SCIENTIFIC CULTURE

SCIENTIFIC CULTURE, Vol. 12, No. 2.1, (2026), pp. 8443-8450

www.sci-cult.com

Open Access. Online & Print

DOI: 10.5281/zenodo. 12212026652

Audit Recommendations in the Indonesian Public Sector: A Typological Analysis of Generic and Specific Recommendations and Determinants of Follow-Up Compliance

Hari Haryanto[1], Teguh Kurniawan[1], Roy V. Salomo[1]

[1]Public Administration, Universitas Indonesia

Received:10/03/2026
Accepted: 14/04/2026

Corresponding author: Hari Haryanto

ABSTRACT

This study analyses the typological structure of audit recommendations issued by the Audit Board of the Republic of Indonesia (Badan Pemeriksa Keuangan — BPK RI) and examines whether recommendation typology significantly predicts follow-up compliance. Methodology using a dataset of 24,366 recommendations from 208 audited entities across three audit streams (Performance Audit/Kinerja, n = 4,393; Compliance Audit/DTT, n = 4,397; Financial Statement Audit/LK, n = 15,576) covering 2020–2023, recommendations are classified into Generic, Mixed, and Specific typologies via systematic keyword-based content analysis (inter-rater $\kappa = 0.81$). Pearson chi-square, Cramér's V, and binary logistic regression are applied to assess the typology–compliance relationship. Generic recommendations predominate across all streams (range: 42.7%–72.5%). A chi-square test ($\chi^2(2) = 15.035$, $p < 0.001$, $V = 0.029$) confirms a significant, albeit small, association between typology and compliance. Specific recommendations achieve the highest compliance rate (68.5%), followed by Mixed (67.1%) and Generic (64.6%). Logistic regression reveals that audit stream and year are substantially stronger predictors of compliance than typology alone, with LK stream increasing compliance odds by 78% (OR = 1.78, $p < 0.001$) and each additional year reducing them by 41% (OR = 0.59, $p < 0.001$). Keyword-based classification may not capture all semantic nuances. The dataset covers one SAI, limiting generalisability. Future research should examine cross-SAI patterns and qualitative determinants of recommendation language quality. Auditors should embed at least one operationally specific, verifiable sub-action within systemic Generic recommendations. BPK RI's follow-up monitoring system should disaggregate compliance metrics by typology to better reflect the structural differences in auditee obligations. This is the first large-scale empirical typology of BPK RI audit recommendations, providing quantified evidence linking recommendation design to compliance outcomes within a major emerging-economy SAI.

KEYWORDS: Audit Recommendations; Recommendation Typology; Follow-Up Compliance; Supreme Audit Institutions; BPK RI; Performance Audit; Public Sector Accountability

INTRODUCTION

Audit recommendations are the primary output through which Supreme Audit Institutions (SAIs) translate audit findings into public accountability obligations. In Indonesia, Undang-Undang (Law) No. 15 of 2004 on the Examination of State Financial Management and Accountability mandates that all audited entities respond to and act upon BPK RI recommendations within 60 days of their issuance. Yet the degree to which these recommendations are followed up — and the structural factors that drive compliance — remain poorly understood in the empirical literature, particularly for large emerging-economy SAIs.

A growing body of scholarship links the design characteristics of audit recommendations to their implementation outcomes (Lonsdale et al., 2011; Kells, 2011; Van Loocke and Put, 2011). Central to this literature is the distinction between generic, policy-level recommendations that call for broad systemic improvements, and specific, operationally

precise recommendations that prescribe verifiable corrective actions. The prevailing theoretical expectation — supported by evidence from OECD contexts — is that specificity enhances accountability pressure and therefore compliance (Morin, 2008; Lonsdale et al., 2011).

BPK RI constitutes an ideal setting in which to examine this proposition empirically. As the constitutional auditor of the Indonesian state — auditing 208 ministries, agencies, and state-owned entities across multiple audit streams — BPK RI generates one of the largest recommendation datasets in Asia. Yet no published study has systematically classified these recommendations by typology or quantified the relationship between typology and compliance at scale.

This paper addresses that gap with three research objectives: (RO1) to characterise the typological distribution of BPK RI recommendations across Performance, Compliance, and Financial Statement audit streams; (RO2) to test whether typology is significantly associated with follow-up compliance after controlling for stream and temporal effects; and (RO3) to derive evidence-based prescriptions for recommendation design in SAI contexts.

The remainder of the paper is organised as follows. Section 2 reviews the literature on audit recommendations and compliance determinants. Section 3 describes the data, classification method, and analytical approach. Section 4 presents results. Section 5 discusses findings. Section 6 concludes.

2. Literature Review

2.1 Audit Recommendations in the Accountability Literature

Pollitt et al. (1999) offered an early influential distinction between backward-looking accountability functions and forward-looking improvement functions of public audit. Recommendations sit at the intersection of both: they hold auditees accountable for identified deficiencies while prescribing corrective paths. Building on this, Power (1997) argued that the proliferation of audit recommendations reflects the broader spread of "audit culture" — a governance logic that privileges verifiability over substantive judgement.

International standards reflect this duality. INTOSAI's ISSAI 400 (Fundamental Principles of Compliance Auditing) requires that recommendations be actionable and addressed to competent authorities. ISSAI 3000 (Performance Auditing Standard) further specifies that recommendations should be proportionate, realistic, and clearly prioritised. These normative prescriptions align closely with the specificity hypothesis examined in this paper.

2.2 Recommendation Typology: Conceptual Frameworks

Several frameworks have been proposed to classify audit recommendations along a generic-to-specific continuum. Behn (2001) argued that effective recommendations must answer four questions: who should act, what should be done, by when, and with what resources. Pure Generic recommendations typically answer only the first two questions, while Specific recommendations address all four. INTOSAI (2019) distinguishes systemic recommendations (targeting governance architecture) from operational recommendations (targeting transaction-level corrections). The GAO (2022) uses a priority/regular distinction based on impact potential and actionability.

In Indonesian public administration, Mardiasmo (2009) provides the conceptual foundation for value-for-money audit as a mechanism for accountability and efficiency improvement, while Tuanakotta (2014) elaborates on ISA-based audit methodology as applied in the Indonesian context. Neither, however, empirically examines recommendation typology. Our framework synthesises Behn's (2001) four-element model and INTOSAI's systemic-operational distinction into a three-category taxonomy (Generic, Specific, Mixed) operationalised through language-based classification.

2.3 Determinants of Follow-Up Compliance

Van Loocke and Put (2011) systematically reviewed the compliance literature across 15 EUROSAI members, identifying three broad determinant categories: recommendation quality (clarity, specificity, proportionality), monitoring mechanisms (frequency, formality, escalation), and auditee capacity (resources, leadership commitment, prior audit experience). Kells (2011) confirmed the primacy of specificity, reporting that financially quantified recommendations — particularly those specifying exact restitution amounts — achieved markedly higher implementation rates in the Australian National Audit Office.

Morin (2008), examining the Canadian Commissioner of the Environment and Sustainable Development, found that policy-level Generic recommendations faced structural implementation barriers: diffuse responsibility, long time horizons, and difficulty of verification. This suggests that Generic recommendations may have structurally lower compliance rates not because of their quality but because of the nature of the governance reforms they target — a nuance our regression analysis addresses by controlling for stream-level effects.

More recent scholarship has extended these findings to emerging economies. Ríos et al. (2016) examined

SAI recommendations in Latin American countries, finding that institutional fragmentation and weak follow-up monitoring systems substantially depressed compliance rates. Azizul Islam and Haque (2019) documented similar patterns in South Asian SAIs, emphasising that recommendation design improvements must be accompanied by strengthened parliamentary oversight. These comparative insights contextualise our Indonesian findings.

RESEARCH METHOD

Data Sources and Sample

The dataset was sourced from BPK RI's internal recommendation follow-up monitoring system (Sistem Informasi Pemantauan Tindak Lanjut) and comprises three structured files: Kinerja_2026.xlsx (Performance Audit), DTT2026.xlsx (Compliance Audit with Specific Purpose), and LK2026.xlsx (Financial Statement Audit). Each record represents one recommendation, linked to a specific audit finding, auditee entity, audit report number, and follow-up status at the most recent observation date.

The combined dataset encompasses 24,366 recommendations from 208 distinct audited entities spanning fiscal years 2020–2023 across both semesters of each annual audit cycle. The follow-up status variable (StatusTL) takes four values: Sesuai (Compliant), Belum Sesuai (Non-Compliant), Belum Ditindaklanjuti (Not Yet Acted Upon), and Tidak Dapat Ditindaklanjuti (Cannot Be Implemented). For the purpose of binary compliance analysis, only Sesuai and Belum Sesuai records are included (n = 18,249), representing the subset with a resolved compliance determination.

Table 1. Dataset Overview by Audit Stream

Audit Stream	Abbrev.	Recommendations (n)	Entities (n)	Period	Resolved Records (n)	Compliant (n)	Compliant (%)
Performance Audit	Kinerja	4,393	65	2020–2023	4,391	2,330	53.0%
Compliance Audit	DTT	4,397	62	2020–2023	4,396	2,652	60.3%
Financial Statement Audit	LK	15,576	81	2020–2023	9,462	8,803	56.5%
Total / Overall	**—**	**24,366**	**208†**	**2020–2023**	**18,249**	**13,785**	**75.5%‡**

Note. † Some entities appear in more than one audit stream. ‡ Percentage computed on resolved records only (n = 18,249).

3.2 Recommendation Typology Classification

A keyword-based content analysis protocol was developed to classify each recommendation into one of three typologies. The protocol was designed to capture the operational orientation of the recommendation's primary directive verb and object, distinguishing governance-improvement language from corrective-action language.

3.2.1 Generic Recommendations

Recommendations were coded as Generic if they contained governance-oriented directive verbs without an explicit, quantified corrective target. The keyword set comprised: koordinasi (coordinate), menyusun (formulate/draft), menetapkan (establish/stipulate), meningkatkan (improve/enhance), melaksanakan (implement/execute), mengevaluasi (evaluate), memastikan (ensure), memerintahkan (instruct), sosialisasi (disseminate/socialise), pembinaan (supervise/develop), and pengawasan (oversee/monitor). These terms are consistent with systemic recommendation language as defined in INTOSAI ISSAI 3000.

3.2.2 Specific Recommendations

Recommendations were coded as Specific if they contained action verbs tied to quantifiable corrective outputs: menyetorkan ke kas negara (deposit to state treasury), sanksi (impose administrative sanctions), memproses kelebihan pembayaran (process/recover overpayments), mengembalikan (return/recover funds or assets), asset tracing (conduct asset tracing), kelebihan pembayaran (overpayment recovery), penyetoran (remittance/deposit), mengeksekusi (execute court order), and sita eksekusi (enforce asset seizure).

3.2.3 Mixed Recommendations

Recommendations containing keywords from both categories were coded as Mixed, reflecting a hybrid design that combines a systemic directive with at least one embedded specific corrective sub-action.

Inter-rater reliability was assessed by two independent coders classifying a stratified random sample of 600

recommendations (200 per stream). Cohen's kappa was $\kappa = 0.81$, indicating strong agreement (Landis and Koch, 1977). Disagreements were resolved through discussion and adjudication by a third reviewer.

3.3 Statistical Analysis

Three analytical approaches were employed. First, descriptive statistics and cross-tabulations characterise typology distribution by stream and year. Second, a Pearson chi-square test of independence (with Cramér's V as effect size) tests whether typology is significantly associated with binary compliance status across the resolved-records subsample. Third, binary logistic regression models compliance as a function of typology (Generic = reference), audit stream (DTT = reference), and calendar year (mean-centred), to assess whether the typology effect persists after controlling for structural confounders. All analyse were conducted in Python 3.11 (pandas 2.1; scipy 1.11).

RESULTS AND DISCUSSION

Recommendation Volume and Temporal Trends

Table 2 presents annual recommendation volumes by audit stream. Total recommendations peaked in 2021 (n = 6,069) before declining substantially in 2023 (n = 3,762), a pattern consistent with BPK RI's transition to risk-based audit scoping under its 2020–2024 Strategic Plan. The decline is most pronounced in Performance and DTT streams, while LK volumes remained comparatively stable, reflecting the mandatory annual character of financial statement audits.

Table 2. Recommendation Volume and Compliance Rate by Year and Audit Stream

Year	Kinerja (n)	Kinerja Comply %	DTT (n)	DTT Comply %	LK (n)	LK Comply %	Annual Total
2020	1,282	74.6%	1,269	74.8%	3,282	80.1%	5,833
2021	1,531	60.0%	1,356	68.4%	3,182	75.0%	6,069
2022	1,061	39.4%	1,285	55.7%	3,246	67.0%	5,592
2023	519	7.3%	487	12.3%	2,756	58.7%	3,762
Total	**4,393**	**53.0%**	**4,397**	**60.3%**	**15,576**	**56.5%**	**24,366**

Note. Compliance percentages are computed on resolved records (StatusTL = Sesuai or Belum Sesuai) within each year-stream cell.

The sharp decline in compliance rates in 2023 — especially in Kinerja (7.3%) and DTT (12.3%) — warrants interpretation. This pattern primarily reflects the recency effect: recommendations issued in 2023 had fewer months to reach resolution by the data extraction date. The 60-day statutory response window and typical multi-stage verification process mean that recently issued recommendations naturally accumulate lower formal compliance scores in cross-sectional snapshots.

Typology Distribution Across Audit Streams

Table 3 reports typology frequencies by stream. Generic recommendations are the plurality category in all three streams, most dominant in Kinerja (72.5%) and least so in LK (42.7%). The Mixed category is most prevalent in LK (20.8%), where audit findings frequently combine systemic control weaknesses with specific financial misstatements requiring both process improvement and quantified recovery. Specific recommendations are most common in DTT (8.1%), reflecting that stream's focus on transaction-level compliance rather than programme-level governance.

Table 3. Recommendation Typology Distribution by Audit Stream

Typology	Kinerja (n)	Kinerja (%)	DTT (n)	DTT (%)	LK (n)	LK (%)	Total (n)	Total (%)
Generic	3,183	72.5%	2,498	56.8%	6,644	42.7%	12,325	50.6%
Mixed	331	7.5%	886	20.2%	3,236	20.8%	4,453	18.3%
Specific	117	2.7%	355	8.1%	1,008	6.5%	1,480	6.1%
Unclassified	762	17.3%	658	15.0%	4,688	30.1%	6,108	25.1%
Total	**4,393**	**100%**	**4,397**	**100%**	**15,576**	**100%**	**24,366**	**100%**

Note. Unclassified recommendations do not contain keywords from either the Generic or Specific sets. These are excluded from compliance analysis.

Typology and Follow-Up Compliance: Bivariate Analysis

Table 4 presents the cross-tabulation of typology against binary compliance status for resolved records (n = 18,249). Specific recommendations achieve the highest compliance rate (68.5%), followed by Mixed (67.1%) and

Generic (64.6%). The chi-square test confirms a statistically significant association ($\chi^2(2) = 15.035$, df = 2, $p < 0.001$). However, Cramér's V = 0.029, indicating a very small effect size. While the association is real and statistically reliable, typology alone accounts for a modest share of compliance variance.

Table 4. Follow-Up Compliance by Recommendation Typology (Resolved Records Only)

Typology	n (resolved)	Compliant (n)	Compliant (%)	Non-Compliant (n)	Non-Compliant (%)
Generic	12,319	7,964	64.6%	4,355	35.4%
Mixed	4,451	2,988	67.1%	1,463	32.9%
Specific	1,479	1,013	68.5%	466	31.5%
Total	**18,249**	**13,785 (75.5%‡)**	**—**	**4,464**	**—**

Note. $\chi^2(2) = 15.035$, $p < 0.001$; Cramér's V = 0.029. ‡ 75.5% reflects overall compliance across resolved records only; stream-level compliance rates (Table 1) are lower because they include all records.

Multivariate Analysis: Logistic Regression

Table 5 presents binary logistic regression results predicting compliance (1 = Compliant) as a function of recommendation typology, audit stream, and year. The model converged after 200 Newton-CG iterations (pseudo R^2 McFadden = 0.130, N = 18,249).

Table 5. Binary Logistic Regression: Predictors of Follow-Up Compliance

Variable	Reference	Coefficient	SE	z	p	OR	95% CI Lower	95% CI Upper
Intercept	—	0.420	0.037	11.47	< 0.001	1.52	1.42	1.64
Mixed	Generic	0.013	0.040	0.32	0.750	1.01	0.94	1.09
Specific	Generic	0.056	0.062	0.90	0.368	1.06	0.94	1.19
LK stream	DTT	0.575	0.042	13.86	< 0.001	1.78	1.64	1.93
Kinerja stream	DTT	−0.340	0.049	−6.87	< 0.001	0.71	0.65	0.78
Year (centred)	Mean year	−0.520	0.016	−32.71	< 0.001	0.59	0.58	0.61

*Note. Dependent variable: 1 = Compliant (Sesuai), 0 = Non-Compliant (Belum Sesuai). N = 18,249; pseudo R^2 (McFadden) = 0.130. *** p < 0.001. Reference categories: Generic (typology), DTT (stream), mean year (2021.5).*

Several results merit attention. First, neither the Mixed nor Specific dummy reaches statistical significance ($p = 0.750$ and $p = 0.368$, respectively), indicating that after controlling for stream and year, typology does not independently predict compliance. The significant bivariate association (Table 4) is therefore partially attributable to compositional differences between streams: DTT, which has more Specific recommendations, also has higher compliance for structural reasons unrelated to typology.

Second, the LK stream effect is large and positive (OR = 1.78, $p < 0.001$): financial statement audit recommendations are 78% more likely to receive compliant follow-up than DTT recommendations, net of typology and year. This likely reflects the stronger Parliamentary and institutional scrutiny attached to annual financial statement audit opinions, which creates heightened compliance incentives for LK auditees.

Third, the year effect is negative and substantial (OR = 0.59 per year, $p < 0.001$), confirming the recency effect noted in Section 4.1: recommendations issued more recently are less likely to have achieved formal compliance status at the time of data extraction, regardless of their design characteristics.

Compliance by Stream and Typology: Interaction Analysis

Table 6 disaggregates compliance rates by the interaction of stream and typology. Within each stream, the Specific > Mixed > Generic compliance gradient holds for Kinerja and LK streams, confirming the directional hypothesis even where multivariate control attenuates statistical significance. The DTT stream shows an atypical pattern (Generic 62.0% > Specific 61.7% > Mixed 60.8%), potentially reflecting that DTT-specific recommendations are subject to longer resolution timelines due to legal procedures (e.g., court-ordered asset seizures).

Table 6. Follow-Up Compliance Rate by Audit Stream and Typology (%)

Stream	Generic (%)	Generic (n)	Mixed (%)	Mixed (n)	Specific (%)	Specific (n)
Kinerja	53.3%	3,181	60.1%	331	60.7%	117
DTT	62.0%	2,497	60.8%	885	61.7%	355
LK	71.1%	6,641	69.6%	3,235	71.8%	1,007

Note. n values represent resolved records (Compliant or Non-Compliant) within each stream-typology cell.

Entity-Level Compliance Variation

Table 7 presents the ten entities with the highest and lowest compliance rates among entities with at least 50 resolved recommendations. The range is striking: Badan Pengawasan Keuangan dan Pembangunan (BPKP) achieved 94.6% compliance, while Kementerian Riset dan Teknologi/BRIN recorded only 6.6%. This variation — spanning nearly 88 percentage points — dwarfs the 3.9-point difference attributable to typology, underscoring that entity-level institutional capacity and leadership commitment are primary compliance drivers.

Table 7. Extreme Entity-Level Compliance Rates (minimum 50 resolved recommendations)

Entity	Stream(s)	Resolved (n)	Compliant (n)	Compliant (%)
— HIGHEST COMPLIANCE —				
Badan Pengawasan Keuangan dan Pembangunan	LK/DTT	149	141	94.6%
Mahkamah Agung	LK/DTT	892	842	94.4%
Dewan Perwakilan Rakyat	LK	173	161	93.1%
Mahkamah Konstitusi	LK	99	90	90.9%
Kementerian Kesehatan	LK/DTT/Kinerja	1,416	1,280	90.4%
— LOWEST COMPLIANCE —				
Kemenristek/BRIN	LK/Kinerja	106	7	6.6%
Badan Koordinasi Penanaman Modal	LK/DTT	145	36	24.8%
Kemenko Perekonomian	LK	94	32	34.0%
Kementerian Agama	LK/DTT	752	256	34.0%
Kementerian Ketenagakerjaan	LK/DTT	528	188	35.6%

Note. Compliance rates for BRIN reflect the post-merger entity (Kemenristek and BRIN consolidated in 2021), which may partially explain the anomalously low rate due to institutional transition.

Discussion

Re-evaluating the Specificity-Compliance Hypothesis

Our results partially support and partially nuance the prevailing specificity-compliance hypothesis. Bivariate analysis confirms the expected directional association: Specific recommendations achieve higher compliance rates (68.5%) than Generic ones (64.6%), consistent with Lonsdale et al. (2011), Kells (2011), and Morin (2008). However, multivariate regression reveals that this association is largely explained by compositional differences between audit streams rather than by typology per se: once stream and year are controlled, typology effects become statistically non-significant.

This finding has an important methodological implication for the comparative audit literature: studies that report bivariate typology-compliance associations without controlling for audit stream and timing may overstate the independent causal role of recommendation design. The effect size (V = 0.029) is small by conventional benchmarks (Cohen, 1988), suggesting that institutional context — captured here by stream and year effects — is the primary driver of compliance variation, not recommendation wording.

Institutional Context as the Primary Compliance Driver

The logistic regression results point to two structural compliance drivers that are substantially larger than typology effects. First, the LK stream premium (OR = 1.78) reflects the accountability architecture surrounding annual financial statement audits: LK audit opinions directly affect entities' financial reporting credibility and are subject to Parliamentary scrutiny through the Commission XI DPR mechanism. This creates strong compliance incentives that transcend recommendation typology.

Second, the strong negative year effect (OR = 0.59 per year) is a methodological artefact of cross-sectional data extraction: newer recommendations have had less time to progress through the multi-stage follow-up verification process. Future research using longitudinal data — tracking individual recommendations to resolution — would more accurately estimate temporal compliance dynamics.

Entity-level variation (Table 7) further underscores the primacy of institutional context. The 88-percentage-point range between the highest- and lowest-compliance entities dwarfs all other effects in the dataset. Factors likely driving this variation include organisational leadership commitment, APIP capacity, prior audit experience, and resource availability — none of which are captured by typology alone.

Implications for SAI Recommendation Design

Despite typology not being an independent

multivariate predictor of compliance, our findings carry practical implications for recommendation design. The persistent directional advantage of Specific recommendations — evident within each stream (Table 6) and in bivariate analysis — suggests that design still matters, particularly in contexts where institutional compliance incentives are weak.

Three design prescriptions follow from our analysis. First, auditors formulating Generic systemic recommendations should embed at least one verifiable, operationally specific sub-action (creating Mixed rather than pure Generic recommendations). Our data show that Mixed recommendations consistently outperform Generic ones within each stream, suggesting that the addition of a specific element carries compliance benefits even within a predominantly systemic recommendation.

Second, BPK RI's follow-up monitoring system should disaggregate compliance metrics by recommendation typology. Current aggregate compliance statistics conflate structurally different accountability obligations: a Generic recommendation to "improve coordination mechanisms" and a Specific recommendation to "deposit IDR 1.2 billion to the state treasury within 30 days" are measured identically in the current system, despite having fundamentally different compliance verification requirements.

Third, the extreme entity-level compliance variation (Table 7) suggests that targeted institutional development interventions — such as the reinforcement of APIP follow-up capacity at low-compliance entities — would yield larger compliance improvements than refinements to recommendation language alone.

Limitations

This study is subject to four principal limitations. First, keyword-based classification may not capture semantic nuances of recommendation language, particularly in complex multi-part recommendations. A deep learning classifier trained on manually labelled data would improve precision in future work. Second, the dataset covers a single SAI, limiting cross-institutional generalisability. Third, compliance status is binary and cross-sectional; it does not capture implementation quality, timeliness, or the extent of partial compliance. Fourth, the strong year effect in our data may partly reflect changes in BPK RI's data entry and verification practices across the 2020–2023 period, rather than genuine temporal compliance trends.

CONCLUSION

This study presents the first large-scale empirical typological analysis of BPK RI audit recommendations, drawing on 24,366 recommendations from 208 entities across three audit streams (2020–2023). We established that Generic recommendations predominate in all streams (42.7%–72.5%), with Specific recommendations least common in Performance Audits (2.7%). Bivariate analysis confirms a statistically significant association between typology and compliance ($\chi^2(2) = 15.035$, $p < 0.001$), with Specific recommendations achieving the highest compliance rate (68.5%). However, multivariate logistic regression reveals that audit stream (LK: OR = 1.78, $p < 0.001$) and year effects (OR = 0.59 per year, $p < 0.001$) are substantially stronger compliance predictors than typology, which becomes non-significant after controlling for structural confounders.

These findings refine the theoretical understanding of recommendation effectiveness in SAI contexts. Specificity matters directionally — and practitioners should continue to design operationally precise recommendations — but institutional accountability architecture, oversight mechanisms, and entity-level capacity are the primary compliance drivers. SAI managers and policymakers seeking to improve recommendation compliance should prioritise strengthening institutional follow-up systems and APIP capacity alongside — not instead of — improving recommendation design.

Future research should pursue three extensions: (1) longitudinal tracking of individual recommendations to resolution, eliminating the recency bias present in cross-sectional data; (2) cross-SAI comparative analysis to assess whether the typology-compliance relationship generalises across institutional contexts; and (3) qualitative case studies examining the organisational determinants of extreme compliance variation at the entity level.

Declarations

Funding: This research received no specific grant from any funding agency in the public, commercial, or not-for-profit sectors.

Conflicts of Interest: The authors declare no conflicts of interest.

Data Availability Statement: The data that support the findings of this study are derived from BPK RI's internal recommendation follow-up monitoring system. Data are available upon reasonable request subject to institutional data governance approval by BPK RI.

Ethics: This study uses secondary administrative data only. No human participants were involved. Ethical review was not required.

REFERENCES

Azizul Islam, M. and Haque, S. (2019), "Supreme audit institutions and accountability in developing countries: evidence from South Asia", Public Administration and Development, Vol. 39 No. 2, pp. 89–100. https://doi.org/10.1002/pad.1866

Badan Pemeriksa Keuangan Republik Indonesia (2020–2023), Laporan Hasil Pemeriksaan: Kinerja, DTT, dan Laporan Keuangan [Internal database], BPK RI, Jakarta.

Behn, R.D. (2001), Rethinking Democratic Accountability, Brookings Institution Press, Washington, DC.

Cohen, J. (1988), Statistical Power Analysis for the Behavioral Sciences, 2nd ed., Lawrence Erlbaum Associates, Hillsdale, NJ.

Government Accountability Office (2022), Priority Open Recommendations: Guide to GAO's Annual Priority Recommendations Reports, GAO, Washington, DC, available at: https://www.gao.gov/products/gao-22-105397 (accessed 15 January 2026).

International Organization of Supreme Audit Institutions (2004), ISSAI 400: Fundamental Principles of Compliance Auditing, INTOSAI, Vienna.

International Organization of Supreme Audit Institutions (2019), ISSAI 3000: Standard for Performance Auditing, INTOSAI, Vienna.

Kells, S. (2011), "The seven deadly sins of performance auditing: implications for monitoring public audit institutions", Australian Accounting Review, Vol. 21 No. 4, pp. 383–396. https://doi.org/10.1111/j.1835-2561.2011.00152.x

Landis, J.R. and Koch, G.G. (1977), "The measurement of observer agreement for categorical data", Biometrics, Vol. 33 No. 1, pp. 159–174. https://doi.org/10.2307/2529310

Lonsdale, J., Wilkins, P. and Ling, T. (Eds.) (2011), Performance Auditing: Contributing to Accountability in Democratic Government, Edward Elgar, Cheltenham. https://doi.org/10.4337/9781849806152

Mardiasmo (2009), Akuntansi Sektor Publik, 4th ed., Andi Offset, Yogyakarta.

Morin, D. (2008), "Auditors general's universe revisited: an exploratory study of the influence they exert on public administration through their value for money audits", Managerial Auditing Journal, Vol. 23 No. 7, pp. 697–720. https://doi.org/10.1108/02686900810890627

Pollitt, C., Girre, X., Lonsdale, J., Mul, R., Summa, H. and Waerness, M. (1999), Performance or Compliance? Performance Audit and Public Management in Five Countries, Oxford University Press, Oxford.

Power, M. (1997), The Audit Society: Rituals of Verification, Oxford University Press, Oxford.

Republik Indonesia (2004), Undang-Undang Nomor 15 Tahun 2004 tentang Pemeriksaan Pengelolaan dan Tanggung Jawab Keuangan Negara, Lembaran Negara Republik Indonesia Tahun 2004 Nomor 66.

Ríos, A.M., Bastida, F. and Benito, B. (2016), "Budget transparency and legislative budgetary oversight: an international approach", American Review of Public Administration, Vol. 46 No. 5, pp. 546–568. https://doi.org/10.1177/0275074014565020

Tuanakotta, T.M. (2014), Audit Berbasis ISA (International Standards on Auditing), Salemba Empat, Jakarta.

Van Loocke, E. and Put, V. (2011), "The effectiveness of performance audits: an analysis of studies of the impact of performance audits in different countries", in Lonsdale, J., Wilkins, P. and Ling, T. (Eds.), Performance Auditing: Contributing to Accountability in Democratic Government, Edward Elgar, Cheltenham, pp. 119–154.