Open Access Indonesia Journal of Social Sciences

Vol. 9 · No. 4 · 2026 · pp. 181–193 · Article ID 331

e-ISSN 2722-4252 · *Journal homepage: journalsocialsciences.com/index.php/oaijss*

ORIGINAL ARTICLE

Digital Government Maturity and Public Service Performance in Indonesia: A Three-Wave Ordinal Panel Study of Level Versus Within-Unit Change

Alex Putra Pratama[1,*], **Aylin Yermekova**[2]

[1] *Department of Civil Law, Enigma Institute, Palembang, Indonesia*
[2] *Department of State Law, Bolatov Research Center, Semey, Kazakhstan*

ARTICLE INFORMATION

Received: June 29, 2026
Revised: July 24, 2026
Accepted: August 22, 2026
Available Online: August 28, 2026
Published: August 30, 2026

Keywords: digital government; Indonesia; ordinal panel analysis; public service performance; technology enactment

***Corresponding author**
Name: Alex Putra Pratama
E-mail: alex.putra.pratama@enigma.or.id
ORCID: 0009-0003-2823-6285
DOI: https://doi.org/10.37275/oaijss.v9i4.331

ABSTRACT

Background: Digital government maturity is assumed to raise public service quality, but longitudinal subnational evidence is scarce and rarely separates whether well-digitised governments simply perform better from whether becoming more digitised makes a government perform better.

Objective: To estimate the association between within-entity change in electronic government maturity and the ordered transition in public service performance predicate among Indonesian local governments, and to distinguish that estimand from the lagged-level association.

Methods: Two Indonesian national open datasets for 2023–2025 were linked on normalised agency names: the Sistem Pemerintahan Berbasis Elektronik (SPBE) maturity index and the Indeks Pelayanan Publik (IPP) performance index. Guided by technology enactment theory, 194 consecutive annual transitions from 135 entities within 665 entity-year observations were modelled with cumulative-link proportional-odds regression adjusted for baseline IPP, interval, and government level, with entity-clustered robust standard errors.

Results: The adjusted odds ratio for a better IPP transition per 0.5-point SPBE gain was 1.400 (95% CI 0.780–2.513; p = 0.261; Benjamini–Hochberg adjusted p = 0.420); proportional odds held for the exposure (p = 0.914). The first difference was null (β = 0.003; 95% CI −0.085–0.091; p = 0.953), whereas SPBE level predicted the following year's IPP score (β = 0.195; 95% CI 0.079–0.312; p = 0.001; adjusted p = 0.006). The design could detect only odds ratios of 2.31 or larger at 80% power.

Conclusion: Digital maturity and assessed service performance appear jointly produced by durable institutional capacity rather than by annual increments in digitisation. Assessment coverage moved by more than 40% between adjacent waves, so Indonesia's evaluation infrastructure must be stabilised before causal claims can be tested.

All authors reviewed and approved the final manuscript.

1. Introduction

Governments across the middle-income world have converted the promise of digital administration into budget lines, ministerial mandates and, increasingly, into scoreboards. Most now assess their own agencies annually against a national electronic government maturity rubric and publish the result, on the premise that a more digitally mature bureaucracy delivers better services. That premise is rarely tested against an independently assessed measure of service performance. Landscaping across the European Union found digital deployment in nearly every administrative domain but attributable outcome evidence in very few,[1] and systematic stock-taking reports a field still dominated by cross-sectional designs with weak identification.[2,3] Reviews of crisis-era transformation reach the same verdict.[4,5] The result is a policy environment in which capability scores circulate as though they were impact evidence, a risk sharpened by experimental evidence that published rankings induce cognitive shortcuts which obscure the data beneath them.[6,7]

Indonesia is an unusually informative setting because it publishes both halves of the equation as open data. The Sistem Pemerintahan Berbasis Elektronik (SPBE) framework, established under Presidential Regulation 95/2018[8] and scored throughout the study window under Ministerial Regulation 59/2020,[9] produces an annual maturity index for every assessed agency on a 0–5 scale with five ordered predicate bands, weighted towards the policy, governance, management and service domains of digital administration. The Indeks Pelayanan Publik (IPP), scored under Ministerial Regulation 29/2022,[10] produces an annual public service performance index across six aspects, reported on the same numerical range but banded into seven ordered predicates. The country administers more than five hundred autonomous regencies, cities and provinces under an explicitly asymmetric decentralisation settlement,[11] alongside a parallel village governance reform[12] and persistent customary institutions,[13] and its local governments differ sharply in fiscal capacity[14,15] and in administrative capacity generally.[16,17] That heterogeneity is precisely the variation an observational design needs.

The organising theory is Fountain's technology enactment framework, which holds that objective information technology — the servers, portals and interoperability layers a maturity rubric can score — becomes enacted technology only after passing through existing organisational forms and institutional arrangements.[18] Contemporary work has both confirmed and operationalised the claim: configurational analysis of government-led projects shows that combinations of organisational conditions, not technology, drive enactment,[19] institutional response rather than technical capability determines whether digital administration retains legitimacy,[20] and integrative frameworks now separate e-government supply, adoption and outcome layers explicitly.[21,22] Public service logic locates value creation at the service encounter rather than in the productive capacity of the organisation,[23,24] which implies a construct gap between an institutional maturity index and a performance index built from sampled service units; proactivity, a defining quality of good digital service, is not what maturity rubrics score.[25] Together these predict a testable asymmetry: the association between the level of maturity and subsequent performance should be robust, because both are downstream of the same durable institutional capacity, while the association between a one-year change in maturity and a change in performance should be weak, because enactment is slow.

The empirical record is consistent with that expectation but has almost never tested it directly. Structural analyses find organisational and managerial factors dominating technological ones in explaining realised transformation;[26] legacy systems act as a structural brake across European administrations;[27] and transformation rhetoric routinely decouples from administrative practice.[28] Whether capability reaches the counter depends on frontline adaptation,[29] employee willingness,[30] user experience,[31,32] co-production with citizens,[33] and identifiable champions.[34] Implementation constraints in public organisations are path-dependent and persist across annual reporting cycles,[35] iterative delivery lengthens rather than shortens the interval before service change is visible,[36] and subnational transformation is constrained by fiscal and staffing conditions that vary widely inside a single country.[37] In lower- and middle-income settings the demand side binds hardest: adoption depends on trust and digital literacy,[38,39] on accessible design,[40] and on platform governance,[41] and ICT improves service delivery in Tanzanian local government only where citizen digital literacy is sufficient.[42] Cross-country panel evidence finds digital diffusion affecting development outcomes only with substantial lags.[43]

A second literature explains why the outcome measure may move slowly even when services improve. Performance measurement systems change organisations only under specific conditions;[44] local governments differ in whether they use performance data at all;[45] measurement invites gaming and distortive behaviour;[46] administrative intensity behaves as capacity rather than burden;[47] structural reform changes performance with a lag;[48] reporting practice shapes what value is recognised;[49] collaboration across organisations is a precondition for translating reform into outcomes;[50] openness affects trust only through institutional capacity;[51] and turbulence in the evaluation environment degrades comparability between successive assessments.[52,53] Composite public-value impacts are, in any case, poorly captured by a single score,[54] and governance of evaluation requires explicit guideline architecture rather than capability scoring alone.[55] Municipal practice converges internationally on assessed dimensions while outcomes remain locally determined,[56] and leadership — a strong candidate mechanism — is only now acquiring validated measurement.[57]

A third constraint is infrastructural, and it turned out to be decisive here. Data stewardship roles vary widely across government data ecosystems and determine whether data can be reused at all;[58] governance design rather than technical capacity determines whether datasets can be combined;[59,60] transparency-oriented datasets frequently break down before they can support accountability;[61] dataset quality frameworks emphasise completeness and consistency as first-order properties;[62] linkage quality must itself be measured when stable identifiers are absent;[63] and reusing linked administrative data for evaluation presupposes institutional arrangements that often do not exist.[64] Indonesian evidence shows local institutional traits predicting disclosure behaviour,[65] resource dependence shaping what capacity can be mobilised,[66] and corruption tracking regional performance[67] — none of which appears in either national index.

This study therefore had four aims: to estimate the direction and magnitude of the association between within-entity change in SPBE maturity and the ordered IPP predicate transition; to separate that within-unit change estimand explicitly from the lagged-level estimand; to assess the sensitivity of both to panel completeness and to the inclusion of provincial governments; and to document the data-quality constraints that determine whether longitudinal evaluation of Indonesian digital government reform is feasible at all. Three hypotheses follow from the enactment framework, and are set out with the conceptual model in Figure 1. H1: the SPBE score at wave t is positively associated with the IPP score at wave t+1 after adjustment for IPP at t. H2: within-entity increases in SPBE are associated with higher odds of a better IPP predicate transition. H3: the level association exceeds the within-unit change association over a one-year horizon. Because H1 and H2 are expressed on different scales, H3 cannot be evaluated by a single contrast statistic; the criterion, fixed in advance, is that H3 is supported if the lagged-level estimate is distinguishable from the null after multiplicity adjustment while the change estimate is not, and that pattern persists across the pre-specified sensitivity analyses.

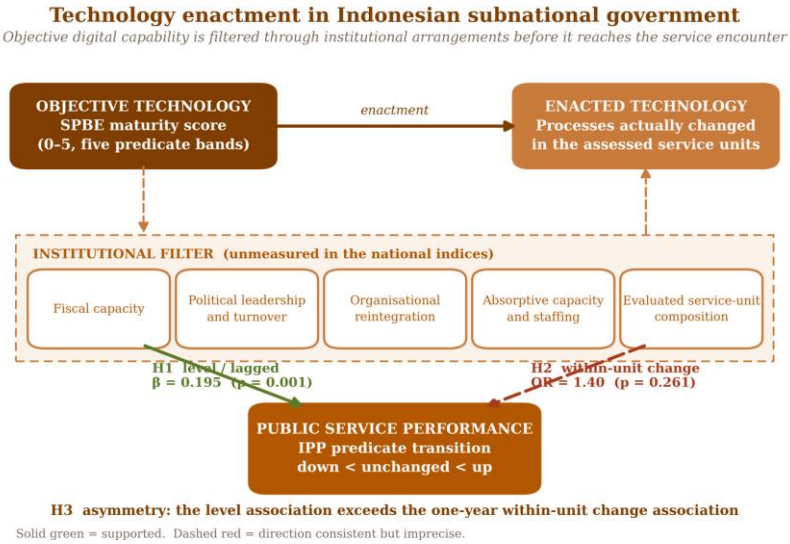

Figure 1. **Conceptual framework: technology enactment in Indonesian subnational government. Objective digital capability, measured by the SPBE maturity index, must pass through an institutional filter of fiscal capacity, leadership, organisational reintegration, absorptive capacity, and evaluated service-unit composition before it can affect the service encounter measured by the IPP predicate. H1 (solid) was supported; H2 (dashed) was not; H3, the asymmetry between them, was supported qualitatively.**

2. Methods

2.1. Study design and setting

This is a retrospective observational analysis of an ecological three-wave subnational panel covering calendar years 2023, 2024 and 2025. The setting is the whole of Indonesia at the subnational tier: provinces, regencies and cities assessed by the Ministry of Administrative and Bureaucratic Reform (Kementerian PANRB). The basic unit is a local government in a single year; the longitudinal unit is a consecutive one-year transition for the same entity. Because the unit of analysis is a government rather than a person, all inferences apply to institutions, and the ecological-inference caution that governs estimation from aggregate units applies throughout.[68] No statement in this paper should be read as a claim about individual citizens, service users or civil servants. Reporting follows the STROBE recommendations for observational research as far as they apply to an ecological design.

2.2. Participants, sampling, and data sources

There are no human participants. The sampling frame is the set of local governments appearing in the two national evaluation datasets published openly through Satu Data Aparatur Negara: the SPBE evaluation dataset[69] and the IPP dataset,[70] both for 2023–2025. Inclusion required that a local government appear in both datasets in the same year with a recoverable numeric score. The public CSV exports omit some numeric values and agency codes; those values were recovered from the server-rendered payload available without an account on the same public pages, so no access restriction was circumvented. Each source file was archived with its byte size and SHA-256 hash; all figures reported here derive from that archived snapshot, retrieved on 3 September 2026. As detailed in Table 2, extraction yielded 1,423 raw SPBE rows and 1,712 raw IPP rows, reduced by deduplication and linkage to 665 matched entity-year observations from 400 unique entities and 194 consecutive transitions from 135 entities.

2.3. Variables, instruments, and procedures

The primary exposure is the change in SPBE score from t to t+1, scaled per 0.5-point increment. The primary outcome is the ordered change in IPP predicate rank, coded down < unchanged < up. Three secondary estimands were pre-specified: a binary indicator that the numeric IPP score rose; the continuous first difference in IPP score; and the IPP score at t+1 regressed on the SPBE score at t with adjustment for IPP at t. Covariates are baseline IPP score, interval (2023–2024 or 2024–2025) and government level. Table 1 gives the full operational definitions. Government level is included for substantive reasons: Indonesian provinces, regencies and cities hold materially different service mandates, revenue bases and relationships to central government under an asymmetric decentralisation settlement,[11,14] so the same numerical index value does not denote the same administrative object across tiers.

Both indices are administratively assessed institutional instruments, scored by evaluator panels against published national rubrics rather than completed by respondents. Their validity therefore rests on rubric provenance and banding consistency: the SPBE domains and weighting are fixed by Ministerial Regulation 59/2020,[9] and the six IPP aspects and service-unit sampling procedure by Ministerial Regulation 29/2022.[10] The two instruments differ in what they observe, and that difference is central to interpreting the results. SPBE weights internal architecture, institutional arrangements and documented process; IPP scores sampled service units on service policy, human-resource professionalism, facilities, service information systems, consultation and complaint handling, and innovation. Only the service information system and complaint-handling aspects are both plausibly and quickly responsive to a change in digital maturity, a construct mismatch that attenuates any within-unit change association before a single model is fitted.[23,25] No inter-assessor reliability evidence has been published for either index, so far as could be established from the public documentation; that absence is itself part of the evaluation-infrastructure problem this study documents.

Two procedural steps were required before analysis. Exact duplicate rows were removed. Agency names were transliterated to ASCII, lower-cased and stripped of all non-alphanumeric characters to form an entity key. Name-based linkage was chosen over code-based linkage because IPP codes 8300 and 8400 are each used for two different provincial names, so joining on code would create many-to-many relations; measuring linkage quality where stable

identifiers are absent is a recognised methodological requirement rather than an improvisation.[58,63] IPP values in the 2024 and 2025 payloads carry inconsistent decimal scaling. A pre-specified reconstruction rule was applied: values above 100 were divided by 100, values above 5 and up to 100 were divided by 10, and values of 5 or below were retained. All 1,016 reconstructed local IPP values were then tested against the official predicate bands, and no conflict was found — an external validation that does not depend on the rule itself.

Table 1. Operational definitions of exposure, outcome, and covariates.

Component	Definition	Scale or estimand
Primary exposure	Change in SPBE score from wave t to wave t+1	Per 0.5-point increment
Primary outcome	Change in ordinal rank of the IPP predicate	Down < unchanged < up
Secondary outcome 1	Numeric IPP score increased between waves	Binary (yes/no)
Secondary outcome 2	First difference in numeric IPP score	IPP points on a 0–5 scale
Secondary outcome 3	IPP score at wave t+1	IPP points, adjusted for IPP at t
Baseline covariate	IPP score at wave t	Continuous
Interval	2023–2024 or 2024–2025	Categorical
Government level	Regency, city, or province	Categorical
Clustering unit	Normalised agency name (entity key)	135 clusters over 194 transitions

Note. SPBE, Sistem Pemerintahan Berbasis Elektronik; IPP, Indeks Pelayanan Publik. SPBE predicates: Kurang (Bad) < Cukup (Average) < Baik (Good) < Sangat Baik (Very Good) < Memuaskan (Excellent). IPP predicates: E < D < C < –B < B < –A < A. Both indices are scored on a 0–5 range by evaluator panels against published national rubrics.

2.4. Data analysis

The primary model is a cumulative-link model with a logit link: a proportional-odds ordinal regression of the IPP transition on SPBE change per 0.5 points, baseline IPP, interval and government level. Standard errors are heteroskedasticity-robust and clustered on the entity key, so that correlation between two transitions contributed by the same government is not ignored. An odds ratio above one indicates increased cumulative odds of occupying a better transition category. Ninety-five per cent confidence intervals are the focus of interpretation and two-sided p values are reported to three decimal places as compatibility measures rather than decision thresholds; α was set at 0.05. The proportional-odds assumption for the exposure was examined with a nominal test, because violations of that assumption distort inference in ordinal models in ways that are not visible from the fitted coefficients alone.[71]

Five pre-specified sensitivity analyses were run: binary logistic regression for whether the numeric IPP score increased; a linear first-difference model; a lagged model of IPP at t+1 on SPBE at t adjusting for IPP at t; an ordinal model restricted to the 59 entities present in all three waves; and an ordinal model restricted to regencies and cities. Because six model families were fitted on overlapping data, the Benjamini–Hochberg procedure was applied across their p values and adjusted values are reported alongside raw ones.[72] The family controlled is the six specifications reported here. These are not six independent tests: four are sensitivity analyses of the primary specification, two address a different estimand, and all six are fitted on the same 194 transitions, so the p values are strongly and positively dependent. Benjamini–Hochberg controls the false discovery rate under positive regression dependence, which is the assumption relied on and is stated rather than left implicit; a narrower family definition would yield a less conservative adjusted value for the primary estimate.

Four quantities were derived analytically from the published interval estimates. Standard errors were recovered as the interval half-width divided by 1.96, on the log scale for odds ratios, which assumes the intervals are Wald-symmetric on the estimation scale. Wald statistics and p values recomputed from those standard errors agreed with the fitted p values to within 0.006 across all six models; that agreement is an internal consistency check, not an independent verification, and would not detect a common upstream error. Odds ratios were converted to a standardised comparability metric using the logistic-to-normal transformation $d = \ln(OR) \times \sqrt{3}/\pi$, reported as a derived quantity and never as a measured group difference; because the outcome has three ordered categories rather than two, the latent-dichotomy assumption behind that mapping is approximate. The minimum detectable effect at 80% power and two-sided $\alpha = 0.05$ was computed as $\exp(2.802 \times SE)$ for odds ratios and $2.802 \times SE$ for linear coefficients, following the argument that a null should be reported against what the design could have found rather than as evidence of no effect.[73] That quantity conditions on an estimated standard error and, in a proportional-odds model, depends on the outcome category distribution, so it is specific to the distribution observed here. Contingency associations were quantified with χ^2 and Cramér's V, and the ordered gradient across tertiles of SPBE change with the Cochran–Armitage trend statistic.

With 135 clusters over 194 transitions the cluster-robust sandwich estimator is adequate but mildly anti-conservative without a finite-sample correction; none was applied, and the consequence is acknowledged.[74] The proportional-odds assumption was examined for the exposure only, not for baseline IPP, interval or government level; baseline IPP is the covariate on which violation is most plausible, and a partial proportional-odds model relaxing that constraint is identified as required further work rather than reported as completed. The primary analysis uses all available consecutive transitions rather than only complete cases, which is consistent under a missing-at-random assumption conditional on the covariates. Analyses were run in R using readr, dplyr, tidyr, ordinal, sandwich, lmtest, ggplot2 and jsonlite; an independent audit written against the Python standard library re-checked source hashes, row counts, duplication, value ranges, predicate bands, key uniqueness, transition composition, model output and figure presence, passing 68 of 68 checks.

2.5. Ethical considerations

The data are aggregated at the level of the local government and are publicly available; they contain no individual-level records, no personal identifiers and no service-encounter data. Ethical review was therefore not

required for this secondary analysis of open, non-personal administrative data, and none was sought. Because no human participants were involved, informed consent was not applicable. The analysis does not substitute for official policy evaluation and does not appraise the performance of individual civil servants. The study was conducted in accordance with the principles of the Declaration of Helsinki as they apply to research not involving human participants.

3. Results

As detailed in Table 2, removing exact duplicates eliminated 542 SPBE rows and 6 IPP rows, leaving 704 and 1,490 unique local-government-year records. Name-based linkage produced 665 matched entity-year observations from 400 unique entities, from which 194 consecutive one-year transitions contributed by 135 entities were derived. Of those 135 entities, 59 contributed two transitions and 76 contributed one; the 59 are precisely the entities assessed in all three waves, and they supply the 177 rows used in the complete-case sensitivity analysis. Across the 194 transitions the outcome was distributed as 22 downward (11.3%), 78 unchanged (40.2%) and 94 upward (48.5%).

Table 2. Construction of the analytic population from the public SPBE and IPP payloads.

Stage	SPBE	IPP	Note
Raw extracted rows	1,423	1,712	All levels of government
Exact duplicate rows removed	542	6	Removed before linkage
Unique local government-year records	704	1,490	Provinces, regencies, and cities
Linked panel observations	665	665	400 unique entities
Consecutive annual transitions	194	194	135 entities; primary outcome
Complete three-wave panel	177 rows	177 rows	59 entities; sensitivity analysis

Note. Source: authors' calculations from the public SPBE and IPP payloads of Satu Data Aparatur Negara; retrieved 3 September 2026. Duplicate removal preceded name-based linkage.

Coverage differed markedly across waves. As Table 3 and Figure 2 show, SPBE assessment covered 295 local governments in 2023, fell to 171 in 2024 and recovered to 238 in 2025 — a decline of 42.0% followed by a rise of 39.2%. IPP coverage was both wider and more stable at 474, 507 and 509, and successful linkage yielded 269, 167 and 229 observations by year. Because the composition of assessed entities changes between waves, annual means cannot be read as a national trend, and the 59 entities observed in all three waves — 14.8% of the 400 linked entities, shown in panel B of Figure 2 — are unlikely to be a random subset. No methodological note accounting for a coverage change of this size could be located in the public documentation. Four administrative mechanisms are plausible: a change in the assessment cycle so that not every agency is evaluated annually; a shift from census to sampled evaluation; a resourcing constraint in the evaluating unit; or deliberate phasing by government tier. These have materially different implications for whether the missingness is ignorable and cannot be distinguished from published material.

Table 3. Assessment coverage and successful linkage of local governments by wave.

Wave	Unique SPBE entities	Unique IPP entities	Successfully linked	Linked as % of SPBE
2023	295	474	269	91.2%
2024	171	507	167	97.7%
2025	238	509	229	96.2%

Note. Unique observations at local-government level after deduplication. SPBE coverage fell 42.0% between 2023 and 2024 and recovered 39.2% in 2025. Differing coverage means annual averages cannot be interpreted as a national trend.

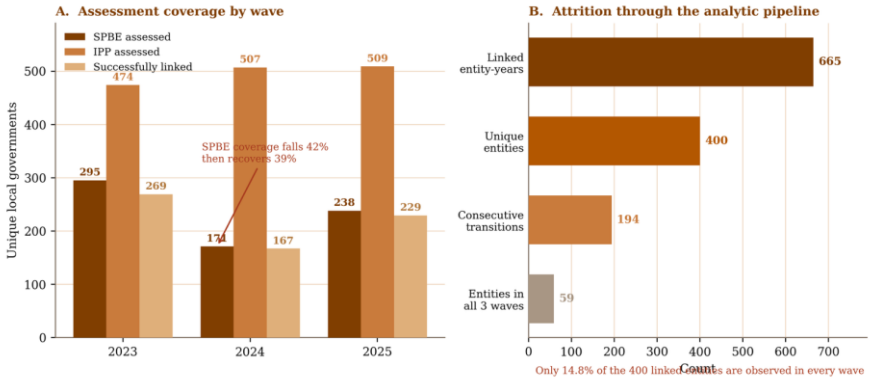

Figure 2. Assessment coverage and analytic attrition. Panel A shows the number of unique local governments assessed on each index by wave with the number successfully linked; SPBE coverage falls 42.0% between 2023 and 2024 and recovers 39.2% in 2025, and no methodological note accounting for that change could be located in the public documentation. Panel B shows attrition from linked observations to the complete three-wave subsample. Source: authors' analysis of SPBE and IPP data, 2023–2025.

Table 4 reports the distribution of both indices among linked observations. Mean SPBE was 2.99 (SD 0.69) in 2023, 3.58 (SD 0.71) in 2024 and 3.45 (SD 0.65) in 2025; mean IPP was 3.51 (SD 0.87), 4.07 (SD 0.60) and 3.89 (SD 0.70)

respectively. The cross-sectional Pearson correlation between the two indices was stable at 0.61, 0.64 and 0.59. A correlation of that size, reproduced in three consecutive waves on partly different entity sets, establishes that governments scoring higher on digital maturity also score higher on assessed service performance. It does not establish that a government which raises its own maturity score subsequently raises its own service score.

Table 4. Distribution of SPBE and IPP scores among successfully linked observations, by wave.

Wave	n	SPBE, mean (SD)	IPP, mean (SD)	Pearson r (SPBE, IPP)
2023	269	2.99 (0.69)	3.51 (0.87)	0.61
2024	167	3.58 (0.71)	4.07 (0.60)	0.64
2025	229	3.45 (0.65)	3.89 (0.70)	0.59

Note. SD, standard deviation. Means describe only observations successfully linked in each wave, and the composition of units differs between waves. The stable cross-sectional correlation establishes that higher-maturity governments also score higher on assessed service performance; it does not establish that raising maturity raises performance within the same government.

The composition of predicate transitions differed substantially between intervals. As set out in Table 5 and panel B of Figure 3, 56.5% of the 115 transitions in 2023–2024 were improvements against 36.7% of the 79 transitions in 2024–2025 (χ^2 = 9.533, df = 2, p = 0.009, Cramér's V = 0.222). Because the two intervals do not draw on the same set of governments, this contrast compares two non-exchangeable samples and cannot be interpreted as national deterioration. Several administrative explanations are compatible with it: normalisation of assessment after pandemic-era disruption of routine evaluation, a change in evaluator strictness or verification practice, a change in which service units were sampled, or genuine improvement concentrated among the governments assessed in the earlier window.

Table 5. Composition of IPP predicate transitions by interval.

Interval	Transition	n	% of interval
2023–2024	Down	8	7.0%
2023–2024	Unchanged	42	36.5%
2023–2024	Up	65	56.5%
2024–2025	Down	14	17.7%
2024–2025	Unchanged	36	45.6%
2024–2025	Up	29	36.7%

Note. χ^2 = 9.533, df = 2, p = 0.009, Cramér's V = 0.222. The 115 and 79 transitions are not contributed by identical sets of governments, so the contrast compares two non-exchangeable samples and does not isolate a national change.

Grouping transitions into tertiles of SPBE change (n = 65, 65 and 64) gave improvement rates of 44.6%, 38.5% and 62.5% from lowest to highest, shown in panel A of Figure 3 (χ^2 = 8.038, df = 2, p = 0.018, Cramér's V = 0.204; Cochran–Armitage trend z = 2.023, p = 0.043). The gradient is not monotone: the middle tertile sits below the lowest. A Cochran–Armitage statistic assumes a monotone trend across ordered groups, so applied here it tests a model the data visibly contradict, and the nominal p value should not be read as evidence of a dose–response relationship. The tertile boundaries were determined from the observed distribution of the 194 transitions rather than pre-specified.

The result that requires explanation is the opposite of the one the trend statistic suggests: this unadjusted association is nominally significant while the adjusted continuous model is not. Three features of the adjusted model account for the difference — adjustment for baseline IPP, interval and government level, all correlated with both indices; the cluster-robust correction for the 59 entities contributing two transitions each; and the information lost by categorising a continuous exposure into tertiles and collapsing a three-category outcome into improved versus not improved.

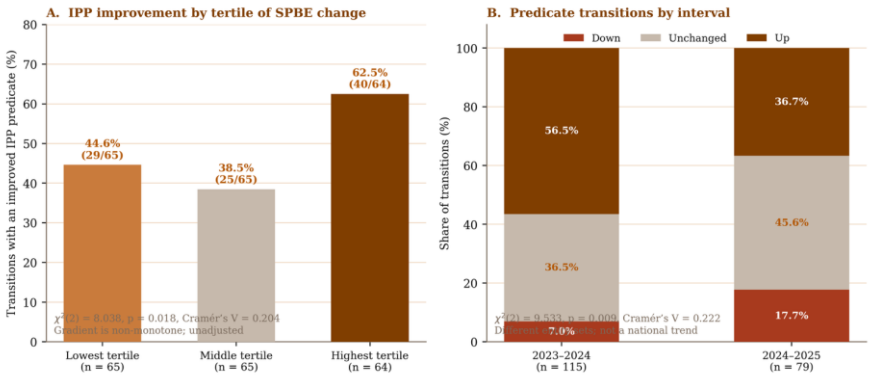

Figure 3. Descriptive composition of IPP predicate transitions. Panel A shows the proportion of transitions with an improved predicate by tertile of SPBE change, with the χ^2 test and Cramér's V; the gradient is non-monotone, so the trend statistic tests a model the data contradict. Panel B shows the full transition composition by interval. Both panels are unadjusted and ignore clustering. Source: authors' analysis of SPBE and IPP data, 2023–2025.

The primary proportional-odds model gave an adjusted odds ratio of 1.400 per 0.5-point increase in SPBE score (95% CI 0.780–2.513; SE on the log scale 0.298; z = 1.127; p = 0.261). Table 6 and Figure 4 present that estimate alongside all five pre-specified sensitivity analyses. The nominal test for the exposure gave p = 0.914, providing no evidence against proportional odds. Expressed as a standardised comparability metric the estimate corresponds to d = 0.185 (95% CI −0.137–0.508), a small effect whose interval spans both a small negative and a moderate positive value.

Table 6. Association between SPBE and IPP across the primary model and five pre-specified sensitivity analyses.

Model	Estimate	95% CI	SE	z	p	BH p	MDE	n
Ordinal transition (primary), OR	1.400	0.780–2.513	0.298	1.127	0.261	0.420	OR 2.31	194
Binary: IPP score increased, OR	1.104	0.568–2.143	0.339	0.292	0.771	0.925	OR 2.58	194
Linear first difference in IPP, β	0.003	−0.085–0.091	0.045	0.067	0.953	0.953	0.126	194
Lagged: IPP at t+1 on SPBE at t, β	0.195	0.079–0.312	0.059	3.281	0.001	0.006	0.166	194
Ordinal, complete three-wave, OR	2.315	0.949–5.645	0.455	1.845	0.068	0.204	OR 3.58	118
Ordinal, regencies and cities, OR	1.511	0.717–3.182	0.380	1.086	0.280	0.420	OR 2.90	154

Note. OR, odds ratio; CI, confidence interval; SE, standard error; BH, Benjamini–Hochberg; MDE, minimum detectable effect at 80% power with two-sided α = 0.05. Estimates are odds ratios for ordinal and binary models and linear coefficients for continuous outcomes. Standard errors are heteroskedasticity-robust and clustered on entity. SE, z, and MDE were derived analytically from the published confidence intervals; recomputed p values agreed with fitted values to within 0.006 across all six models, which is an internal consistency check rather than independent verification. Nominal test of proportional odds for the exposure: p = 0.914.

Sensitivity analyses are reported in Table 6. The binary model for an increase in numeric IPP score gave OR 1.104 (95% CI 0.568–2.143; p = 0.771). The linear first-difference model was null (β = 0.003 IPP points per 0.5-point SPBE change; 95% CI −0.085–0.091; p = 0.953). The lagged model gave β = 0.195 IPP points per one-point SPBE score at the previous wave, after adjustment for baseline IPP (95% CI 0.079–0.312; SE 0.059; z = 3.281; p = 0.001). The complete three-wave model on 59 entities gave OR 2.315 (95% CI 0.949–5.645; p = 0.068), and the model restricted to regencies and cities gave OR 1.511 (95% CI 0.717–3.182; p = 0.280). After Benjamini–Hochberg adjustment only the lagged estimate remained below the conventional threshold (adjusted p = 0.006); the primary ordinal model adjusted to p = 0.420 and the complete-case model to p = 0.204.

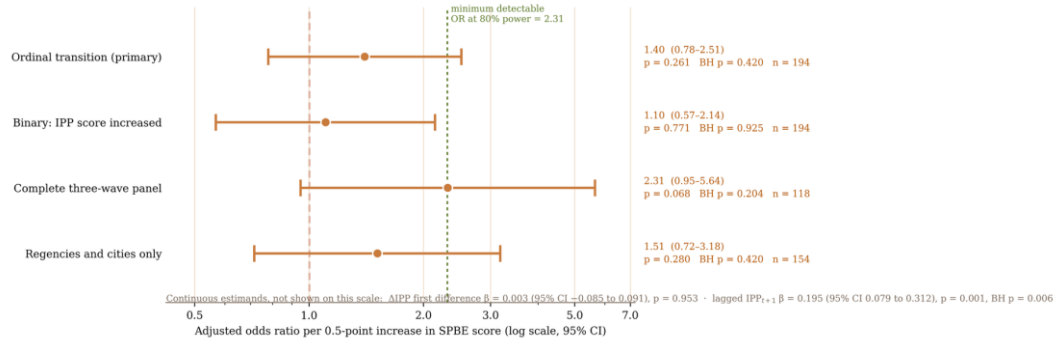

Figure 4. Forest plot of the primary ordinal model and three sensitivity analyses, showing adjusted odds ratios with 95% confidence intervals on a logarithmic scale. The dashed vertical reference line marks the null at OR = 1; the dotted line marks the minimum odds ratio the primary design could detect at 80% power. Raw and Benjamini–Hochberg adjusted p values are given for each model. Source: authors' analysis of SPBE and IPP data, 2023–2025.

Design sensitivity was computed from the standard errors recovered from the published intervals. At 80% power and two-sided α = 0.05 the primary model could detect an odds ratio of 2.31 or larger; the binary model 2.58 or larger; the complete-case model 3.58 or larger; and the regency-and-city model 2.90 or larger. The first-difference model could detect a change of 0.126 IPP points per 0.5-point SPBE increment. No sensitivity analysis in this study was therefore powered to detect a moderate effect, and each should be read as a consistency check on the direction of the primary estimate rather than as independent corroboration. Two consequences follow, and Figure 5 places these floors alongside the estimates. The primary null is uninformative about effects smaller than an odds ratio of 2.31, so it should be reported as imprecision rather than as evidence of absence. And the complete-case point estimate of 2.315 coincides almost exactly with the primary model's detection floor, which is what a selected subsample looks like when an apparent gain in effect size is produced by the selection rather than by the exposure.

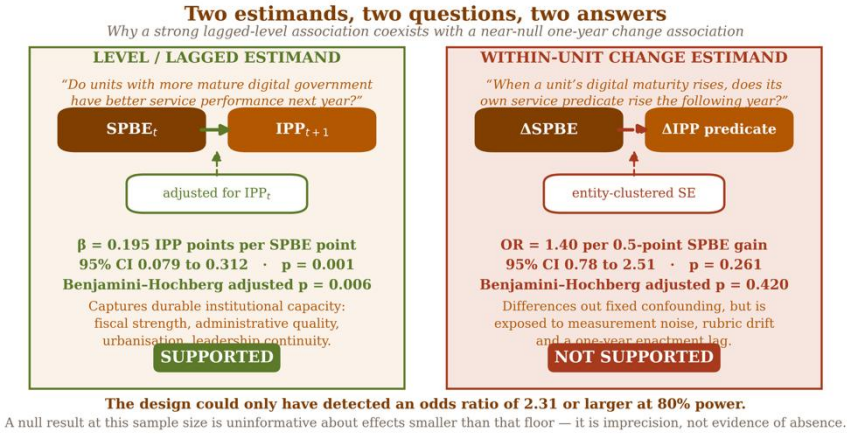

Figure 5. Contrast of the two estimands. The lagged-level question and the within-unit change question are answered by different models fitted to the same 194 transitions from the same 135 entities, and yield different conclusions; the detection floor of the design is stated beneath both panels. Source: authors' analysis of SPBE and IPP data, 2023–2025.

Table 7 sets the four estimands, as operationalised in Table 1, side by side, ordered from weakest to strongest evidential support. The odds that a numeric IPP score rose showed no association with SPBE change (OR 1.10; 95% CI 0.57–2.14). The within-unit change relationship was indistinguishable from zero (β 0.003; 95% CI −0.085–0.091). Excluding provinces left the ordinal direction intact (OR 1.51; 95% CI 0.72–3.18). The level relationship was the only one supported: entities with a higher SPBE score at one wave had a higher IPP score at the next, by 0.195 points per SPBE point (95% CI 0.079–0.312). The ordering in Table 7 is by evidential support, not by effect size. Overall the pattern is summarised in Figure 5 and is the study's central empirical result: H1 was supported, H2 was not, and H3 — that the level association exceeds the one-year within-unit change association — was supported under the criterion fixed in Section 1.

Table 7. Interpretation of the four estimands, ordered from weakest to strongest evidential support.

Estimand	Result	Correct interpretation
Odds that the numeric IPP score rose	OR 1.10 (95% CI 0.57–2.14)	No evidence of a consistent increase in the numeric score
Change in IPP per 0.5-point change in SPBE	β 0.003 (95% CI −0.085–0.091)	The within-unit change relationship is indistinguishable from zero
Ordinal model, regencies and cities only	OR 1.51 (95% CI 0.72–3.18)	The positive direction persists once provinces are excluded
IPP at t+1 per one SPBE point at t	β 0.195 (95% CI 0.079–0.312)	Entities with higher maturity at one wave have higher service scores at the next

Note. OR, odds ratio; CI, confidence interval. All four estimands were pre-specified. Only the lagged-level estimand survives Benjamini–Hochberg adjustment across the six model families.

4. Discussion

4.1. Principal findings

This study makes two claims, and the second is the more consequential. The first is that the level and the within-unit change estimands relating digital maturity to assessed service performance are empirically distinct, that only the former is supported in these data, and that a literature reporting them interchangeably is substituting one quantity for another. Higher digital government maturity was reliably associated with better assessed service performance in the following year (β 0.195; 95% CI 0.079–0.312; adjusted p = 0.006), while year-on-year increases in maturity within the same government were not reliably associated with improvement in its own service predicate (OR 1.400; 95% CI 0.780–2.513; adjusted p = 0.420). The second claim is that Indonesia cannot at present evaluate its own flagship digital government reform longitudinally: two agency codes each serve two provinces, decimal scaling is inconsistent across payload years, assessment coverage moves by more than 40% between adjacent waves without published explanation, and only 14.8% of linked entities appear in all three waves. On the data as published, no credible longitudinal estimate of SPBE's effect on service performance is obtainable — a remediable administrative choice rather than an inherent limitation of the domain.

Judged against the criterion set out in Section 1, H3 is supported: the lagged-level estimate is distinguishable from the null after multiplicity adjustment, the within-unit change estimate is not, and the pattern holds across the pre-specified sensitivity analyses. That support is qualitative. A difference in statistical significance between two estimates is not a test of the difference between them, the two quantities are on different scales, and they are computed on the same 194 transitions, so a formal contrast would require a joint model this analysis does not fit.

4.2. Comparison with previous studies

The asymmetry is the pattern technology enactment theory predicts.[18] Objective technology enters an organisation that already has a fiscal position, a staffing profile, a political leadership and a set of legal routines, and those arrangements determine whether a new platform changes what happens at a service counter; configurational evidence from government-led projects shows exactly this, with combinations of organisational conditions rather than

technology driving enactment.[19] A maturity score measures the objective layer at the moment of assessment; a service index measures, imperfectly and with a lag, some of what has been enacted. On that reading the lagged-level association is not a weak proxy for a causal effect but a measurement of capability stock, consistent with evidence that institutional capacity mediates the effects of governance reform on outcomes.[47,51] The finding aligns with structural evidence that organisational and managerial factors dominate technological ones,[26] with the decoupling of transformation rhetoric from practice,[28] and with legacy constraints that persist across reporting cycles.[27,35] The stability of the cross-sectional correlation across all three waves in Table 4 — 0.61, 0.64 and 0.59 — is the descriptive signature of that shared stock, while the paths in Figure 1 identify where the two quantities diverge.

Five mechanisms plausibly attenuate the within-unit change association, and they are not equally consistent with what was observed. Implementation lag is the first: interoperability gains, process redesign and user adoption take longer than one reporting cycle, and iterative delivery lengthens rather than shortens that interval.[35,36] Construct mismatch is the second: SPBE scores institutional maturity whereas IPP aggregates six aspects of sampled service units, only two of which are directly and quickly responsive to digitisation.[10,23,25] Both predict precisely the pattern reported — a preserved level association alongside a null change association — and neither can be distinguished from the other in these data. Ceiling effects are the third, predicting attenuation concentrated among high-baseline units, which is partially testable here and was not tested. Instrument and coverage drift is the fourth, and is the only one for which this study supplies direct documentary evidence; turbulence of this kind degrades comparability between successive assessments.[52,53] The interval contrast in Table 5 and Figure 3 is consistent with drift: improvement rates of 56.5% and 36.7% in adjacent intervals exceed any plausible one-year change in administrative capability. Time-varying confounding is the fifth: budget shifts, changes of regional head, restructuring and demand shocks are absent from both indices, and Indonesian evidence shows fiscal settlement,[14] budget-process capacity,[16] participatory practice[17] and corruption[67] all moving service outcomes independently of any digitisation programme.

A numeric comparison against specific published coefficients is deliberately not offered. The closest studies use different outcomes, exposure scalings and designs, and placing their coefficients beside an ordinal transition odds ratio would manufacture a comparability that does not exist. The comparison here is directional. A standardised equivalent of d = 0.185 sits below effects typically recovered from cross-sectional digital government studies, which is expected when fixed institutional differences are differenced out, but the interval spans d = −0.137 to d = 0.508 and is compatible with the cross-sectional literature, with a null, and with a modest reverse effect. This design was never capable of distinguishing between them: at 80% power it could detect only odds ratios of 2.31 or larger. Reporting the detection floor alongside the estimate is what separates an honest null from a claim of no effect,[73] and it also explains the complete-case result, whose OR of 2.315 on 59 entities is numerically almost identical to that floor — the signature of selection rather than of a stronger underlying relationship, since entities assessed in all three waves are plausibly those with the most stable administrative capacity.[37] Table 6 and Figure 4 report the detection floor beside every estimate for this reason.

Demand-side evidence from comparable settings suggests why supply-side maturity may be the wrong lever to measure. Adoption in lower- and middle-income countries depends on trust and digital literacy,[38,39] on accessible design[40] and on platform governance,[41] and in Tanzanian local government ICT improved service delivery only where citizen digital literacy was sufficient.[42] Cross-country panel work finds digital diffusion affecting development outcomes only with substantial lags,[43] and maturity has been theorised as a sustainability property rather than a short-run performance driver.[22] None of these mediators is measured in either Indonesian index.

4.3. Strengths and limitations

The study has four principal strengths. It uses real observed values from open government sources rather than simulated or imputed data, and archives every source file with a cryptographic hash. It separates cross-sectional, lagged, first-difference and ordinal-transition estimands explicitly instead of collapsing them. Its inference accommodates the repeated contribution of entities through entity-clustered robust standard errors, and the proportional-odds assumption for the exposure was tested rather than assumed.[71] And the entire pipeline passed 68 of 68 independent validation checks covering source integrity, row counts, duplication, value ranges, predicate-band consistency, key uniqueness, transition composition and model output.

Nine limitations bound the conclusions. First, as Table 2, Table 3 and Figure 2 together show, SPBE coverage changed by more than 40% between adjacent waves and the missingness is unlikely to be at random; missingness driven by assessment-cycle decisions correlated with unobserved administrative capacity would violate the missing-at-random assumption the primary analysis relies on. Second, only 59 entities have three complete waves, so the complete-case sensitivity analysis is itself selection-prone; the direct test would compare those 59 with the 76 single-transition entities on baseline SPBE, baseline IPP and government level, a comparison that requires no additional data and is the first analysis a revision should add. Third, name-based linkage can fail where an agency name changes substantively between years, although one-to-one uniqueness was verified after normalisation.[63] Fourth, the decimal reconstruction of IPP values was necessary because of payload inconsistency, and predicate-band validation reduces but cannot eliminate the risk of source error.[62] Fifth, IPP is an aggregate index partly built from sampled service units and does not measure individual citizen satisfaction. Sixth, important time-varying covariates — fiscal capacity, demography, leadership turnover and service-mix composition — are unavailable in both datasets.[57,66] Seventh, a one-year horizon is probably too short to capture enactment effects. Eighth, standard errors were reconstructed from published interval estimates rather than read from the fitted model objects, and the proportional-odds assumption was tested for the exposure only. Ninth, because the unit of analysis is a government rather than a person, no finding here supports any inference about individuals; treating an institutional association as a citizen-level one would be an ecological fallacy.[68]

Association is not causation, and no adjustment available in these data closes that gap. Better scores on both indices may arise from shared institutional capacity, from evaluation strategy, or from the selection of which service

units are assessed — and performance measurement systems are known to invite distortive responses.[44,46] Even the lagged model, which adjusts for the previous service score, leaves residual confounding and measurement error untouched.

4.4. Implications

The distributional consequence of the level interpretation follows from the results rather than from speculation. If digital maturity is largely an expression of durable institutional capacity rather than a lever that can be pulled within a budget year, then an evaluation system that scores maturity rewards prior capacity. In a country whose inter-regional divergence is well documented,[11,15] the predictable effect is that resource-rich urban administrations accumulate recognition while thinly-staffed peripheral governments are recorded as underperforming for structural reasons largely outside their control, and if those records inform budgetary or reputational allocation the gap widens rather than closes. Published rankings compound this, because they displace attention from the data beneath.[6,7] The design implication is specific: an instrument intended to drive reform should be built around indicators that can move within the evaluation cycle, and the evidence assembled here suggests SPBE as currently constituted is not such an indicator. It is a good measure of where capacity already sits and a poor measure of where reform effort is being expended.

Four practical implications follow, and Table 8 states each with the actor responsible and the indicator by which implementation could be verified. First, a single SPBE score should not be used as a proxy for assessed service performance in planning documents, budget submissions or performance contracts; it should be read alongside IPP and, where available, direct user-experience indicators, none of which exists in the two datasets analysed here.[32,54] Second, the governments most meriting attention are those whose SPBE score rose while their IPP predicate stagnated or fell, because that divergence localises where enactment is failing and is a better trigger for a process audit than a low score alone.[28,29] Third, and most urgently, the evaluation infrastructure needs repair: a stable national agency identifier, an annually published roster of assessed governments and service units, versioned rubrics with documented crosswalks, and published domain-level and aspect-level subscores. Each recommendation has an addressee, and naming it matters, because cross-organisational data standardisation fails at stewardship rather than at schema design.[58,59,60] The identifier reform belongs jointly to the ministry that scores SPBE, the unit that scores IPP and the custodian of the national agency register; the roster and rubric versioning belong to the evaluating units; the phased-rollout recording belongs to programme owners in each sector. Fourth, where reform is rolled out in phases, that phasing should be recorded prospectively, because it is the only route by which the causal question could be answered; no observational reanalysis of these two indices will settle it.[2,3] The verification indicators in Table 8 are deliberately auditable: each can be checked from published material without privileged access to ministry systems, and reusing linked administrative data for evaluation presupposes exactly such arrangements.[61,64]

Table 8. Priorities for evaluation infrastructure derived from the analysis.

Priority	Recommended action	Implementing actor	Verification indicator
National agency identifier	Adopt a stable agency identifier used consistently across datasets and years	Ministry scoring SPBE, unit scoring IPP, and custodian of the national agency register, jointly	No duplicate codes; one-to-one linkage without name normalisation
Minimum viable panel	Publish annually the list of local governments and service units actually assessed	The evaluating unit for each index	Coverage and attrition reconcilable between waves
Instrument metadata	Version the rubric, weighting, verification process, and every methodological change	The evaluating unit for each index	Year-to-year comparison supported by a published crosswalk
Published subscores	Release domain-level SPBE and aspect-level IPP scores, not only the composite	The evaluating unit for each index	Targeted recommendations checkable against the aspect they name
Targeted reform	Link SPBE recommendations to the specific IPP aspects they are expected to move	Programme owners in each sector, with the evaluating units	Action plans name a service outcome, not only a capability score
Impact evaluation	Record phased rollout prospectively so quasi-experimental designs become feasible	Programme owners in each sector	Causal estimand and pre-trends documented before rollout

Note. Priorities follow from the data-quality constraints documented in Table 2 and Table 3 and from the limits on inference established in Table 6. Implementing actors are named because cross-organisational data standardisation typically fails at stewardship rather than at schema design.

One regulatory development bounds the forward relevance of these estimates. From 2026, Ministerial Regulation 8/2026 introduces digital government performance evaluation and repeals Regulation 59/2020.[9,75] Three consequences follow. The series analysed here is the last complete evaluation window under the superseded instrument, giving it an archival value it will not have again. The crosswalk study required to link the two regimes has an effective deadline, because it must be designed before the first assessment under the new instrument is published. And the defects documented in this paper constitute a design brief for the new instrument rather than a retrospective complaint — each is cheapest to implement at the moment an instrument is being replaced, and governance of evaluation requires explicit guideline architecture rather than capability scoring alone.[50,55]

One boundary on external validity should be stated explicitly: the comparative evidence drawn on here comes predominantly from European and East Asian settings, and other large decentralised middle-income systems operating national e-government scoring regimes are under-represented, because comparable published panel analyses

could not be identified. High-income systems have more stable evaluation infrastructure and shorter administrative chains than the setting studied here,[37,56] so their reported associations are likely to be upper bounds rather than benchmarks. The theoretical contribution is to convert enactment theory from an interpretive frame into a testable empirical asymmetry: by estimating both estimands in the same population, on the same indices, under the same adjustment set, the study shows they answer different questions and give different answers. The distinction generalises to any policy domain evaluated through an annually scored institutional index.[24,45,48,49] Future work should extend the panel beyond three waves, add SPBE domain and IPP aspect detail with fiscal and demographic covariates and entity fixed effects, incorporate the demand-side measures that adoption research shows to be decisive,[1,4,5,12,13,20,21,30,31,33,34,65,72,74] and exploit phased reform rollouts for quasi-experimental identification.

5. Conclusion

In an open panel of Indonesian local governments observed between 2023 and 2025, increases in electronic government maturity were associated with better public service predicate transitions in the hypothesised direction, but imprecisely and not significantly (OR 1.400; 95% CI 0.780–2.513; p = 0.261; adjusted p = 0.420), against a design that could detect only odds ratios of 2.31 or larger. The level of digital maturity predicted the following year's service score far more consistently than within-unit change did (β 0.195; 95% CI 0.079–0.312; p = 0.001; adjusted p = 0.006). The theoretical contribution is to render technology enactment testable as an empirical asymmetry between the level and change estimands, and to show that Indonesian evidence exhibits it. For policy, the results argue against treating SPBE scores as a proxy for assessed service quality and in favour of stabilising the national evaluation infrastructure — a persistent agency identifier, published assessment coverage and versioned rubrics — before further causal claims are made. The most consequential single recommendation is that phased reform rollouts be recorded prospectively, because that is the only route by which the causal question can be answered.

6. Declarations

Acknowledgements

The authors thank the Ministry of Administrative and Bureaucratic Reform (Kementerian PANRB) for making the SPBE and IPP evaluation data publicly available, without which this analysis would not have been possible.

Statement on the Use of Artificial Intelligence

The authors confirm that no generative artificial intelligence or AI-assisted technologies were used in the drafting, editing, or preparation of this manuscript.

Author Contribution Statement

APP: Conceptualization, methodology, software, data curation, formal analysis, visualization, and writing—original draft. AY: Validation, interpretation, supervision, and writing—review and editing. APP and AY approved the final manuscript and agree to be accountable for all aspects of the work.

Funding Statement

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

Conflict of Interest

The authors declare that they have no financial or non-financial competing interests.

Data Availability Statement

The source SPBE and IPP datasets are publicly available through Satu Data Aparatur Negara (https://satudata.menpan.go.id). Every source file used in the analysis was archived with its byte size and SHA-256 hash. Analysis scripts, intermediate panel and transition files, model output, and figure source files are available from the corresponding author upon reasonable request.

Ethics Approval and Consent to Participate

Ethical review was not required because this secondary study analysed aggregate, publicly released institutional performance data containing no individual-level records, personal identifiers, or service-encounter data. No human participants were involved; therefore, consent to participate was not applicable.

Consent for Publication

Not applicable. The manuscript contains no identifiable individual data.

7. References

1. van Noordt C, Misuraca G. Artificial intelligence for the public sector: results of landscaping the use of AI in government across the European Union. Gov Inf Q 2022;39(3):101714. doi: 10.1016/j.giq.2022.101714
2. Medaglia R, Gil-Garcia JR, Pardo TA. Artificial intelligence in government: taking stock and moving forward. Soc Sci Comput Rev 2023;41(1):123–140. doi: 10.1177/08944393211034087
3. Bharosa N. The rise of GovTech: trojan horse or blessing in disguise? A research agenda. Gov Inf Q 2022;39(3):101692. doi: 10.1016/j.giq.2022.101692
4. Eom SJ, Lee J. Digital government transformation in turbulent times: responses, challenges, and future direction. Gov Inf Q 2022;39(2):101690. doi: 10.1016/j.giq.2022.101690
5. Moser-Plautz B, Schmidthuber L. Digital government transformation as an organizational response to the COVID-19 pandemic. Gov Inf Q 2023;40(3):101815. doi: 10.1016/j.giq.2023.101815
6. Hohensinn L, Willems J, George B, et al. Performance rankings reduce cognitive processing of underlying performance information. Public Manag Rev 2025;28(2):498–520. doi: 10.1080/14719037.2025.2464761
7. Meng X, Li C. Does public sector performance information impact stakeholders? Evidence from a meta-analysis. Public Adm Rev 2025;85(4):1086–1097. doi: 10.1111/puar.13924
8. Republik Indonesia. Peraturan Presiden Nomor 95 Tahun 2018 tentang Sistem Pemerintahan Berbasis Elektronik [Presidential Regulation 95/2018 on the Electronic-Based Government System]. Jakarta: Sekretariat Kabinet Republik Indonesia; 2018.
9. Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi. Peraturan Menteri PANRB Nomor 59 Tahun 2020 tentang Pemantauan dan Evaluasi SPBE [Ministerial Regulation 59/2020 on SPBE monitoring and evaluation]. Jakarta: Kementerian PANRB; 2020.
10. Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi. Peraturan Menteri PANRB Nomor 29 Tahun 2022 tentang Pemantauan dan Evaluasi Kinerja Penyelenggaraan Pelayanan Publik [Ministerial Regulation 29/2022 on public-service performance evaluation]. Jakarta: Kementerian PANRB; 2022.
11. Lele G. Asymmetric decentralization, accommodation and separatist conflict: lessons from Aceh and Papua, Indonesia. Territ Polit Gov 2021;11(5):972–990. doi: 10.1080/21622671.2021.1875036
12. Syukri M. Indonesia's new developmental state: interrogating participatory village governance. J Contemp Asia 2022;54(1):2–23. doi: 10.1080/00472336.2022.2089904
13. Sandi JRA, Hadi K, Haridison A. Reframing decentralisation: the persistence of customary governance in Indonesia. Territ Polit Gov 2026:1–19. doi: 10.1080/21622671.2026.2658039
14. Lewis BD. Indonesia's new fiscal decentralisation law: a critical assessment. Bull Indones Econ Stud 2023;59(1):1–28. doi: 10.1080/00074918.2023.2180838
15. Hummel D, Kusumasari B. Local government financial performance and decentralization in Indonesia. State Local Gov Rev 2025;58(2):206–227. doi: 10.1177/0160323x251364379
16. Vidyattama Y, Sutiyono W, Sugiyarto S. Factors of budget delay: local government capacity, bargaining and political interaction in Indonesia. Bull Indones Econ Stud 2022;58(1):55–77. doi: 10.1080/00074918.2020.1812515

17. Musadat A. The effects of participatory budgeting on local government service delivery: evidence from Sumedang. Bull Indones Econ Stud 2023;59(1):29–60. doi: 10.1080/00074918.2021.1931031
18. Fountain JE. Building the virtual state: information technology and institutional change. Washington, DC: Brookings Institution Press; 2001.
19. Mu R, Haershan M, Wu P. What organizational conditions, in combination, drive technology enactment in government-led smart city projects?. Technol Forecast Soc Change 2022;174:121220. doi: 10.1016/j.techfore.2021.121220
20. Grimmelikhuijsen S, Meijer A. Legitimacy of algorithmic decision-making: six threats and the need for a calibrated institutional response. Perspect Public Manag Gov 2022;5(3):232–242. doi: 10.1093/ppmgov/gvac008
21. Malodia S, Dhir A, Mishra M, et al. Future of e-government: an integrated conceptual framework. Technol Forecast Soc Change 2021;173:121102. doi: 10.1016/j.techfore.2021.121102
22. Feleke H, Lessa L. Maturity as a critical sustainability factor for e-government: toward a conceptual framework. Electron J Inf Syst Dev Ctries 2023;90(3):e12308. doi: 10.1002/isd2.12308
23. Osborne SP, Nasi G, Powell M. Beyond co-production: value creation and public services. Public Adm 2021;99(4):641–657. doi: 10.1111/padm.12718
24. Osborne SP, Powell M, Cui T, et al. Value creation in the public service ecosystem: an integrative framework. Public Adm Rev 2022;82(4):634–645. doi: 10.1111/puar.13474
25. Scholta H, Lindgren I. Proactivity in digital public services: a conceptual analysis. Gov Inf Q 2023;40(3):101832. doi: 10.1016/j.giq.2023.101832
26. Tangi L, Janssen M, Benedetti M, et al. Digital government transformation: a structural equation modelling analysis of driving and impeding factors. Int J Inf Manage 2021;60:102356. doi: 10.1016/j.ijinfomgt.2021.102356
27. Irani Z, Abril RM, Weerakkody V, et al. The impact of legacy systems on digital transformation in European public administration: lesson learned from a multi case analysis. Gov Inf Q 2023;40(1):101784. doi: 10.1016/j.giq.2022.101784
28. Crusoe J, Magnusson J, Eklund J. Digital transformation decoupling: the impact of willful ignorance on public sector digital transformation. Gov Inf Q 2024;41(3):101958. doi: 10.1016/j.giq.2024.101958
29. Afzal M, Panagiotopoulos P. Coping with digital transformation in frontline public services: a study of user adaptation in policing. Gov Inf Q 2024;41(4):101977. doi: 10.1016/j.giq.2024.101977
30. Ahn MJ, Chen YC. Digital transformation toward AI-augmented public administration: the perception of government employees and the willingness to use AI in government. Gov Inf Q 2022;39(2):101664. doi: 10.1016/j.giq.2021.101664
31. Chen T, Guo W, Gao X, et al. AI-based self-service technology in public service delivery: user experience and influencing factors. Gov Inf Q 2021;38(4):101520. doi: 10.1016/j.giq.2020.101520
32. Wang C, Teo TSH, Janssen M. Public and private value creation using artificial intelligence: an empirical study of AI voice robot users in Chinese public sector. Int J Inf Manage 2021;61:102401. doi: 10.1016/j.ijinfomgt.2021.102401
33. Scupola A, Mergel I. Co-production in digital transformation of public administration and public value creation: the case of Denmark. Gov Inf Q 2022;39(1):101650. doi: 10.1016/j.giq.2021.101650
34. Wilson C, Mergel I. Overcoming barriers to digital government: mapping the strategies of digital champions. Gov Inf Q 2022;39(2):101681. doi: 10.1016/j.giq.2022.101681
35. Maragno G, Tangi L, Gastaldi L, et al. Exploring the factors, affordances and constraints outlining the implementation of artificial intelligence in public sector organizations. Int J Inf Manage 2023;73:102686. doi: 10.1016/j.ijinfomgt.2023.102686
36. Mergel I, Ganapati S, Whitford AB. Agile: a new way of governing. Public Adm Rev 2021;81(1):161–165. doi: 10.1111/puar.13202
37. Kuhlmann S, Heuberger M. Digital transformation going local: implementation, impacts and constraints from a German perspective. Public Money Manag 2023;43(2):147–155. doi: 10.1080/09540962.2021.1939584
38. AbdulKareem AK, Oladimeji KA. Cultivating the digital citizen: trust, digital literacy and e-government adoption. Transform Gov People Process Policy 2024;18(2):270–286. doi: 10.1108/tg-11-2023-0196
39. Sabani A, Thai V, Hossain MA. Factors affecting citizen adoption of e-government in developing countries. J Glob Inf Manag 2023;31(1):1–23. doi: 10.4018/jgim.318131
40. Abdurahman S, Kabanda S. Factors influencing the design and implementation of accessible e-government services in South Africa. Electron J Inf Syst Dev Ctries 2024;90(4):e12317. doi: 10.1002/isd2.12317
41. Abdul-Salam A, Adam IO, Preko M. Examining platform governance and user behavior in the adoption of e-government services in a developing country. Electron J Inf Syst Dev Ctries 2026;92(2):e70063. doi: 10.1002/isd2.70063
42. Ngowi N, Kissoka G, Pastory P. Information and communication technology and public service delivery in Tanzania local government authorities: the moderating role of digital literacy. Electron J Inf Syst Dev Ctries 2026;92(2):e70060. doi: 10.1002/isd2.70060
43. Lechman E, Popowska M. Harnessing digital technologies for poverty reduction: evidence for low-income and lower-middle income countries. Telecomm Policy 2022;46(6):102313. doi: 10.1016/j.telpol.2022.102313
44. Castelo SL, Gomes CF. The role of performance measurement and management systems in changing public organizations: an exploratory study. Public Money Manag 2023;44(5):399–406. doi: 10.1080/09540962.2023.2204400
45. Rivenbark WC, Fasiello R, Tassi F. The performance function in local government: does location impact performance data use?. Public Adm Rev 2023;83(5):1123–1135. doi: 10.1111/puar.13651
46. Choi S, Park S. Exploring performance paradox in public organizations: analyzing the predictors of distortive behaviors in performance measurement. Int Rev Adm Sci 2021;89(2):501–518. doi: 10.1177/00208523211054876
47. Kim J, Kang H. Bureaucratic burden or administrative capacity? Administrative intensity and fiscal performance of local governments. Public Manag Rev 2024;28(3):580–601. doi: 10.1080/14719037.2024.2447314
48. Veronesi G, Kirkpatrick I, Altanlar A, et al. Corporatization, administrative intensity, and the performance of public sector organizations. J Public Adm Res Theory 2022;33(4):701–715. doi: 10.1093/jopart/muac048
49. Cohen S, Karatzimas S. Public value co-creation and sense of belonging: evidence from local government annual reports. Int Public Manag J 2025;28(6):769–786. doi: 10.1080/10967494.2025.2479034
50. Bianchi C, Nasi G, Rivenbark WC. Implementing collaborative governance: models, experiences, and challenges. Public Manag Rev 2021;23(11):1581–1589. doi: 10.1080/14719037.2021.1878777
51. Schmidthuber L, Ingrams A, Hilgers D. Government openness and public trust: the mediating role of democratic capacity. Public Adm Rev 2021;81(1):91–109. doi: 10.1111/puar.13298
52. Ansell C, Sørensen E, Torfing J. The COVID-19 pandemic as a game changer for public administration and leadership? The need for robust governance responses to turbulent problems. Public Manag Rev 2021;23(7):949–960. doi: 10.1080/14719037.2020.1820272
53. Wirtz BW, Müller WM, Weyerer JC. Digital pandemic response systems: a strategic management framework against Covid-19. Int J Public Adm 2021;44(11–12):896–906. doi: 10.1080/01900692.2020.1858316
54. Chen YC, Ahn MJ, Wang YF. Artificial intelligence and public values: value impacts and governance in the public sector. Sustainability 2023;15(6):4796. doi: 10.3390/su15064796
55. Wirtz BW, Weyerer JC, Kehl I. Governance of artificial intelligence: a risk and guideline-based integrative framework. Gov Inf Q 2022;39(4):101685. doi: 10.1016/j.giq.2022.101685
56. Manoharan AP, Ingrams A, Kang D, et al. Globalization and worldwide best practices in e-government. Int J Public Adm 2021;44(6):465–476. doi: 10.1080/01900692.2020.1729182
57. van Roekel H, Branderhorst M, Tummers L, et al. Digital transformation leadership: a public value-centered measurement scale. Gov Inf Q 2025;42(4):102091. doi: 10.1016/j.giq.2025.102091
58. van Donge W, Bharosa N, Janssen MFWHA. Data-driven government: cross-case comparison of data stewardship in data ecosystems. Gov Inf Q 2022;39(2):101642. doi: 10.1016/j.giq.2021.101642
59. Ruijer E. Designing and implementing data collaboratives: a governance perspective. Gov Inf Q 2021;38(4):101612. doi: 10.1016/j.giq.2021.101612
60. Wang F, Zhang Z, Ma X, et al. Paths to open government data reuse: a three-dimensional framework of information need, data and government preparation. Inf Manage 2023;60(8):103879. doi: 10.1016/j.im.2023.103879
61. Martinez TM, Whitley EA. Open data as an anticorruption tool? Using distributed cognition to understand breakdowns in the creation of transparency data. Data Policy 2023;5:e13. doi: 10.1017/dap.2023.10
62. Wu D, Xu H, Yongyi W, et al. Quality of government health data in COVID-19: definition and testing of an open government health data quality evaluation framework. Libr Hi Tech 2021;40(2):516–534. doi: 10.1108/lht-04-2021-0126
63. Goode H, Hampton J, Huck R. Measuring linkage quality with hashed identifiers: an example using PAYE-ASHE. Int J Popul Data Sci 2023;8(2):2236. doi: 10.23889/ijpds.v8i2.2236
64. Pilkington R, Montgomerie A, Malvaso C, et al. Linked administrative data can inform policy and practice: an academic-government research partnership in South Australia. Int J Popul Data Sci 2022;7(3):1965. doi: 10.23889/ijpds.v7i3.1965
65. Ferza R, Lee DS. Online accessibility of local public enterprise financial statements in Indonesia: can it be predicted by institutional traits of parent local governments?. Public Adm Dev 2025;46(1):7–20. doi: 10.1002/pad.70025

66. Hummel D, Kusumasari B. Power dynamics and resource dependence: NGO-government collaboration in Yogyakarta, Indonesia. Public Adm Dev 2023;44(1):32–42. doi: 10.1002/pad.2034
67. Yunan ZY, Freyens B, Vidyattama Y. Corruption and Indonesia's regional economic performance. Bull Indones Econ Stud 2025;61(2):205–227. doi: 10.1080/00074918.2024.2437826
68. Pavía JM, Romero R. Improving estimates accuracy of voter transitions: two new algorithms for ecological inference based on linear programming. Sociol Methods Res 2022;53(3):1491–1533. doi: 10.1177/00491241221092725
69. Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi. Satu Data Aparatur Negara: Sistem Pemerintahan Berbasis Elektronik (SPBE) evaluation dataset, 2023–2025. Jakarta: Kementerian PANRB; 2026. Available from: https://satudata.menpan.go.id.
70. Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi. Satu Data Aparatur Negara: Indeks Pelayanan Publik (IPP) dataset, 2023–2025 [Internet]. Jakarta: Kementerian PANRB; 2026. Available from: https://satudata.menpan.go.id.
71. Edlinger M, van Smeden M, Alber HF, et al. Risk prediction models for discrete ordinal outcomes: calibration and the impact of the proportional odds assumption. Stat Med 2021;41(8):1334–1360. doi: 10.1002/sim.9281
72. Liu G, Shiraito Y. Multiple hypothesis testing in conjoint analysis. Polit Anal 2023;31(3):380–395. doi: 10.1017/pan.2022.30
73. Riesthuis P. Simulation-based power analyses for the smallest effect size of interest: a confidence-interval approach for minimum-effect and equivalence testing. Adv Methods Pract Psychol Sci 2024;7(2):25152459241240722. doi: 10.1177/25152459241240722
74. Huang FL, Li X. Using cluster-robust standard errors when analyzing group-randomized trials with few clusters. Behav Res Methods 2021;54(3):1181–1199. doi: 10.3758/s13428-021-01627-0
75. Kementerian Pendayagunaan Aparatur Negara dan Reformasi Birokrasi. Peraturan Menteri PANRB Nomor 8 Tahun 2026 tentang Evaluasi Kinerja Pemerintah Digital [Ministerial Regulation 8/2026 on digital-government performance evaluation]. Jakarta: Kementerian PANRB; 2026.