

SEI Journal
Science, Education and Innovations in the Context of Modern Problems
ISSN (Print): 2790-0169
ISSN (Online): 2790-0177
DOI Prefix: 10.56334/sei
International Multidisciplinary • Peer-Reviewed • Open Access Journal

RESEARCH ARTICLE

Evaluation of internal control systems: operational and financial processes, weaknesses and practical improvements

Chibalonza Mpenda
Researcher
Universidad Internacional Iberoamericana (UNINI)
Mexico
https://orcid.org/0001-0009-3571-6186
E-mail: Chibalonza.mpenda@pesquisa.unic.co.ao

Keywords internal control; COSO; internal audit; risk management; Monte Carlo simulation

Abstract

Internal control systems (ICS) protect an organisation only when their weaknesses are measured, located and corrected in order of importance, yet in medium-sized organisations control is often designed as a formality and assessed impressionistically. This study proposes and demonstrates a quantitative protocol for evaluating ICS that analyses operational and financial processes, identifies weaknesses and prioritises practical improvements, grounded in COSO (2013) and the IIA (2024) Global Internal Audit Standards. All data were simulated (random seed 20261112) for a medium-sized organisation and do not describe any real entity. The protocol combines attribute testing of 48 key controls across eight processes (1,794 items tested), a COSO maturity survey (n = 96), cluster-robust logistic regression, Benford and threshold analysis of 4,252 payables invoices, a risk matrix with Monte Carlo simulation (100,000 runs) and a three-year cost-benefit analysis. The overall deviation rate was 6.7% (95% CI 5.6%–7.9%), and four of eight processes were classified as ineffective. The COSO maturity index was 49.0 out of 100, with monitoring the weakest component (2.28 out of 5). Segregation-of-duties conflicts raised the odds of a deviation almost threefold (OR = 2.95; 95% CI 1.90–4.58), while documented evidence lowered them (OR = 0.48). Invoice analytics revealed clustering just below the approval limit and 52 duplicate pairs worth MZN 9.0 million. Expected annual loss was MZN 7.8 million; eight actions costing MZN 4.25 million would cut it by 64.6% under full effectiveness (38.9% under a conservative scenario), with a three-year net present value of MZN 5.77 million. Weaknesses concentrate in segregation of duties, supplier master data, privileged access and monitoring. Multi-criteria prioritisation and quantification of residual risk allow improvements to be ranked and the investment to be justified to management.

Citation

Mpenda, C. (2026). Evaluation of internal control systems: operational and financial processes, weaknesses and practical improvements. *Science, Education and Innovations in the Context of Modern Problems, 9*(11), 1–22.
https://doi.org/10.56334/sei/9.11.6

Licensed

© 2026 The Author(s). Published by *Science, Education and Innovations in the Context of Modern Problems (SEI)*, under the auspices of IMCRA – International Meetings and Conferences Research Association (Azerbaijan).

This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original work is properly cited.

http://creativecommons.org/licenses/by/4.0/

Received: August 14, 2026 **Accepted**: November 06, 2026 **Published Online**: November 19, 2026

1. INTRODUCTION

An internal control system (ICS) protects an organisation only when its weaknesses are measured, located and corrected in order of importance. The literature and professional practice converge on this point, yet what is frequently observed in medium-sized organisations is internal control designed as a formality and assessed impressionistically (Musikavanhu & Nqala, 2024). The cost of this gap is high. The global study by the Association of Certified Fraud Examiners, based on 1,921 cases from 138 countries and losses of more than USD 3.1 billion, attributed more than half of occupational frauds to a lack of controls (32%) or to their override by the people who should have complied with them (19%), and showed that 43% of cases were detected by a tip (ACFE, 2024). This distance between formal design and real operation reinforces the need for assessments based on measurable evidence.

The problem addressed in this article is both methodological and practical: how to assess, in a replicable and quantified way, an internal control system that spans operational and financial processes, so as to identify the relevant weaknesses and propose improvements that management can prioritise and fund. The usual answer, the qualitative checklist, has three limitations. It does not indicate the margin of uncertainty of its conclusions, it does not separate the design of a control from its effective operation, and it does not link deficiencies to expected losses, which makes the investment decision harder.

Objectives of the study

The general objective is to demonstrate a quantitative protocol for assessing internal control, applied to an organisation X, which starts from test evidence and ends in a prioritised portfolio of improvements. The specific objectives are:

1. to measure the operating effectiveness of key controls by process, with confidence intervals;
2. to assess ICS maturity against the five COSO components, based on employees' perceptions;
3. to identify the control attributes associated with a higher probability of failure;
4. to detect anomalies in accounts-payable transaction data using analytical techniques;
5. to quantify the residual risk and the expected benefit of the proposed improvements.

Research questions

The study answers four research questions. **RQ1:** Which processes show deviation rates incompatible with a 5% tolerance? **RQ2:** Which COSO components show the lowest maturity, and are there differences between departments? **RQ3:** Which control characteristics (automation, preventive nature, documented evidence, segregation-of-duties conflict) are associated with deviations? **RQ4:** Which portfolio of actions maximises the reduction in expected loss per unit of cost?

Hypotheses

On the basis of the literature, the following hypotheses are formulated and tested in the Findings:

- **H1.** Automated controls show a lower deviation rate than manual controls.
- **H2.** Segregation-of-duties (SoD) conflicts increase the probability of deviation, controlling for the other characteristics.
- **H3.** COSO maturity differs between departments.
- **H4.** Monitoring is the component with the lowest maturity, below the target level of 3.5.

The association between entity-level control weaknesses and financial reporting fraud (Donelson et al., 2017) and the African evidence that internal control and internal audit contribute to financial accountability (Kolawole, 2026) support, respectively, H2 and the relevance of monitoring in H4.

Theoretical framework

The COSO framework

The most widely used framework for designing and assessing internal control is the COSO *Internal Control – Integrated Framework*, in its 2013 version. It defines internal control as a process, effected by an entity's board of directors, management and other personnel, designed to provide reasonable assurance regarding the achievement of objectives relating to operations, reporting and compliance (COSO, 2013). It is structured in five components and 17 principles. For the system to be considered effective, each component and each relevant principle must be present and functioning, and the five components must operate together (COSO, 2013). This condition has a practical consequence for assessment: a serious deficiency in one component, for example monitoring, is not offset by good results in another, for example the control environment.

Table 1. COSO (2013) components and their correspondence with the protocol of this study

Component	Principles	Question the assessment seeks to answer	Evidence used in this study
Control environment	1–5	Does management set the example, define responsibilities and demand competence?	Survey (5 items)
Risk assessment	6–9	Are relevant risks, including fraud risk, identified and analysed?	Survey (4 items); risk register
Control activities	10–12	Are there actions that mitigate risks and are they performed as designed?	Attribute tests; regression; invoice analytics
Information and communication	13–15	Is relevant information produced, shared and usable?	Survey (4 items)
Monitoring	16–17	Are deficiencies evaluated, communicated and corrected in time?	Survey (4 items); improvement plan

Source: prepared by the author, based on COSO (2013).

Enterprise risk management and sustainability reporting

The *Enterprise Risk Management – Integrating with Strategy and Performance* framework (COSO, 2017) has a broader scope than internal control, because it links risk to strategy and performance, and is organised in five components and 20 principles. In March 2023, COSO published supplementary guidance on internal control over sustainability reporting, which applies the 17 principles of 2013 to non-financial information (COSO, 2023). For the subject of this article, the relevance is twofold: the assessment of an ICS should start from objectives and risks (and not from controls), and the same testing logic can be extended to new types of information.

Three Lines Model and Global Internal Audit Standards

The Three Lines Model, updated by the IIA in 2020, replaced the image of "lines of defence" with six principles that emphasise governance, the roles of the governing body, management and internal audit, the independence of the third line and value creation (IIA, 2020). Management and the risk and compliance functions (first and second lines) are responsible for designing and operating controls; internal audit provides independent assurance and advice. This separation guides the protocol: attribute tests assess the operation of first-line controls, whereas the maturity assessment and the portfolio of improvements are addressed to management and the governing body.

The Global Internal Audit Standards were published on 9 January 2024 and became applicable on 9 January 2025, replacing the 2017 standards (IIA, 2024). They include requirements on the use of technology by audit functions, which supports the incorporation of data analytics in the protocol (see Accounts-payable transaction analytics in the Methods).

Risk assessment in the audit of financial statements

ISA 315 (Revised 2019), applicable to audits of periods beginning on or after 15 December 2021, requires a more robust identification and assessment of the risks of material misstatement, with separate assessment of inherent risk and control risk, and introduces inherent risk factors such as complexity, subjectivity, change and uncertainty (IAASB, 2019). The separation between inherent and residual risk, used in the risk matrix presented in the Findings, follows this logic.

Fraud, opportunity and control deficiencies

The fraud literature rests on three conditions, usually described as pressure, opportunity and rationalisation. Internal control acts on opportunity. Donelson et al. (2017) showed that material control weaknesses precede the disclosure of financial reporting fraud and that the effect is mainly due to entity-level weaknesses rather than to specific processes. This distinction justifies the protocol's combination of process-level controls (attribute tests) and entity-level controls (maturity survey).

In this article, a **control deficiency** is a statistically relevant deviation between the control as designed and the control as practised; a **significant deficiency** is one whose observed deviation rate exceeds 10% or whose residual exposure is classified as critical in the risk matrix. The Methods operationalise these definitions.

Literature review

Internal control, fraud and the quality of financial reporting

The relationship between the quality of internal control and the quality of financial information is the most investigated. In French SBF 120 companies, with 686 firm-year observations between 2012 and 2018, the disclosure of control weaknesses was negatively

associated with discretionary accruals and positively associated with real activities manipulation, which indicates that a good control system limits earnings management through accruals but not necessarily through operating decisions (Boulhaga et al., 2022). Manry et al. (2023) relate financial statement fraud litigation to the existence of material weaknesses and to board characteristics. On the prevention side, Mandal and Amilan (2023) found that fraud awareness partially mediates the relationship between the internal control system and the prevention of financial statement fraud, which suggests that controls produce an effect when employees understand the risk. The review by Hazzaa et al. (2022) on corporate governance and internal control systems treats both as determinants of firms' financial performance, and the public-sector literature highlights the role of the control environment in preventing fraud (Harahap & Isgiyarta, 2023).

Internal audit and accountability in developing countries

Recent African evidence is consistent on the importance of internal control and internal audit but heterogeneous on the factors that make them effective. In Nigerian local governments (n = 125, ordered logit), internal audit and the internal control system had positive and significant effects on financial accountability, with a larger effect of the latter and signs of complementarity between the two (Kolawole, 2026). In Cameroonian public limited companies, coordination between internal and external auditors and the outsourcing of internal audit increased effectiveness, whereas team size, the existence of an internal control system, board independence and CEO duality showed no significant effect (Kueda et al., 2026). The second result draws attention to a point that this article takes up: the mere formal existence of a system does not guarantee effectiveness. In Mozambique, an analysis of ICS indicators in three public universities, using ANOVA, detected significant differences between institutions in all COSO dimensions (Graziano, 2023), which motivates the comparison between units carried out in the Findings.

Small and medium-sized enterprises

SMEs are the context closest to the illustrative organisation. In Cape Town, many SMEs design and implement control systems as a formality, with no link to strategic objectives, and show a gap in understanding of their use (Musikavanhu & Nqala, 2024). In Uganda, a cross-sectional study of 361 firms found a positive and significant relationship between internal audit (controls, procedures and practices) and SME sustainability (Apon et al., 2024), and another, in Mbale, found no significant association between internal controls and financial performance (Otai & Birungi, 2025). In Nigeria, most SME owners set up internal control reactively, after losses from employee fraud, and the number of frauds differed significantly between proactive and reactive strategies (Ololade, 2024). In Mozambique, research on SME survival describes firms that are poorly structured for sustained strategic management and that adapt to the market reactively (Bila, 2023), a pattern compatible with equally reactive controls.

Data analytics, Benford's law and artificial intelligence

Data analytics moves auditing from sampling to the analysis of complete populations. In anomaly detection, Benford's law remains the entry-level technique: Sylwestrzak (2023) applied it to the detection of earnings management in Poland by splitting the sample between firms with and without fraud, and Sushkov et al. (2023) combine Benford tests with clustering to reduce the limitations of the isolated test. A systematic review of 100 peer-reviewed studies (2015–2025) on artificial intelligence in auditing identified machine learning (58%), natural language processing (31%) and robotic process automation (24%) as the dominant techniques (Anwar & Akeel, 2026).

Table 2. Selected recent studies

Study	Context and method	Main result
Donelson et al. (2017)	US firms; material weaknesses and fraud	Strong association with future fraud disclosure; effect linked to entity-level weaknesses
Boulhaga et al. (2022)	France, 686 firm-year observations; OLS and GMM	Disclosed weaknesses associated with fewer discretionary accruals and more real manipulation
Mandal & Amilan (2023)	Prevention of reporting fraud; mediation model	Fraud awareness partially mediates the effect of the ICS
Kolawole (2026)	Nigeria, local governments, n = 125; ordered logit	Internal control and audit improve accountability; larger effect of the ICS
Apon et al. (2024)	Uganda, 361 SMEs; cross-sectional	Internal audit positively associated with sustainability
Ololade (2024)	Nigeria, SMEs	Reactive control predominates; fewer frauds with a proactive strategy

Study	Context and method	Main result
Graziano (2023)	Mozambique, three universities; ANOVA	Significant differences between institutions in all COSO dimensions
Kueda et al. (2026)	Cameroon, public limited companies; OLS	Internal–external coordination and outsourcing increase effectiveness; team size does not
Sylwestrzak (2023)	Poland; Benford's law	Application to earnings management with a sample split between fraud and control

Source: prepared by the author, based on the cited literature.

Gaps and positioning

Three gaps emerge. Most studies measure internal control by perception or by disclosure, and few measure it through tests of control operation. The result of tests is rarely linked to an expected monetary loss that supports investment decisions. And quantitative evidence on medium-sized organisations in Southern Africa is scarce. The protocol presented fills the first two gaps by construction and provides a transferable method for the third.

Potential contributions of the present study

The article contributes in three ways. First, it integrates into a single protocol techniques usually treated separately (attribute sampling, maturity survey, regression, digital analysis and risk simulation). Second, it links the result of the tests to a monetary estimate of residual risk and of the return on improvements, which is the language of management decisions. Third, it documents the data design and parameters, so that the protocol can be adapted to real data. The Methods and Materials, which describe the protocol and characterise the organisation, follow the Introduction; the Findings present the results and the proposed improvements; the Discussion places them in the literature and sets out limitations and future research; and the final section concludes.

2. METHODS AND MATERIALS

Study design and nature of the data

The study is cross-sectional and of mixed quantitative methods: it combines tests of control operation (objective evidence), a perception survey (subjective evidence), transaction analysis (complete population) and risk modelling. The data were generated by simulation, with seed 20261112, in Python (*numpy*, *scipy*, *statsmodels*), from plausible parameters for a medium-sized organisation. The parameters were fixed before the analysis. Because the data are simulated, the results demonstrate how the protocol works and how its statistics are interpreted; they do not constitute empirical evidence on any entity or on the population of Mozambican organisations.

Attribute tests and operating effectiveness

The population of key controls is 48 controls in eight processes. The sample size per control depends on the frequency of operation: 12 items for monthly controls (the full population for the year), 25 for weekly, 40 for daily and 60 for transaction-level controls. A **deviation** is the failure of a control attribute in a tested item (for example, a missing approval or an unreviewed reconciliation). The deviation rate is $\hat{p} = k / n$. To express uncertainty, the 95% Wilson interval (Wilson, 1927) and the one-sided 95% Clopper–Pearson upper limit (UCL), which is the maximum plausible value of the true rate, are used.

Wilson interval (1):

$$\tilde{p}_{\pm} = \frac{\hat{p} + \frac{z^2}{2n} \pm z\sqrt{\frac{\hat{p}(1-\hat{p})}{n} + \frac{z^2}{4n^2}}}{1 + \frac{z^2}{n}} \quad (1)$$

Clopper–Pearson upper limit (2):

$$UCL_{95} = B^{-1}(0.95;\ k+1,\ n-k) \quad (2)$$

The tolerable deviation rate (TDR) is 5%. Each process is classified as **effective** if UCL ≤ 5%, **effective with reservations** if the observed rate is ≤ 5% but the UCL exceeds 5%, and **ineffective** if the observed rate exceeds 5%. At the level of the individual control, the classification is: effective (≤ 5%), deficiency (> 5% and ≤ 10%) and significant deficiency (> 10%).

Process control risk index

To rank processes by priority, the process control risk index (PCRI) is calculated. It combines the uncertainty of the deviation rate with three control design attributes:

$$PCRI_j = 100\,[\,0.50\,m(UCL_j) + 0.20\,m(M_j) + 0.15\,m(D_j) + 0.15\,m(S_j)\,] \tag{3}$$

where M is the proportion of manual controls, D the proportion without documented evidence, S the proportion with an SoD conflict, and $m(\cdot)$ the min–max normalisation across processes (0 to 1). The weights reflect the primacy of operating evidence and were tested in the sensitivity analysis (see Strengths and limitations).

COSO maturity survey

The instrument has 22 items on a Likert scale from 1 (strongly disagree) to 5 (strongly agree), distributed across the five components (Table 3). It was administered to 96 employees from four departments (finance, 28; operations, 30; purchasing and logistics, 20; human resources and IT, 18). The score of each component is the mean of its items; the COSO maturity index (CMI) is the mean of the five components, converted to 0–100 by $100 \times (mean - 1)/4$. The target level is 3.5 ("managed"). Internal reliability is measured by Cronbach's alpha:

$$\alpha = \frac{k}{k-1}\left(1 - \frac{\sum_{i=1}^{k}\sigma_i^2}{\sigma_T^2}\right) \tag{4}$$

Table 3. Structure of the survey instrument

Component	Items	Examples of statements
Control environment	5	Management communicates and practises ethical values; responsibilities are assigned in writing
Risk assessment	4	The risks of each process are identified and reviewed at least once a year
Control activities	5	Approvals and checks are carried out by a person other than the one who performs the operation
Information and communication	4	Management reports arrive on time and with sufficient quality to support decisions
Monitoring	4	Identified deficiencies are followed up until resolved (M2)

Source: prepared by the author.

Differences between the components and the target level are tested with the one-sample t-test; differences between departments with the Kruskal–Wallis test (H statistic, effect size $\varepsilon^2 = (H - k + 1) / (N - k)$) and, as a check, ANOVA; associations between components with Spearman's coefficient.

Logistic model of deviations

The level of analysis is the tested item (n = 1,794). The dependent variable is the occurrence of a deviation (1/0). The regressors are: automated control (*Auto*), preventive control (*Prev*), documented evidence (*Doc*) and SoD conflict (*SoD*). Because items are clustered within 48 controls, standard errors are cluster-robust by control, which avoids underestimating uncertainty. Odds ratios (OR) with 95% CI, McFadden's R^2 and the area under the ROC curve (AUC) are reported:

$$\ln\frac{P(Y=1)}{1-P(Y=1)} = \beta_0 + \beta_1\,Auto + \beta_2\,Prev + \beta_3\,Doc + \beta_4\,SoD \tag{5}$$

Accounts-payable transaction analytics

The population of 4,252 invoices is analysed in full. Four tests are applied. The first is Benford's law for the first digit, with expected probability $P(d) = \log_{10}(1 + 1/d)$, the chi-square test with 8 degrees of freedom, the mean absolute deviation (MAD) and the z-test per digit. Following Nigrini (2012), a first-digit MAD up to 0.006 indicates close conformity, 0.006 to 0.012 acceptable conformity, 0.012 to 0.015 marginal conformity and above 0.015 nonconformity. The second is concentration near the approval limit (MZN 50,000): the number of invoices between MZN 45,000 and 50,000 is compared with the mean of the adjacent bands (MZN 40,000–45,000 and 55,000–60,000) through a Poisson z-test. The third is duplicate detection (same supplier, same amount, interval of up to 7 days). The fourth is the proportion of amounts that are exact multiples of MZN 1,000.

$$P(d) = \log_{10}\left(1 + \frac{1}{d}\right), \qquad MAD = \frac{1}{9}\sum_{d=1}^{9}|\hat{p}_d - P(d)| \tag{6}$$

Risk register and Monte Carlo simulation

Fourteen weaknesses were identified. Each was assigned a likelihood (L) and an impact (I) on a scale of 1 to 5; the score is $L \times I$, with bands: critical (≥ 15), high (10–14), medium (5–9) and low (< 5). To quantify exposure, each weakness i is modelled with an annual number of events following a Poisson distribution with mean λ_i and a lognormal severity with median m_i and shape parameter σ_i. The analytical expected loss is:

$$E[L_i] = \lambda_i\, m_i\, e^{\sigma_i^2/2}, \quad L = \sum_{i=1}^{14}\sum_{j=1}^{N_i} X_{ij}, \quad N_i \sim Poisson(\lambda_i),\ X_{ij} \sim LogN(\ln m_i, \sigma_i^2) \tag{7}$$

A total of 100,000 years are simulated for the current scenario, for the scenario with all improvements at full effectiveness (frequency reduced by r_i, between 50% and 75%) and for a conservative scenario in which only 60% of the expected reduction is realised. The mean, the median, the 95th percentile (value at risk, VaR95) and the conditional mean loss above the VaR (CVaR95) are reported.

Prioritisation and cost-benefit analysis

The eight proposed actions are ranked by a priority index with four criteria normalised by min–max: expected annual benefit (40%), cost (20%, sign reversed), ease of implementation (20%) and regulatory relevance (20%). Economic viability is assessed by the payback period and the three-year net present value (NPV), with an illustrative discount rate i = 15% and an annual maintenance cost M = 15% of the initial investment C_0:

$$NPV = -C_0 + \sum_{t=1}^{3}\frac{B - M}{(1+i)^t} \tag{8}$$

Hypotheses, tests and decision criteria

Table 4. Hypotheses, statistical tests and decision criteria

Hypothesis	Test	Decision criterion ($\alpha = 0.05$)
H1: automated controls fail less	2×2 chi-square and OR; coefficient of *Auto* in the logit	$p < 0.05$ and the CI of the OR excludes 1
H2: SoD conflict increases deviation	Coefficient of *SoD* in the logit with robust errors	CI of the OR excludes 1
H3: maturity differs between departments	Kruskal–Wallis; ANOVA as a check	$p < 0.05$; $\varepsilon^2 \geq 0.14$ (large effect)
H4: monitoring is the weakest component, below 3.5	One-sample t-test against 3.5	Lowest mean and $p < 0.05$

Source: prepared by the author.

Reproducibility

The simulation design and parameters are described in this article. With the same seed, the results are exactly reproducible.

The illustrative organisation

"Empresa Alfa, S.A." is a fictitious medium-sized organisation in the commercial and services sector, with 420 employees, six operating centres and an annual turnover of MZN 850 million. It uses an integrated ERP system for purchasing, accounts payable, sales and inventories, and processes payroll in a separate module. It has a small internal audit function, with two staff members, which reports to the chief financial officer. This reporting line does not meet the independence requirement of the third line (IIA, 2020) and is treated as an entity-level weakness in the Discussion.

The assessment covered the full 2025 financial year. The accounts-payable perimeter handled MZN 411.6 million in 4,252 invoices (mean of MZN 96.8 thousand; median of MZN 15.0 thousand), a strongly skewed distribution, typical of portfolios with few high-value documents.

Processes and key controls

The analysis of processes followed the transaction cycle: from purchase order to payment, from sale to collection, from receipt to inventory count, from hiring to payroll, and from journal entry to reporting. Forty-eight key controls were identified in eight processes. Table 5 summarises the composition of each process in terms of the number of controls, items tested and design attributes.

Table 5. Processes assessed, samples and control design attributes

Code	Process	Key controls	Items tested	Automated (%)	With documented evidence (%)	With SoD conflict (%)
P1	Purchasing and procurement	7	186	14	57	43
P2	Accounts payable and treasury	8	281	50	88	75
P3	Sales and collections	6	320	67	67	0
P4	Inventory and warehouses	6	235	0	83	33
P5	Payroll and human resources	5	210	20	80	0
P6	Fixed assets	4	61	75	100	0
P7	Closing and financial reporting	6	229	33	83	17
P8	IT and access (ITGC)	6	272	17	83	17
Total		**48**	**1,794**	**33**	**79**	**27**

Source: prepared by the author, with simulated illustrative data.

Of the 48 controls, 16 (33%) are automated, 25 (52%) are preventive, 38 (79%) have documented evidence and 13 (27%) present some segregation-of-duties conflict, that is, the same person can initiate and approve, or record and pay, within the same flow. The conflict is concentrated in accounts payable (75% of the process's controls), purchasing (43%) and inventories (33%), the processes in which the risk of misappropriation is most direct.

Survey participants

96 employees answered the survey: 28 from finance, 30 from operations, 20 from purchasing and logistics and 18 from human resources and IT. Choosing departments, rather than processes, as the unit of comparison makes it possible to identify where the control culture is weakest and to guide training and communication.

Findings

The results follow the order of the research questions: operating effectiveness of controls (RQ1 and RQ3), COSO maturity (RQ2), anomalies in accounts payable, residual risk and simulation (RQ4). All monetary values are in millions of meticais (M MZN) unless otherwise indicated, and **all of them result from simulated data**.

Operating effectiveness of controls by process

The overall deviation rate was 6.7% (120 deviations in 1,794 items; 95% Wilson CI: 5.6%–7.9%), above the 5% tolerance. Of the 48 controls, 24 (50%) were classified as effective, 14 (29%) as deficiencies and 10 (21%) as significant deficiencies. Table 6 and Figure 1 detail the results by process.

Table 6. Attribute test results and process control risk index by process

Process	Items (n)	Deviations (k)	Rate (%)	Wilson 95% CI (%)	UCL 95% (%)	Conclusion	PCRI	Rank
P2 Accounts payable and treasury	281	36	12.8	9.4–17.2	16.6	Ineffective	76.0	1
P1 Purchasing and procurement	186	16	8.6	5.4–13.5	12.8	Ineffective	73.2	2
P4 Inventory and warehouses	235	19	8.1	5.2–12.3	11.6	Ineffective	61.0	3
P8 IT and access (ITGC)	272	20	7.4	4.8–11.1	10.5	Ineffective	48.3	4
P5 Payroll and human resources	210	9	4.3	2.3–7.9	7.4	Effective with reservations	31.5	5
P6 Fixed assets	61	2	3.3	0.9–11.2	10.0	Effective with reservations*	21.2	6
P7 Closing and financial reporting	229	6	2.6	1.2–5.6	5.1	Effective with reservations	20.3	7
P3 Sales and collections	320	12	3.8	2.2–6.4	6.0	Effective with reservations	17.8	8
Total	**1,794**	**120**	**6.7**	**5.6–7.9**				

Source: prepared by the author, with simulated illustrative data. Note: * A sample of 61 items is insufficient to conclude; increasing it is recommended.

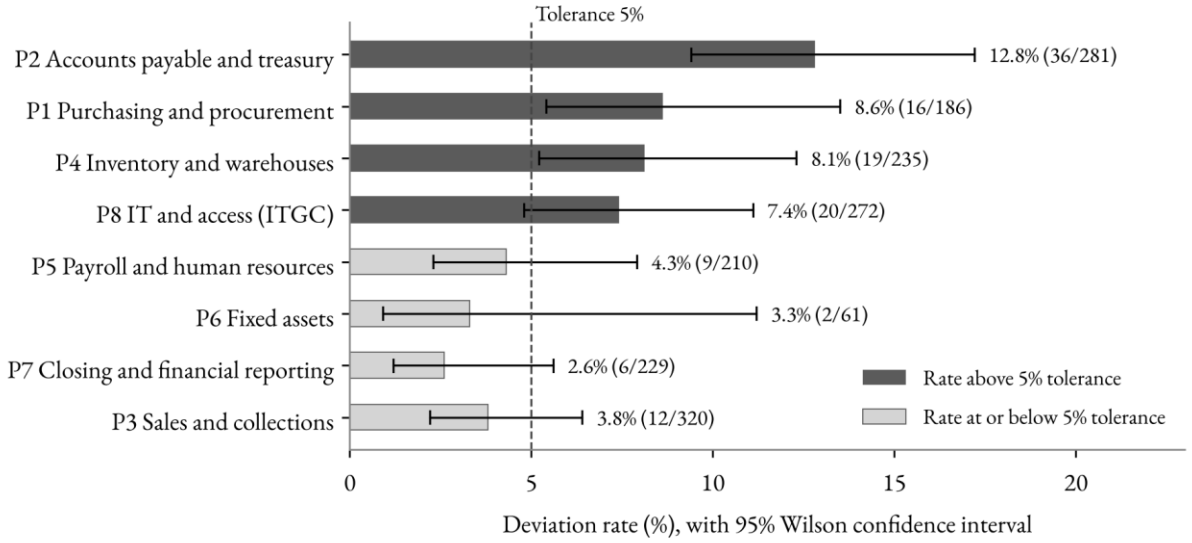


Figure 1. Deviation rate by process. Simulated data (seed 20261112); 1,794 items tested across 48 controls.

Accounts payable shows the worst result: 12.8% deviations, with the lower limit of the CI (9.4%) well above the tolerance. Purchasing (lower limit 5.4%) and inventories (5.2%) also exceed 5% with statistical significance. In IT, the CI (4.8%–11.1%) includes 5%, so the classification as ineffective follows from the prudent decision rule and not from conclusive statistical proof. No process is fully effective: closing and financial reporting has a rate of 2.6%, but its UCL of 5.1% is marginally above the tolerance. The individual control with the worst result is a per-transaction payment approval in accounts payable, with 18 deviations in 60 items (30%), manual, without documented evidence and with an SoD conflict.

The PCRI ranks the processes consistently with the classification: accounts payable (76.0), purchasing (73.2), inventories (61.0) and IT (48.3) lead. The ranking is robust. With equal weights, with 70% weight on the UCL or with 70% weight on the design attributes, the same four processes remain at the top, and the Spearman correlation with the baseline ranking varies between 0.905 and 0.952.

The heterogeneity between processes at the level of individual controls, however, is not statistically conclusive (Kruskal–Wallis: H = 12.38; df = 7; p = 0.089), which is explained by the small number of controls per process and by the dispersion within each.

Determinants of deviations

In the simple comparison, automated controls had 18 deviations in 527 items (3.4%), against 102 in 1,267 (8.1%) for manual controls (Pearson chi-square: $\chi^2(1) = 12.81$; $p < 0.001$; OR = 0.40; 95% CI: 0.24–0.67). This test, however, treats the 1,794 items as independent and ignores that they are clustered in 48 controls. The logistic model with cluster-robust standard errors (Table 7) corrects this limitation and controls for the other characteristics.

Table 7. Logistic model of the probability of deviation (cluster-robust standard errors by control)

Variable	Coefficient	Std. error	OR	95% CI	p
Constant	−2.21	0.20	0.11	0.07–0.16	< 0.001
Automated	−0.38	0.27	0.68	0.40–1.16	0.161
Preventive	−0.45	0.24	0.64	0.40–1.02	0.059
Documented evidence	−0.74	0.24	0.48	0.30–0.76	0.002
SoD conflict	1.08	0.22	2.95	1.90–4.58	< 0.001

Source: prepared by the author, with simulated illustrative data. Note: n = 1,794 items; 48 controls (clusters); McFadden's R^2 = 0.060; AUC = 0.68.

Two results stand out. The segregation-of-duties conflict multiplies the odds of a deviation by 2.95, and documented evidence reduces them to less than half (OR = 0.48). In terms of predicted probability, a manual, detective control, without evidence and with an SoD conflict has a 24.4% probability of failing on each item, against 2.2% for an automated, preventive, documented control without conflict, a ratio of about 11 to 1. The advantage of automation, strong in the simple comparison, weakens and ceases to be significant after adjustment (OR = 0.68; p = 0.161). A plausible explanation is that automated controls are, in this organisation, more often documented and less exposed to conflicts, so part of their apparent effect is due to those characteristics. The overall explanatory power is modest (McFadden's R^2 of 0.060; AUC of 0.68), which indicates that most of the variation in deviations depends on unobserved factors, such as the competence of those who perform the control and workload.

COSO maturity

The COSO maturity index was 2.96 out of 5 (SD = 0.36), equivalent to 49.0 out of 100, below the target level of 3.5. The reliability of the full scale is good (Cronbach's alpha = 0.86; 22 items), but that of the individual components is low (0.52–0.58, with four or five items each), so component scores should be read as descriptive indicators and not as rigorous psychometric measures (see Strengths and limitations).

Table 8. Maturity by COSO component (n = 96)

Component	Items	Mean (SD)	95% CI	Index (0–100)	Below 3.5 (%)	t (df = 95)	p	Alpha
Control environment	5	3.62 (0.44)	3.53–3.71	65.5	37.5	2.66	0.009	0.58
Control activities	5	3.29 (0.42)	3.20–3.38	57.2	68.8	−4.86	< 0.001	0.57
Information and communication	4	2.97 (0.45)	2.88–3.06	49.2	81.3	−11.48	< 0.001	0.52
Risk assessment	4	2.64 (0.47)	2.54–2.73	40.9	97.9	−18.18	< 0.001	0.57
Monitoring	4	2.28 (0.44)	2.19–2.37	32.0	99.0	−27.13	< 0.001	0.55
CMI (mean of the five)	**22**	**2.96 (0.36)**		**49.0**				**0.86**

Source: prepared by the author, with simulated data.

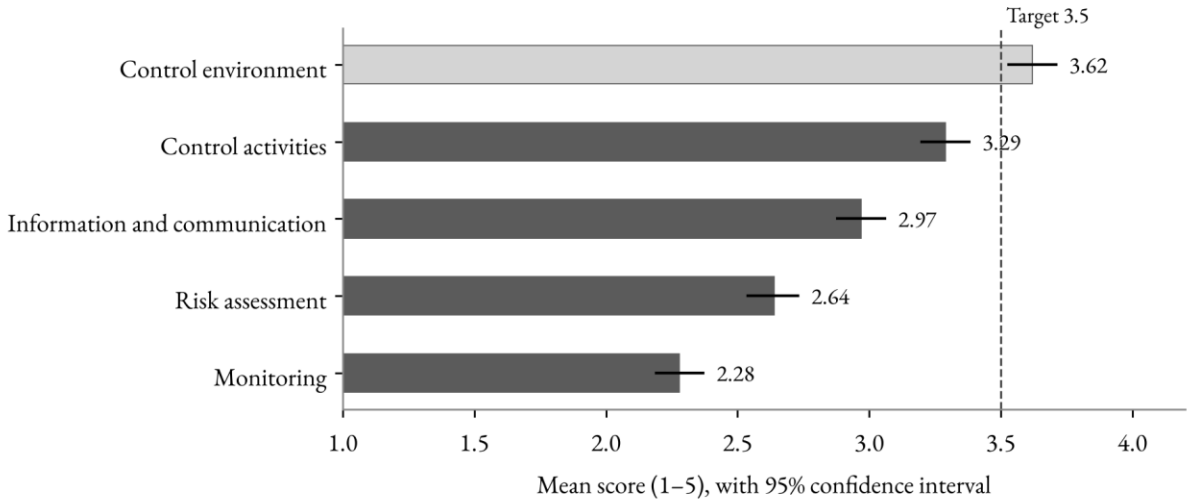


Figure 2. Maturity by COSO component. Simulated data; survey of 96 employees, 22 items.

Only the control environment exceeds the target (3.62; p = 0.009). Monitoring is the weakest component (2.28; t = −27.13; p < 0.001), with 99% of respondents below 3.5, which confirms H4. The lowest item on the whole scale is the follow-up of deficiencies until resolution (M2, mean 1.96), followed by the items on the measurement of monitoring indicators (2.33) and on reporting to management (2.34). The pattern describes an organisation that declares values and assigns responsibilities (control environment) but does not close the loop: it identifies little risk (2.64), follows up little on deficiencies and, therefore, does not learn from them. The components correlate positively with each other (Spearman between 0.40 and 0.65; the strongest between control activities and monitoring, 0.65), which is consistent with the COSO premise that the components function in an integrated way.

Table 9. COSO maturity index by department

Department	n	Mean CMI (SD)
HR and IT	18	3.19 (0.25)
Finance	28	3.16 (0.29)
Operations	30	2.81 (0.32)
Purchasing and logistics	20	2.70 (0.30)

Source: prepared by the author, with simulated data.

The differences between departments are statistically significant and large (Kruskal–Wallis: H(3) = 34.43; p < 0.001; ε^2 = 0.34; ANOVA check: F(3, 92) = 15.59; p < 0.001), which confirms H3. The distance between the most mature department (HR and IT, 3.19) and the least mature (purchasing and logistics, 2.70) is 0.49 points, or 12 points on the 0–100 index. No department reaches the target. The least mature department coincides with the two worst-performing operational processes linked to its activity (purchasing and inventories), which suggests that employees' perceptions and the evidence from the tests point in the same direction. This convergence between subjective and objective evidence reinforces the validity of the assessment.

Accounts-payable data analysis

The analysis covered the complete population of 4,252 invoices, worth MZN 411.6 M. The global Benford test rejects conformity ($\chi^2(8)$ = 53.88; p < 0.001), although the MAD of 0.0076 lies within Nigrini's (2012) acceptable-conformity range. The apparent contradiction is resolved by looking at the individual digits (Table 10): only digit 4 exceeds the critical z value (1.96), with z = 6.81. A total of 544 invoices beginning with 4 were observed (12.8%), against about 412 expected (9.7%), an excess of approximately 132 invoices. The other eight digits are within the expected variation.

Table 10. First digit of invoice amounts: expected (Benford) and observed

Digit	Expected (%)	Observed (%)	Invoices (n)	z
1	30.1	30.2	1,284	0.12
2	17.6	16.7	711	1.50

Digit	Expected (%)	Observed (%)	Invoices (n)	z
3	12.5	11.8	501	1.38
4	**9.7**	**12.8**	**544**	**6.81**
5	7.9	7.3	310	1.49
6	6.7	6.8	291	0.36
7	5.8	5.2	222	1.58
8	5.1	4.5	191	1.81
9	4.6	4.7	198	0.22

Source: prepared by the author, with simulated illustrative data.

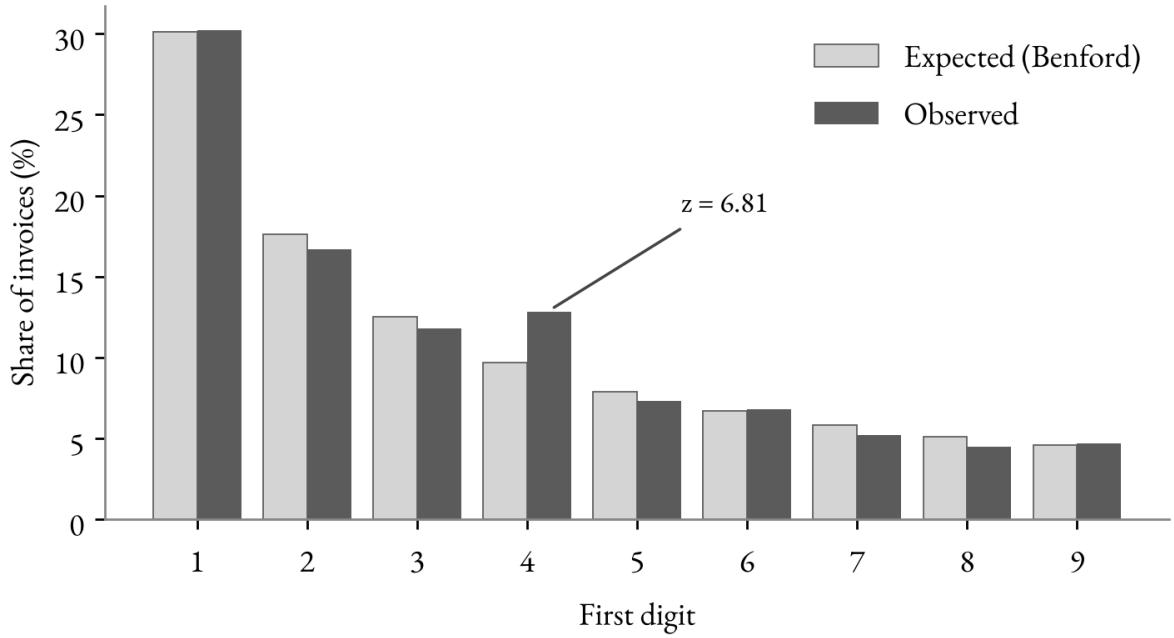


Figure 3. Benford test on the first digit. Simulated data; population of 4,252 payable invoices.

The reason for the excess of digit 4 becomes clear in the threshold test. Between MZN 45,000 and 49,999, immediately below the approval limit of MZN 50,000, 193 invoices were recorded, against 77 expected from the mean of the adjacent bands (98 between MZN 40,000 and 45,000 and 56 between MZN 55,000 and 60,000), that is, 2.5 times more ($z = 13.2$; $p < 0.001$). Between MZN 47,000 and 49,999 there are 153 invoices, totalling MZN 7.40 M (1.8% of the volume). The pattern is compatible with the splitting of purchases to avoid the higher approval level and corresponds to weakness W02 in the risk register. It would be a mistake to conclude fraud from this test: the pattern is an indication that requires documentary verification of contracts, suppliers and purchase orders.

The remaining tests confirm control weaknesses at the payment stage. Fifty-two pairs of duplicate invoices were detected (same supplier, same amount, interval of up to seven days), worth MZN 9.00 M, that is, 1.2% of invoices and 2.2% of the volume. In addition, 71 invoices (1.7%) with amounts that are exact multiples of MZN 1,000 were identified, a low proportion that warrants monitoring but does not in itself constitute an indication. The result illustrates the value of analysing complete populations: a random sample of 60 invoices would have only about a 52% probability of including at least one of the 52 duplicate invoices and would not make it possible to quantify the phenomenon.

Risk register and risk matrix

The integration of the tests, the survey and the invoice analysis made it possible to identify 14 weaknesses (Table 11). Five were classified as critical (score ≥ 15), five as high and four as medium. The weaknesses are anchored in the preceding evidence: the lack of segregation between recording and payment (W01) stems from the SoD conflicts in 75% of accounts-payable controls; splitting (W02) and duplicates are supported by the invoice analysis; privileged access (W04) and inventories (W03) correspond to processes P8 and P4, with deviation rates of 7.4% and 8.1%.

Table 11. Risk register: weaknesses, score, exposure and expected reduction

ID	Proc.	Weakness	L	I	L×I	Band	λ (events/year)	Median (M MZN)	Expected loss (M MZN)	Expected reduction (%)
W06	P2	Supplier master data changes without independent validation	3	5	15	Critical	0.8	0.60	0.99	75
W01	P2	No segregation between recording and payment of invoices	4	5	20	Critical	1.6	0.35	0.92	70
W04	P8	Privileged access without review and shared accounts	4	4	16	Critical	1.2	0.40	0.88	70
W03	P4	Inventory discrepancies without periodic reconciliation	5	3	15	Critical	3.0	0.18	0.74	60
W05	P3	Credit granted without limits or customer risk review	3	4	12	High	1.5	0.30	0.74	55
W12	P8	Backups without restore testing	2	5	10	High	0.3	1.20	0.74	60
W02	P1	Splitting of purchases below the approval limit	4	4	16	Critical	2.2	0.20	0.66	65
W11	P1	Supplier selection without documented competition	3	3	9	Medium	1.3	0.22	0.43	55
W07	P7	Manual journal entries without second-level review	3	4	12	High	1.0	0.25	0.41	65
W10	P3	Invoicing and cash receipt processed by the same person	3	4	12	High	0.9	0.28	0.38	70
W13	P7	Bank reconciliations more than 30 days late	4	3	12	High	2.5	0.08	0.28	70
W14	P4	Open access to warehouses and no movement records	3	3	9	Medium	1.1	0.16	0.26	55
W08	P5	Payroll not checked against attendance records	2	3	6	Medium	1.4	0.12	0.23	55
W09	P6	Fixed assets without annual physical count	3	2	6	Medium	1.0	0.10	0.13	50
Total									**7.79**	

Source: prepared by the author, with simulated data.

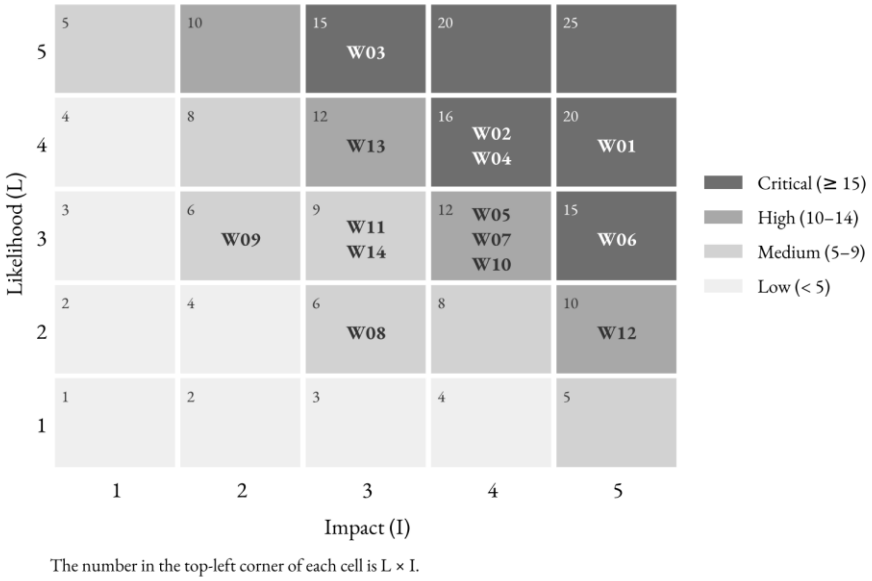


Figure 4. 5 × 5 risk matrix. Ordinal scores assigned to 14 weaknesses.

The ordinal score $L \times I$ and the expected loss in monetary terms agree to a large extent (Spearman correlation = 0.80; $p < 0.001$) but diverge at the extremes. Weakness W12 (backups without restore testing) has a score of 10, only "high", but its expected loss (MZN 0.74 M) is equal to that of weakness W03, classified as critical. This is because the event is rare (0.3 per year) but of high severity and heavy-tailed (median of MZN 1.2 M; $\sigma = 1.2$). The ordinal matrix tends to underestimate this type of risk, which justifies complementing it with monetary quantification.

The aggregate annual expected loss is MZN 7.79 M, or 0.9% of turnover. The concentration is moderate: the first three weaknesses (master data, segregation of duties and privileged access) account for 35.8% of the expected loss and the first five for 54.9%, but nine of the 14 (64%) are needed to exceed 80% (Figure 5). The 80/20 Pareto rule does not hold, which implies that improvement has to be broad-based. This is also why the prioritisation presented below uses several criteria and not only the expected loss.

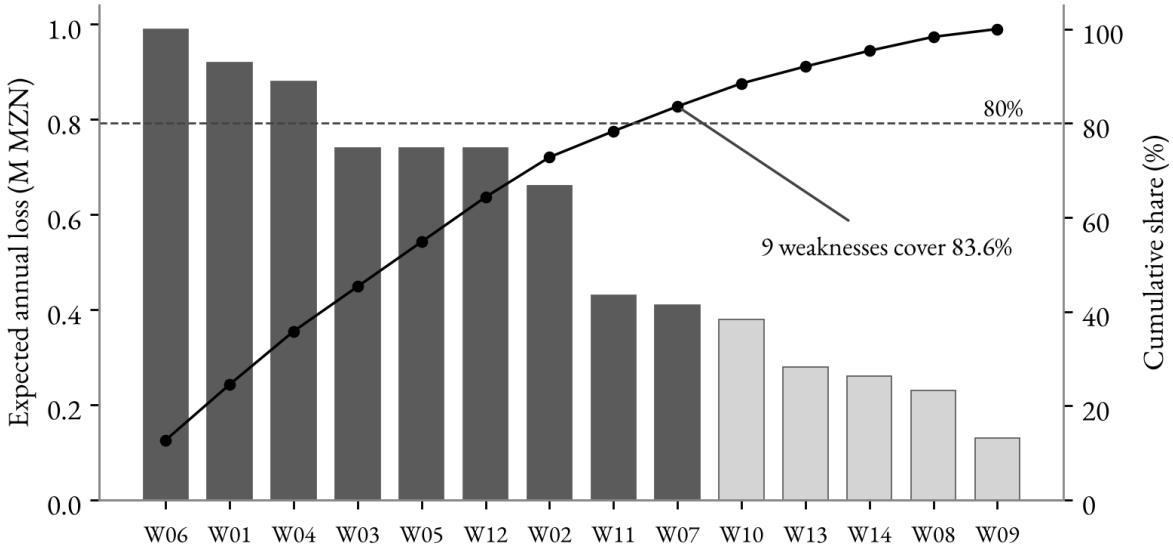


Figure 5. Pareto chart of annual expected loss (λ × median × e^(σ^2/2)). Parameters from the risk register.

Monte Carlo simulation of residual risk

The simulation of 100,000 years reproduces the analytical expected loss with negligible error (MZN 7.80 M simulated against MZN 7.79 M analytical), which validates the implementation. Table 12 compares the current scenario with the scenario in which the improvements reach full effectiveness and with a conservative scenario in which only 60% of the expected reduction is realised.

Table 12. Simulated annual loss (M MZN) and reduction relative to the current scenario

Metric	Current	Full effectiveness	Conservative (60%)	Reduction (full)	Reduction (conservative)
Mean	7.80	2.76	4.77	−64.6%	−38.9%
Median	6.89	2.19	4.05	−68.2%	−41.3%
Standard deviation	4.44	2.55	3.37	−42.4%	−24.0%
VaR 95%	15.13	6.55	10.15	−56.7%	−32.9%
VaR 99%	23.27	11.65	16.55	−49.9%	−28.9%
CVaR 95%	20.76	10.19	14.65	−50.9%	−29.4%

Source: prepared by the author, with simulated data.

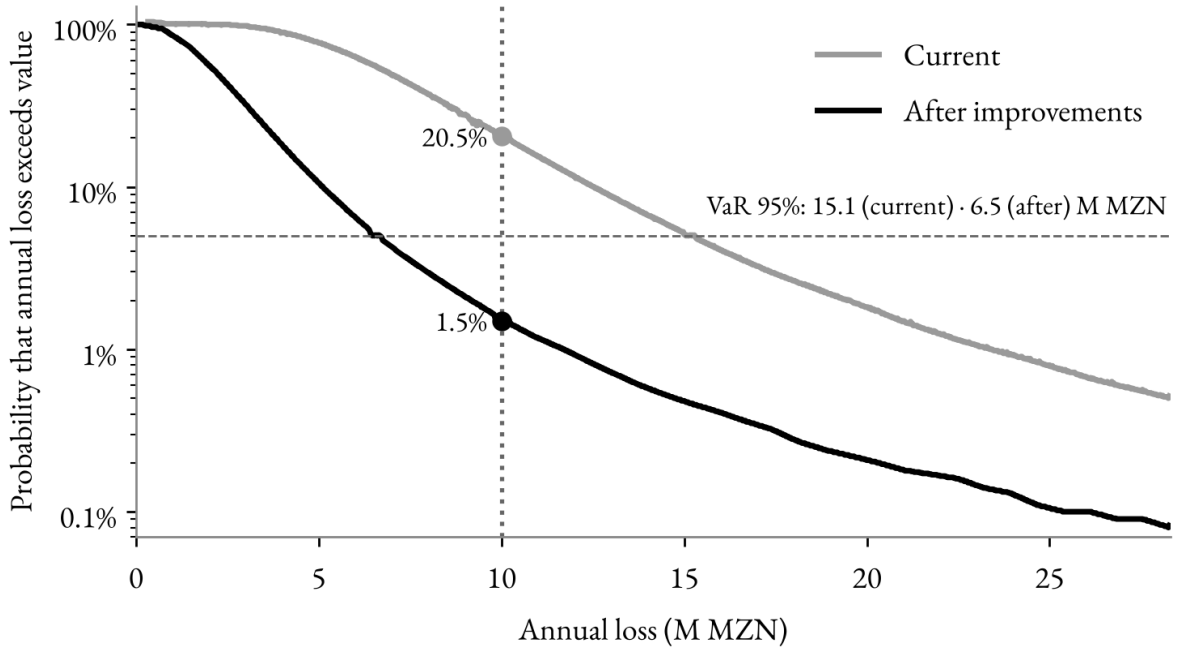


Figure 6. Loss exceedance curve. Monte Carlo simulation, 100,000 years, simulated data.

With full effectiveness, the mean loss falls from MZN 7.80 M to 2.76 M (−64.6%) and the VaR95 from 15.13 to 6.55 (−56.7%). In the conservative scenario, the reductions are 38.9% and 32.9%. In both cases, the reduction in the tail is smaller than in the mean: the conditional mean loss above the VaR95 falls by 50.9% with full effectiveness, against 64.6% for the mean. The reason lies in the design of the actions, which mainly reduce the frequency of events and not their severity. Rare and severe events, such as a data recovery failure (W12), continue to occur, and it is in them that the extreme loss is concentrated. This asymmetry recommends complementing the plan with impact-mitigation measures, such as a continuity plan, and not only prevention.

Summary of hypothesis tests

Table 13. Summary of hypothesis tests

Hypothesis	Evidence	Decision
H1: automated controls fail less	Simple: 3.4% vs 8.1% ($\chi^2 = 12.81$; $p < 0.001$; OR = 0.40). Adjusted: OR = 0.68 (95% CI: 0.40–1.16; $p = 0.161$)	Partially supported
H2: SoD conflict increases deviation	OR = 2.95 (95% CI: 1.90–4.58; $p < 0.001$)	Supported
H3: maturity differs between departments	$H(3) = 34.43$; $p < 0.001$; $\varepsilon^2 = 0.34$	Supported
H4: monitoring is the weakest component and is below 3.5	Mean of 2.28; $t(95) = -27.13$; $p < 0.001$	Supported

Source: prepared by the author, with simulated data.

In summary, RQ1 identifies purchasing, accounts payable, inventories and IT as ineffective processes. RQ2 shows that monitoring and risk assessment are the weakest components and that purchasing and logistics is the least mature department. RQ3 points to the segregation-of-duties conflict and the lack of documented evidence as the attributes most associated with failures. RQ4 is answered below, which ranks the actions by benefit, cost, ease and regulatory relevance.

Proposed improvements: portfolio and prioritisation

For each weakness in the register, a practical corrective action was designed. The actions were grouped into eight packages, by affinity of process and solution, and ranked by the priority index described in the Methods. Table 14 presents the portfolio.

Table 14. Prioritised portfolio of actions (values in M MZN)

Rank	Action	Weaknesses addressed	Initial cost	Expected annual benefit	3-year NPV	Payback (months)	Ease (0–1)	Regulatory relevance (1–5)	Priority (0–100)
1	A1. Segregate recording, approval and payment in the ERP (role matrix); independent validation of master data changes	W01, W06, W10	1.10	1.65	2.29	8.9	0.9	5	80.0
2	A4. Privileged access management and semi-annual user review	W04	0.50	0.62	0.73	11.1	0.9	5	65.0
3	A6. Second-level review of manual journal entries and bank reconciliations within 10 working days	W07, W13	0.25	0.46	0.72	7.1	0.7	4	46.3
4	A2. Two-level electronic approval and blocking of purchase splitting	W02, W11	0.90	0.67	0.31	20.4	0.8	4	42.4
5	A3. Monthly inventory reconciliation and cycle counts	W03, W14	0.65	0.59	0.48	15.8	0.7	4	41.0
6	A8. Semi-annual restore test and IT continuity plan	W12	0.35	0.44	0.54	10.7	0.5	4	33.6
7	A5. Credit policy, customer limits and quarterly review	W05	0.30	0.41	0.53	9.9	0.6	3	28.7
8	A7. Payroll check against attendance and physical count of assets	W08, W09	0.20	0.19	0.17	14.9	0.6	3	25.0
Total		**4.25**	**5.03**	**5.77**	**11.6**				

Source: prepared by the author, with simulated data. Note: The payback period discounts the annual maintenance cost (15% of the initial investment). NPV at a 15% rate. The total NPV is the sum of the individual NPVs.

Action A1 leads clearly, with 80.0 points. On its own, it addresses three weaknesses (including the two with the highest expected loss), costs MZN 1.10 M and generates MZN 2.29 M of NPV. It is also the action that most reduces the factor the logistic model identified as the most determinant of deviations, the segregation-of-duties conflict. Action A4 ranks second, because it addresses a critical weakness at low cost and with high regulatory relevance. Action A6 has the shortest payback period (7.1 months): it is cheap and resolves closing and reporting failures.

The ranking by priority does not coincide with the ranking by financial return (Spearman correlation = 0.31). Action A2 (purchase splitting) ranks fourth in priority but has the longest payback period (20.4 months), because the expected loss from splitting is modest (MZN 0.66 M) compared with the cost of blocking the purchasing flow. It remains in the portfolio because it addresses an indication confirmed by the invoice analysis and because its value is not only monetary: it reduces the risk of non-compliance with procurement rules.

Proposed improvements: economic viability and sensitivity

The total investment is MZN 4.25 M, the expected annual benefit is MZN 5.03 M (a 64.6% reduction in expected loss) and the three-year NPV is MZN 5.77 M, with payback in 11.6 months. The result depends on the effectiveness of the actions. Table 15 varies the realised effectiveness between 60% and 120% of that expected.

Table 15. Sensitivity of the return to the realised effectiveness of the actions (M MZN)

Realised effectiveness	Annual benefit	3-year NPV
60%	3.02	1.18
80%	4.02	3.47
100%	5.03	5.77
120%	6.03	8.06

Source: prepared by the author, with simulated data.

The NPV is positive across the whole range, and the break-even threshold is reached with 50% of the expected effectiveness. This means that the plan continues to pay off even if the actions deliver only half of what is expected, which is a relevant argument for management. The conclusion requires caution: the frequency and severity parameters are hypothetical and, with real data, should be estimated from the organisation's loss history and from external databases.

Proposed improvements: governance and monitoring

Actions A1 to A8 address process weaknesses. The survey results show that monitoring (2.28) and risk assessment (2.64) are the weakest components, so a second set of low-cost, organisational measures is necessary:

1. **Independence of internal audit.** Have the function report functionally to the supervisory body or audit committee, and administratively to management, in line with the Three Lines Model (IIA, 2020).
2. **Central register of deficiencies.** Record each deficiency with an owner, a deadline and a status, and report quarterly the percentage resolved on time (indicator K1 in Table 16).
3. **Annual fraud risk assessment** by process, with the participation of operational managers, and revision of the risk register whenever there is a material change in a process or system.
4. **Independent whistleblowing channel.** ACFE (2024) shows that tips are the main means of fraud detection, accounting for 43% of cases.
5. **Targeted training.** Prioritise purchasing and logistics and operations, the departments with the lowest maturity (2.70 and 2.81), with content on approvals, segregation of duties and handling of exceptions.
6. **Continuous data analysis.** Run the Benford, threshold, duplicate and master-data tests monthly on the full population of invoices, with results sent to internal audit (Sylwestrzak, 2023; Sushkov et al., 2023).

Proposed improvements: performance indicators and targets

Table 16. Internal control indicator dashboard (baseline values of the illustrative organisation and 12-month targets)

ID	Indicator	Baseline	12-month target	Frequency	Owner
K1	Deficiencies resolved on time (%)	Not measured	≥ 85	Quarterly	Internal audit
K2	Overall deviation rate in key controls (%)	6.7	≤ 4.0	Semi-annual	Internal audit
K3	Processes with UCL ≤ 5% (no. of 8)	0	≥ 5	Semi-annual	Internal audit
K4	Controls with segregation-of-duties conflict (%)	27	≤ 5	Quarterly	CFO and IT
K5	Controls with documented evidence (%)	79	≥ 95	Quarterly	Process owners
K6	COSO maturity index (0–100)	49.0	≥ 62.5 (mean 3.5)	Annual	Audit committee
K7	Monitoring component (mean, 1–5)	2.28	≥ 3.2	Annual	Audit committee
K8	Invoices between 90% and 100% of the approval limit (per quarter)	48	≤ 20	Monthly	Purchasing
K9	Duplicate invoice pairs detected (per quarter)	13	0	Monthly	Accounts payable

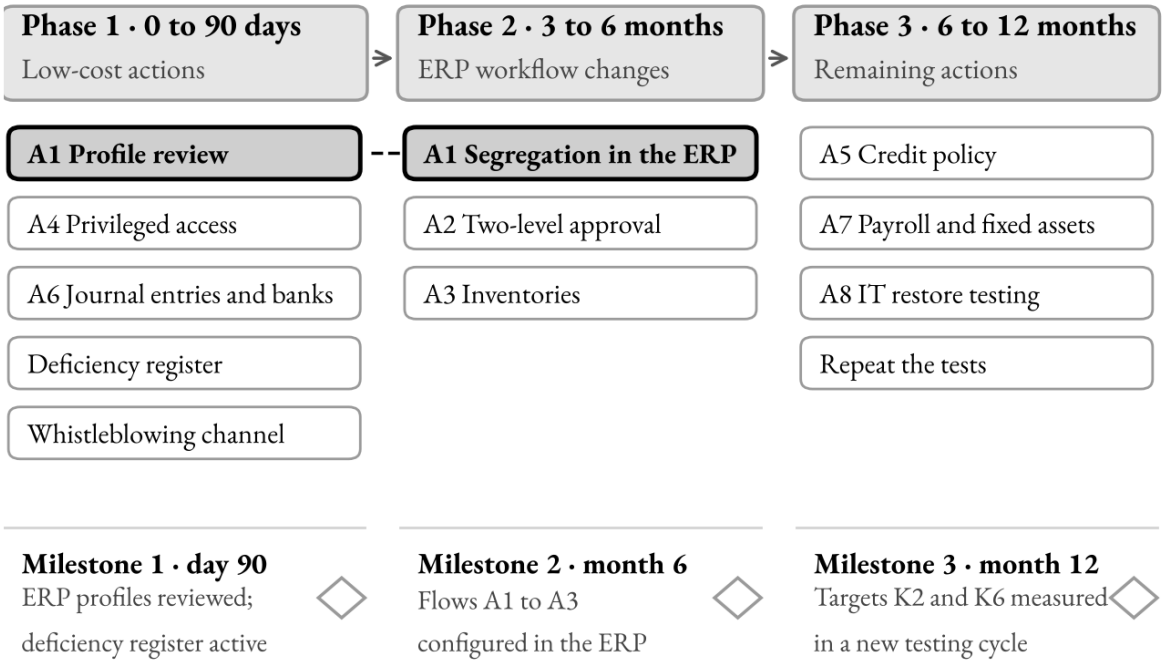
ID	Indicator	Baseline	12-month target	Frequency	Owner
K10	Annual expected loss (M MZN)	7.79	≤ 4.0	Annual	Risk management

Source: prepared by the author, with simulated data.

The baseline values of K8 and K9 result from dividing by four the annual results of the accounts-payable analysis (193 invoices in the MZN 45,000–50,000 band and 52 duplicate pairs). The K10 target is prudent, as it lies above the residual loss of the full-effectiveness scenario (MZN 2.76 M) and below that of the conservative scenario (MZN 4.77 M).

Proposed improvements: implementation roadmap

Implementation takes place in three phases, ordered by priority and technical dependency. The first, from 0 to 90 days, brings together the low-cost, immediate-effect actions (A1 in its profile-review part, A4 and A6), the creation of the deficiency register and the design of the whistleblowing channel. The second, from 3 to 6 months, introduces the workflow changes that require ERP configuration (A1 in full, A2 and A3). The third, from 6 to 12 months, completes the lower-priority actions (A5, A7 and A8), repeats the attribute tests with the same sampling and measures the progress of the indicators.



In grey, action A1 (priority 1): it starts in phase 1 and is completed in phase 2.

Figure 7. Three-phase implementation roadmap, not drawn to scale; milestones with decision criteria.

3. DISCUSSION

A caveat precedes the discussion: because the data are simulated, the comparison with the literature serves to check the coherence of the patterns and the interpretability of the protocol, and not to empirically validate any conclusion.

Main messages

The first message is that the problem of the illustrative organisation lies not in the formal existence of controls but in their operation and supervision. There are 48 documented controls and a control environment above target (3.62) and, even so, 50% of controls show deficiencies and four processes are ineffective. This distance between design and practice is what checklist-based assessments do not capture, and it is compatible with the observation that many systems are implemented as a formality (Musikavanhu & Nqala, 2024).

The second message is that two simple design attributes, segregation of duties and documented evidence, explain more of the probability of failure than technology alone. Automation reduces deviation in the simple comparison, but the effect disappears after adjustment. This result argues against the idea that automating is, in itself, the remedy: an automated control with a segregation-of-

duties conflict, without a record and without monitoring continues to fail, only faster and more silently. The priority becomes reconfiguring profiles and approval flows in the ERP and only then expanding automation.

The third message is that monitoring is the deepest failure, and it is cross-cutting. At 2.28 out of 5, the component is the weakest, and the deficiency follow-up item has the lowest mean of the whole instrument (1.96). An organisation that does not follow up on what it finds tends to repeat the same failures, which is why the plan includes follow-up indicators and a quarterly reporting cycle to the governing body.

Relation to the literature

The dominant role of entity-level weaknesses is consistent with Donelson et al. (2017), who associated the future disclosure of fraud mainly with weaknesses at that level and not with specific processes. In the illustrative organisation, the entity-level components (risk assessment and monitoring) are the weakest, and the internal audit function reports to the chief financial officer, which limits its independence (IIA, 2020). The similarity with the Cameroonian result, in which the mere presence of an internal control system did not increase the effectiveness of internal audit (Kueda et al., 2026), reinforces the need to assess quality and not existence.

The occupational fraud data are consistent with the profile of weaknesses. ACFE (2024) attributed 32% of cases to a lack of controls and 19% to their override, and the highest-exposure weaknesses in the register, supplier master data (W06) and segregation between recording and payment (W01), are precisely the points at which controls can be circumvented by a single user with privileged access. The result of Mandal and Amilan (2023), in which fraud awareness mediates the effect of the control system, helps to interpret the difference in maturity between departments: where the control culture is weakest, the same system produces less protection. The significant difference between units is similar to that observed in three Mozambican universities (Graziano, 2023).

The result for the tail of the risk, in which the improvements reduce extreme loss less than mean loss, has a parallel in the results of Boulhaga et al. (2022), according to which good control limits manipulation through accruals but does not control manipulation through real decisions. In both cases, preventive control reduces the frequency of events but does not eliminate high-severity events. Finally, the reactive nature of internal control in SMEs, in which mechanisms emerge after fraud losses (Ololade, 2024), is a warning that the investment in the proposed improvements should be made before and not after the event.

Implications for practice

For management, the protocol replaces the question "does the control exist?" with the question "at what rate and with what uncertainty does it fail, and how much does that failure cost?". For the governing body, a matrix accompanied by the expected loss in monetary value makes risks of different natures comparable. For internal audit, the protocol is a way of meeting the requirements of the 2024 Global Standards (IIA, 2024) on risk-based planning, use of technology and communication of results. The analysis of complete populations, illustrated in accounts payable, shows an effectiveness gain that samples do not allow, in line with the literature on data analytics in auditing (Sylwestrzak, 2023; Sushkov et al., 2023).

Robustness of the assessment

Three elements reinforce robustness. First, the convergence between the objective evidence (operating tests) and the subjective evidence (survey): the least mature department corresponds to the processes with the worst performance. Second, the stability of the PCRI under alternative weighting schemes. Third, the reproduction of the analytical expected loss by the simulation. Conversely, three results call for caution: the conclusion of ineffectiveness in IT rests on a decision rule and not on statistical proof; the reliability measures of the survey components are low; and the explanatory power of the logistic model is modest.

Strengths and limitations

The main limitation is the simulated nature of the data. The parameters (baseline deviation rates, loss frequencies and severities, survey weights) were chosen by the author, and the results reflect those assumptions. The values should therefore be read as a demonstration of the method. They cannot be extrapolated to real organisations or used as a benchmark for deviation rates or losses.

There are other methodological limitations that remain even with real data:

- **Sampling by process.** With 61 items, the fixed-assets process does not allow a conclusion; the power of the tests is low for rates close to the tolerance. It is recommended to size the sample by the desired upper limit.
- **Independence of items.** The items of each control are correlated; the use of cluster-robust standard errors mitigates the problem, but with only 48 clusters the standard-error estimates may be unstable.
- **Survey reliability.** The Cronbach's alpha of the components (0.52–0.58) is low, in part because of the small number of items. With real data, the instrument should be validated with factor analysis and with more items per component.

- **Loss model.** The Poisson distribution assumes independent events, and the lognormal may underestimate very heavy tails. Frequencies and severities should be calibrated with historical losses, structured expert opinion and, if possible, external databases.
- **Index weights.** The PCRI and the priority index depend on weights chosen by the researcher. The PCRI ranking proved stable (Spearman correlation of 0.905 to 0.952 with alternative schemes), but the priority index was not subjected to an equal test.
- **Benefits.** The estimated benefit considers only the expected loss avoided and ignores reputational, regulatory and cost-of-capital effects.
- **Interdependencies.** The actions were assessed independently, but segregation of duties (A1) and access control (A4) reinforce each other, so the joint benefit may differ from the sum.

Future research can follow four directions: (i) applying the protocol to real organisations, with test and transaction data, and comparing sectors; (ii) estimating multilevel models that separate the variation between processes, controls and items; (iii) integrating machine learning and process mining techniques in anomaly detection, following the frameworks proposed for the internal audit cycle (Anwar & Akeel, 2026); and (iv) assessing, with a quasi-experimental design, whether the improvements actually reduce deviation rates in subsequent testing cycles.

4. CONCLUSION AND RECOMMENDATIONS

This article demonstrated a quantitative protocol for assessing internal control, which runs from evidence of control operation to the investment decision. Applied to a fictitious organisation, the protocol concluded that the overall deviation rate (6.7%) exceeds the 5% tolerance, that four of the eight processes are ineffective (accounts payable, purchasing, inventories and IT) and that COSO maturity (49.0 out of 100) is below the target level, with monitoring as the weakest component.

The design attributes most associated with failures are the segregation-of-duties conflict (OR = 2.95) and the absence of documented evidence (evidence reduces the odds to OR = 0.48). Automation, advantageous in the simple comparison, does not withstand adjustment, which shifts the priority from technology to the configuration of profiles and flows. The analysis of 4,252 invoices indicated a concentration near the approval limit (2.5 times the expected) and 52 duplicate pairs (MZN 9.0 M), which are indications to be verified and not proof of fraud.

The annual expected loss is MZN 7.8 M. The portfolio of eight actions, costing MZN 4.25 M, would reduce this loss by 64.6% with full effectiveness and by 38.9% in the conservative scenario, with a positive NPV even if only half of the expected effectiveness is realised. The highest-priority action is segregation of duties in the ERP, which addresses the two weaknesses with the highest expected loss.

Three final recommendations summarise the article. First, measure the operation of controls with uncertainty, and not merely verify their existence. Second, convert weaknesses into expected loss, so that management can compare risks and costs in the same unit. Third, close the loop with monitoring: record deficiencies, follow them up until resolution and repeat the tests. All numerical results are illustrative, and their main usefulness is to allow managers, internal auditors and researchers to adapt the protocol to their own data.

Declaration of interest statement

The author reports that there are no competing interests to declare.

Funding

This research did not receive any specific funding or support from grant-awarding bodies.

Data availability statement

All data analysed in this study were generated by simulation (seed 20261112) and do not describe any real entity. The simulation parameters are reported in the tables of this article.

5. REFERENCES

Anwar, A., & Akeel, M. O. (2026). Integrating artificial intelligence in audit workflow: Opportunities, architecture, and challenges: A systematic review. *Accounting and Auditing, 2*(1), Article 4. https://doi.org/10.3390/accountaudit2010004

Apon, L. I., Kintu, I., & Nuwagaba, G. (2024). Internal auditing and sustainability of small and medium enterprises in Uganda. *Vallis Aurea, 10*(2), 34–50. https://doi.org/10.62598/JVA.10.2.3.8

Association of Certified Fraud Examiners. (2024). *Occupational fraud 2024: A report to the nations*. ACFE. https://www.acfe.com/en-us/-/media/files/acfe/pdfs/rttn/2024/2024-report-to-the-nations.pdf

Bila, S. (2023). Determinantes de sobrevivência das pequenas e médias empresas (PMEs) em Moçambique [Determinants of the survival of small and medium-sized enterprises (SMEs) in Mozambique]. *Revista Internacional Académica da Lusofonia, 11*(2), 277–287. https://doi.org/10.25746/ruiips.v11.i2.32804

Boulhaga, M., Bouri, A., & Elbardan, H. (2022). The effect of internal control quality on real and accrual-based earnings management: Evidence from France. *Journal of Management Control, 33*(4), 545–567. https://doi.org/10.1007/s00187-022-00348-5

Committee of Sponsoring Organizations of the Treadway Commission. (2013). *Internal control -- integrated framework*. COSO.

Committee of Sponsoring Organizations of the Treadway Commission. (2017). *Enterprise risk management -- integrating with strategy and performance*. COSO. https://www.scribd.com/document/854173331/Coso-Erm-Standard-2017

Committee of Sponsoring Organizations of the Treadway Commission. (2023). *Achieving effective internal control over sustainability reporting (ICSR): Building trust and confidence through the COSO internal control -- integrated framework*. COSO. https://www.coso.org/_files/ugd/3059fc_a3a66be7a48c47e1a285cef0b1f64c92.pdf

Donelson, D. C., Ege, M. S., & McInnis, J. M. (2017). Internal control weaknesses and financial reporting fraud. *Auditing: A Journal of Practice & Theory, 36*(3), 45–69. https://doi.org/10.2308/ajpt-51608

Graziano, N. (2023). Análise dos indicadores/dimensões do sistema de controlo interno nas Universidades Pedagógica de Maputo, Eduardo Mondlane e Unilicungo, no período de 2018 a 2020 [Analysis of the indicators/dimensions of the internal control system at the Pedagogical University of Maputo, Eduardo Mondlane and Unilicungo, 2018–2020]. *Revista Internacional Académica da Lusofonia, 11*(2), 44–59. https://doi.org/10.25746/ruiips.v11.i2.32785

Harahap, L., & Isgiyarta, J. (2023). The role of control environment (CE) in the public sector in preventing fraud: A literature study. *JRAP (Jurnal Riset Akuntansi dan Perpajakan), 10*(2), 264–279. https://doi.org/10.35838/jrap.2023.010.02.22

Hazzaa, O. T., Abdullah, D. F., & Dhahebi, A. M. (2022). Review on the role of corporate governance and internal control system on firms' financial performance. *Asian Journal of Accounting Perspectives, 15*(1), 1–28. https://doi.org/10.22452/AJAP.vol15no1.1

International Auditing and Assurance Standards Board. (2019). *ISA 315 (Revised 2019), Identifying and assessing the risks of material misstatement*. IAASB. https://www.irba.co.za/upload/ISA-315-Full-Standard-and-Conforming-Amendments-2019.pdf

Institute of Internal Auditors. (2020). *The IIA's three lines model: An update of the three lines of defense*. IIA.

Institute of Internal Auditors. (2024). *Global internal audit standards*. IIA. https://www.theiia.org/globalassets/site/standards/globalinternalauditstandards_2024january9.pdf

Kolawole, K. D. (2026). Internal audit, internal controls, and fiscal discipline: Empirical insights from strengthening control systems as drivers of financial accountability in Nigerian local governments. *International Journal of Economic Performance, 9*(1), 72–92. https://doi.org/10.54241/2065-009-001-004

Kueda, W. B., Njong, A. M., Ngassa, M., & Nkinen, M. (2026). The internal audit effectiveness of public limited companies in Cameroon: A study of its determinants. *Journal of Business Management and Accounting, 16*(1), 13–30. https://doi.org/10.32890/jbma2026.16.1.2

Mandal, A., & Amilan, S. (2023). Preventing financial statement fraud in the corporate sector: Insights from auditors. *Journal of Financial Reporting and Accounting, 23*(1), 56–80. https://doi.org/10.1108/JFRA-02-2023-0101

Manry, D., Huang, H.-W., & Yan, Y.-C. (2023). Financial statement fraud litigation, material weaknesses, and board characteristics. *Accounting Research Journal, 36*(4/5), 349–368. https://doi.org/10.1108/ARJ-08-2022-0218

Musikavanhu, T., & Nqala, L. (2024). Exploring challenges in designing and implementing internal control systems: A case study of small and medium enterprises in the Cape Metropole, South Africa. *Business and Finance Journal, 9*(1). https://doi.org/10.33086/bfj.v9i1.5721

Nigrini, M. J. (2012). *Benford's law: Applications for forensic accounting, auditing, and fraud detection*. Wiley. https://www.wiley.com/en-us/shop/general-introductory-accounting/forensic-analytics-methods-and-techniques-for-forensic-accounting-investigations-2nd-edition-p-9781119585909

Ololade, B. M. (2024). Internal control systems of SMEs in Nigeria: A proactive or reactive strategy against employees' fraud? *Journal of Financial Crime, 32*(1), 49–63. https://doi.org/10.1108/JFC-10-2023-0259

Otai, I. P., & Birungi, B. D. E. (2025). Internal controls and financial performance of small and medium enterprises in Mbale City, Uganda. *CECCAR Business Review, 6*(3), 73–82. https://doi.org/10.37945/cbr.2025.03.07

Sushkov, V. M., Leonov, P. Y., Nadezhina, O. S., & Blagova, I. Y. (2023). Integrating data mining techniques for fraud detection in financial control processes. *International Journal of Technology, 14*(8). https://doi.org/10.14716/ijtech.v14i8.6830

Sylwestrzak, M. (2023). Applying Benford's law to detect earnings management. *Journal of Economics and Management, 45*(1), 216–236. https://doi.org/10.22367/jem.2023.45.10

Wilson, E. B. (1927). Probable inference, the law of succession, and statistical inference. *Journal of the American Statistical Association, 22*(158), 209–212. https://doi.org/10.1080/01621459.1927.10502953

This article has been assigned the following Digital Object Identifier (DOI). The journal will make the article available through its website and submit the article or its bibliographic metadata, as applicable, to participating academic library discovery services, digital archives, and indexing databases serving researchers in Europe, America, Africa and Asia. Inclusion in each service is subject to its selection and coverage policies. The manuscript was screened using *StrikePlagiarism* and *Turnitin*. The editorial office reviewed the matches identified in both reports. Based on that review, no instances of plagiarism were identified. The author is responsible for the accuracy of the information presented, the views expressed, the originality of the work, and the proper attribution of sources. The editorial office bears primary responsibility for conducting and documenting the journal's plagiarism screening and for assessing the reported matches before publication.