. Advanced Analytics, LLC (2014)
4. ISO Central Secretary: ISO 31000:2018 risk management—guidelines. Standard, International Organization for Standardization, Geneva, CH (2018)
5. ISO Central Secretary: ISO 31010:2019 risk management—risk assessment techniques. Standard, International Organization for Standardization, Geneva, CH (2019)
6. Kupper, L.L., Hafner, K.B.: On assessing interrater agreement for multiple attribute responses. Biometrics, 957–967 (1989)
7. Landis, J.R., Koch, G.G.: The measurement of observer agreement for categorical data. Biometrics, 159–174 (1977)
8. Linstone, H.A., Turoff, M., et al.: The Delphi Method. Addison-Wesley, Reading (1975)