International Journal of Financial, Administrative, and Economic Sciences (IJFAES)

IJFAES

المجلة الدولية للعلوم المالية والإدارية والاقتصادية

Vol. (5), No. (8) August 2026 الإصدار (5)، العدد (8)

Integrating Program and Performance Budgeting with Artificial Intelligence Technologies to Improve the Quality of Financial Decision-Making

Mothana Mohammed Khudher
Master's Degree in Accounting, Faculty of Business Administration and Economics, American University of Science and Technology (AUST), Lebanon
mothanaaloubaid1@gmail.com

Jaafar Ibrahim Mohsin
Master's Degree in Accounting, Faculty of Business Administration and Economics, American University of Science and Technology (AUST), Lebanon
Jaferalkabi2@gmail.com

Abstract

This study examines whether integrating program and performance budgeting (PPB) with artificial intelligence (AI) technologies improves the quality of financial decision-making. PPB links appropriations to programmes, objectives, outputs, outcomes, and performance indicators, whereas AI can strengthen forecasting, anomaly detection, reporting, scenario analysis, and decision support. The proposed model treats integration as an organisational, informational, and analytical capability rather than the mere acquisition of software. Five dimensions constitute the independent variable: data readiness and integration, intelligent analytics, AI-supported financial forecasting, budget-cycle automation, and programme evaluation and intelligent decision support. Financial decision-making quality is assessed through timeliness, accuracy, evidence use, allocation efficiency, expenditure rationalisation, risk detection, traceability, transparency, and responsiveness. A quantitative, cross-sectional, descriptive-analytical design was used with 64 eligible respondents and a 45-item questionnaire. The findings show a strong positive relationship between overall integration and financial decision-making quality, while the integration dimensions overlap substantially. The results support treating integration as a coherent capability bundle while retaining human oversight and accountability.

Keywords: Program and Performance Budgeting, Artificial Intelligence, Public Financial Management, Financial Forecasting, Decision Quality, Programme Evaluation.

Chapter One: Introduction and Research Framework

1.1 Introduction:

Public financial management has increasingly shifted from verifying compliance with expenditure items toward demonstrating public value and measurable results. This move is underpinned by programme and performance budgeting, which organises appropriations in programmes and uses performance information to inform, rather than mechanically drive, budget decisions. The OECD Performance Budgeting Framework emphasizes relevant performance information, accountability and openness, an enabling institutional framework and the actual use of information in budget choices. (Tryggvadottir & Bambalaite, 2024).

AI systems are more than automation. They evaluate historical and near-real-time data to find aberrant patterns and anticipate future trends and compare possibilities. Use cases in public financial management include forecasting, spending analysis, budget monitoring, document classification, invoice matching, internal controls and reporting. The current standard is that AI will be an aid to public authorities, not a replacement for professional judgment. (OECD, 2025).

Technology does not make better decisions on its own. ‘The results are driven by the quality of the data, clear use cases, the right models, the capacity of the users to interpret the outputs, and strong governance. The NIST AI Risk Management Framework identifies seven core features of trustworthy AI: validity and reliability, safety, security and resilience, accountability and transparency, explainability and interpretability, privacy enhancement, and mitigation of harmful bias (Tabassi, 2023). Accordingly, this study defines integration as the connection of programme structures, financial and performance data, analytical tools, workflows, and human accountability throughout the budget cycle.

1.2 Research Problem:

Despite expanding digital transformation and the availability of financial systems, many organizations still experience a gap between producing data and using it in timely, accurate decisions. Expenditure data may remain fragmented, performance indicators may arrive late, forecasts may rely on simple historical extrapolation, and the relationship between programme cost and results may be unclear. These conditions limit PPB’s ability to influence resource allocation.

Research also shows that performance budgeting does not automatically improve results. Its success depends on sound programme design, consistent indicators, competent staff and actual utilization of performance information. Budget processes, implementation and use of performance information were recognized as persistent concerns in a recent thematic evaluation (Mohd Gharib et al., 2024). The deployment of AI brings more issues such as data fragmentation, legacy systems, lack of expertise, opaque models, privacy, security risk, and bias (OECD, 2024; Tabassi, 2023). Therefore, the central problem is the uncertain level and effect of integrating PPB with AI technologies on financial decision-making quality in the organization under study.

Main research question: What is the effect of integrating program and performance budgeting with artificial intelligence technologies on the quality of financial decision-making in the organization under study?

https://doi.org/10.59992/IJFAES.2026.v5n8p9 E-ISSN 2977-1498

1.3 Research Questions:

1. What is the level of availability of the requirements for integrating PPB with AI technologies?
2. What is the perceived level of financial decision-making quality?
3. Is total integration significantly associated with financial decision-making quality?
4. What are the individual effects of data readiness, intelligent analytics, AI-supported forecasting, budget-cycle automation, and programme evaluation and decision support?
5. Do perceptions differ according to relevant demographic and occupational characteristics?

1.4 Research Objectives:

1. Explain the theoretical foundations of PPB and relevant financial applications of AI.
2. Measure organisational readiness across data, analytics, forecasting, automation, and decision-support dimensions.
3. Measure financial decision-making quality through operational, allocative, risk, and accountability indicators.
4. Test the relationships and predictive effects between the five integration dimensions and decision quality.
5. Present a practical model that connects the budget cycle to intelligent analysis, human review, and feedback.

1.5 Significance of the Study:

Theoretical significance. The study connects two fields often examined separately: programme- and results-oriented budget reform and AI-enabled analytical transformation. It proposes a multidimensional measure of the value created when AI is embedded within a budget architecture that links money to objectives, outputs, and outcomes.

Practical significance. The instrument can diagnose institutional readiness, identify the dimensions most strongly related to decision quality, and guide priorities for investment in data, platforms, skills, workflow redesign, and governance. The findings may help managers move from isolated digital tools to a controlled decision-support environment.

1.6 Hypotheses:

H1: Integration of PPB with AI technologies has a positive, statistically significant effect on financial decision-making quality.

H1a: Data readiness and integration have a positive, statistically significant effect on financial decision-making quality.

H1b: Intelligent analytics have a positive, statistically significant effect on financial decision-making quality.

H1c: AI-supported financial forecasting has a positive, statistically significant effect on financial decision-making quality.

H1d: Budget-cycle automation has a positive, statistically significant effect on financial decision-making quality.

H1e: Programme evaluation and intelligent decision support have a positive, statistically significant effect on financial decision-making quality.

H2: Respondents' assessments differ significantly according to analytically relevant occupational and demographic characteristics.

1.7 Conceptual Model:

Table (1) Conceptual model of the study

Independent-variable dimensions	Expected direction	Dependent variable
Data readiness and integration	Positive	Financial decision-making quality
Intelligent analytics	Positive	Financial decision-making quality
AI-supported financial forecasting	Positive	Financial decision-making quality
Budget-cycle automation	Positive	Financial decision-making quality
Programme evaluation and intelligent decision support	Positive	Financial decision-making quality

The dimensions are related but conceptually distinct. Data integration provides the infrastructure; analytics and forecasting provide interpretive and forward-looking capacity; automation embeds tools in routine workflows; and programme evaluation converts evidence into resource-allocation alternatives. Human review, auditability, and feedback surround the model.

1.8 Scope and Operational Definitions:

Subject scope. The study addresses the integration of PPB and AI technologies and its relationship with perceived decision quality; it does not measure final financial performance directly from audited statements. Human scope. Participants are employees and managers involved in budgeting, finance, planning, performance management, internal audit, information systems, or digital transformation. Spatial scope. [Insert organisation(s)]. Time scope. Data collection will occur in 2026; exact opening and closing dates will be stated after fieldwork.

https://doi.org/10.59992/IJFAES.2026.v5n8p9 E-ISSN 2977-1498

Table (2) Operational definitions of study terms

Term	Operational definition
Program and performance budgeting	A budgeting system that links resources to programmes, objectives, outputs, outcomes, and indicators and uses performance information in planning, monitoring, and evaluation.
Artificial intelligence technologies	Computational methods that generate predictions, classifications, recommendations, anomaly alerts, or text-based assistance, including machine learning, predictive analytics, NLP, and decision-support systems.
Integration	The extent to which PPB data are connected to analytical, forecasting, and automation tools and their outputs are embedded in budget preparation, monitoring, programme evaluation, and decision procedures.
Financial decision-making quality	The extent to which a financial decision is timely, accurate, evidence-based, efficient in allocating resources, risk-aware, traceable, transparent, and responsive.
Data readiness	The availability of accurate, timely, standardised, integrated, secure, and analytically usable financial and performance data.

Chapter Two: Theoretical Framework and Previous Studies

2.1 Program and Performance Budgeting:

PPB is an institutional arrangement where appropriations are arranged around programmes with stated objectives and accompanied by information on outputs, outcomes, and indicators. PPB, unlike line-item budgeting, poses not just the question of what money will be spent on, but also why expenditure is undertaken, and what results are expected. It keeps input control, and adds a results chain linking inputs, activities, outputs, outcomes and longer-term influence.

PPB techniques vary from the provision of performance information along with the budget, to performance-informed budgeting, where information informs dialogue and allocation, to managerial models, which allow more flexibility but also responsibility for results. There isn’t one design to suit all organizations. Legal structure, decentralization, service characteristics, information maturity, and managerial capability all influence the right model (Tryggvadottir & Bambalaite, 2024; Begg et al., 2024).

Its main objectives are fiscal discipline, allocative efficiency, operational efficiency, strategic alignment, transparency and learning. Sound design is characterized by solid programmed classifications, unambiguous responsibility, limited and relevant indicators, baselines, targets, documented sources and connections between financial and non-financial information. Indicators must balance quantity, quality, cost and timing to minimize gaming and overemphasis on what is easiest to measure.

https://doi.org/10.59992/IJFAES.2026.v5n8p9 E-ISSN 2977-1498

Table (3) Program and performance budgeting cycle requirements

Stage	Core requirement
Priority setting	Translate strategy into objectives, programmes, and target results.
Programme design	Define scope, beneficiaries, governance, and accountability.
Activities and costing	Link resources and activities to outputs and allocate direct and shared costs.
Indicator design	Specify definition, baseline, target, source, frequency, and owner.
Budget estimates	Estimate revenue, expenditure, demand, and alternative funding scenarios.
Review and approval	Test consistency between priorities, appropriations, and expected results.
Execution and monitoring	Update expenditure and performance data and analyse variances.
Evaluation and feedback	Interpret results and use lessons in the next budget cycle.

PPB can make resource debates more substantive by comparing programmes rather than applying uniform increases or cuts. It can distinguish underfunding from design or implementation failure and improve dialogue between financial and service units. However, benefits are conditional. Common obstacles include poor data quality, unclear indicator ownership, missing baselines, unstable programme structures, excessive measures, weak analytical skills, separated planning and finance functions, and reporting that becomes ceremonial rather than decision-relevant.

Behavioural risks also matter. Units may select easy targets, optimise the indicator instead of the intended result, or conceal weak performance if information is used punitively. Evidence from performance-based funding in higher education illustrates that indicator-linked funding does not reliably produce intended outcomes across contexts and designs (Shin et al., 2023). Performance information should therefore support structured dialogue, verification, learning, and proportional accountability rather than automatic allocation formulas.

2.2 Artificial Intelligence in Financial Management:

AI in financial management is not a single system. It includes rule-based decision support, machine learning, deep learning, natural language processing, anomaly detection, predictive modelling, and generative tools. Automation executes predefined rules; descriptive analytics explains what happened; predictive analytics estimates what may happen; and prescriptive analytics compares alternatives. A financial platform may combine these levels, but each requires different data, controls, and skills.

During budget preparation, predictive models may estimate baseline revenue, expenditure, service demand, and financing needs. Scenario analysis can replace a single point estimate with ranges based on different assumptions. During allocation, integrated cost, performance, and priority data can support programme comparison. During execution, frequent updates enable variance monitoring and early-warning alerts. At year-end, analytics can support programme evaluation and feed lessons into the next cycle (OECD, 2025).

AI applications in public financial management currently tend to reinforce existing systems rather than redesign them entirely. Key uses include improving forecasts, facilitating spending decisions, supporting budget planning and monitoring, automating management and oversight, and improving

International Journal of Financial, Administrative and Economic Sciences, London Vol (5), No (8), 2026
https://doi.org/10.59992/IJFAES.2026.v5n8p9 E-ISSN 2977-1498

user engagement. Adoption is constrained by legacy systems, fragmented data, legal uncertainty, skill shortages, and the need for transparency and human oversight (OECD, 2025).

Four requirements are central. First, data governance must standardise definitions, codes, metadata, quality rules, access rights, and lineage between accounts, programmes, and indicators. Second, model governance must establish ownership, purpose, documentation, testing, monitoring, escalation, and stopping rules. Third, multidisciplinary capability is needed so financial staff understand assumptions and technical staff understand budgeting and accountability. Fourth, infrastructure and security must cover systems integration, identity management, encryption, audit logs, continuity, vendor management, and lawful storage.

Table (4) Artificial intelligence risks and proposed controls

Risk	Possible effect	Control
Low-quality data	Unstable or biased outputs	Data dictionary, quality tests, and independent verification.
Opacity	Weak trust and accountability	Interpretable models and explanation of influential factors.
Bias	Unfair resource distribution	Representative data, fairness tests, and human review.
Privacy breach	Disclosure of sensitive information	Data minimisation, access control, encryption, and retention rules.
Automation bias	Uncritical acceptance of recommendations	Training, explicit human approval, and challenge procedures.
Model drift	Declining accuracy as conditions change	Monitoring, recalibration, retraining, and stopping thresholds.
Cybersecurity threat	Manipulation, leakage, or disruption	Security testing, logging, access monitoring, and incident response.

2.3 Financial Decision-Making Quality:

A financial decision selects an alternative for obtaining, allocating, using, or controlling resources under objectives, constraints, and uncertainty. In public organisations, quality includes legality, equity, public value, transparency, and accountability, not merely financial return. This study measures quality through timeliness, accuracy, integrated evidence, allocation efficiency, expenditure rationalisation, early risk detection, comparison of alternatives, traceability, responsiveness, and transparency.

Decision quality depends on timely and reliable data, clear objectives, explicit decision criteria, professional expertise, feasible alternatives, coordination, time, uncertainty, and legal controls. Dashboards and analytical tools can reduce information-processing burden, but poorly designed systems may increase complexity. Useful decision support should direct attention to exceptions, permit drill-down to sources, communicate assumptions and uncertainty, and preserve a record of analysis, recommendations, human revisions, and approval.

2.4 Integration Logic and Proposed Operating Model:

PPB supplies the conceptual architecture of programmes, objectives, indicators, and costs. AI supplies the capacity to process the volume, variety, and velocity of relevant data. Without a programme architecture, analytics may remain disconnected from objectives; without analytical capacity, performance information may remain delayed and descriptive. Integration creates value when trusted data are converted into forecasts, alerts, programme comparisons, and transparent alternatives within controlled workflows.

Table (5) Proposed operating model for PPB–AI integration

Step	Process	Main output
1	Collect financial, operational, performance, and relevant external data.	Integrated input set.
2	Clean, standardize, validate, and link accounts, programmes, and indicators.	Traceable analytical dataset.
3	Describe baselines, trends, costs, and variances.	Current-state diagnosis.
4	Forecast revenue, expenditure, demand, and risk under scenarios.	Forward-looking estimates and uncertainty.
5	Compare programme cost, outputs, outcomes, priority, and risk.	Programme evidence profile.
6	Generate feasible allocation alternatives, impacts, and limitations.	Decision options.
7	Apply authorised human review, challenge, and approval.	Accountable decision.
8	Compare actual with target and forecast; update data, indicators, and models.	Feedback and continuous improvement.

The model is cyclical. Actual results are compared with forecasts and targets, differences are explained, and data definitions, indicators, assumptions, or models are revised. Practical safeguards include starting with a defined use case and baseline; maintaining a common data dictionary; retaining the human decision-maker; logging automated recommendations and human changes; monitoring accuracy, stability, bias, privacy, and security; separating test and production environments; and training users not to confuse correlation with causation or forecasts with certainty.

2.5 Previous Studies and Research Gap:

Table (6) Summary of previous studies and research gap

Source	Verified contribution and relevance
Tryggvadottir & Bambalaite (2024)	Developed the OECD framework around tools and methods, accountability and transparency, the enabling environment, and use of performance information. Supports the study's information-use logic.
Mohd Gharib et al. (2024)	Thematically reviewed 61 studies from 2018–2023 and identified practices, implementation, and use of performance information as the principal themes. Supports the implementation focus.
Shin et al. (2023)	Reviewed 36 studies of US performance-based funding and found largely inconclusive or mixed effects across designs. Warns against assuming automatic results from indicator-linked funding.
OECD (2024)	Assessed opportunities, risks, and readiness for public-sector AI, emphasising productivity, responsiveness, accountability, enabling conditions, and safeguards.

Source	Verified contribution and relevance
OECD (2025)	Examined 200 AI use cases across 11 government functions and described forecasting, spending support, monitoring, automation, and oversight applications in public financial management.
Tabassi (2023)	Presented the NIST AI RMF 1.0 and its Govern, Map, Measure, and Manage functions. Supports model governance and trustworthy-AI controls.
Bahoo et al. (2024)	Reviewed AI in finance through bibliometric and content analysis, highlighting forecasting, classification, early warning, data mining, and multiple financial application streams.
Begg et al. (2024)	Analysed performance budgeting in EU spending frameworks and demonstrated that designs, benefits, and drawbacks vary by institutional setting.

The literature on performance budgeting concentrates on indicators, institutional settings, implementation, and actual information use, whereas the AI literature concentrates on forecasting, automation, analytical capability, and trustworthy governance. Field-based measurement of how these domains jointly relate to financial decision quality remains limited, particularly in Arab public-sector settings. This study addresses that gap by linking specific AI-enabled capabilities to specific points in the PPB cycle and by defining decision quality more broadly than speed or accuracy alone. Because the design is cross-sectional and perception-based, the eventual findings will be interpreted as associations and explanatory statistical effects, not definitive causation.

Chapter Three: Methodology

3.1 Research Design, Population, and Sampling:

The study adopts a quantitative, descriptive-analytical, cross-sectional survey design. The descriptive component measures the level of integration and decision quality; the analytical component tests associations and predictive effects at one point in time. The target population comprises employees and managers connected with budget preparation, execution, monitoring, finance, strategic planning, performance measurement, internal audit, information systems, or digital transformation. The unit of analysis is an individual employee with practical knowledge of budget, data, or financial decisions.

Where a complete sampling frame exists, proportionate stratified random sampling by department or job level is preferred. A census may be used for a small population. If a sampling frame is unavailable, criterion-based purposive sampling may be used, provided that the limitation on generalizability is reported. The final sample should be calculated when the population size and expected response rate are known. As a practical minimum, at least 150 valid responses are desirable for multiple regression and exploratory factor analysis, with 200 or more preferred when feasible.

3.2 Instrument:

The closed-ended questionnaire contains informed-consent and eligibility items, general information, 35 independent-variable statements across five dimensions, and 10 dependent-variable statements, for 45 measurement items. Responses use a five-point Likert scale: 5 = Strongly agree, 4 = Agree, 3

= Neutral, 2 = Disagree, and 1 = Strongly disagree. All substantive items are positively worded to simplify coding; item order may be varied within each section in the electronic form.

Table (7) Questionnaire constructs and scoring

Code	Construct	Items	Score
DR	Data readiness and integration	1–7	Mean
IA	Intelligent analytics	8–14	Mean
FF	AI-supported financial forecasting	15–21	Mean
BA	Budget-cycle automation	22–28	Mean
PD	Programme evaluation and intelligent decision support	29–35	Mean
FDQ	Financial decision-making quality	36–45	Mean

3.3 Validity and Reliability:

Content and face validity will be assessed by five to seven specialists in accounting, public finance, statistics, information systems, or research methods. They will evaluate relevance, clarity, dimensional fit, duplication, and cultural appropriateness. A pilot test with approximately 20–30 eligible individuals who are excluded from the main sample will examine comprehension, completion time, and electronic-form usability.

After fieldwork, corrected item-total correlations and correlations between each dimension and the total scale will assess internal consistency. If the sample is adequate, exploratory factor analysis will be conducted after the Kaiser–Meyer–Olkin measure and Bartlett's test. Loadings, cross-loadings, and theoretical coherence will be considered jointly. Cronbach's alpha will be reported for each dimension and the complete instrument, but items will not be deleted solely to maximise alpha when they are essential to content validity.

3.4 Data Collection and Ethics:

Create the questionnaire in Microsoft Forms or Google Forms without collecting names or email addresses unless institutionally required and explicitly disclosed.

Present the research purpose, voluntariness, confidentiality, estimated duration, and academic contact information before consent.

Use an eligibility question and terminate the form for respondents whose work is unrelated to the defined fields.

Test the form on desktop and mobile devices and conduct the pilot before launch.

Announce a defined collection period and send no more than reasonable, non-coercive reminders.

Export responses to Excel, preserve an unedited raw file, and use a separate working copy for cleaning and coding.

Participation is voluntary and based on informed consent. Data will be used only for research, reported in aggregate, stored securely, and accessed only by the research team. Unnecessary identifiers will not be collected. Any required administrative or ethics approval must be obtained before distribution.

3.5 Statistical Analysis Plan:

Table (8) Statistical analysis plan

Purpose	Procedure
Initial screening	Frequency checks, missing values, valid ranges, duplicate indicators, response patterns, and outliers.
Sample description	Frequencies and percentages for categorical variables.
Scale description	Means, standard deviations, ranks, and confidence intervals where appropriate.
Reliability	Cronbach's alpha and corrected item-total correlations.
Construct validity	KMO, Bartlett's test, and exploratory factor analysis if sample size permits.
Relationships	Pearson correlation or a non-parametric alternative when assumptions are not met.
Effect testing	Simple regression for overall integration and multiple regression for the five dimensions.
Group differences	Independent-samples t-test or ANOVA with post-hoc tests; non-parametric alternatives when required.
Diagnostics	Linearity, independence of errors, homoscedasticity, residual normality, tolerance, VIF, and influential cases.
Reporting	Unstandardised and standardised coefficients, 95% confidence intervals, p-values, R^2, adjusted R^2, and effect sizes.

Proposed model: $FDQ = \beta 0 + \beta 1DR + \beta 2IA + \beta 3FF + \beta 4BA + \beta 5PD + \varepsilon$. Statistical significance will be evaluated at $\alpha = .05$, but substantive interpretation will also consider coefficient magnitude, confidence intervals, model fit, diagnostics, and study limitations.

Table (9) Interpretation of mean scores

Mean range	Descriptive interpretation
1.00–1.80	Very low
1.81–2.60	Low
2.61–3.40	Moderate
3.41–4.20	High
4.21–5.00	Very high

These intervals are descriptive only. Means and standard deviations will still be reported, and small numerical differences will not be exaggerated. A single Likert item is ordinal, while a coherent multi-item mean is commonly treated as a composite score in applied analysis.

Chapter 4: Results

4.1 Data Screening and Analytical Sample:

The raw workbook contained 71 submitted records. All respondents consented, but seven respondents answered that their duties were not related to budgeting, finance, planning, performance measurement, audit, information systems, or digital transformation. In accordance with the pre-specified eligibility rule, these seven records were excluded. The final analytical sample therefore comprised 64 eligible respondents. All 45 Likert items were complete, so no missing-value imputation was required.

A response-pattern check identified two pairs of records with identical demographic and item-response profiles but different timestamps. Because the dataset contained no personal identifier and identical profiles may arise legitimately, the records were retained in the primary analysis. A sensitivity analysis retaining only the first record in each pair produced virtually unchanged conclusions: the overall integration model remained strong ($n = 62$, $R^2 = .788$, $b = .993$, $p < .001$). This supports the robustness of the main result while preserving transparent reporting.

4.2 Respondent Characteristics:

Table (10) Respondent characteristics

Characteristic	Category	n	%
Gender	Male	48	75.0
	Female	16	25.0
Education	Bachelor’s	16	25.0
	Master’s	45	70.3
	Doctorate	3	4.7
Work area	Finance	33	51.6
	Planning and performance	14	21.9
	Audit and control	11	17.2
	Budgeting	5	7.8
	Digital transformation	1	1.6

The sample was concentrated in finance and held relatively high academic qualifications. The small numbers in the doctorate, budgeting, and digital-transformation categories limit the precision of subgroup comparisons; consequently, subgroup findings are interpreted cautiously rather than as population estimates.

4.3 Reliability and Measurement Checks:

Table (11) Reliability and internal consistency results

Scale	Items	Cronbach’s α	Corrected item–total r
DR	7	0.920	0.617–0.846
IA	7	0.924	0.699–0.804
FF	7	0.942	0.770–0.866
BA	7	0.949	0.728–0.868
PD	7	0.963	0.824–0.916
FDQ	10	0.966	0.769–0.926
All items	45	0.987	0.550–0.887

Internal consistency was excellent for every construct, with alpha values ranging from .920 to .966, while the complete instrument yielded $\alpha = .987$. Corrected item–total correlations were all above .55. These findings support score reliability. The very high total-scale alpha and strong inter-construct correlations also indicate substantial conceptual overlap, which is important when interpreting unique regression coefficients.

https://doi.org/10.59992/IJFAES.2026.v5n8p9 E-ISSN 2977-1498

The Kaiser–Meyer–Olkin statistic was .848 and Bartlett's test of sphericity was significant, χ^2 (990) = 3,771.71, $p < .001$. Nevertheless, the case-to-item ratio was only 1.42:1 (64 respondents for 45 items), which is inadequate for stable exploratory factor analysis. Therefore, factor extraction was not used to claim structural validity. The study relies on the theory-based six-scale structure, content review, item–total evidence, and reliability results, while recommending confirmatory validation in a substantially larger independent sample.

4.4 Descriptive Results:

Table (12) Descriptive statistics for study constructs

Construct	Mean	SD	95% CI	Level
Data readiness and integration	3.701	0.824	[3.495, 3.907]	High
Intelligent analytics	3.681	0.927	[3.449, 3.912]	High
AI-supported financial forecasting	3.650	0.951	[3.412, 3.887]	High
Budget-cycle automation	3.743	0.967	[3.502, 3.985]	High
Programme evaluation and decision support	3.746	0.983	[3.500, 3.991]	High
Overall integration	3.704	0.856	[3.490, 3.918]	High
Financial decision-making quality	3.792	0.951	[3.555, 4.030]	High

All construct means fell within the pre-specified "high" interval of 3.41–4.20. Financial decision-making quality had the highest mean (M = 3.792, SD = .951), followed by programme evaluation and intelligent decision support (M = 3.746) and budget-cycle automation (M = 3.743). AI-supported financial forecasting was the lowest dimension, although it remained high (M = 3.650, SD = .951). The 95% confidence intervals were relatively wide, reflecting the modest sample and heterogeneous perceptions.

https://doi.org/10.59992/IJFAES.2026.v5n8p9 E-ISSN 2977-1498

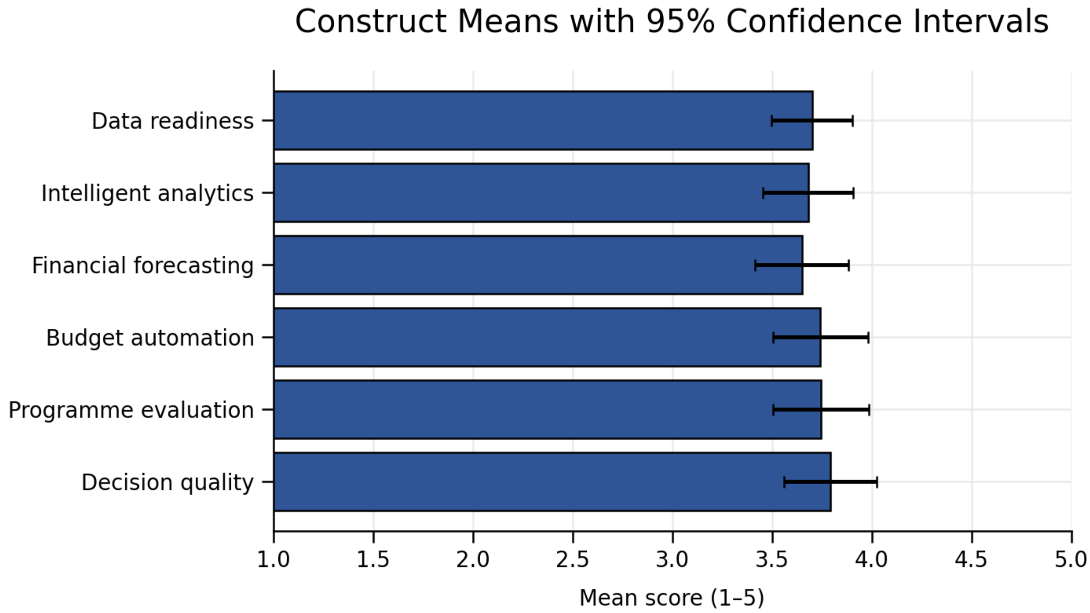

Figure (1) Construct means and 95% confidence intervals.

At item level, the highest-rated integration statement was that automation reduces errors and time in the budget cycle (Item 28, M = 4.031, SD = 1.069). The next highest were regular data accuracy and completeness checks (Item 4, M = 3.969) and intelligent ranking of programmes by priority and performance (Item 30, M = 3.953). The lowest-rated integration item concerned combining data from different systems without substantial manual intervention (Item 5, M = 3.453), followed by scenario development rather than reliance on a single forecast (Item 18, M = 3.516). These results identify interoperability and scenario capability as the clearest improvement priorities.

4.5 Correlation Analysis:

Table (13) Pearson correlation matrix

Variable	DR	IA	FF	BA	PD	FDQ
DR	1.000	0.755	0.779	0.777	0.825	0.751
IA	0.755	1.000	0.844	0.842	0.839	0.844
FF	0.779	0.844	1.000	0.747	0.806	0.804
BA	0.777	0.842	0.747	1.000	0.863	0.839
PD	0.825	0.839	0.806	0.863	1.000	0.851
FDQ	0.751	0.844	0.804	0.839	0.851	1.000

All five integration dimensions were strongly and positively related to financial decision-making quality: DR, $r = .751$; IA, $r = .844$; FF, $r = .804$; BA, $r = .839$; and PD, $r = .851$; all $p < .001$. Correlations among the five predictors ranged from .747 to .863. Thus, each dimension was associated with decision quality when considered separately. Still, the strong predictor intercorrelations show that the dimensions largely move together and should not be interpreted as statistically independent organizational capabilities.

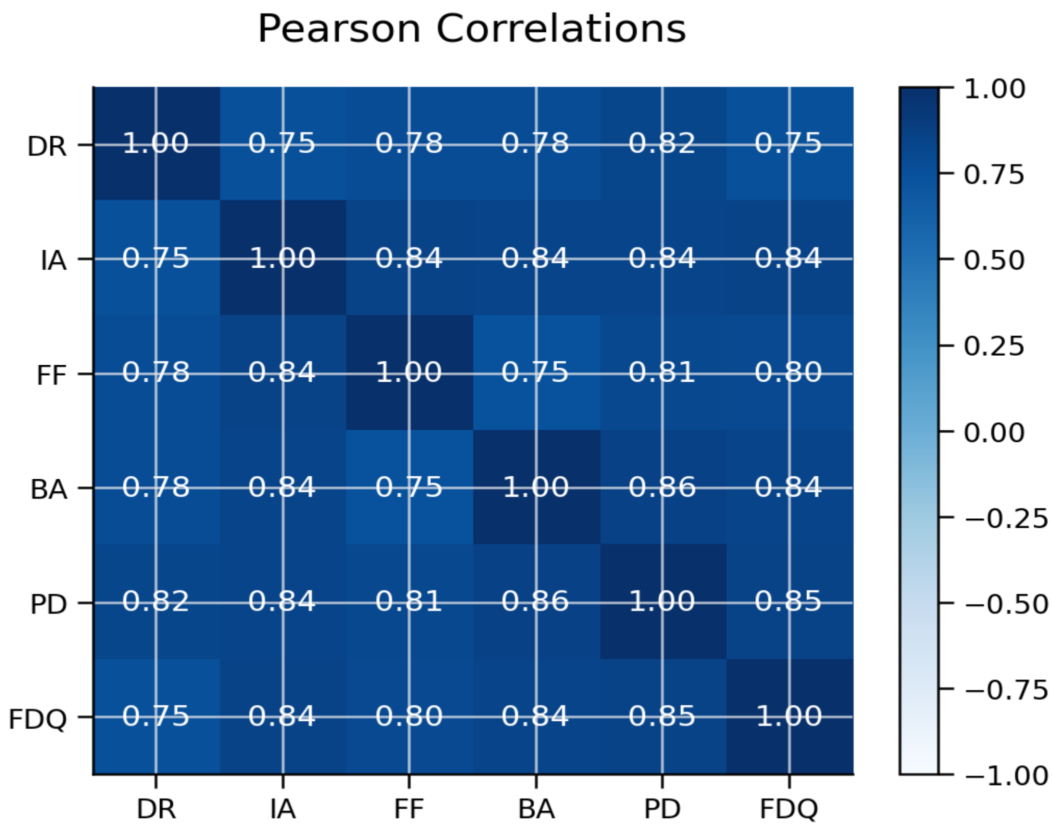

Figure (2) Pearson correlation matrix for the study constructs.

4.6 Regression Analysis and Hypothesis Testing:

Table (14) Simple regression of overall integration on decision quality

Predictor	b	SE	β	t	p	95% CI
Overall integration	0.989	0.064	0.891	15.430	< .001	[0.861, 1.117]

Overall integration significantly predicted financial decision-making quality, $F(1, 62) = 238.09$, $p < .001$, $R^2 = 0.793$, adjusted $R^2 = 0.790$. A one-point increase in the integration score was associated with an estimated .989-point increase in decision quality (95% CI [.861, 1.117]). The very strong standardised effect ($\beta = .891$) supports H1 at the aggregate level. Given the cross-sectional self-report design, this coefficient is interpreted as a strong explanatory association, not proof of causality.

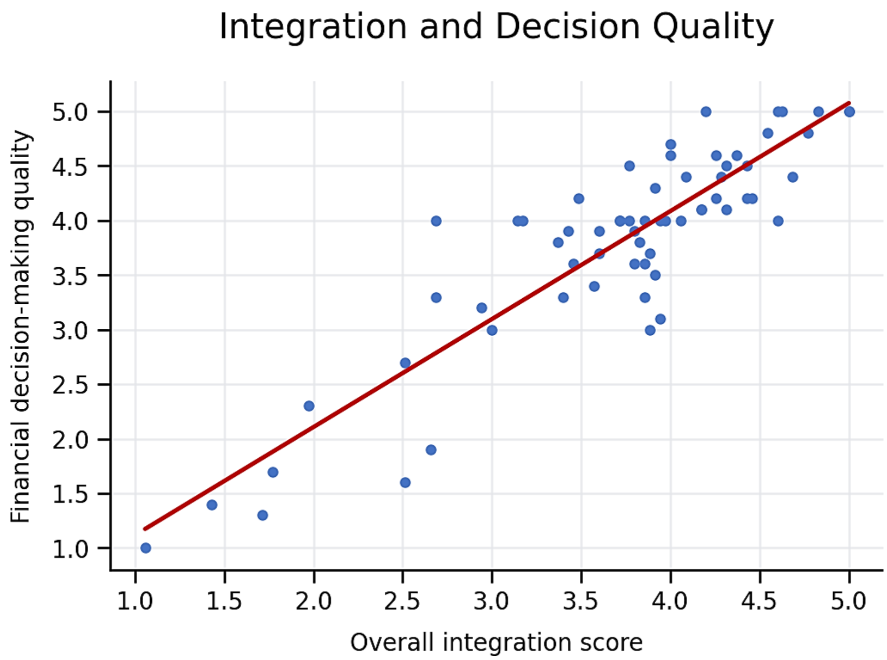

Figure (3) Overall integration score and financial decision-making quality.

Table (15) Multiple regression results with HC3 robust standard errors

Predictor	b	HC3 SE	β	p	95% robust CI	VIF
DR	-0.017	0.237	-0.014	0.944	[-0.482, 0.449]	3.67
IA	0.228	0.221	0.223	0.302	[-0.205, 0.662]	5.55
FF	0.196	0.221	0.196	0.374	[-0.237, 0.629]	4.33
BA	0.269	0.229	0.273	0.240	[-0.180, 0.717]	5.00
PD	0.273	0.247	0.282	0.269	[-0.212, 0.758]	5.93

The simultaneous five-predictor model was significant, $R^2 = 0.804$, adjusted $R^2 = 0.787$, with the conventional omnibus $F (5, 58) = 47.65521971854085592212868505157530307769775390625$, $p < .001$. The Breusch–Pagan test indicated heteroscedasticity ($p = .029$), so HC3 robust standard errors and confidence intervals are reported. With robust inference, none of the five overlapping dimensions had a statistically significant unique coefficient. VIF values ranged from 3.67 to 5.93, confirming moderate-to-high multicollinearity. Therefore, H1a–H1e are supported as separate bivariate relationships, but their unique incremental effects are not confirmed when all dimensions are entered simultaneously. The appropriate substantive conclusion is that integrated capability as a coherent bundle, rather than one isolated dimension, is strongly associated with decision quality.

Residuals were approximately normal (Shapiro–Wilk $W = .982$, $p = .459$), and the Durbin–Watson statistic was 1.95. Eight cases exceeded the simple Cook's distance threshold of $4/n$, with a maximum of .845. Because no case exceeded 1.0 and the sensitivity analysis was stable, observations were retained. These diagnostics, together with evidence of heteroscedasticity and a modest sample size, justify cautious estimation and the use of robust standard errors.

4.7 Group Differences:

Table (16) Group differences in financial decision-making quality

Grouping variable	Welch statistic	Df	p	Conclusion
Gender	F = 0.361	1, 28.74	.552	Not significant
Education	F = 1.218	2, 6.38	.356	Not significant
Work area*	F = 0.245	3, 15.07	.863	Not significant

No statistically significant differences in financial decision-making quality were found by gender, education, or work area. The digital-transformation category contained only one eligible respondent and was excluded from the work-area inferential test, although retained in descriptive reporting. H2 was therefore not supported. These null findings should not be taken as evidence of equivalence because several groups were small and statistical power was limited.

4.8 Hypothesis Decisions:

Table (17) Summary of hypothesis decisions

Hypothesis	Decision	Evidence
H1	Supported	Overall integration: $\beta = .891$, $R^2 = .793$, $p < .001$.
H1a	Supported bivariate; unique effect not confirmed	DR–FDQ $r = .751$, $p < .001$; robust partial $p = .944$.
H1b	Supported bivariate; unique effect not confirmed	IA–FDQ $r = .844$, $p < .001$; robust partial $p = .302$.
H1c	Supported bivariate; unique effect not confirmed	FF–FDQ $r = .804$, $p < .001$; robust partial $p = .374$.
H1d	Supported bivariate; unique effect not confirmed	BA–FDQ $r = .839$, $p < .001$; robust partial $p = .240$.
H1e	Supported bivariate; unique effect not confirmed	PD–FDQ $r = .851$, $p < .001$; robust partial $p = .269$.
H2	Not supported	No significant subgroup differences; all $p \geq .356$.

Chapter Five: Discussion, Conclusions, and Recommendations

5.1 Discussion:

The central finding is a strong positive relationship between integrated PPB–AI capability and financial decision-making quality. The result is consistent with the logic of the OECD Performance Budgeting Framework, which places the actual use of meaningful performance information within an enabling and accountable environment at the centre of effective performance budgeting. It also aligns with evidence that AI in government can improve forecasting, monitoring, spending support, and organisational responsiveness when embedded in established public-management processes rather than deployed as a stand-alone tool (Tryggvadottir & Bambalaite, 2024; OECD, 2025).

The high mean for programme evaluation and intelligent decision support suggests that respondents recognise value at the point where cost, performance, priorities, and alternatives are brought together. Its bivariate correlation with decision quality was the strongest among the five dimensions. This supports the study's proposition that data and models create public-management value only when their outputs inform programme comparison and accountable resource-allocation choices. The finding also reinforces the distinction between presenting indicators and actually using them, a continuing implementation issue identified by Mohd Gharib et al. (2024).

Budget-cycle automation was also scored well. The statement on automation reducing error and cycle time received the highest item mean. This implies that the immediate organisational benefits are likely to be more obvious in workflow efficiency, in regular reporting and in less manual processing. However, automation is not the same as intelligence. Routine automation can increase timeliness, but decision quality also requires credible data, interpretability, scenario assessment and human challenge. This differentiation is compatible with the NIST emphasis on governance, measurement, transparency and continuous risk management (Tabassi, 2023).

The lowest-rated construct was financial forecasting, and scenario development was among the lowest-rated constructs. This pattern suggests that firms are potentially further along in digital workflow and dashboarding than in probabilistic and multi-scenario forecasting. To forecast maturity, you need historical and present data, specific assumptions, tests on out-of-sample accuracy, communication about uncertainty, and updating the model. Building these competencies would shift the firm from retrospective reporting to forward-looking financial management.

Overall data integration was graded positively but the lowest item meaning was for cross-system combination without extensive manual involvement. This suggests a practical bottleneck: advanced analytics can't be dependable or scalable if programme, accounting, operational and performance records are not compatible. Thus, the theoretical model is validated in its view of data readiness as foundational, even though data readiness did not have a distinct coefficient in the simultaneous model.

The lack of unique robust effects for the five dimensions should not be construed as proof that the dimensions are useless. Each measure was highly associated to decision quality, but the dimensions were also highly related to each other. Statistically, shared variance makes partial coefficients unstable and increases uncertainty. In essence, this pattern demonstrates that effective integration is systemic: data, analytics, forecasting, automation and programme assessment are mutually reinforcing skills. A phased portfolio strategy is the preferred technique compared to investing in one single technology.

The findings also require caution. The design is cross-sectional and uses a common self-report questionnaire, which can inflate associations through common-method variance. Strong alphas and high construct correlations may reflect coherent organizational capability, item similarity, global positive or negative response tendencies, or a combination of these mechanisms. Consequently, publication claims should use "association," "relationship," or "predictive statistical effect" rather than causal language such as "AI causes better decisions."

5.2 Conclusions:

1. The eligible respondents perceived both PPB–AI integration and financial decision-making quality at a high level.

2. Overall integration was strongly associated with decision quality and explained approximately 79.3% of its observed variance in the simple cross-sectional model.
3. Every integration dimension was strongly associated with decision quality when considered separately.
4. The dimensions overlapped substantially, and no dimension demonstrated a robust unique effect in the simultaneous model. Integration should therefore be understood as a capability bundle.
5. Interoperability and scenario-based forecasting were the clearest operational weaknesses, while decision-support use and automation were comparatively strong.
6. No statistically significant differences were detected across the measured demographic and occupational groups, but small subgroup sizes limit inference.

5.3 Practical Recommendations:

1. Establish a PPB–AI governance committee with financial, programme, performance, data, audit, legal, cybersecurity, and technical representation.
2. Create a common programme and performance data model linking chart-of-accounts codes, activities, outputs, outcomes, baselines, targets, data owners, and source systems.
3. Prioritise interoperability by using controlled identifiers, APIs or governed data pipelines, metadata catalogues, validation rules, and documented data lineage.
4. Develop scenario-based forecasting for revenue, expenditure, demand, and fiscal risk. Report assumptions, uncertainty ranges, forecast errors, and model revisions.
5. Implement automation selectively for reconciliations, repeated entry, exception alerts, report refreshes, and workflow approvals, while retaining reviewable audit trails.
6. Design programme-evaluation dashboards that combine cost, output, outcome, priority, risk, and distributional considerations rather than ranking programmes through a single opaque score.
7. Adopt model-risk controls based on defined purpose, validation, performance monitoring, bias and privacy checks, cybersecurity testing, human override, escalation, and stopping rules.
8. Provide role-based training so financial managers can interpret uncertainty and model limitations and technical teams can understand budget rules and public accountability.
9. Measure implementation benefits against baselines, including reporting lead time, manual corrections, forecast error, early risk detection, programme-review quality, and user confidence.
10. Pilot one high-value case, such as expenditure forecasting or variance alerts, evaluate it independently, and scale only after demonstrating accuracy, usability, security, and governance.

5.4 Limitations and Future Research:

The study has five principal limitations. First, the analytical sample was modest (n = 64), which limits precision, subgroup power, and factor analysis. Second, respondents were selected through an eligibility-based field survey rather than a documented probability sample, so generalisation should be cautious. Third, all variables were measured through one questionnaire at one time, creating possible common-method bias and preventing causal inference. Fourth, several constructs and item wordings were conceptually close, producing high intercorrelations and multicollinearity. Fifth, the dataset did not contain age, experience, job level, or training variables that had been proposed in the original methodology, so those comparisons could not be performed.

Future studies should use a larger multi-organisation sample and validate the measurement model through confirmatory factor analysis or partial least squares structural equation modelling with adequate power. Longitudinal or quasi-experimental designs should compare decision outcomes before and after implementation. Survey evidence should be triangulated with objective measures such as forecast error, budget-cycle time, number of manual adjustments, audit exceptions, programme cost effectiveness, and documented use of analytical recommendations. Marker variables, temporal separation, multiple informants, and objective outcomes can also reduce common-method risk.

References

1. Abbasov, R. (2025). The effectiveness of performance-based budgeting in the public sector: An empirical analysis and policy implications. iBusiness, 17(1), 56–76. https://doi.org/10.4236/ib.2025.171003.
2. Autio, C., Schwartz, R., Dunietz, J., Jain, S., Stanley, M., Tabassi, E., Hall, P., & Roberts, K. (2024). Artificial intelligence risk management framework: Generative artificial intelligence profile (NIST AI 600-1). National Institute of Standards and Technology. https://doi.org/10.6028/NIST.AI.600-1.
3. Dener, C., Watkins, J. A., & Dorotinsky, W. L. (2011). Financial management information systems: 25 years of World Bank experience on what works and what doesn't. World Bank. https://doi.org/10.1596/978-0-8213-8750-4.
4. Jain, V., & Kulkarni, P. A. (2023). Integrating AI techniques for enhanced financial forecasting and budgeting strategies. SSRG International Journal of Economics and Management Studies, 10(9), 9–15. https://doi.org/10.14445/23939125/IJEMS-V10I9P102.
5. Jumardi. (2025). Assessing the effectiveness of performance-based budgeting in achieving public sector financial goals. The Journal of Academic Science, 2(4). https://doi.org/10.59613/zezb3994.
6. Khadivar, A., & Abdian, S. (2026). AI-enabled robust performance-based budgeting: A hybrid neural–optimization approach for smart governance. Emerging Technologies and Governance, 1(1), 41–65. https://doi.org/10.47176/ETG.2026.1005.

7. Mohd Gharib, N. S., Saleh, Z., & Isa, C. R. (2024). A thematic review on performance-based budgeting in the public sector organisation. IPN Journal of Research and Practice in Public Sector Accounting and Management, 14(2), 1–31. https://doi.org/10.58458/ipnj.v14.02.01.0104.
8. National Institute of Standards and Technology. (2023). Artificial intelligence risk management framework (AI RMF 1.0) (NIST AI 100-1). U.S. Department of Commerce. https://doi.org/10.6028/NIST.AI.100-1.
9. Organisation for Economic Co-operation and Development. (2015). Recommendation of the Council on Budgetary Governance (OECD/LEGAL/0410). https://legalinstruments.oecd.org/en/instruments/OECD-LEGAL-0410.
10. Organisation for Economic Co-operation and Development. (2019a). Budgeting and public expenditures in OECD countries 2019. OECD Publishing. https://doi.org/10.1787/9789264307957-en.
11. Organisation for Economic Co-operation and Development. (2019b). OECD good practices for performance budgeting. OECD Publishing. https://doi.org/10.1787/c90b0305-en.
12. Organisation for Economic Co-operation and Development. (2023). OECD performance budgeting framework (GOV/SBO (2023)1). https://one.oecd.org/document/GOV/SBO(2023)1/en/.
13. Organisation for Economic Co-operation and Development. (2024). Using artificial intelligence in public financial management (GOV/SBO (2024)14). https://one.oecd.org/document/GOV/SBO(2024)14/en/.
14. Rivero del Paso, L., Pattanayak, S., Uña, G., & Tourpe, H. (2023). Digital solutions guidelines for public financial management (Technical Notes and Manuals No. 2023/007). International Monetary Fund. https://doi.org/10.5089/9798400251634.005.
15. Wang, D., Li, C., Wang, Y., Ren, X., Hu, N., & Ooi, K. L. (2023). The role of artificial intelligence and machine learning in financial forecasting and decision-making. International Journal of Accounting, Finance and Business, 8(50), 275–289.