© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1719533

From Tax Mismatches to Revenue Protection: A Data-Driven Framework for Audit Prioritization, Voluntary Compliance, and SME Compliance Education

LUCY GANYANI[1], SABELO NARE[2], CATHERINE DANDA[3], LAST CHINGEZI[4], MUNASHE NAPHTALI MUPA[5]

[1]*Babson College*

[2]*William Paterson University*

[3,4]*University of Northern Iowa*

[5]*Hult International Business School*

Abstract- Tax administrations and compliance-sensitive businesses face the same operational dilemma: revenue risk is concentrated in a minority of returns, yet audit resources are limited and many small and medium-sized enterprise (SME) errors arise from weak recordkeeping, filing complexity and digital capability gaps rather than deliberate evasion. This article develops a data-driven framework that converts taxpayer mismatches into a structured revenue-protection system. The framework integrates third-party data matching, tax-type risk features, estimated revenue-at-risk ranking, audit exception dashboards and targeted SME education. It is grounded in tax-compliance theory, compliance risk management, evidence on third-party reporting, behavioral compliance research and recent work on explainable machine learning for tax and audit planning. Using an illustrative synthetic dataset of 5,400 taxpayer-period records, the article demonstrates how risk heat maps, revenue-at-risk matrices, a risk-decile capture curve and education-need segmentation can support audit prioritization while preserving voluntary compliance. The results show how compliance teams can distinguish cases requiring enforcement from cases better suited to correction notices, desk review or education. The framework contributes a practical operating model for revenue authorities, tax-compliance units and advisory teams seeking to protect public revenue without imposing unnecessary compliance costs on lower-risk taxpayers.

Keywords: Tax Compliance, Audit Prioritization, Taxpayer Mismatches, Voluntary Compliance, SME Education, Revenue Protection, Explainable Analytics, Compliance Risk Management

I. INTRODUCTION

Tax compliance is increasingly a data problem as much as it is a legal, administrative and behavioral problem. Modern revenue authorities receive tax returns, payroll filings, customs declarations, third-party information, bank-payment indicators, electronic invoices, point-of-sale summaries and registry data. The policy challenge is no longer merely whether data exists, but whether the data can be converted into fair, explainable and operationally useful compliance decisions. Where data are fragmented, tax officers may pursue low-yield cases while high-risk taxpayers remain under-reviewed. Where analytics are not explainable, automated selection may weaken trust or create fairness concerns. Where enforcement ignores taxpayer capability, SMEs may face unnecessary penalties for errors that could have been corrected through education and better controls.

The revenue stakes are substantial. The IRS projects a gross federal tax gap of USD 696 billion for tax year 2022 and a voluntary compliance rate of 85.0 percent, showing the continuing importance of both enforcement and self-assessment compliance (IRS, 2026). At the operational level, the IRS Data Book reports that in fiscal year 2025 the IRS closed 497,621 tax return audits and recommended USD 26.8 billion in additional tax (IRS, 2025). These figures are jurisdiction-specific, but the underlying problem is universal: administrations must select a small subset of returns for review, while the wider

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1719533

system depends on most taxpayers complying voluntarily.

The article makes three contributions. First, it presents a mismatch-to-treatment framework that links tax data quality, risk scoring and revenue-at-risk estimation to specific compliance treatments. Second, it demonstrates the framework using a transparent synthetic dataset designed for practitioner replication. Third, it integrates audit prioritization with SME education, recognising that compliance improvement requires both deterrence and capability-building. The aim is not to replace tax officers with algorithms, but to provide a governed, explainable and evidence-based workflow that helps tax and business-compliance teams decide which cases require monitoring, education, desk review or field audit.

II. LITERATURE REVIEW

2.1 Tax compliance, deterrence and voluntary compliance

The classic economic model of tax evasion frames compliance as a function of the probability of detection, penalty severity and taxpayer risk preferences (Allingham and Sandmo, 1972). Subsequent research shows that this deterrence model is necessary but incomplete. Compliance also depends on third-party reporting, withholding, trust, perceived fairness, administrative simplicity, tax morale and the taxpayer's ability to comply (Kirchler, Hoelzl and Wahl, 2008; Alm, 2019; Slemrod, 2019).

Voluntary compliance is therefore not passive. It is produced by a system in which taxpayers understand their obligations, have the capability to keep adequate records, receive credible signals that non-compliance can be detected, and trust that the enforcement burden is allocated fairly. The slippery-slope framework developed by Kirchler, Hoelzl and Wahl (2008) is useful because it treats compliance as a function of both power and trust. A revenue authority that relies only on power may generate resistance, while one that relies only on trust may leave deliberate non-compliance undeterred.

The behavioral literature adds that messages and taxpayer communications can affect compliance behavior. Hallsworth et al. (2017) find that social-norm messages increased payment rates in large-scale UK tax experiments. This does not mean that messages replace audit enforcement; rather, it suggests that data-driven segmentation should identify which taxpayers can be moved into compliance through reminders, education and correction opportunities, and which taxpayers require stronger enforcement.

2.2 Third-party reporting, matching and taxpayer mismatches

Third-party reporting is one of the most important compliance mechanisms in modern tax administration. Kleven et al. (2011) show in a Danish tax-audit experiment that evasion is very low where income is subject to third-party reporting but substantially higher for self-reported income. Pomeranz (2015) demonstrates that the VAT's paper trail can create self-enforcing incentives, while Carrillo, Pomeranz and Singhal (2017) show that firms may shift misreporting to margins that remain less observable even after enforcement letters are issued. The implication is that data matching is powerful, but taxpayers may adapt when only one margin is monitored.

Mismatch detection therefore needs to be multi-dimensional. A sales-tax or VAT mismatch is not merely a mathematical difference between reported sales and third-party transactions. It may involve e-invoice gaps, bank-deposit discrepancies, unusual input-output ratios, inconsistent product classifications, payroll-to-headcount anomalies, customs import values that do not reconcile to domestic sales, or a registry profile that no longer matches actual activity. The practical challenge is to rank these mismatches by likelihood, materiality and treatment suitability rather than treating every exception as an audit case.

Recent work on explainable tax analytics supports this approach. Nayo et al. (2025), in a study co-authored by Munashe Naphtali Mupa, develop an explainable gradient-boosting approach to SME sales/use-tax compliance risk using registry information, point-of-sale summaries, seasonality and audit findings. Their emphasis on interpretability is important because tax officers and taxpayers must be able to understand why a case was flagged.

2.3 Compliance risk management and audit selection

Compliance risk management (CRM) provides the administrative structure for using analytics responsibly. Risk-based audit selection is designed to allocate scarce audit resources to cases where non-compliance is more probable or more material, while limiting unnecessary burdens on compliant taxpayers (Khwaja, Awasthi and Loeprick, 2011). The IMF's technical guidance on essential analytics for compliance risk management presents analytics as a tool for compliance planning, risk intelligence, case selection and evaluation (Aslett et al., 2024). The OECD similarly reports that tax administrations are increasingly using automated checks, cross-matching and data science to assess accuracy and completeness of taxpayer reporting (OECD, 2024).

A good risk model therefore has to perform several functions at once. It must rank cases, identify the reason for risk, estimate the revenue at risk, route cases to a proportionate compliance treatment and generate feedback for future model improvement. A model that identifies likely non-compliance but produces an unmanageable number of cases may fail operationally. A model that maximises immediate audit yield but concentrates audits unfairly on lower-income or easier-to-audit taxpayers may undermine equity and trust. Black et al. (2022) show that the design of tax audit models raises important questions of vertical equity, target definition and performance trade-offs. This is why human review, fairness monitoring and transparent governance are essential.

2.4 SMEs, education and digital compliance capability

SMEs require special attention because they often have limited administrative capacity, fragmented records and weaker access to professional tax advice. Technology may reduce compliance costs, but it can also create uneven benefits where smaller or less digitally prepared firms cannot adopt new systems effectively. Okunogbe and Pouliquen (2022) show that electronic filing can reduce compliance costs and corruption opportunities, but the benefits depend on adoption conditions and firm capabilities. World Bank work on digital tax administration similarly emphasizes that digitalisation must be accompanied by institutional capacity, taxpayer support and risk-based administration (Junquera-Varela et al., 2022; World Bank, 2025).

Compliance education should therefore be treated as part of the revenue-protection system, not as an afterthought. Where mismatches arise from poor recordkeeping, misunderstanding of VAT classification, payroll mistakes or filing calendar problems, education may prevent recurrence at lower cost than repeated audits. This logic is consistent with Nhemachena et al. (2026a), who propose integrating tax compliance, internal controls and standard operating procedures in community-serving and growth-stage organizations, and with Nhemachena et al. (2026b), who emphasise cash-flow visibility and accounting analytics for U.S. small businesses. Although those studies are not tax-audit studies, they support the broader proposition that compliance improves when operational controls become routine, documented and measurable.

The framework developed in this article therefore combines enforcement and education. It does not assume that all mismatches are fraudulent; it distinguishes probable evasion from administrative weakness. This distinction is vital for both fairness and efficiency.

Table 1. Literature synthesis linking tax-compliance evidence to the proposed framework

Evidence stream	Representative sources	Implication for the proposed framework
Third-party reporting and matching	Kleven et al. (2011); Pomeranz (2015); Carrillo, Pomeranz and Singhal (2017)	Use third-party, bank, payroll, customs, registry and e-invoice records to detect declared-return mismatches.
Risk-based audit selection	Khwaja, Awasthi and Loeprick (2011); Aslett et al. (2024); OECD (2024, 2025)	Prioritise cases by likelihood, materiality and treatment suitability rather than by blanket audit coverage.
Voluntary compliance	Allingham and Sandmo (1972); Kirchler,	Treat taxpayers differently: inform those willing to

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1719533

and behavioral responses	Hoelzl and Wahl (2008); Hallsworth et al. (2017); Slemrod (2019)	comply, enforce against deliberate non-compliance and evaluate behavioral nudges.
SME education and compliance capacity	Okunogbe and Pouliquen (2022); Junquera-Varela et al. (2022); Nayo et al. (2025)	Combine risk scoring with education because many SME errors arise from capability gaps, digital filing difficulties and weak recordkeeping.
Explainable analytics and audit dashboards	Black et al. (2022); Homwe et al. (2025); Nhemachena et al. (2026a)	Use interpretable indicators and heat maps so audit teams can explain, review and govern automated risk scores.

III. CONCEPTUAL FRAMEWORK

The proposed framework converts tax mismatches into revenue-protection decisions through six connected stages: data ingestion, mismatch detection, explainable risk scoring, revenue-at-risk estimation, treatment routing and feedback learning. The framework is designed for revenue authorities, internal compliance teams, tax advisory practices and businesses that need to monitor tax compliance across large volumes of taxpayer or client records.

At the core of the framework is a simple principle: not all mismatches deserve the same treatment. A low-value filing error by a first-time small taxpayer may require education and correction. A repeated high-value VAT mismatch by a taxpayer with prior audit adjustments may require enforcement review. A customs-sales mismatch may require reconciliation between import records, inventory and sales. A payroll/headcount mismatch may require employer education or PAYE audit. The framework turns these distinctions into operational rules that can be monitored through dashboards.

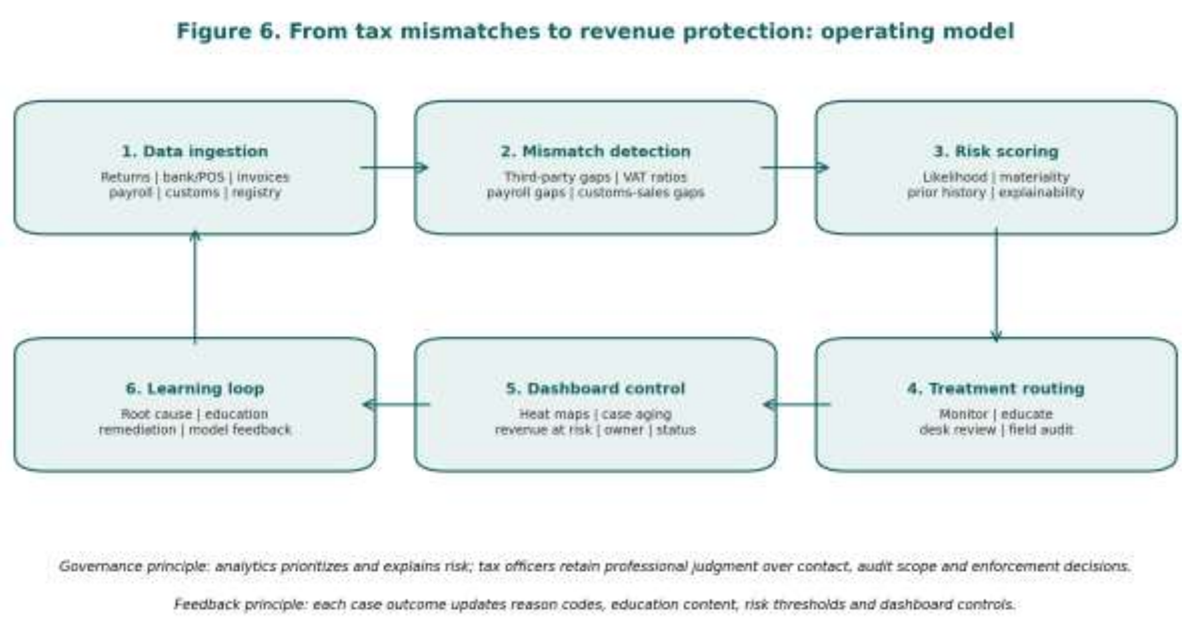

Figure 6. Integrated analytics and compliance-treatment cycle

Source: Author's conceptual framework based on compliance risk management and tax analytics literature.

Table 2. Operational data dictionary for mismatch-driven tax-risk analytics

Risk feature	Likely data source	Mismatch logic	Operational response
Return vs third-party reporting gap	Tax return, bank/POS/e-invoice/withholding records	Declared sales or income differ materially from third-party reported transactions.	Audit selection, advisory notice or return correction.
VAT input-output ratio anomaly	VAT returns, purchase ledgers, sales records	Input claims are unusual relative to output tax, sector norm or trading pattern.	VAT desk review, documentation request or classification education.
Payroll/headcount mismatch	PAYE returns, payroll records, social security or HR data	Reported payroll does not align with reported headcount, sales activity or sector labour norms.	Payroll compliance review and employer education.
Customs imports vs sales gap	Customs declarations, domestic sales, inventory records	High imports without corresponding sales, stock movement or customs-to-VAT reconciliation.	Customs-linked tax audit and inventory reconciliation.
Late filing or debt recurrence	Filing logs, payment ledger, arrears records	Repeated late filing, partial payment or tax debt signals liquidity or compliance risk.	Payment plan, education or enforcement escalation.
Registry/activity mismatch	Business registration, sector code, licences	Taxpayer activity in returns differs from registered activity or licence profile.	Registry update, sector classification review or audit.
Education need flag	Call-centre, outreach, filing errors, digital access	Errors appear capability-based rather than deliberately evasive.	Targeted SME education before punitive enforcement.

Source: Author's framework. The data sources listed should be adapted to each jurisdiction's legal and data-governance environment.

The framework uses three governing rules. The first is proportionality: case treatment should reflect both risk and materiality. The second is explainability: every selected case should have a reason code such as third-party mismatch, VAT ratio anomaly, payroll gap, customs-sales discrepancy or late filing recurrence. The third is learning: after every intervention, the outcome should update the taxpayer profile, improve risk definitions and refine education content.

This framework also recognizes the difference between revenue protection and revenue extraction. Revenue protection aims to collect legally due taxes, improve the integrity of future filings and reduce repeated errors. It should not create arbitrary audit pressure. The use of dashboards and heat maps therefore has to be coupled with governance: model validation, fairness review, audit trail controls, privacy protections and periodic evaluation of whether education cases are actually reducing repeated mismatches.

IV. METHODOLOGY AND DATA DESIGN

The empirical section uses a synthetic dataset of 5,400 taxpayer-period records. The dataset is illustrative and does not contain any real taxpayer, ZIMRA, IRS, Babson, client or government data. It is designed to demonstrate how a revenue authority or compliance team could structure real data once appropriate legal authority, privacy controls and data-governance arrangements are in place. The choice of 5,400 records was selected for illustrative purposes. The design is informed by consulting experience using SQL and Excel to analyze more than 3,000 small-business records across five U.S. cities and identify target accounts. The risk variables draw conceptually on experience with tax mismatches, high-risk portfolios, audit exception reports and compliance recovery (Ganyani, 2026).

Each record represents a taxpayer-period observation and contains sector, taxpayer segment, tax type, region, reported sales, third-party coverage, filing delay, mismatch flags, prior audit signal, education-need indicator, estimated revenue at risk and recommended treatment. The risk score is deliberately transparent. It combines mismatch flags,

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1719533

third-party coverage, filing delay, taxpayer segment and risk typology weights. The aim is not to claim predictive superiority, but to show an auditable scoring logic that can be replaced by jurisdiction-specific statistical or machine-learning models where real labelled data exist.

The methodology follows a design-science logic. It identifies an operational problem, develops a practical artifact, demonstrates the artifact using simulated records and evaluates whether the framework produces decision-useful outputs such as heat maps, risk tiers, revenue-at-risk concentration and treatment queues. This is appropriate for practitioner-oriented tax analytics because many administrations cannot publish taxpayer-level records, but can still test frameworks on synthetic data before deploying them on protected internal systems.

Table 3. Summary of the illustrative taxpayer-period dataset

Metric	Value
Illustrative taxpayer-period records	5,400
Total reported sales in simulation (USD)	485,446,002
Estimated revenue at risk (USD)	10,654,340
Average risk score	68.4/100
High-risk records	3,746 (69.4%)
Records with at least one mismatch flag	91.5%
Simulated audit-adjustment rate	57.7%
Records recommended for education before enforcement	58 (1.1%)
Top risk decile share of revenue at risk	9.5%

Source: Author's synthetic dataset. Figures are illustrative and should not be interpreted as actual taxpayer or jurisdictional results.

Table 4. Mismatch typologies ranked by estimated revenue at risk

Mismatch / signal	Flagged records	Share of records	Average risk score	Estimated revenue at risk (USD)
Return vs third-party reporting gap	1565	29.0%	79.9	3,645,412
VAT input-output ratio anomaly	1466	27.1%	78.4	3,322,035
Expense or margin outlier	1260	23.3%	77.6	3,151,131
SME education need flag	1657	30.7%	72.8	3,097,002
Customs imports vs domestic sales gap	1226	22.7%	78.1	2,908,280
Late filing or debt recurrence	1432	26.5%	77.0	2,839,281
Payroll/headcount mismatch	1145	21.2%	75.4	2,377,423
Prior audit adjustment signal	1004	18.6%	76.0	2,333,604
Registry/activity mismatch	1024	19.0%	76.3	1,925,370

Source: Author's synthetic dataset. Estimated revenue at risk combines materiality and risk score for demonstration purposes.

V. DATA ANALYSIS AND FINDINGS

5.1 Sector-level mismatch concentration

The first analytical output is a sector-by-mismatch heat map. It allows managers to see whether risk is concentrated in a particular sector or whether a specific mismatch type is widespread across the taxpayer base. In practice, this is useful because sector concentration may indicate the need for sector-specific guidance, while cross-sector concentration may indicate a systemic data-quality issue or a broadly misunderstood filing requirement.

Figure 1 shows that construction, agriculture and food, hospitality and retail carry relatively high levels of mismatch flags in the illustrative dataset. This is plausible in a compliance-risk sense because these sectors can combine cash transactions, inventory movement, subcontracting, seasonal activity and variable documentation quality. The point is not that these sectors are inherently non-compliant; rather, the heat map demonstrates how compliance teams can identify where to look and what type of intervention to design.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1719533

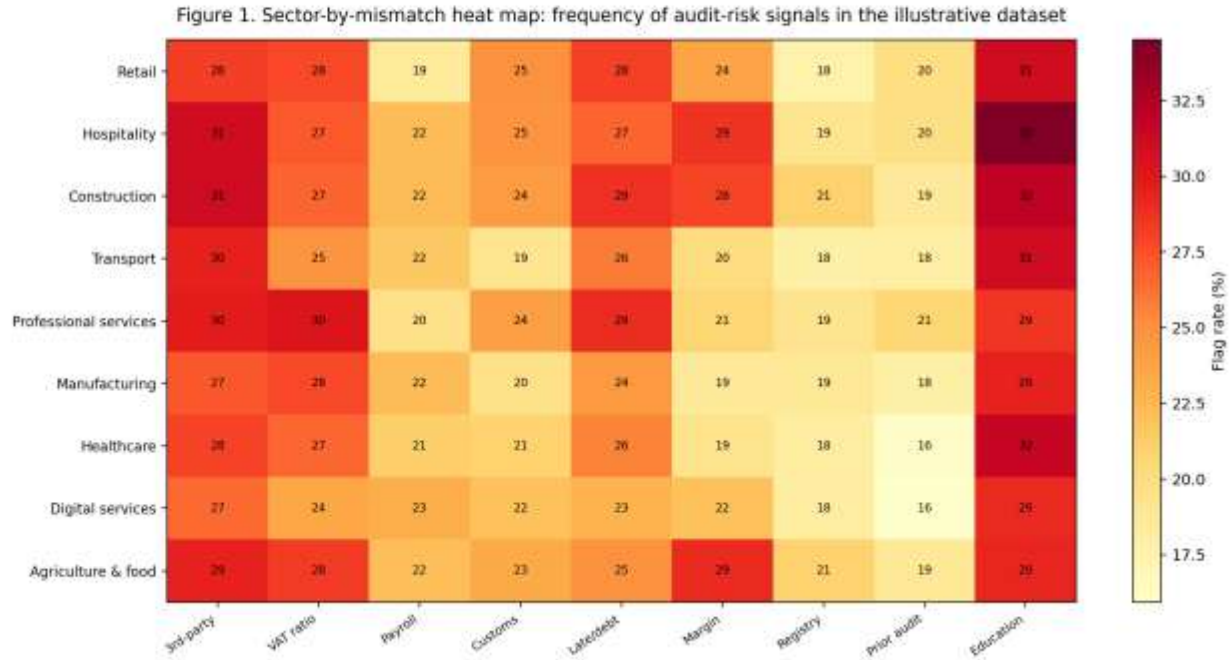

Figure 1. Sector-by-mismatch heat map

Table 5. Sector risk profile and recommended education-treatment share

Sector	Records	Average risk score	High-risk share	Revenue at risk (USD)	Education-treatment share
Construction	677	71.5	77.3%	2,045,356	0.4%
Agriculture & food	502	70.7	73.7%	622,822	0.8%
Hospitality	513	70.6	73.1%	653,127	0.8%
Retail	917	68.9	69.5%	1,949,701	1.2%
Professional services	654	68.3	68.7%	906,298	1.2%
Transport	538	66.9	67.3%	915,793	0.6%
Manufacturing	666	66.4	66.4%	2,008,420	1.1%
Digital services	483	66.1	62.9%	905,231	1.9%
Healthcare	450	65.3	63.1%	647,591	2.0%

5.2 Revenue-at-risk prioritization by tax type

Risk alone is not enough for audit planning. A case may have a high probability of adjustment but very low revenue impact; another may have moderate probability but large fiscal exposure. Figure 2 therefore cross-tabulates estimated revenue at risk by risk band and tax type. This helps audit managers allocate scarce skilled staff to the cases where the expected public value of review is highest.

In the illustrative results, high-risk VAT/sales-tax, customs-linked and corporate income tax records account for a large share of estimated revenue at risk. This is consistent with the literature on information reporting and VAT trails: mismatches in observable tax bases can be detected earlier, but firms may shift non-compliance to less observable margins if compliance strategies are too narrow (Pomeranz, 2015; Carrillo, Pomeranz and Singhal, 2017). The recommended response is to combine cross-tax matching with sector knowledge and case-level explanation.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1719533

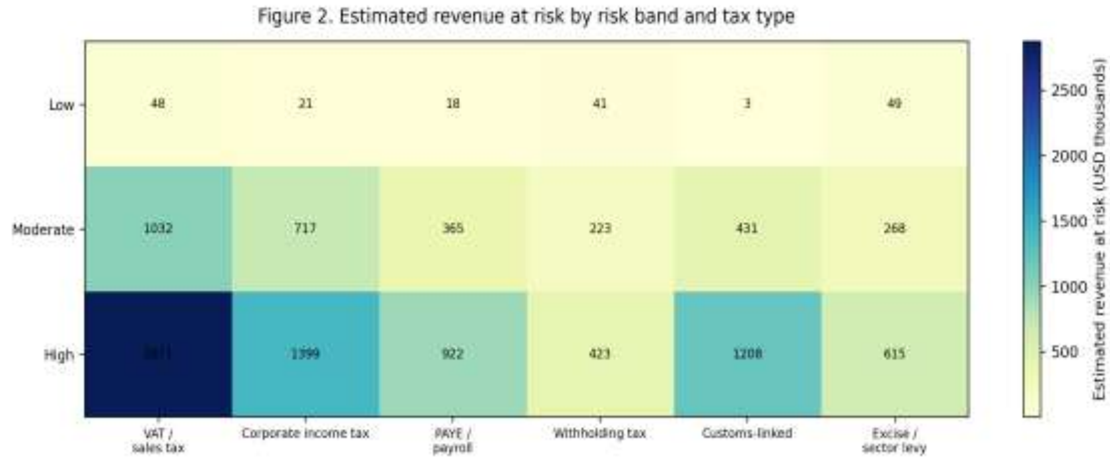

Figure 2. Revenue-at-risk heat map by risk band and tax type

5.3 Risk-decile capture and audit efficiency

A good audit-prioritization model should concentrate revenue-at-risk in the highest risk deciles without excluding fairness and random-audit needs. Figure 4 orders records from highest to lowest risk and compares the share of records reviewed with the cumulative share of estimated revenue at risk captured. This is a practical audit-planning tool because it helps managers test the marginal value of reviewing additional cases.

The capture curve shows why risk-based selection can outperform random selection when the scoring model is informative. The top risk decile captures a disproportionate share of estimated revenue at risk, and the highest three deciles capture substantially more than their numerical share of records. However, the figure should not be used to justify eliminating random audits. Random audits are important for estimating the tax gap, validating models and preventing blind spots in risk rules (Slemrod, 2019; Aslett et al., 2024).

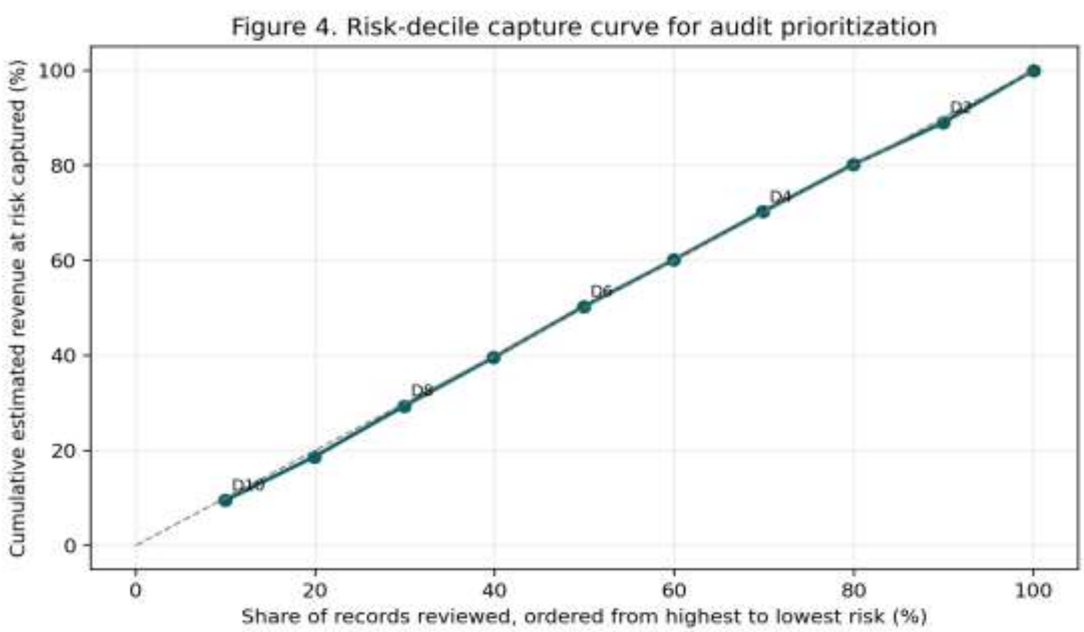

Figure 4. Risk-decile capture curve for audit prioritization

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1719533

5.4 Education need and voluntary compliance segmentation

The framework treats SME education as a revenue-protection instrument. Figure 3 maps education-need flags by taxpayer segment and learning barrier. This kind of heat map is useful because it separates non-compliance that may be caused by capability gaps from non-compliance that requires enforcement. For example, micro and small taxpayers with recordkeeping and digital filing challenges may benefit from structured support, whereas high-risk repeat cases should be escalated.

This segmentation is particularly important for administrations operating under self-assessment. If every mismatch produces a punitive response, taxpayers may become defensive and trust may decline. If every mismatch produces only education, deliberate non-compliance may persist. A balanced system uses data to match the treatment to the taxpayer's likely behavior and capacity.

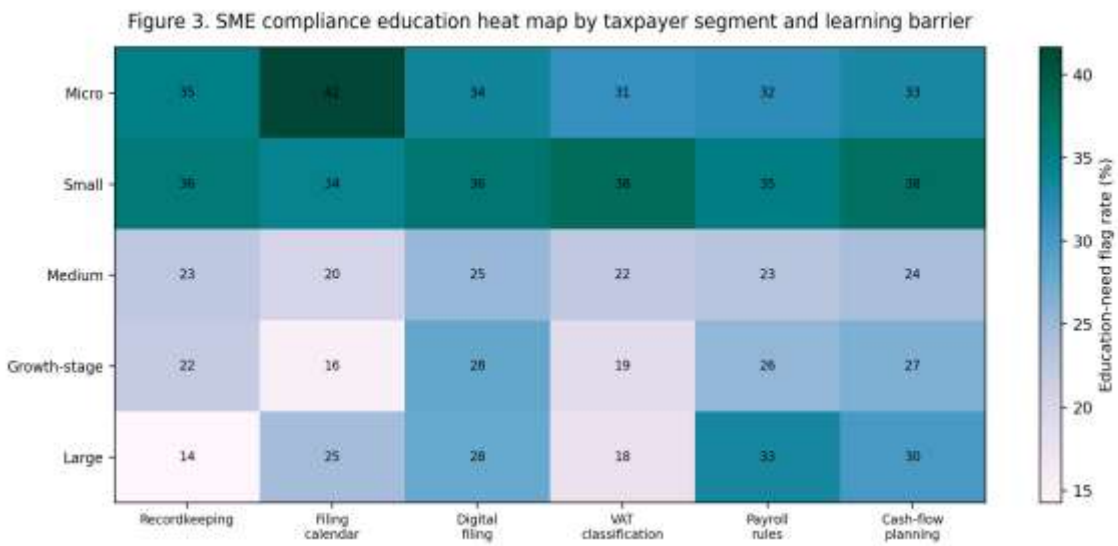

Figure 3. SME compliance education heat map

5.5 Threshold performance and human review

Figure 5 and Table 6 summarise the performance of a simple high-risk threshold in the illustrative dataset. The threshold should be understood as a management rule, not as a final legal decision. A tax officer should still review the facts, legal basis, documentation, taxpayer history and proportionality of the proposed intervention before any enforcement action is taken.

The confusion matrix shows the trade-off between identifying adjustment cases and avoiding unnecessary audits. A more aggressive threshold may increase revenue capture but also increase false positives and taxpayer burden. A stricter threshold may reduce burden but miss revenue risk. This trade-off is why dashboards should include precision, recall, case capacity, revenue-at-risk captured, fairness indicators and post-review outcomes rather than only reporting total assessed tax.

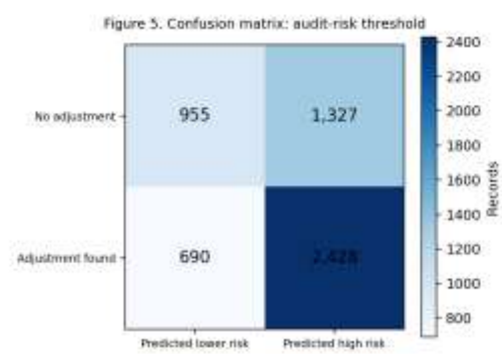

Figure 5. Confusion matrix for illustrative high-risk threshold

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1719533

Table 6. Illustrative threshold performance and revenue-at-risk capture

Measure	Value
High-risk threshold	Risk score >= 60
Precision	64.7%
Recall	77.9%
F1 score	0.71
Selected records	69.5%
Revenue-at-risk captured	70.0%

Source: Author's synthetic dataset. Metrics evaluate an illustrative risk score threshold and should be recalibrated when real audit labels are available.

VI. AUDIT PRIORITIZATION, EDUCATION AND DASHBOARD OPERATING MODEL

The proposed operating model is built around a treatment queue. Low-risk cases remain under automated monitoring; lower-risk education cases receive advisory nudges or structured training; moderate-risk cases enter desk review; high-risk and high-materiality cases enter field audit or enforcement review. The practical value of this routing logic is that it prevents the audit function from being used as the only compliance tool.

Table 7 summarises the recommended treatment logic. In practice, the thresholds should be set by each administration or compliance team based on staff capacity, taxpayer rights, legal requirements, data quality and model performance. The treatment categories should also be reviewed periodically because taxpayer behavior changes after enforcement campaigns, digital-filing changes or major sector shocks.

Table 7. Risk-based treatment matrix linking mismatches to compliance actions

Risk situation	Recommended treatment	Operational actions	Compliance objective
Low risk / low materiality	Monitor / automated validation	System validation, confirmation message, dashboard watchlist	Prevent errors without over-auditing compliant taxpayers.
Low-to-moderate risk / high education need	SME education + advisory nudge	Plain-language guidance, filing calendar, recordkeeping template, webinar or one-to-one support	Increase voluntary compliance and reduce repeated errors.
Moderate risk / medium materiality	Desk review / document request	Targeted document request, return amendment opportunity, data-matching notice	Resolve mismatches efficiently before field audit.
High risk / high materiality	Field audit / enforcement review	Formal audit, detailed reconciliation, penalty assessment where warranted	Protect revenue and deter deliberate non-compliance.
Repeat non-compliance	Escalated compliance management	Case owner, remediation plan, payment monitoring, repeat-risk score adjustment	Break cycles of recurring mismatches and arrears.

The dashboard prototype in Figure 7 translates the framework into a management view. It contains summary KPIs, average sector risk, a treatment queue and a trend panel for mismatch and education conversion. The important design feature is that the dashboard does not only rank cases; it links each case to an action owner and a treatment rationale. This reduces the risk of analytical outputs becoming unused reports.

The dashboard should be supported by a case log. Each case should include taxpayer identifier, period, tax type, mismatch reason code, risk score, estimated revenue at risk, evidence source, treatment category, officer, status, taxpayer response, adjustment amount, education outcome and closure notes. This case log can later support model validation, staff training, quality assurance and policy evaluation.

© JUL 2026 | IRE Journals | Volume 10 Issue 1 | ISSN: 2456-8880
DOI: https://doi.org/10.64388/IREV10I1-1719533

Figure 7. Audit-prioritization and SME compliance-education dashboard prototype

VII. IMPLEMENTATION ROADMAP

Table 8. Six-phase implementation roadmap for mismatch-driven tax compliance analytics

Phase	Main activities	Outputs
Phase 1: Data readiness	Map available return, third-party, payroll, customs, registry and payment data; define legal basis and access controls.	Data inventory, data dictionary, privacy assessment, governance owner.
Phase 2: Mismatch rules	Build transparent mismatch rules before advanced models; test with historic closed cases.	Reason-code library, exception definitions, data-quality checks.
Phase 3: Scoring and treatment routing	Combine likelihood, materiality and education indicators; assign treatment categories.	Risk score, revenue-at-risk model, treatment matrix, case queue.
Phase 4: Dashboard and workflow	Create heat maps, case queues, escalation rules and KPI reporting.	Management dashboard, officer workbench, case log, aging report.
Phase 5: Pilot and evaluation	Pilot with selected sectors or tax types; compare adjustment yield, case time, taxpayer burden and education outcomes.	Pilot report, calibration changes, fairness review, model validation.
Phase 6: Scale and feedback	Integrate outcomes into the model; update education content and risk rules.	Model monitoring, learning loop, updated guidance, annual evaluation.

A practical implementation should start modestly. A revenue authority or compliance team does not need a complex machine-learning system on day one. The first maturity stage is reliable matching: returns to third-party records, VAT outputs to inputs, payroll filings to headcount and customs import values to sales. The second maturity stage is prioritization: scores, thresholds and materiality. The third maturity stage is learning: model feedback, fairness review, random-audit calibration and education-outcome measurement.

Data governance must be built into the roadmap. Missing identifiers, inconsistent sector codes, duplicated taxpayer records and poor audit-outcome capture can weaken model reliability. The governance model should define who owns the data dictionary, who approves risk-feature changes, how access is controlled, how taxpayer privacy is protected, and how officers challenge model outputs. These controls are consistent with the broader governance and internal-control logic found in Mupa-linked research on audit analytics, operational controls and tax compliance frameworks (Homwe et al., 2025; Nhemachena et al., 2026a).

VIII. IMPLICATIONS FOR REVENUE AUTHORITIES AND COMPLIANCE-SENSITIVE BUSINESSES

For revenue authorities, the framework offers a way to improve audit selection without abandoning voluntary compliance. It can help compliance planning teams identify sector risks, estimate revenue exposure, protect audit capacity and design better taxpayer education. It can also help managers move away from purely reactive audits toward continuous compliance monitoring. This is consistent with OECD and IMF guidance on using analytics to support compliance risk management rather than treating analytics as a separate technical exercise (OECD, 2024; Aslett et al., 2024).

For SMEs, the framework highlights the importance of recordkeeping, reconciliation and timely filing. Many SMEs can reduce risk by maintaining a tax calendar, reconciling sales to bank/POS/e-invoice records, reviewing VAT input-output ratios, reconciling payroll filings to staff records, aligning customs imports to inventory and sales, and retaining documentation for sector-specific deductions or exemptions. This aligns with research on SME financial literacy, budgeting and control systems in underserved business environments, including Mupa-linked work on financial literacy and accounting analytics for small businesses (Hlahla, Mupa and Danda, 2025; Nhemachena et al., 2026b).

For compliance-sensitive businesses, the framework can be adapted into a pre-audit self-review model. Instead of waiting for a revenue authority notice, businesses can run internal mismatch dashboards monthly or quarterly. This can support tax governance, internal audit planning, board reporting and remediation. It also gives advisers a structured way to separate honest errors from high-risk patterns before engagement with the tax authority.

IX. LIMITATIONS AND ETHICAL SAFEGUARDS

This study has limitations. The dataset is synthetic, so the numerical results are illustrative rather than empirical evidence of actual tax behavior. Real deployment would require validated taxpayer-level data, legal authority to match data sources, data-protection safeguards, bias testing, random-audit calibration and officer training. The risk weights used in this article should be replaced with jurisdiction-specific estimates once historic audit outcomes are available.

There are also ethical risks. Algorithmic audit selection can create fairness problems if models optimise only revenue yield, rely on biased labels, or concentrate enforcement on taxpayers who are easier to audit rather than taxpayers who pose the greatest legal risk. Black et al. (2022) show that audit-model design choices can affect vertical equity. Therefore, revenue authorities should monitor audit selection by income group, sector, region, taxpayer type, prior compliance history and protected characteristics where legally permitted. The objective should be fair and effective compliance, not automated punishment.

Human-in-the-loop governance is essential. Analytics should prioritise, explain and monitor; it should not replace legal judgment. Taxpayers should have access to clear notices, correction opportunities, appeal rights and understandable explanations of the mismatch where disclosure is legally permissible. Education interventions should be evaluated not only by attendance but by subsequent reduction in repeated errors.

X. CONCLUSION

This article developed a data-driven framework for moving from tax mismatches to revenue protection. The framework integrates mismatch detection, risk scoring, revenue-at-risk ranking, treatment routing,

DOI: https://doi.org/10.64388/IREV10I1-1719533

audit exception dashboards and SME education. It responds to a central problem in tax administration: audit resources are limited, non-compliance risks are unevenly distributed, and many taxpayers need support to comply voluntarily.

The illustrative analysis shows how heat maps, typology tables, risk-decile capture curves, confusion matrices and dashboard prototypes can turn fragmented tax data into actionable compliance management. The strongest use of analytics is not simply to find more audits, but to choose the right intervention: monitoring for low-risk cases, education for capability gaps, desk review for resolvable mismatches and field audit for high-risk, high-materiality cases.

The framework reflects practical experience across tax audits, customs enforcement, compliance monitoring, revenue recovery, automated quota management, SQL/Excel analytics and business consulting. It presents a practical contribution that can support revenue authorities, SMEs, tax advisers and compliance-sensitive organizations seeking stronger, fairer and more data-driven tax compliance systems.

REFERENCES

[1] Allingham, M.G. and Sandmo, A. (1972) 'Income tax evasion: A theoretical analysis', Journal of Public Economics, 1(3-4), pp. 323-338.

[2] Alm, J. (2019) 'What motivates tax compliance?', Journal of Economic Surveys, 33(2), pp. 353-388.

[3] Aslett, J., Beer, S., Coelho, M. and Leduc, S. (2024) Tax Administration: Essential Analytics for Compliance Risk Management. Washington, DC: International Monetary Fund.

[4] Black, E., Elzayn, H., Chouldechova, A., Goldin, J. and Ho, D.E. (2022) 'Algorithmic fairness and vertical equity: Income fairness with IRS tax audit models', Proceedings of the 2022 ACM Conference on Fairness, Accountability, and Transparency, pp. 1479-1503.

[5] Carrillo, P., Pomeranz, D. and Singhal, M. (2017) 'Dodging the taxman: Firm misreporting and limits to tax enforcement', American Economic Journal: Applied Economics, 9(2), pp. 144-164.

[6] Ganyani, L. (2026) Professional Resume. Unpublished resume, Boston, MA.

[7] Hallsworth, M., List, J.A., Metcalfe, R.D. and Vlaev, I. (2017) 'The behavioralist as tax collector: Using natural field experiments to enhance tax compliance', Journal of Public Economics, 148, pp. 14-31.

[8] Hlahla, V., Mupa, M.N. and Danda, C. (2025) 'Advancing financial literacy in underserved communities: Building sustainable budgeting models for small businesses and nonprofits'. Available via ResearchGate.

[9] Homwe, T., Mupa, M.N., Matope, A., Mlambo, N., Chingezi, L. and Chihota, T.A. (2025) 'Interpretable machine learning for audit planning: Improving misstatement and compliance risk detection in financial services', World Journal of Advanced Research and Reviews, 28(2), pp. 925-933. doi: 10.30574/wjarr.2025.28.2.3779.

[10] Internal Revenue Service (IRS) (2025) Data Book, 2025. Washington, DC: Internal Revenue Service.

[11] Internal Revenue Service (IRS) (2026) The Tax Gap. Washington, DC: Internal Revenue Service.

[12] Junquera-Varela, R.F., Lucas-Mas, C.O., Krsul, I. and Calderon, J.R. (2022) Digital Transformation of Tax and Customs Administrations. Washington, DC: World Bank.

[13] Khwaja, M.S., Awasthi, R. and Loeprick, J. (eds.) (2011) Risk-Based Tax Audits: Approaches and Country Experiences. Washington, DC: World Bank.

[14] Kirchler, E., Hoelzl, E. and Wahl, I. (2008) 'Enforced versus voluntary tax compliance:

The slippery slope framework', Journal of Economic Psychology, 29(2), pp. 210-225.

[15] Kleven, H.J., Knudsen, M.B., Kreiner, C.T., Pedersen, S. and Saez, E. (2011) 'Unwilling or unable to cheat? Evidence from a tax audit experiment in Denmark', Econometrica, 79(3), pp. 651-692.

[16] Mupa, M.N. (2026) Google Scholar profile. Available at: https://scholar.google.com/citations?hl=en&user=JmqqLHQAAAAJ (Accessed: 8 June 2026).

[17] Mupa, M.N. (2026) ResearchGate profile. Available at: https://www.researchgate.net/profile/Munashe-Naphtali-Mupa/research (Accessed: 8 June 2026).

[18] Nayo, D., Mupa, M.N., Imene, F., Danda, C. and Mukwata, N.A. (2025) 'Detecting SME sales/use-tax compliance risk with explainable gradient boosting: Evidence from Midwestern retailers', World Journal of Advanced Research and Reviews, 28(2), pp. 131-140. doi: 10.30574/wjarr.2025.28.2.3687.

[19] Nhemachena, T.L., Shambare, A.M., Kufandada, D., Chingezi, E., Taanisa, T., Yelduora, P.G., Chawatama, B. and Mupa, M.N. (2026a) 'Integrating tax compliance, internal controls, and standard operating procedures in community-serving and growth-stage organizations: Building a unified operational-control framework', World Journal of Advanced Research and Reviews, 30(1), pp. 2624-2631. doi: 10.30574/wjarr.2026.30.1.1164.

[20] Nhemachena, T.L., Taanisa, T., Yelduora, P.G., Chawatama, B., Kufandada, D., Chingezi, E., Shambare, A.M. and Mupa, M.N. (2026b) 'Data-driven budget control, cash-flow visibility, and receivables optimization for U.S. small businesses: A practical accounting-analytics framework', World Journal of Advanced Research and Reviews, 30(1), pp. 2632-2640. doi: 10.30574/wjarr.2026.30.1.1163.

[21] OECD (2024) Tax Administration 2024: Comparative Information on OECD and Other Advanced and Emerging Economies. Paris: OECD Publishing.

[22] OECD (2025) Tax Administration 2025: Comparative Information on OECD and Other Advanced and Emerging Economies. Paris: OECD Publishing.

[23] Okunogbe, O. and Pouliquen, V. (2022) 'Technology, taxation, and corruption: Evidence from the introduction of electronic tax filing', American Economic Journal: Economic Policy, 14(1), pp. 341-372.

[24] Pomeranz, D. (2015) 'No taxation without information: Deterrence and self-enforcement in the value added tax', American Economic Review, 105(8), pp. 2539-2569.

[25] Slemrod, J. (2019) 'Tax compliance and enforcement', Journal of Economic Literature, 57(4), pp. 904-954.

[26] World Bank (2025) Leveraging Digitalization to Improve Tax Compliance and Revenue Mobilization. Washington, DC: World Bank.