Future of Work and Digital Management Journal

Article type:
Original Research

Article history:
Received 20 June 2026
Revised 21 September 2026
Accepted 24 September 2026
Initial Publish 27 October 2026
Published online 01 March 2027

Teymoor. Bidari[1], Hossien. Shafiei[2*], Mohammad Reza. Khezripour[2]

1 PhD Student, Department of Accounting, Si.C., Islamic Azad University, Sirjan, Iran.
2 Department of Accounting, Si.C., Islamic Azad University, Sirjan, Iran.

Corresponding author email address:
hosseinshafiee@iau.ac.ir

How to cite this article:
Bidari, T., Shafiei, H., & Khezripour, M. R. (2027). Empirical Validation and Testing of the Human Performance Evaluation Model in the Context of Implementing the New Tax Law. *Future of Work and Digital Management Journal,* 5(3), 1-20. https://doi.org/10.61838/fwdmj.356

Empirical Validation and Testing of the Human Performance Evaluation Model in the Context of Implementing the New Tax Law

ABSTRACT

Fundamental changes in the tax system and the imperative to transition from traditional paradigms to data-driven management require a high degree of adaptability in human capital. Nevertheless, the evident gap between auditors' individual competencies and the structural inertia of the organization impedes the effective implementation of new policies. The present study aimed to validate and empirically test a human performance evaluation model in the context of implementing modern tax laws. In terms of purpose, this study was applied, and in terms of methodological approach, it employed a quantitative, descriptive-survey, correlational design. The statistical population consisted of employees and managers of the Tax Administration, from whom a sample of 170 participants was selected using convenience sampling. Data were collected using a standardized questionnaire based on a five-point Likert scale. Data analysis and hypothesis testing were conducted through partial least squares structural equation modeling (PLS-SEM) using SmartPLS, supplemented by one-sample t-tests and the Friedman test. Evaluation of the measurement model confirmed convergent validity, with average variance extracted (AVE) values ranging from 0.52 to 0.57, as well as composite reliability (CR), with values ranging from 0.79 to 0.90. The structural model also demonstrated satisfactory goodness-of-fit indices (SRMR = 0.075; NFI = 0.90) and explained 50% of the variance in the outcome variable (R^2 = 0.50). Path analysis indicated that contextual conditions, with an effect coefficient of β = 0.36 and the highest ranking in the Friedman test (mean rank = 4.39), constituted the most influential moderating determinant within the system. Moreover, the strongest direct effect was observed for the path from strategies to outcomes (β = 0.71). Descriptive statistical analysis further revealed that auditors' digital skills (μ = 4.10) represented an organizational strength, whereas the transition toward data-driven analytical practices (μ = 3.58) constituted the organization's principal weakness. Based on the empirical findings, the successful implementation of tax laws depends on reengineering contextual and organizational structures and overcoming organizational inertia so that existing human-capital capabilities can be aligned with data-driven requirements and translated into desirable performance outcomes.

Keywords: empirical validation, structural equation modeling, human performance, tax laws, PLS modeling.

Introduction

Tax systems have undergone profound structural and technological transformation in recent years, shifting from predominantly document-based, procedure-oriented, and manually controlled administrative arrangements toward digitally integrated, data-intensive, and increasingly automated models of governance. This transition is not merely a technological upgrade; rather, it represents a broader reconfiguration of how tax administrations collect information, assess risk, interact with taxpayers, allocate organizational resources, and evaluate the performance of their personnel. In traditional tax administration, the effectiveness of employees and auditors was often assessed through relatively visible quantitative

indicators, such as the number of cases processed, the amount of assessed tax, or the volume of completed files. However, contemporary tax administration increasingly depends on more complex competencies, including data interpretation, interaction with intelligent systems, professional judgment under algorithmically generated recommendations, digital adaptability, and the capacity to reconcile legal, ethical, and organizational requirements. The growing digitalization of taxation therefore creates a need to reconsider the adequacy of existing human performance evaluation systems and to determine whether they remain aligned with the realities of contemporary tax administration. Internationally, the notion of Tax Administration 3.0 reflects this transition toward seamless, digitally embedded, and data-driven tax systems in which taxation becomes increasingly integrated into broader economic and administrative processes [1]. Within such an environment, the role of human resources does not disappear; rather, it becomes more specialized, interpretive, and strategically important.

The digital transformation of tax administration has increasingly been associated with artificial intelligence, automated risk assessment, digital reporting, large-scale data processing, and integrated information systems. These technologies are expected to improve efficiency, strengthen compliance monitoring, and reduce administrative costs; nevertheless, they also reshape the institutional roles of tax professionals and challenge traditional forms of expertise and authority. Sonnerfeldt emphasizes that tax technology has significant implications for systems of expertise and legitimacy because professional authority is increasingly shared with or mediated by technological infrastructures [2]. This issue is especially important for tax auditors, whose work has historically depended on technical expertise, discretionary judgment, and interpretation of complex legal and financial evidence. As digital systems increasingly identify anomalies, flag risks, and recommend actions, auditors are required not simply to follow system outputs but to understand, interpret, validate, and, where necessary, challenge them. Consequently, the ability to combine human judgment with technologically generated information becomes a central element of high-quality tax administration. Belahouaoui and Attak likewise show that digital taxation and artificial intelligence are becoming defining features of emerging tax administration models and that these changes influence both institutional performance and taxpayer compliance behavior [3]. Thus, the quality of implementation depends not only on the technical sophistication of digital systems but also on the extent to which organizational personnel possess the knowledge, skills, autonomy, and evaluative criteria required to use those systems effectively.

The increasing role of technology in tax administration is also closely connected with questions of compliance, inequality, and enforcement effectiveness. Alm argues that technological developments can influence tax evasion by changing the informational capacity of tax authorities, the probability of detection, and the distributional effects of enforcement mechanisms [4]. This observation highlights an important organizational implication: technological modernization can produce desirable outcomes only when it is accompanied by appropriate managerial and human-capital arrangements. Digital tools may increase access to information, but poor interpretation, resistance to technological change, lack of analytical competence, or excessive reliance on automated outputs can undermine their effectiveness. Therefore, technological reform requires parallel changes in performance expectations and personnel evaluation. A system that continues to reward employees solely for processing volume or numerical output may fail to recognize the increasingly important competencies associated with data analysis, error detection, professional judgment, system interpretation, and responsible decision-making. In this sense, performance evaluation becomes not merely an administrative mechanism but a strategic tool for aligning employee behavior with the broader requirements of digital tax reform.

Tax reform itself is a multidimensional organizational process involving legal change, institutional redesign, technological modernization, and behavioral adjustment. Aribowo et al. conceptualize tax reform as an ongoing journey toward more effective tax administration rather than as a single legislative or technical intervention [5]. From this perspective, the successful implementation of new tax laws depends on the compatibility of reforms with administrative capacity, organizational structures, technological readiness, and employee behavior. This is particularly relevant in environments where legal and procedural changes are introduced rapidly while organizational systems and human-resource practices adapt more slowly. New tax laws may impose different reporting standards, compliance mechanisms, digital processes, or assessment requirements, yet their operational success ultimately depends on the employees responsible for implementation. Purnomo and Gunadi similarly stress that policy implementation in taxation is shaped by multiple interacting components and that implementation quality has direct consequences for enforcement effectiveness and revenue outcomes [6]. Accordingly, evaluating human performance in the context of new tax legislation requires a framework capable of capturing not only individual effort but also the contextual and institutional conditions that influence how policies are translated into practice.

The importance of legal and policy design is particularly evident in the complexity of modern tax systems. New legislation may improve administrative capacity, broaden the tax base, or address emerging economic realities, but it may also generate ambiguity, implementation difficulties, and adjustment costs. Dashti, in examining the impact of new tax laws on municipal accounting and financial systems, highlights the operational consequences that legislative changes can have for accounting practices, financial processes, and administrative systems [7]. Such effects inevitably extend to the human resources responsible for applying revised procedures. When regulations change, employees must interpret new requirements, modify established routines, learn new systems, and manage uncertainty in situations where rules may not yet be fully institutionalized. The quality of employee performance under these circumstances cannot be assessed fairly through static measures developed for earlier administrative environments. Instead, evaluation systems should account for adaptability, learning, legal comprehension, professional judgment, and the capacity to operate effectively under conditions of regulatory transition.

The Iranian tax system provides a particularly relevant context for examining these issues because it has experienced substantial efforts toward modernization, electronic integration, and institutional reform. Earlier studies on the Iranian National Tax Administration have already emphasized the significance of technological preparedness and organizational readiness. Najari and Ahangari developed an e-readiness assessment model for the Iranian National Tax Administration, demonstrating that readiness for electronic transformation depends on a combination of technological, organizational, managerial, and human factors [8]. This is a critical point because the presence of digital infrastructure alone does not guarantee effective implementation. The ability of employees to use, interpret, and trust technological systems is equally important. Likewise, the modernization of the tax system can create new expectations regarding transparency, accountability, speed, and analytical precision, all of which alter the nature of human performance. As a result, personnel evaluation systems must evolve in parallel with technological transformation if they are to remain meaningful and behaviorally effective.

At the same time, implementation challenges in taxation are not exclusively technological. Legal coherence, administrative clarity, and the practical design of tax laws play equally important roles. Hashemian et al., in examining the value added tax

law and the challenges associated with its implementation, identified difficulties that may emerge when legal requirements encounter administrative realities [9]. Such challenges can create inconsistencies between formal rules and practical implementation, potentially increasing uncertainty for both employees and taxpayers. In a data-driven tax environment, these inconsistencies may become even more consequential because automated systems require relatively clear rules, standardized procedures, and reliable data structures. Where regulations are contradictory or difficult to interpret, tax employees may be forced to exercise substantial professional judgment. Therefore, performance evaluation systems should not penalize justified professional discretion simply because an employee deviates from a mechanically generated output. Instead, they should distinguish between arbitrary deviation and evidence-based expert judgment.

The behavioral dimension of taxation must also be considered because tax administration is fundamentally embedded in relationships of trust, enforcement, legitimacy, and compliance. Kirchler's economic psychology of tax behavior demonstrates that taxpayer behavior is influenced not only by deterrence and enforcement but also by psychological, social, and institutional factors, including trust in authorities and perceptions of fairness [10]. This perspective is directly relevant to employee performance because the conduct of tax auditors and administrators can shape taxpayers' perceptions of fairness, professionalism, and institutional legitimacy. An evaluation framework focused solely on numerical productivity may unintentionally encourage behaviors that undermine procedural justice, such as overly rapid case closure, rigid treatment of complex cases, or excessive emphasis on collection targets. In contrast, a more comprehensive system can incorporate qualitative indicators related to professional conduct, consistency, accuracy, fairness, and taxpayer interaction. Ehteshami and Abdi similarly emphasize the role of tax culture in strengthening the tax system, suggesting that sustainable compliance requires more than formal enforcement and depends on broader behavioral and institutional conditions [11]. Consequently, employee performance should be understood not only in operational terms but also in relation to its effects on institutional trust and tax culture.

Performance management in tax organizations therefore requires a systemic rather than narrowly individual perspective. Mansor and Tayib proposed an integrated and open systems model for tax administration performance management, emphasizing the need to assess performance through interconnected organizational dimensions rather than isolated metrics [12]. This systems perspective is especially applicable to contemporary tax administrations, where employee performance is shaped by technology, legal rules, managerial support, organizational structure, monitoring mechanisms, and external oversight. A tax auditor who possesses strong technical skills may nevertheless perform poorly in a context characterized by fragmented information systems, contradictory regulations, inadequate managerial support, or excessive fear of supervisory sanctions. Conversely, a supportive organizational context can amplify the value of professional expertise. Thus, human performance evaluation should recognize the interaction between individual capabilities and contextual conditions.

The ethical dimension of tax administration further reinforces the limitations of purely quantitative performance systems. Vaezi et al. designed a performance evaluation model for the national tax system with an emphasis on ethical principles, highlighting the importance of integrating ethical considerations into the evaluation of organizational performance [13]. Tax administration involves the exercise of public authority over individuals and organizations, and therefore fairness, accountability, integrity, and professional responsibility are central to legitimate performance. An evaluation system that rewards only output volume may create incentives for superficial compliance or statistical inflation rather than substantive quality. In contrast, an ethically informed system should recognize the quality of professional reasoning, the fairness of

decision-making, the accuracy of assessments, and the responsible use of discretion. This is particularly important in technologically mediated environments, where employees must balance system recommendations with legal obligations and professional judgment.

Decision-making quality is another important concern in tax administration because public-sector decisions often require balancing multiple and potentially competing criteria. Leontyev and Ketners show that tax decision-making may involve tensions among reliability, equity, and efficiency, and that such criteria are not always fully compatible [14]. This implies that employee performance cannot be reduced to a single quantitative outcome. A decision that maximizes speed may compromise fairness; a decision that prioritizes risk avoidance may reduce efficiency; and strict adherence to system-generated recommendations may not always produce the most legally defensible outcome. Therefore, a sophisticated performance evaluation model should recognize multidimensionality and allow for balanced assessment across different performance criteria. Such an approach is especially relevant when employees operate under conditions of technological uncertainty and legal complexity.

The relationship between legal design and taxpayer circumstances also illustrates the importance of context-sensitive tax administration. Mohammed and Hammadi examined legal exemptions and their compatibility with taxpayers' incomes and their effects on tax revenues, demonstrating how policy design can influence both taxpayer treatment and fiscal outcomes [15]. Although their focus was not employee performance directly, their findings underscore a broader principle: tax administration operates at the intersection of legal provisions, economic realities, and administrative judgment. Employees responsible for implementing tax rules must frequently interpret formal regulations in relation to diverse taxpayer conditions. Accordingly, effective human performance is partly reflected in the ability to apply laws accurately while maintaining consistency, equity, and responsiveness to contextual differences.

Taken together, the existing literature indicates that modern tax administration is characterized by a convergence of digital transformation, legal reform, organizational restructuring, behavioral considerations, ethical demands, and changing professional roles. However, an important practical gap remains in translating these developments into empirically validated systems for evaluating human performance. Traditional measures may no longer adequately capture the competencies required in environments where auditors interact continuously with intelligent systems, analyze large volumes of digital information, exercise professional judgment over automated recommendations, and operate under complex oversight arrangements. At the same time, the effectiveness of any performance evaluation model depends on its empirical validity within the specific organizational context in which it is applied. The implementation of new tax laws creates a particularly demanding setting because employees must simultaneously adapt to regulatory change, technological systems, new procedural expectations, and heightened accountability.

Against this background, there is a need for a performance evaluation framework that recognizes causal conditions related to technological and professional capacity, contextual conditions associated with organizational and regulatory infrastructure, intervening conditions such as resistance and oversight pressures, a core evaluative phenomenon centered on professional judgment, strategies for implementing a balanced evaluation system, and performance-related outcomes. Empirical testing of such a model can clarify which dimensions have the greatest influence, whether the proposed constructs demonstrate acceptable measurement properties, and how the structural relationships among them operate in practice. Such validation is important not only for academic model development but also for managerial decision-making because a

scientifically grounded evaluation system can help tax administrations align employee incentives with the demands of digital transformation, enhance implementation quality, and reduce the gap between technological capacity and actual organizational performance.

Accordingly, the aim of this study was to empirically validate and test a human performance evaluation model in the context of implementing new tax laws and to examine the structural relationships among its causal conditions, contextual conditions, intervening conditions, core phenomenon, strategies, and outcomes.

Methodology

This study adopted a positivist paradigm and a deductive research approach and was designed to empirically validate and test a previously developed conceptual model of human performance evaluation in the context of implementing new tax legislation. In terms of research purpose, the study was applied, as it sought to provide empirically supported evidence that could be used to improve human performance management and the implementation of modern tax policies within the Tax Administration. With respect to methodological strategy, the research employed a quantitative, descriptive-correlational, cross-sectional design based on structural equation modeling. No exploratory, qualitative, or theory-generating phase was incorporated into the present study. Instead, the hypothesized relationships among the latent constructs were specified a priori on the basis of the theoretical framework and previous empirical literature and were subsequently subjected to quantitative testing. The study specifically examined the structural relationships among the principal dimensions of the proposed human performance evaluation model, including organizational and contextual conditions, human and professional competencies, implementation strategies, and performance-related outcomes associated with the execution of new tax laws.

The statistical population consisted of professional employees, senior tax auditors, and middle-level managers directly involved in modern taxation processes within the Tax Administration. Eligibility for participation required direct professional involvement in tax assessment, auditing, implementation, supervision, management, or other operational activities associated with the new tax system. Because the target population was characterized by both hierarchical diversity and organizational distribution, proportionate stratified random sampling was applied. Relevant occupational and organizational strata were represented in the sample proportionally to their size in the target population, thereby reducing within-group sampling variability and improving the representativeness and external validity of the findings. Sample-size adequacy was determined on the basis of statistical power analysis, which is recommended for partial least squares structural equation modeling rather than reliance exclusively on simplified rules of thumb (Hair et al., 2022). Considering a maximum of eight structural paths directed toward an endogenous construct, a medium effect size of ($f^2 = 0.15$), a Type I error probability of ($\alpha = .05$), and statistical power of approximately ($1-\beta = .80$), the minimum required sample size was estimated at 119 participants. To compensate for potential nonresponse, incomplete responses, and unusable questionnaires, the instrument was distributed to a larger number of eligible respondents. Following data screening and exclusion of incomplete or invalid questionnaires, 170 valid questionnaires were retained and entered into the final analyses. The final sample therefore exceeded the minimum sample-size requirement and provided adequate statistical power for estimating the proposed PLS-SEM model.

Quantitative data were collected using a structured, researcher-developed questionnaire designed specifically to operationalize the latent constructs incorporated into the conceptual model. Development of the instrument was grounded

in a systematic examination of the theoretical and empirical literature concerning human performance evaluation, organizational adaptability, implementation of modern tax regulations, digital and analytical competencies, organizational resilience, performance agility, and the quality of tax-law implementation. Items identified from the relevant literature were adapted to the organizational and regulatory context of the Tax Administration to ensure conceptual relevance and contextual appropriateness. The final questionnaire consisted of closed-ended statements measured on a five-point Likert scale ranging from 1 = strongly disagree to 5 = strongly agree. Each group of items was designed to serve as an observable indicator of a theoretically specified latent construct rather than as an exploratory set of variables. Accordingly, the measurement instrument was developed to permit empirical assessment of the relationships among human capabilities, organizational and contextual conditions, strategic responses, and performance outcomes under the requirements imposed by the new tax legislation.

The psychometric adequacy of the instrument was evaluated through content validity, convergent validity, discriminant validity, and internal consistency reliability. Prior to the main data collection, the content relevance, representativeness, clarity, and appropriateness of the questionnaire items were examined by a panel of subject-matter experts familiar with taxation, human resource management, organizational performance, and quantitative management research. Content validity was assessed using the content validity ratio (CVR) proposed by Lawshe (1975) and the content validity index (CVI), with a CVI value greater than .79 considered indicative of satisfactory item-level content validity (Polit & Beck, 2006). Measurement-model validation was subsequently conducted according to contemporary PLS-SEM recommendations (Hair et al., 2019). Indicator reliability was examined through the standardized outer loadings of the measurement items, with loadings of approximately .708 or higher regarded as desirable because they indicate that the construct explains at least 50% of the indicator variance. Convergent validity was assessed using the average variance extracted (AVE), with values of .50 or higher considered acceptable (Fornell & Larcker, 1981). Discriminant validity was primarily evaluated using the heterotrait-monotrait ratio of correlations (HTMT), and values below the conservative threshold of .85 were interpreted as evidence that the latent constructs possessed adequate empirical distinctiveness (Henseler et al., 2015). Internal consistency reliability was assessed using Cronbach's alpha ((\alpha)), composite reliability (CR), and Dijkstra–Henseler's rho ((\rho_A)), with coefficients of .70 or greater considered satisfactory indicators of construct reliability (Dijkstra & Henseler, 2015). Taken together, these procedures were intended to ensure that the questionnaire provided reliable and valid measurement of the theoretical constructs before interpretation of the structural relationships.

Data analysis was conducted at both descriptive and inferential levels. Preliminary screening and descriptive analyses were performed using IBM SPSS Statistics version 28. Descriptive statistics, including the mean, standard deviation, skewness, and kurtosis, were calculated to characterize the distribution of responses and to identify potential irregularities in the dataset. Skewness and kurtosis coefficients were inspected as indicators of univariate distributional characteristics, while the dataset was also screened for incomplete responses, coding errors, and potentially problematic response patterns before estimation of the structural model. In addition to descriptive analysis, one-sample *t*-tests were used where appropriate to determine whether observed mean scores differed significantly from the theoretical midpoint of the five-point measurement scale. The Friedman test was applied to compare and rank the relative importance of the principal dimensions of the human performance evaluation model on the basis of respondents' assessments. These supplementary statistical procedures

provided information regarding the empirical status and relative priority of the model dimensions beyond the examination of their structural relationships.

Hypothesis testing and validation of the conceptual model were performed using partial least squares structural equation modeling (PLS-SEM) in SmartPLS version 4. The analysis followed the two-stage procedure commonly associated with the Anderson and Gerbing framework, whereby the measurement model was evaluated before interpretation of the structural model. In the first stage, indicator loadings, internal consistency reliability, convergent validity, and discriminant validity were assessed to establish the adequacy of construct measurement. In the second stage, the structural model was examined through standardized path coefficients ((\beta)), coefficients of determination ((R^2)), effect sizes ((f^2)), predictive relevance measures where applicable, and model-fit indicators. Collinearity among predictor constructs was evaluated using the variance inflation factor (VIF) to ensure that structural estimates were not distorted by excessive multicollinearity. The explanatory power of the model was assessed primarily through (R^2) values for endogenous constructs, while the magnitude of individual predictor contributions was evaluated using (f^2). Overall model fit was additionally examined using the standardized root mean square residual (SRMR) and normed fit index (NFI). Statistical significance of the hypothesized structural paths was determined through the nonparametric bootstrapping procedure implemented in SmartPLS, from which bootstrap standard errors, *t* statistics, *p* values, and confidence intervals for the standardized path coefficients were obtained. A two-tailed significance level of (\alpha = .05) was used as the principal criterion for hypothesis testing. The combined use of measurement-model assessment, structural-model evaluation, statistical power analysis, and supplementary inferential tests provided a comprehensive empirical basis for evaluating the validity of the proposed human performance evaluation model in the context of new tax-law implementation.

Findings and Results

The demographic profile of the respondents indicated that 61.8% were male and 38.2% were female. Most respondents were married (65.3%). In terms of educational attainment, the most frequent category was a master's degree (50.6%), followed by a bachelor's degree (29.4%) and a doctoral degree (20.0%). The respondents' ages ranged from 25 to 65 years, with the 36–45-year age group representing the largest proportion of the sample and the 56–65-year age group representing the smallest proportion.

At the item level, the descriptive findings showed that the mean score of every item across all constructs was above the theoretical midpoint of 3.00, indicating an overall tendency toward agreement. For the causal conditions construct, the item stating that current organizational procedures rely more on system-based data analysis than on physical document review had a mean of 3.89; the item stating that intelligent systems play a decisive role in the organization's tax auditing process had a mean of 4.06; the item indicating that the organization's auditors possess sufficient skills to work with new digital tools had the highest mean in this construct (4.10); the item concerning the shift from traditional auditing to data-driven analytics had the lowest mean in the construct (3.58); the item referring to big-data analysis as an essential competency had a mean of 3.96; and the item on the adequacy of current training programs for enhancing software and analytical skills had a mean of 3.98. Overall, these results indicated that although respondents perceived digital competence among auditors as a clear strength, the actual transition from traditional auditing to data-driven analytical practice was viewed less favorably than the other dimensions of causal conditions.

For intervening conditions, all items also scored above 3.00. The item stating that employees usually resist replacing traditional methods with new systems had a mean of 3.89. The item indicating that the complexity of intelligent tax systems creates confusion for users during task performance had the lowest mean in this construct (3.78). The item referring to difficulty in understanding the logic of information processing in new systems had a mean of 3.99. The statement that some employees exploit shortcomings in the new systems to reduce their workload had a mean of 3.82. The item reporting that loopholes in the new systems have made it possible to circumvent certain rules had the highest mean in this construct (4.33). The item suggesting that fear of accountability by oversight bodies prevents employees from making independent decisions had a mean of 4.31. The item indicating that auditors act in a highly conservative manner in their reports because of fear of legal consequences had a mean of 3.98, and the statement that strict supervision has deprived employees of the courage to exercise professional judgment had a mean of 4.02. These findings suggest that respondents perceived the major intervening barriers to be legal and procedural pressures, loopholes in newly introduced systems, and the discouraging effects of strict monitoring on professional autonomy.

The contextual conditions items also all exceeded the midpoint. The statement that current quantitative indicators, such as the number of processed cases, are not appropriate criteria for assessing the quality of an auditor's work had a mean of 4.21. The item stating that the current evaluation system encourages employees more toward producing statistics than toward performing high-quality work had a mean of 4.11. The item indicating that the new tax laws and circulars possess sufficient clarity for accurate implementation in the system had a mean of 4.02. The statement that different information systems in the organization are well connected and integrated had the lowest mean in this construct (3.90). The item indicating that the speed and accuracy of the organization's current hardware and networks adequately meet the requirements of intelligent systems had a mean of 4.28. The item stating that contradictions in some regulations disrupt the data-driven assessment process had the highest mean in this construct (4.35). Thus, respondents largely agreed that the contextual environment plays a major role, especially through the adequacy of infrastructure and the disruptive consequences of regulatory inconsistencies.

Within the core phenomenon construct, all three items were above the midpoint as well. The item stating that, in the evaluation system, the quality of a tax auditor's reasoning and professional judgment in complex tax cases should be the main basis for allocating performance scores had a mean of 4.05. The item stating that the tax auditor's success in identifying and correcting false alerts or computational errors in intelligent tax systems should be considered a positive indicator in performance evaluation had the highest mean in the construct (4.38). The item indicating that rapidly closing minor cases solely to increase performance statistics should directly lead to score deductions in the final performance evaluation had the lowest mean in the construct (3.79). These results show strong support for a professional-judgment-based evaluation logic, particularly when auditors meaningfully improve the outputs of intelligent systems.

For the strategies construct, all items again had means above 3.00. The item stating that the final evaluation score of a tax auditor should be derived from a composite function of quantitative indicators, such as the number of cases, and qualitative indicators, such as diagnostic accuracy, had the highest mean in the construct (4.22). The item proposing that part of the annual performance score should be based on expert peer review in major cases had the lowest mean in the construct (3.75). The item stating that taxpayers' survey-based feedback regarding procedural justice and professional conduct should have a specific weight in the auditor's evaluation had a mean of 3.86. The item indicating that senior managers support employees'

expert and evidence-based decisions in cases of disagreement had a mean of 3.99. The statement that the organization has given auditors sufficient autonomy to reject or confirm the outputs of intelligent systems had a mean of 3.91. The item suggesting that the rate at which auditors' decisions are not overturned in tax dispute resolution boards should be one of the main criteria for performance scoring and promotion had a mean of 3.98. Finally, the item indicating that the organizational structure requires fundamental reform in order to accept professional judgment instead of absolute dependence on machines had a mean of 4.05. Collectively, these findings indicate broad support for hybrid and reform-oriented strategies combining quantitative and qualitative indicators, professional review, taxpayer feedback, and organizational reform.

The outcomes construct also showed mean scores above the midpoint for all items. The item stating that the establishment of an evaluation system based on professional judgment would guarantee employees' job security vis-à-vis oversight institutions had the lowest mean in this construct (3.81). The item indicating that the new approach would enable auditors to issue decisions with greater psychological and legal confidence had a mean of 4.07. The item stating that proper use of the evaluation model would reduce the rate at which tax decisions are overturned in tax dispute resolution boards had a mean of 4.01. The statement that combining human expertise with system precision increases consistency in tax decisions had a mean of 4.08. The item indicating that continuous interaction between auditors and intelligent systems has enhanced mutual knowledge and organizational learning had the highest mean in the construct (4.14). The item stating that feedback derived from employees' professional judgments contributes to improving and correcting system errors had a mean of 4.05. Overall, the respondents believed that the model would produce favorable organizational and professional consequences, particularly in terms of learning, decision consistency, and more confident decision-making.

The descriptive statistics of the main study variables are presented in Table 1. As shown, the mean values of all constructs were higher than the midpoint of the scale, indicating relatively high levels of agreement with all dimensions of the model. Among the six variables, contextual conditions had the highest mean (4.15), followed by the core phenomenon (4.07), whereas causal conditions had the lowest mean (3.93).

Table 1

Descriptive Statistics of the Main Study Variables

Variable	Mean	SD	Minimum	Maximum
Causal conditions	3.93	0.49	2.83	5.00
Contextual conditions	4.15	0.38	3.00	5.00
Intervening conditions	4.01	0.39	2.63	5.00
Core phenomenon	4.07	0.42	3.00	5.00
Strategies	3.97	0.39	2.86	5.00
Outcomes	4.02	0.44	2.83	5.00

The normality of the variable distributions was then assessed using skewness and kurtosis coefficients. The results presented in Table 2 show that all skewness and kurtosis values fell within the acceptable range of -2 to +2, and most were even within the more conservative range of -1 to +1. Accordingly, the univariate distributions of the study variables did not deviate meaningfully from normality, and the use of parametric procedures, including Pearson correlation and one-sample *t* tests, was considered appropriate.

Table 2

Skewness and Kurtosis of the Main Study Variables

Variable	Skewness	Kurtosis
Causal conditions	0.051	-0.220
Contextual conditions	-0.281	0.501
Intervening conditions	-0.466	1.360
Core phenomenon	-0.109	-0.239
Strategies	-0.099	1.100
Outcomes	-0.719	0.691

At the inferential level, the measurement model was first evaluated by confirmatory factor analysis using SmartPLS. Factor loadings, *t* values, average variance extracted, composite reliability, Cronbach's alpha, and rho were examined. The minimum acceptable factor loading for retaining an item was set at 0.50. The results presented in Table 3 indicate that all items met this criterion. The lowest loading belonged to Item 32 (0.52), and all *t* values were greater than 1.96, indicating statistically significant loadings at the 95% confidence level. The AVE values ranged from 0.52 to 0.57, confirming satisfactory convergent validity. Cronbach's alpha ranged from 0.67 to 0.88, composite reliability ranged from 0.79 to 0.90, and rho ranged from 0.71 to 0.88, all of which supported the internal consistency and reliability of the instrument. In summary, the results supported both the construct validity and reliability of the questionnaire.

Table 3

Confirmatory Factor Analysis Results: Validity and Reliability of the Study Variables

Construct	Item	Standardized Loading	*t*	AVE	CR	Cronbach's α	rho
Causal conditions	1	0.77	13.95	0.54	0.88	0.83	0.86
	2	0.74	14.91				
	3	0.67	10.11				
	4	0.58	8.21				
	5	0.81	18.75				
	6	0.83	23.16				
Intervening conditions	7	0.73	21.12	0.52	0.90	0.88	0.87
	8	0.77	22.71				
	9	0.76	18.57				
	10	0.77	18.35				
	11	0.72	14.06				
	12	0.69	13.90				
	13	0.65	8.91				
	14	0.65	9.66				
Contextual conditions	15	0.73	18.32	0.52	0.87	0.82	0.82
	16	0.68	14.04				
	17	0.67	13.76				
	18	0.73	18.31				
	19	0.77	22.19				
	20	0.76	18.59				
Core phenomenon	21	0.70	13.23	0.56	0.79	0.67	0.71
	22	0.80	16.18				
	23	0.75	19.96				
Strategies	24	0.58	8.69	0.57	0.90	0.87	0.88
	25	0.78	25.22				
	26	0.78	20.71				
	27	0.71	13.42				
	28	0.78	17.62				
	29	0.83	24.58				
	30	0.79	24.48				
Outcomes	31	0.68	15.04	0.53	0.87	0.82	0.83
	32	0.52	5.07				
	33	0.74	16.06				
	34	0.81	23.70				
	35	0.81	26.85				
	36	0.77	18.64				

The Pearson correlation matrix is reported in Table 4. All correlations among the study variables were positive and statistically significant. The correlation coefficients ranged from 0.44 to 0.73, indicating moderate to strong relationships. The weakest association was between causal conditions and the core phenomenon (r = 0.44), whereas the strongest was between contextual conditions and intervening conditions (r = 0.73). With respect to the dependent variable, outcomes showed the strongest correlation with strategies (r = 0.70), followed closely by intervening conditions (r = 0.69), contextual conditions (r = 0.66), causal conditions (r = 0.61), and the core phenomenon (r = 0.58).

Table 4

Pearson Correlation Matrix of the Study Variables

Variable	1	2	3	4	5	6
1. Causal conditions	1.00					
2. Contextual conditions	0.57***	1.00				
3. Intervening conditions	0.66***	0.73***	1.00			
4. Core phenomenon	0.44***	0.59***	0.58***	1.00		
5. Strategies	0.61***	0.65***	0.63***	0.66***	1.00	
6. Outcomes	0.61***	0.66***	0.69***	0.58***	0.70***	1.00

The conceptual model was then tested using partial least squares structural equation modeling. The statistical significance version of the model showed that all paths among the main variables were statistically significant, because all path-specific t values exceeded 1.96. The measurement component of the model was likewise acceptable because all item loadings remained above 0.50 and all loading t values were significant. T

Figure 1

Research model in the t-value state (statistical significance)

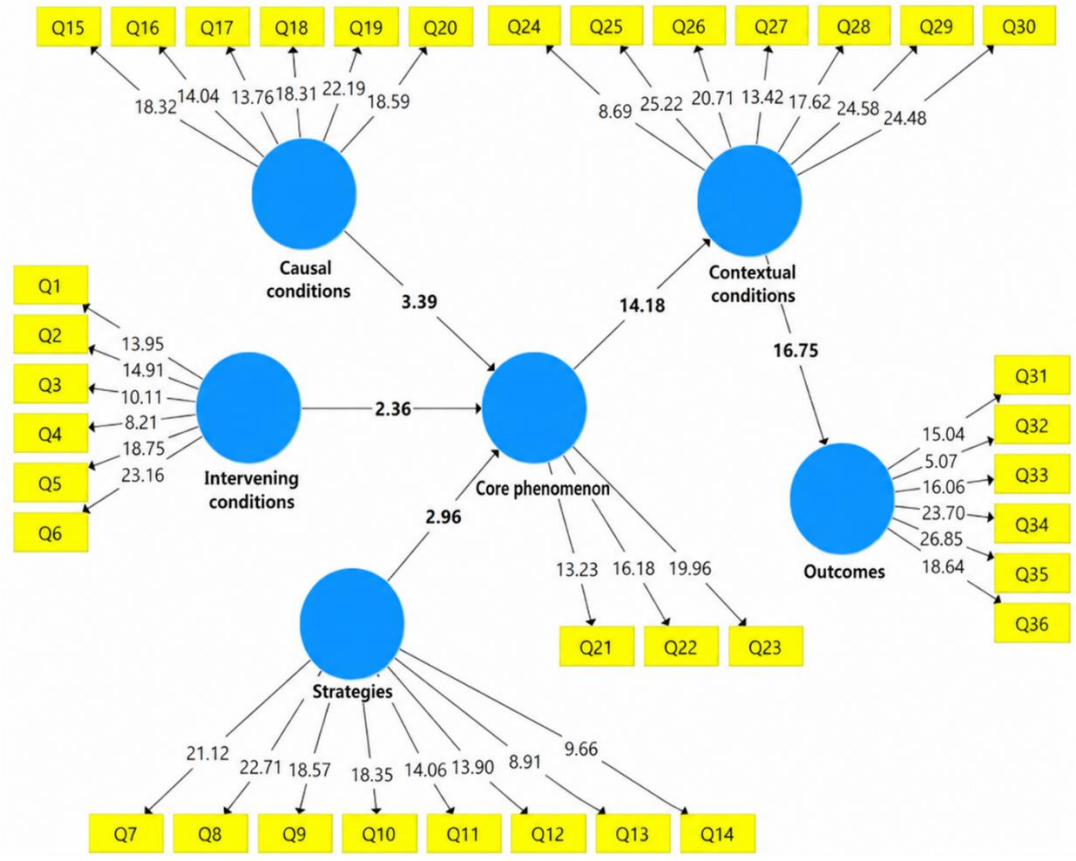
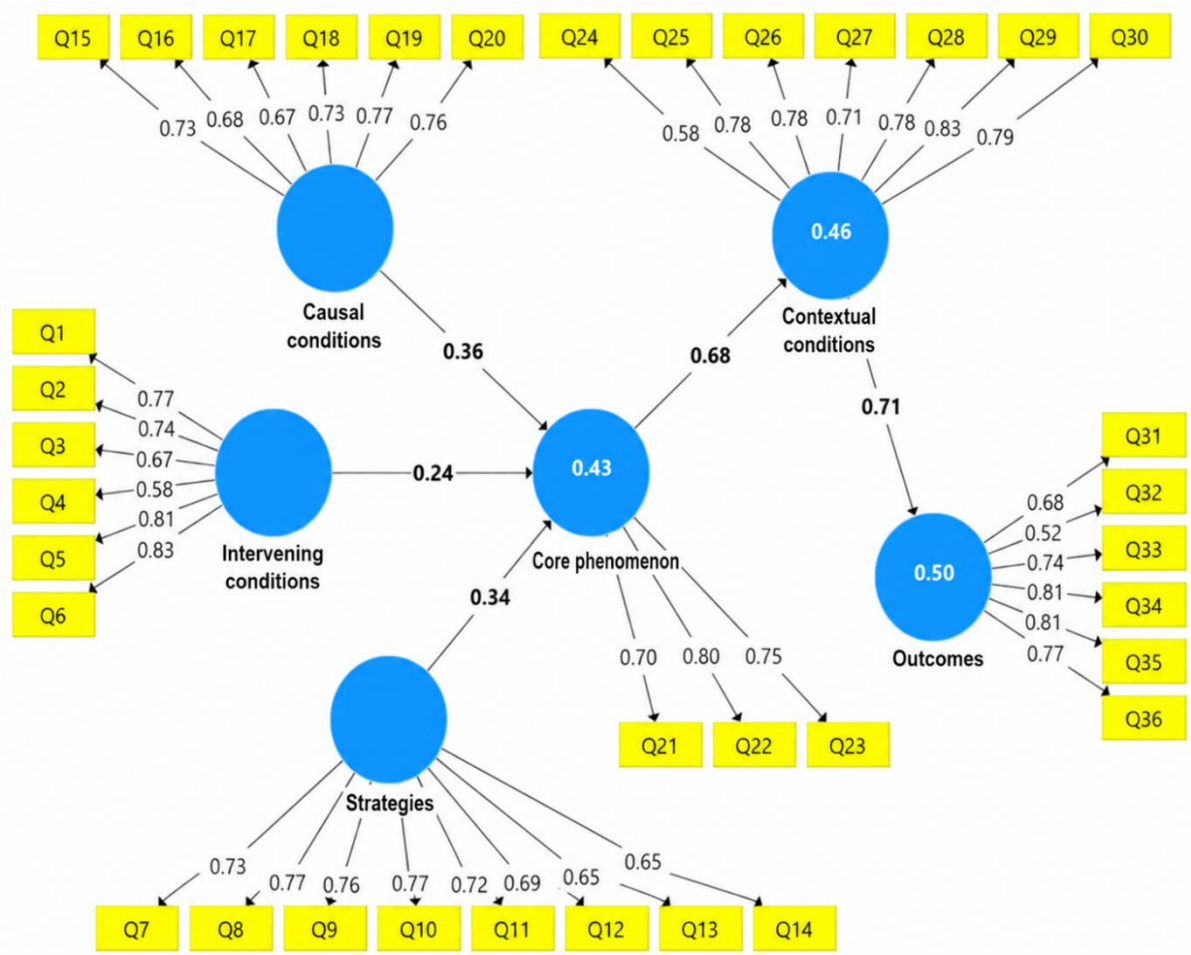


Figure 2

Research model in the standardized path coefficient state



The standardized-coefficient version of the model further showed that all relationships were positive. Among the exogenous predictors of the core phenomenon, contextual conditions had the strongest effect (β = 0.36), followed by intervening conditions (β = 0.34) and causal conditions (β = 0.24). The effect of the core phenomenon on strategies was strong (β = 0.68), and the effect of strategies on outcomes was even stronger (β = 0.71).

The model-fit indices are presented in Table 5. The normed fit index (NFI) was 0.90, indicating a borderline yet acceptable level of overall fit. The standardized root mean square residual (SRMR) was 0.075, which is below the threshold of 0.08 and therefore indicates satisfactory fit. The coefficients of determination were 0.43 for the core phenomenon, 0.46 for strategies, and 0.50 for outcomes, demonstrating acceptable explanatory power. The Q^2 values were all positive and above zero, showing that the model possessed satisfactory predictive relevance. Overall, the fit indices supported the adequacy of the empirical model.

Table 5

Model Fit and Predictive Power Indices

Endogenous Variable	Q^2	R^2	NFI	SRMR
Core phenomenon	0.22	0.43	0.90	0.075
Strategies	0.23	0.46	0.90	0.075
Outcomes	0.24	0.50	0.90	0.075

The structural path coefficients are summarized in Table 6. All five hypothesized paths were supported. Contextual conditions exerted a positive and statistically significant effect on the core phenomenon (β = 0.36, *t* = 3.39, *p* = .001). Causal conditions also had a positive and significant effect on the core phenomenon (β = 0.24, *t* = 2.36, *p* = .019). Intervening conditions positively affected the core phenomenon as well (β = 0.34, *t* = 2.96, *p* = .003). The effect of the core phenomenon on strategies was strong and highly significant (β = 0.68, *t* = 14.18, *p* < .001), and the effect of strategies on outcomes was the strongest direct path in the model (β = 0.71, *t* = 16.75, *p* < .001).

Table 6

Structural Equation Modeling Results for the Relationships Among Variables

Path	Standardized Coefficient (β)	SE	*t*	*p*
Contextual conditions → Core phenomenon	0.36	0.105	3.39	.001
Causal conditions → Core phenomenon	0.24	0.102	2.36	.019
Intervening conditions → Core phenomenon	0.34	0.114	2.96	.003
Core phenomenon → Strategies	0.68	0.048	14.18	< .001
Strategies → Outcomes	0.71	0.042	16.75	< .001

To compare the relative standing of the six study variables, the Friedman ranking test was performed. The test was statistically significant, $\chi^2(5) = 84.29$, $p < .001$, indicating a significant difference in the ranks of the variables. As shown in Table 7, contextual conditions received the highest mean rank (4.39), followed by the core phenomenon (3.85) and outcomes (3.51). Causal conditions received the lowest mean rank (2.85). These findings indicate that respondents assigned the greatest relative importance to contextual conditions and the least to causal conditions.

Table 7

Friedman Test Results for Ranking the Study Variables

Rank	Variable	Mean Rank	Arithmetic Mean
1	Contextual conditions	4.39	4.15
2	Core phenomenon	3.85	4.07
3	Outcomes	3.51	4.02
4	Intervening conditions	3.36	4.01
5	Strategies	3.04	3.97
6	Causal conditions	2.85	3.93

Finally, one-sample *t* tests were conducted to compare the mean of each variable against the theoretical midpoint of 3.00. The results in Table 8 show that the mean score of every study variable was significantly above the midpoint, with *p* values lower than .001 in all cases. Therefore, all variables can be interpreted as being at a relatively high level. The highest mean belonged to contextual conditions (4.15), followed by the core phenomenon (4.07), while the lowest mean belonged to causal conditions (3.93). Although all variables were above average, the differences among their means were not large, suggesting that respondents expressed relatively high agreement with all dimensions of the model.

Table 8

One-Sample t Test for the Current Status of the Study Variables

Variable	Test Value	Mean	Mean Difference	*t*	*p*	Importance Level
Causal conditions	3.00	3.93	0.93	24.94	< .001	High
Contextual conditions	3.00	4.15	1.15	39.16	< .001	High
Intervening conditions	3.00	4.01	1.01	33.71	< .001	High
Core phenomenon	3.00	4.07	1.07	33.51	< .001	High
Strategies	3.00	3.97	0.97	32.75	< .001	High
Outcomes	3.00	4.02	1.02	30.63	< .001	High

In summary, the findings showed that respondents reported relatively high agreement with all components of the proposed model. At the descriptive level, contextual conditions emerged as the most highly rated construct, whereas causal conditions had the lowest mean, although still above the scale midpoint. At the measurement level, the questionnaire demonstrated satisfactory validity and reliability. At the relational level, all study variables were positively and significantly correlated, and the structural model supported all hypothesized paths. The strongest direct path was from strategies to outcomes, while among the antecedents of the core phenomenon, contextual conditions exerted the largest effect. The Friedman and one-sample *t* test results further confirmed the salience of contextual conditions and the generally high standing of all variables in the respondents' evaluations.

Discussion and Conclusion

The present study empirically validated and tested a human performance evaluation model in the context of implementing new tax laws and demonstrated that the proposed model had acceptable psychometric properties, satisfactory explanatory power, and statistically significant structural relationships among its principal constructs. The descriptive findings showed that all six dimensions of the model—causal conditions, contextual conditions, intervening conditions, the core phenomenon, strategies, and outcomes—had mean scores above the theoretical midpoint of 3.00, indicating a generally favorable assessment by respondents. Among these constructs, contextual conditions had the highest mean score (M = 4.15), followed by the core phenomenon (M = 4.07), outcomes (M = 4.02), intervening conditions (M = 4.01), strategies (M = 3.97), and causal conditions (M = 3.93). The Friedman ranking test confirmed that these differences were statistically significant, with contextual conditions obtaining the highest mean rank (4.39) and causal conditions the lowest (2.85). This pattern indicates that employees and managers regarded the institutional, regulatory, infrastructural, and organizational environment as the most salient component affecting human performance under the new taxation regime. This result is consistent with the systemic perspective proposed by Mansor and Tayib, who argued that tax administration performance cannot be understood solely through individual-level indicators because it is shaped by the interaction among organizational structures, systems, environmental inputs, and management processes [12]. It is also consistent with Aribowo et al., who described tax reform as a broad administrative transformation requiring alignment among institutional structures, administrative capacity, and implementation mechanisms rather than merely the enactment of new rules [5].

At the item level, one of the most important findings concerned causal conditions. Auditors' digital skills received the highest mean score within this construct (M = 4.10), whereas the item reflecting the organization's transition from traditional auditing to data-driven analytics received the lowest mean (M = 3.58). This discrepancy suggests that individual digital capability has developed more rapidly than the organizational transformation required to fully utilize that capability. In other words, employees appear relatively prepared to work with digital tools, but the institutional architecture and work processes have not yet completely shifted toward a genuinely data-driven model. This finding closely aligns with Najari and Ahangari's emphasis on organizational e-readiness, according to which effective digital transformation within the Iranian National Tax Administration depends not only on technological infrastructure but also on managerial readiness, organizational processes, and human capability [8]. The finding is also compatible with the broader Tax Administration 3.0 framework, which conceptualizes digital transformation as an integrated redesign of tax administration processes rather than the simple

introduction of new technologies [1]. Sonnerfeldt similarly argues that tax technologies change the nature of professional expertise and institutional legitimacy because employees must operate within increasingly technologically mediated systems of judgment and decision-making [2]. Therefore, the gap observed in the present study between individual digital competence and organizational transition can be interpreted as evidence that technological modernization is advancing faster than the redesign of organizational routines, performance systems, and decision structures.

The results concerning intervening conditions further demonstrated that the implementation environment includes significant behavioral, procedural, and supervisory constraints. The highest-scoring item in this dimension indicated that loopholes in new systems may make it possible to circumvent certain rules (M = 4.33), while fear of accountability by oversight bodies also received a very high mean (M = 4.31). In addition, respondents agreed that strict supervision can restrict employees' willingness to exercise professional judgment. These findings suggest that the digitalization of tax administration does not automatically eliminate implementation problems; rather, it may generate new forms of operational ambiguity, strategic behavior, and risk aversion. Purnomo and Gunadi emphasize that tax policy implementation is influenced by the practical operation of enforcement mechanisms and the interaction among institutional actors, which may either support or hinder effective implementation [6]. The present findings also resonate with Leontyev and Ketners, who showed that tax decision-making frequently involves tension among reliability, equity, and efficiency, especially when administrators must balance formal procedures against contextual judgment [14]. Accordingly, excessive supervisory pressure may unintentionally encourage highly conservative behavior, reducing the willingness of auditors to apply informed discretion even when such discretion is necessary to correct system errors or respond to exceptional circumstances.

Contextual conditions emerged as both the highest-rated construct and the strongest predictor of the core phenomenon. The structural model showed that contextual conditions had the largest direct effect on the core phenomenon ($\beta = 0.36$, $p = .001$), followed by intervening conditions ($\beta = 0.34$, $p = .003$) and causal conditions ($\beta = 0.24$, $p = .019$). At the item level, the strongest contextual concern was the existence of contradictions in regulations that disrupt data-driven assessment processes (M = 4.35), while respondents also expressed strong concern about the adequacy of current quantitative evaluation indicators and the tendency of existing systems to emphasize numerical output rather than work quality. These findings reinforce the argument that human performance in modern tax administration cannot be separated from regulatory coherence and organizational design. Hashemian et al. highlighted the practical implementation challenges associated with tax legislation and showed that legal reforms may generate operational difficulties when administrative procedures and institutional capacities are not sufficiently aligned [9]. Dashti likewise demonstrated that new tax laws can produce substantial consequences for accounting and financial systems, requiring administrative structures to adapt to legislative change [7]. Thus, the present finding that contextual conditions exerted the strongest influence on the core evaluative phenomenon indicates that improvements in individual competence alone are insufficient unless regulatory, technological, and organizational environments are simultaneously reengineered.

The core phenomenon was centered on the recognition of professional judgment and the quality of auditors' analytical reasoning. The highest-scoring item within this construct concerned the inclusion of auditors' success in identifying and correcting false alerts or computational errors in intelligent tax systems as a positive component of performance evaluation (M = 4.38). Respondents also supported emphasizing the quality of professional reasoning in complex cases. This result is highly relevant in contemporary digital tax administration because it reflects a shift from a purely compliance-based

understanding of employee performance toward a hybrid model in which human expertise complements automated systems. Belahouaoui and Attak's review of digital taxation and artificial intelligence indicates that intelligent technologies increasingly shape compliance and administrative processes but continue to require human interpretation and governance [3]. Similarly, Sonnerfeldt emphasizes that technological systems do not eliminate expertise; instead, they alter the location and expression of expertise within tax administration [2]. The present finding therefore suggests that a valid human performance model should reward not only conformity with system outputs but also the capacity to identify system weaknesses, exercise defensible professional judgment, and improve the quality of automated decision support.

The strong structural effect of the core phenomenon on strategies ($\beta = 0.68$, $p < .001$) further indicates that when professional judgment is accepted as a central principle of performance evaluation, it translates into specific managerial and evaluative strategies. Respondents strongly supported the use of a composite performance score combining quantitative indicators, such as the number of processed cases, with qualitative indicators, such as diagnostic accuracy ($M = 4.22$). They also endorsed structural reform aimed at reducing excessive dependence on automated systems. These findings are consistent with Vaezi et al., who emphasized that performance evaluation in the national tax system should incorporate ethical and qualitative dimensions rather than relying exclusively on numerical measures [13]. They also support the integrated performance-management approach proposed by Mansor and Tayib, in which tax administration performance is assessed through multiple interrelated dimensions rather than narrow productivity indicators [12]. From a behavioral perspective, this finding is also consistent with Kirchler's argument that tax administration effectiveness is closely associated with fairness, trust, and the quality of institutional interactions, not simply deterrence or output maximization [10]. Therefore, a more balanced performance system may not only improve internal management but also contribute to more legitimate and procedurally fair tax administration.

The strongest direct structural relationship in the model was the effect of strategies on outcomes ($\beta = 0.71$, $p < .001$). This finding indicates that the practical design and implementation of evaluation strategies are decisive in determining favorable organizational consequences. The outcomes construct included greater legal and psychological confidence among auditors, lower rates of overturned tax decisions, greater consistency in decision-making, enhanced organizational learning, and improved correction of system errors. Among these items, the highest mean concerned continuous interaction between auditors and intelligent systems as a mechanism for organizational learning ($M = 4.14$). This result suggests that the relationship between humans and intelligent systems should be conceptualized as iterative rather than substitutive: auditors learn from system outputs while the system is improved through feedback derived from professional judgment. This interpretation is consistent with the Tax Administration 3.0 perspective, which emphasizes integration, continuous information flows, and digitally enabled administrative processes [1]. It also corresponds with Belahouaoui and Attak's emphasis on the transformative role of artificial intelligence in tax administration and the importance of integrating technological capabilities into compliance management [3]. The strong strategies-to-outcomes path therefore demonstrates that organizational benefits are not produced automatically by technology; rather, they are mediated through appropriate performance-management strategies.

The correlation analysis also supported the structural findings. All study variables were positively and significantly related, with coefficients ranging from 0.44 to 0.73. Strategies had the strongest correlation with outcomes ($r = 0.70$), followed closely by intervening conditions ($r = 0.69$), while contextual conditions were also strongly related to outcomes ($r = 0.66$). These

associations reinforce the systemic character of the model and suggest that improvements in performance outcomes require coordinated changes across multiple organizational dimensions. The finding is particularly consistent with Alm's argument that technological development can strengthen tax enforcement and reduce evasion only when institutional capacity and administrative systems are able to effectively use new information and technologies [4]. It is likewise compatible with Mohammed and Hammadi's observation that tax outcomes are affected by how legal provisions interact with taxpayers' circumstances and administrative implementation [15]. The implication is that high-quality human performance requires alignment among individual competency, organizational infrastructure, regulatory clarity, evaluative mechanisms, and implementation strategies.

The overall model demonstrated acceptable empirical performance. The measurement model showed satisfactory convergent validity, with AVE values ranging from 0.52 to 0.57, while composite reliability ranged from 0.79 to 0.90 and rho coefficients ranged from 0.71 to 0.88. All factor loadings exceeded 0.50 and were statistically significant. The structural model also showed satisfactory fit, with SRMR = 0.075 and NFI = 0.90. Furthermore, the model explained 43% of the variance in the core phenomenon, 46% of the variance in strategies, and 50% of the variance in outcomes, while all Q^2 values were positive. These results indicate that the proposed model has both measurement adequacy and meaningful explanatory capacity. The significance of tax culture and organizational behavior emphasized by Ehteshami and Abdi provides additional support for this multidimensional interpretation, because the effectiveness of tax systems depends not only on formal legislation but also on the broader institutional and behavioral environment in which tax administration operates [11]. Taken together, the findings show that an effective human performance evaluation system under new tax laws must move beyond conventional output metrics and incorporate technological competence, contextual support, professional judgment, balanced evaluation strategies, and learning-oriented outcomes.

This study was subject to several limitations that should be considered when interpreting the findings. First, the research employed a cross-sectional design, meaning that all variables were measured at a single point in time; therefore, although the structural model was theoretically specified in causal terms, temporal causality cannot be definitively established. Second, the data were based on self-reported questionnaire responses, which may be influenced by social desirability, common-method variance, or respondents' subjective interpretations of organizational conditions. Third, although the sample size was statistically adequate for PLS-SEM, the study was limited to employees, senior auditors, and middle managers within the Tax Administration, and the findings may not fully represent other stakeholder groups such as taxpayers, external auditors, dispute-resolution authorities, or regulatory bodies. Fourth, the study focused on a specific organizational and legal context, and some constructs—particularly those related to regulatory ambiguity, oversight pressure, and system maturity—may operate differently in other jurisdictions or administrative environments. Finally, despite the acceptable explanatory power of the model, approximately half of the variance in the final outcome construct remained unexplained, indicating that other organizational, technological, behavioral, or institutional determinants may also be relevant.

Future studies should examine the proposed model longitudinally in order to determine how the relationships among technological adaptation, professional judgment, evaluation strategies, and performance outcomes evolve as new tax laws and intelligent systems become more institutionalized. Researchers could also employ multi-source designs by combining employee assessments with objective indicators such as audit accuracy, processing time, appeal rates, frequency of overturned decisions, taxpayer satisfaction, system-error correction records, and training performance. Comparative studies

across different regional tax offices, organizational levels, or national tax systems would be useful for evaluating the external validity and contextual stability of the model. Future research could also investigate potential mediating and moderating mechanisms, including organizational trust, digital literacy, perceived supervisory support, psychological safety, technological acceptance, regulatory clarity, and perceived fairness of performance evaluation. Multi-group analyses may be conducted to test whether the model differs according to managerial level, work experience, educational level, or degree of exposure to digital taxation systems. Finally, future studies could extend the model by incorporating technological maturity, data quality, algorithmic transparency, digital ethics, and organizational learning capacity as additional explanatory variables.

Tax administrators should redesign employee performance evaluation systems so that quantitative productivity indicators are supplemented by qualitative measures of professional judgment, diagnostic accuracy, case complexity, procedural fairness, and the ability to critically evaluate intelligent-system outputs. Organizational leaders should prioritize the clarification and harmonization of regulations because regulatory contradictions can weaken the effectiveness of even highly skilled personnel and advanced digital systems. Training programs should move beyond basic software instruction and place greater emphasis on data analytics, interpretation of algorithmic outputs, error detection, legal reasoning, and defensible professional decision-making. Management should also create a supportive environment in which well-reasoned professional judgments are protected from disproportionate supervisory risk, provided that decisions are adequately documented and evidence-based. Performance systems should incorporate structured feedback mechanisms through which auditors can report system errors, identify misleading alerts, and contribute to continuous improvement of intelligent platforms. Finally, the Tax Administration should adopt a balanced evaluation architecture that integrates human expertise and technological precision rather than treating them as competing alternatives, thereby allowing existing digital capabilities to be translated more effectively into higher-quality decisions, organizational learning, and improved implementation of new tax laws.

Acknowledgments

We would like to express our appreciation and gratitude to all those who cooperated in carrying out this study.

Authors' Contributions

All authors equally contributed to this study.

Declaration of Interest

The authors of this article declared no conflict of interest.

Ethical Considerations

The study protocol adhered to the principles outlined in the Helsinki Declaration, which provides guidelines for ethical research involving human participants. Written consent was obtained from all participants in the study.

Transparency of Data

In accordance with the principles of transparency and open research, we declare that all data and materials used in this study are available upon request.

Funding

This research was carried out independently with personal funding and without the financial support of any governmental or private institution or organization.

References

[1] C.-o. Organisation for Economic and Development, "Tax Administration 3.0 and the Digital Transformation of Tax Administration," OECD Publishing, Paris, 2023.

[2] A. Sonnerfeldt, "Navigating the Digital Frontier: The Implications of Tax Technology on Systems of Expertise and Legitimacy in Taxation," in *Taxation in the Digital Era: Economic, Legal, and Policy Challenges*. Cham: Springer Nature Switzerland, 2025, pp. 153-176.

[3] R. Belahouaoui and E. H. Attak, "Digital Taxation, Artificial Intelligence and Tax Administration 3.0: Improving Tax Compliance Behavior-A Systematic Literature Review Using Textometry (2016-2023)," *Accounting Research Journal*, vol. 37, no. 2, pp. 172-191, 2024, doi: 10.1108/ARJ-12-2023-0372.

[4] J. Alm, "Tax Evasion, Technology, and Inequality," *Economics of Governance*, vol. 22, no. 4, pp. 321-343, 2021, doi: 10.1007/s10101-021-00247-w.

[5] I. Aribowo, N. N. Kamilah, J. Kumar, and L. Rofi'ah, "Assessing Tax Reform as a Journey to Tax Administration," *KnE Social Sciences*, pp. 503-524, 2022, doi: 10.18502/kss.v7i5.10574.

[6] C. A. J. Purnomo and G. Gunadi, "Policy Implementation Process and Its Component on Tax Crime Investigation Boundary to Promote an Optimal Tax Revenue: A Literature Review," *Nagari Law Review*, vol. 6, no. 2, pp. 72-85, 2023, doi: 10.25077/nalrev.v.6.i.2.p.72-85.2023.

[7] M. Dashti, "Evaluating the Impact of New Tax Laws on Municipal Accounting and Financial Systems," in *Fourth International Conference on Political Science, Management, Economics, and Accounting*, Hamedan, Iran, 2024. [Online]. Available: https://civilica.com/doc/2189529.

[8] R. Najari and J. Ahangari, "Designing and Explaining an E-Readiness Assessment Model for the Iranian National Tax Administration," *Tax Research Journal*, vol. 24, no. 30, pp. 107-136, 2016. [Online]. Available: http://taxjournal.ir/article-1-917-fa.html.

[9] R. Hashemian, M. Kaski Langroudi, and R. Hashemian, "Examining the Value Added Tax Law and the Challenges of Its Implementation," in *National Conference on Management and Humanities Research in Iran*, 2017: SID. [Online]. Available: https://sid.ir/paper/896049/fa.

[10] E. Kirchler, *The Economic Psychology of Tax Behaviour*. Cambridge University Press, 2007.

[11] S. Ehteshami and F. Abdi, "Promoting Tax Culture and Its Impact on the Tax System," in *International Conference on Modern Research in Civil Engineering, Structure, and Urban Planning*, 2016: SID. [Online]. Available: https://sid.ir/paper/870959/fa.

[12] M. Mansor and M. Tayib, "Integrated and Open Systems Model: An Innovative Approach to Tax Administration Performance Management," *The Innovation Journal*, vol. 18, no. 3, p. 2, 2013.

[13] S. K. Vaezi, A. Afri, A. Nargesiyan, and A. Shams, "Designing a Performance Evaluation Model for the National Tax System with an Emphasis on Ethical Principles," *Public Administration Perspective*, vol. 10, no. 2, pp. 121-158, 2019, doi: 10.48308/jpap.2019.96500.

[14] A. Leontyev and K. Ketners, "The Improvement of Decision-Making in the Latvian Tax System: Cases of Irreducible Incompatibility Taking into Account Reliability, Equity and Efficiency Criteria," *Intellectual Economics*, vol. 17, no. 2, pp. 308-329, 2023.

[15] K. S. Mohammed and S. M. Hammadi, "Legal Exemptions and Their Compatibility with Taxpayers' Incomes and Their Impact on Tax Revenues: Applied Research on a Sample of the Faculty of Human Education," *Academy of Strategic Management Journal*, vol. 20, no. 65, pp. 1-18, 2021.