Management Research Spectrum
ISSN: 2659-1960 (Online)
ISSN: 2659-1871 (Print)

Validation of an Integrated Measurement Scale for GSCM, Industry 4.0, and Organizational Performance in the Energy Sector

Abba Adam[1,*], Halima A. A Yusuf[1], Salisu Alhaji Uba[1], Rajeh Bati Almasradi[2], Alharthi Rami Hashem E[3]

[1] *Department of Business Administration, College of Arts, Management and Social Sciences, Al-Ansar University, Maiduguri, Nigeria.*
[2] *Department of Business Administration, College of Business, University of Bisha, Kingdom of Saudi Arabia.*
[3] *Department of Financial and Administrative Sciences, Ranyah University College, Taif University, Kingdom of Saudi Arabia.*

* Corresponding author: abbadamuna@yahoo.com

Abstract

This study presents the psychometric validation of a newly developed survey instrument designed to evaluate the relationship between Green Supply Chain Management (GSCM), Industry 4.0 integration, and Organizational Performance within the petroleum downstream sector. The instrument was synthesized from established literature and refined through a two-stage qualitative process involving expert review and face validity assessment by industry practitioners. Subsequently, a pilot study was conducted with 40 key personnel from NNPC business units to assess internal consistency, item reliability, and factorial structure. This sample size ($N = 40$) is robust for pre-testing purposes, following the methodological precedents in supply chain research established by Lavaei Adaryani *et al.* (2024). The results of the reliability analysis led to significant scale refinement: the GSCM Practice scale was optimized from an initial 29 items to 27 items, while the Industry 4.0 scale was pruned from 40 items to 34 items following the deletion of items with negative or poor item-total correlations. Exploratory Factor Analysis (EFA) using Principal Axis Factoring confirmed the theoretical dimensionality, with a KMO of 0.842 and 64.85% of total variance explained. Furthermore, Harman's Single-Factor Test indicated that Common Method Bias was not a significant threat (32.45% variance). Post-refinement, all latent constructs achieved Cronbach's alpha and Composite Reliability (CR) coefficients exceeding the 0.70 threshold, with Average Variance Extracted (AVE) values above 0.50. The study concludes that the refined instrument is psychometrically sound, statistically distinct, and suitable for full-scale field administration within the industrial value chain context.

Keywords: *GSCM, Organizational Performance, Industry 4.0, Scale Validation, Pilot Study, Reliability Analysis, Exploratory Factor Analysis.*

Article Information:

Received: 22 April 2026
Revised: 25 June 2026
Accepted: 30 June 2026
Published: 2026

Vol. 16, No. 1, 2026

Introduction

The global push for sustainable development has forced a major reassessment of operational strategies within high-impact industrial sectors. Firms increasingly adopt Green Supply Chain Management (GSCM) practices to harmonize economic growth with environmental stewardship. The emergence of Industry 4.0 technologies—specifically the Internet of Things (IoT) and Cyber-Physical Systems (CPS)—provides a transformative framework for enhancing transparency and precision in these green initiatives. However, scholars and practitioners in the Nigerian petroleum downstream sector face significant challenges due to a lack of psychometrically validated instruments that simultaneously measure the synergy between digital transformation and green practices.

This study fills this methodological gap by developing and validating a comprehensive research instrument tailored for the energy sector value chain. This research follows the developmental sequence for supply chain management (SCM) established by Eng (2004), which emphasizes an intensive expert review phase to ensure content validity. The development process integrates established theoretical frameworks, such as the green service construct protocols of Chan *et al.* (2016) and the Industry 4.0 digital integration architectures proposed by Ghadge *et al.* (2022).

The finalized instrument is structured into three primary sections (A–C) to provide a multidimensional view of industrial operations. Section A evaluates the implementation of Green Supply Chain Management (GSCM) practices across five validated dimensions: internal management, green purchasing, customer cooperation, eco-design, and investment recovery. Section B measures Organizational Performance using the environmental, economic, and operational metrics identified by Yu *et al.* (2014) and Yeung *et al.* (2013). Section C assesses Industry 4.0 application, focusing on the integration of digital systems and cyber-security. This validation study ensures that the resulting scale is psychometrically robust, statistically distinct, and suitable for full-scale field administration in complex industrial environments.

Theoretical Nexus and Rationale for Integration

The integration of Green Supply Chain Management (GSCM), Industry 4.0, and Organizational Performance into a single validation framework is necessitated by the inherent functional interdependencies observed in modern industrial environments. Industry 4.0 technologies—specifically the Internet of Things (IoT) and Cyber-Physical Systems (CPS)—act as critical enablers for green practices by providing the digital infrastructure necessary for real-time monitoring and data transparency. As argued by Ghadge *et al.* (2022), digital integration is a primary driver of green innovation, suggesting that any measurement scale focused on sustainable supply chains is conceptually incomplete without accounting for the underlying technological capabilities that facilitate these initiatives. Furthermore, the ultimate strategic objective of both digital transformation and sustainable sourcing is the optimization of multidimensional Organizational Performance. Validating these scales together acknowledges the synergistic effects that technology and green strategy exert on environmental, economic, and operational outcomes, aligning with the performance frameworks established by Yu *et al.* (2014) and Yeung *et al.* (2013). From a methodological standpoint, this integrated approach ensures that the measurement items are semantically and contextually synchronized for the Nigerian energy sector. Utilizing a singular expert panel and a unified pilot group ($N = 40$) eliminates potential measurement artifacts and cultural biases that typically arise when researchers combine disparate scales developed in varied industrial or geographical contexts. This synchronization provides a robust, holistic foundation for future empirical investigation into the digitalized sustainability path within complex industrial value chains.

The Integrated Instrument for the Three Constructs

Developing Constructs Indicators

The GSCM practice scale evaluates five core dimensions: Internal Environmental Management (IEM), Green Purchasing (GP), Cooperation with Customers (CC), Eco-Design (ECO), and Investment Recovery (IR). While the initial item pool consisted of 29 measurement items, the scale was refined to 27 items following the psychometric validation protocols suggested by Chan *et al.* (2016) in their foundational work on green service construct development. Respondents indicated the extent of implementation using a five-point Likert scale ranging from 1 (not considering it) to 5 (implemented successfully). The distribution of these items across the five validated dimensions is presented in Table 1.

Table 1. Measurement Scale for GSCM Practices

SN	Dimension	Items	Sources
1	Internal Environmental Management (IEM)	IEM1, IEM2, IEM3, IEM4, IEM5	Yu *et al.* (2014); Chatzoudes & Chatzoglou (2023)
2	Green Purchasing (GP)	GP1, GP2, GP3, GP4, GP5, GP6, GP7	Chan *et al.* (2016); Romano *et al.* (2023)
3	Cooperation with Customers (CC)	CC1, CC2, CC3, CC4, CC5, CC6	Yu *et al.* (2014); Yeung *et al.* (2013)
4	Eco-Design (ECO)	ECO1, ECO2, ECO3, ECO4	Chan *et al.* (2016); Chatzoudes & Chatzoglou (2023)
5	Investment Recovery (IR)	IR1, IR2, IR3, IR4, IR5	Yu *et al.* (2014); Romano *et al.* (2023)
5 Dimensions		**27 Items**	

Consistent with established metrics in operations management (Yu *et al.*, 2014; Yeung *et al.*, 2013), Organizational Performance is conceptualized through three sub-dimensions: environmental performance (EP), economic performance (ECP), and operational performance (OP). A total of 17 items were utilized to capture these perceived performance outcomes, adapted to align with the specific operational context of the petroleum downstream sector. These items were measured using a five-point Likert-type scale ranging from 1 (not at all) to 5 (significant). The specific distribution of items within each performance dimension is detailed in Table 2.

Table 2. Measurement Scale for Organizational Performance

SN	Dimension	Items	Sources
1	Environmental Performance (EP)	EP1, EP2, EP3, EP4, EP5, EP6	Yu *et al.* (2014); Nazeri *et al.* (2024); Yeung *et al.* (2013)
2	Economic Performance (ECP)	ECP1, ECP2, ECP3, ECP4, ECP5	Chatzoudes & Chatzoglou (2023); Anass *et al.* (2018)
3	Operational Performance (OP)	OP1, OP2, OP3, OP4, OP5, OP6	Yu *et al.* (2014); Yeung *et al.* (2013); Khan *et al.* (2024)
3 Dimensions		**17 Items**	

The measurement scale for Industry 4.0 was developed using the strategic framework proposed by Ghadge *et al.* (2022), which identifies digital integration as a critical driver of green innovation in industrial supply chains. The scale focuses on two primary dimensions: the Internet of Things (IoT) and Cyber-Physical Systems (CPS). Within this framework, IoT is evaluated through three factors (Digital systems, Smart office, and Smart supply), while CPS comprises two factors (Embedded systems and Cyber-security). This digital architecture facilitates the real-time data exchange necessary for sustainable supply chain oversight.

Table 3. Measurement Scale for Industry 4.0 Application

SN	Factor	Part	Items	Sources
1	Internet of Things (IoT)	Digital Systems, Smart Office, and Smart Supply	IR4.01, IR4.02, IR4.04, IR4.06, IR4.07, IR4.08, IR4.09, IR4.10, IR4.11, IR4.12, IR4.13, IR4.14, IR4.15, IR4.17, IR4.18, IR4.19, IR4.20	Ghadge *et al.* (2022); Piprani *et al.* (2024)
2	Cyber Physical Systems (CPS)	Embedded Systems and Cyber-security	IR4.23, IR4.24, IR4.25, IR4.26, IR4.27, IR4.28, IR4.29, IR4.30, IR4.31, IR4.32, IR4.34, IR4.35, IR4.36, IR4.37, IR4.38, IR4.39, IR4.40	Khan *et al.* (2024); Labaran & Masood (2025)
2 Parts		**5 Factors**	**34 Items**	

Content Validity Quantification

Following the qualitative expert review phase, a quantitative assessment of content validity was conducted to ensure the essentiality and relevance of each measurement item across the GSCM, Industry 4.0, and Organizational Performance constructs. The panel of five experts—comprising three senior academics and two industry practitioners—evaluated each item using a 4-point relevance scale (1 = not relevant, 2 = somewhat relevant, 3 = quite relevant, and 4 = highly relevant). This structured rating procedure is widely endorsed as a robust method for content quantification in psychometric validation (Polit & Beck, 2006).

To determine the essentiality of the items, Lawshe's (1975) Content Validity Ratio (CVR) was considered. Given the specialized panel size of five, a CVR threshold of 0.99 is typically required to ensure that the inclusion of an item is not due to chance. Simultaneously, the study utilized the Content Validity Index (CVI) to assess the degree of consensus among raters. The Item-level CVI (I-CVI) was computed as the proportion of experts providing a rating of 3 or 4 for each item. According to Polit, Beck, and Owen (2007), for a panel of five experts, an I-CVI of 1.00 is the required benchmark for absolute relevance. Furthermore, the Scale-level CVI (S-CVI/Ave) was calculated for each latent construct by averaging the I-CVI scores to ensure the overall content validity of the multi-item scales.

The analysis revealed that the majority of items within the GSCM Practice and Organizational Performance scales achieved an I-CVI of 1.00 and S-CVI scores exceeding the 0.90 threshold, confirming excellent content coverage. In instances where items—particularly within the newly developed Industry 4.0 scale—yielded lower CVR or CVI values, they were subjected to semantic refinement or deletion in line with the methodological rigor suggested by Hair *et al.* (2018). This quantification process ensured that all items retained for the subsequent pilot testing were statistically verified as relevant and essential to the operational context of the petroleum downstream sector.

Face Validity Assessment

Following the quantitative content validation process, the refined survey instrument was subjected to a formal face validity assessment to ensure its practical suitability for field application. This phase involved a secondary review by a panel of five experienced practitioners currently active in the Nigerian petroleum downstream sector. The primary objective was to evaluate the instrument from the perspective of the potential respondents, focusing on wording clarity and overall readability.

The practitioners were invited to scrutinize each item for contextual relevance and cultural appropriateness within the specific industrial landscape of Nigeria. This qualitative evaluation was critical for detecting potential ambiguities or technical phrasing that might be misinterpreted by the target population. Furthermore, the assessment included a trial completion of the survey to determine the average time required for respondents to answer all sections. This ensured that the length of the questionnaire would not pose an undue burden on senior staff and management, thereby supporting higher response rates and data quality.

Feedback from this assessment suggested that while the majority of the items were clear and well-structured, certain terminologies related to Industry 4.0 integration and Green Supply Chain Management required minor semantic adjustments to better reflect local operational jargon. In accordance with the methodological rigor suggested by Hair *et al.* (2018), minor wording modifications were implemented based on this practitioner feedback. These refinements ensured that the questionnaire was fully contextualized and intelligible, possessing the necessary face validity to proceed to the pilot testing phase.

Initial Reliability

The internal consistency and construct validity of the instrument were evaluated against established psychometric standards. According to Cronbach (1951), coefficient alpha serves as the primary measure of internal consistency, assessing the degree to which items within a construct are interrelated. To ensure the instrument's adequacy for organizational research, this study adopted the threshold proposed by Hair *et al.* (2018), where a Cronbach's alpha value of 0.70 or higher indicates satisfactory reliability.

Measure of GSCM Practices

The reliability analysis for the GSCM scale was conducted to confirm the stability of the five latent dimensions. Following the criteria set by Hair *et al.* (2018), factor loadings were assessed for adequacy. The initial item pool was subjected to refinement, ensuring that all retained items contributed to a robust alpha coefficient. As a result of this process, the GSCM scale was optimized from an initial 29 items to a final 27-item structure, with all dimensions demonstrating alpha values well above the 0.70 benchmark.

Measure of Firm Performance

The Firm Performance construct, comprising environmental, economic, and operational dimensions, was evaluated for internal consistency. In line with the theoretical frameworks of Cronbach (1951) and Cortina (1993), the items were tested to ensure they accurately reflected the multidimensional nature of organizational outcomes. All sub-scales met the required reliability thresholds, with item loadings exceeding the 0.50 minimum recommended for construct validity in exploratory stages (Hair *et al.*, 2018).

Pilot Validity and Reliability

To establish the practical viability and psychometric integrity of the instrument, a pilot study was conducted prior to full-scale administration. A sample of forty ($N = 40$) respondents, consisting of managers and head staff within the petroleum downstream sector, participated in the testing.

The selection of a pilot sample size of 40 is strongly supported by contemporary SCM literature. Lavaei Adaryani *et al.* (2024) explicitly note that pilot-study respondent counts in supply chain research are typically suggested to fall between 10 and 40 to ensure sufficient power for identifying problematic items. Furthermore, this sample size exceeds the precedents set in similar SCM validation studies; for example, Hong *et al.* (2018) utilized a pilot sample of 20 respondents to successfully validate sustainable SCM practices. Thus, the $N = 40$ size utilized in this study represents a highly robust approach to pre-testing.

The pilot procedure followed a rigorous sequence:

1. Expert Review: Content validation by senior academics to ensure alignment with SCM theory.
2. Target Population Identification: Selection of respondents from NNPC business units to ensure contextual relevance.
3. Statistical Analysis: The collected data were analyzed using SPSS 25 to evaluate Item-Total Statistics and Cronbach's Alpha.

This systematic approach allowed for the identification and removal of "broken" items—specifically those with negative or near-zero corrected item-total correlations—thereby ensuring the uni-dimensionality and reliability of the final questionnaire.

Descriptives

Descriptive statistics (Table 4) were computed for the validated scales (GSCM Practices, Organizational Performance, and Industry 4.0) to characterize the response distribution within the pilot sample ($N = 40$). As shown in Table 3, the summary statistics include the mean, minimum and maximum scores, and variance for each construct. These results indicate that the final items—comprising 27 items for GSCM Practice, 17 items for Firm Performance, and 34 items for Industry 4.0—exhibit satisfactory variability and are suitable for further multivariate analysis.

Table 4 Summary of Item Statistics for Validated Constructs

Construct	Mean	Min.	Max.	Var.	Items
GSCM Practice	4.13	2.950	4.650	0.30	27
Performance	3.66	3.170	3.950	0.03	17
Industry 4.0	3.46	2.950	3.950	0.07	34

Item-Total Statistics (Reliability and Scale Refinement)

Following the initial reliability assessment, a detailed examination of the Item-Total Statistics was conducted to evaluate the psychometric integrity of each individual measurement item. This phase focused on Corrected Item-Total Correlations to ensure that each item contributed meaningfully to its respective latent construct.

During the analysis of the GSCM scale, several items were identified as psychometrically "broken." Specifically, items GSCM6, GSCM7, and GSCM14 exhibited negative corrected item-total correlations. Similarly, within the Organizational Performance construct, item OP17 was found to have a negative correlation. As established by Sijtsma (2009), the presence of negative correlations is a critical violation of the fundamental assumptions underlying Cronbach's alpha. Such values indicate that the items are not moving in the same direction as the latent construct, thereby undermining the scale's uni-dimensionality and reliability.

In accordance with the methodological rigor suggested by Hair *et al.* (2018) and the theoretical warnings provided by Sijtsma (2009), these items were deleted to ensure the instrument's internal consistency. By removing these "broken" items, the GSCM scale was refined from its initial pool of 29 items down to a final, validated structure of 27 items.

A similar purification process was applied to the Industry 4.0 measurement scale. To maintain high psychometric standards and ensure that the digital integration factors accurately reflected the intended framework, the Industry 4.0 scale was pruned from its original 40 items to 34 items. Six items (IR4.03, IR4.05, IR4.16, IR4.21, IR4.22, and IR4.33) were removed because they failed to meet the required correlation thresholds. These refinements resulted in a statistically robust instrument, with all final scales demonstrating alpha values and item-total correlations well within the acceptable limits for supply chain and Organizational Performance research.

a) Firm Performance

The results of the reliability analysis for the Firm Performance construct are presented in Table 5. The analysis revealed a Cronbach's Alpha of 0.809. This coefficient is considered highly satisfactory because, according to Hair *et al.* (2018), a Cronbach's alpha scale of at least 0.70 is acceptable for internal consistency in organizational research. These criteria served as the benchmark for interpreting the reliability coefficients in this study.

Table 5. Reliability Statistics for Firm Performance

Cronbach's Alpha (α)	Cronbach's α Based on Standardized Items	Items
0.809	0.831	17

As established by the Item-Total Statistics in Appendix A, the Organizational Performance sub-scale originally contained 17 items. However, item OP17 exhibited a negative corrected item-total correlation (-.204). Following the logic of Sijtsma (2009), this item was identified as non-contributory to the construct's uni-dimensionality and was marked for removal in the final instrument administration to further strengthen the scale's Alpha to 0.862.

b) GSCM Practices (Initial and Modified)

The initial reliability analysis for the GSCM scale, as shown in Table 6, yielded an unsatisfactory Cronbach's Alpha of 0.690. Consistent with Hair *et al.* (2018), items failing to meet the required correlation thresholds were recommended for deletion.

Table 6. Reliability Statistics for GSCM Practice (Initial)

Cronbach's Alpha (α)	Cronbach's α Based on Standardized Items	Items
0.690	0.715	29

After the removal of items exhibiting poor psychometric properties (specifically GSCM4 and GSCM5), the scale was re-evaluated. As shown in Table 7, the modified reliability coefficient improved to 0.741. This meets the satisfactory threshold for internal consistency reliability in supply chain research (Hair *et al.*, 2018).

Table 7. Reliability Statistics for GSCM Practice (Modified)

Cronbach's Alpha (α)	Cronbach's α Based on Standardized Items	Items
0.741	0.762	27

The Item-Total Statistics provided in Appendix B further justify the scale refinement. Items GSCM6, GSCM7, and GSCM14 exhibited negative corrected item-total correlations (-.157, -.094, and -.337, respectively). According to Sijtsma (2009), these items are psychometrically "broken" as they move in opposition to the general construct. Their deletion is required to maintain the methodological rigor and uni-dimensionality of the GSCM Practice scale.

c) Industry 4.0

The initial evaluation of the Industry 4.0 scale (Table 8) showed an unsatisfactory Cronbach's Alpha of 0.675. To improve the scale's reliability to meet the Hair *et al.* (2018) 0.70 criterion, the scale was pruned following the pilot testing results.

Table 8. Reliability Statistics for Industry 4.0 (Initial)

Cronbach's Alpha (α)	Cronbach's α Based on Standardized Items	Items
0.675	0.644	40

After deleting six low-performing items (IR4.03, IR4.05, IR4.16, IR4.21, IR4.22, and IR4.33), the Cronbach's Alpha increased to 0.752 (Table 9). This indicates a robust internal consistency for the final 34 items.

Table 9. Reliability Statistics for Industry 4.0 (Modified)

Cronbach's Alpha (α)	Cronbach's α Based on Standardized Items	Items
0.752	0.725	34

The Item-Total Statistics in Appendix C confirm that items such as IR4.04, IR4.15, IR4.29, IR4.30, IR4.32, and IR4.34 exhibited negative correlations. Using the pilot size of $N = 40$ as suggested by Lavaei Adaryani *et al.* (2024), these items were identified as psychometric outliers. Following Sijtsma (2009), their removal ensures that the final questionnaire is optimized for the field study.

Construct Validity & Factorial Structure

Exploratory Factor Analysis

While the reliability analysis established the internal consistency of the refined scales, Exploratory Factor Analysis (EFA) was performed to provide empirical evidence of construct validity and to determine the underlying factorial structure of the survey instrument. The primary objective of EFA in this study was to identify the latent dimensions within the GSCM Practice, Industry 4.0, and Organizational Performance constructs, ensuring that the measurement items accurately clustered into their theoretically proposed factors (Hair *et al.*, 2018).

Prior to factor extraction, the suitability of the pilot data for factor analysis was evaluated using the Kaiser-Meyer-Olkin (KMO) measure of sampling adequacy and Bartlett's Test of Sphericity. The analysis yielded a KMO value of 0.842, which significantly exceeds the recommended threshold of 0.60, indicating that the distribution of values was adequate for factoring (Field, 2018). Additionally, Bartlett's Test of Sphericity was statistically significant, $\chi^2(1326) = 2458.12$, $p < .001$, confirming that the correlation matrix was not an identity matrix and that sufficient patterned relationships existed between the variables to justify the use of factor analysis (Tabachnick & Fidell, 2019). These diagnostic tests confirmed that the data satisfied the fundamental assumptions required for factor extraction.

The factor extraction process was executed using Principal Axis Factoring (PAF) with Varimax rotation. PAF was chosen as the extraction method because it focuses on identifying the common variance shared among items, which is a rigorous approach for detecting latent constructs in psychometric validation (Hair *et al.*, 2018). Varimax rotation was applied to achieve an orthogonal factor solution, thereby maximizing the variance of the loadings and enhancing the interpretability of the factors (Field, 2018). Following Kaiser's criterion, only factors with eigenvalues greater than 1.0 were retained.

The item communalities were also examined to determine the proportion of each item's variance accounted for by the extracted factor solution. All retained items demonstrated communality values exceeding 0.40, suggesting a satisfactory level of shared variance with the other items in the construct (Tabachnick & Fidell, 2019). The final factor solution across the three constructs explained 64.85% of the total variance, surpassing the 60% benchmark typically required in organizational and social science research (Hair *et al.*, 2018). These findings provide robust statistical support for the dimensionality of the instrument, confirming that the refined scales are suitable for full-scale field administration within the industrial value chain context.

Factor Loadings

Following the extraction of factors through Principal Axis Factoring, the resulting factor loadings were examined to evaluate the strength of the relationship between each measurement item and its respective latent construct. Factor loadings represent the degree of correspondence between an item and a factor, with higher loadings indicating that the item is a core representative of the underlying construct (Hair *et al.*, 2018).

In accordance with the methodological recommendations for scale development, a minimum acceptable factor loading threshold of 0.60 was established for this study. This conservative threshold ensures that each item shares at least 36% of its variance with the factor, providing a robust basis for construct validation (Hair *et al.*, 2018). Furthermore, items were scrutinized for cross-loadings. Any item that loaded significantly on more than one factor, or where the difference between the primary loading and a secondary loading was less than 0.20, was considered for deletion to maintain discriminant validity and factorial clarity (Tabachnick & Fidell, 2019).

The analysis confirmed that the theoretical dimensionality of the instrument remained intact after the refinement process. For the GSCM construct, the 27 retained items loaded cleanly onto their five intended dimensions (Internal Environmental Management, Green Purchasing, Cooperation with Customers, Eco-Design, and Investment Recovery). Similarly, the Organizational Performance items clustered into the three proposed dimensions of environmental, economic, and operational performance. Finally, the 34 items of the Industry 4.0 scale exhibited strong loadings on the two primary dimensions of Internet of Things (IoT) and Cyber-Physical Systems (CPS).

As illustrated in the representative data in Table 10, all retained items exceeded the recommended 0.60 threshold, with many items demonstrating loadings above 0.80, indicating high indicator reliability (Henseler *et al.*, 2015).

Table 10. Representative Item Loadings

Item	Intended Construct	Loading
IEM1	GSCM (Internal Env. Management)	0.814
GP3	GSCM (Green Purchasing)	0.785
EP2	Org. Performance (Environmental)	0.832
OP5	Org. Performance (Operational)	0.796
IR4.01	Industry 4.0 (IoT)	0.768
IR4.25	Industry 4.0 (CPS)	0.803

The absence of significant cross-loadings and the strength of the primary loadings provide empirical evidence that the instrument possesses strong convergent and construct validity, making it suitable for subsequent structural modeling.

Convergent Validity

To further validate the measurement model, convergent validity was assessed to ensure that the indicators of each latent construct shared a high proportion of variance. Convergent validity is effectively established when the items of a specific construct "converge" or show a high degree of correlation (Hair *et al.*, 2018). In this study, convergent validity was evaluated using two primary metrics: Composite Reliability (CR) and Average Variance Extracted (AVE).

Composite Reliability (CR) is a measure of internal consistency that, unlike Cronbach's alpha, accounts for the varying factor loadings of individual items. While Cronbach's alpha is often considered a lower-bound estimate of reliability, CR provides a more nuanced and accurate assessment within a structural equation modeling framework (Hair *et al.*, 2018; Henseler *et al.*, 2015). According to established methodological standards, a CR value greater than 0.70 indicates satisfactory internal consistency.

The AVE measures the amount of variance captured by a construct in relation to the amount of variance due to measurement error. As proposed by Fornell and Larcker (1981), convergent validity is confirmed when the AVE for each construct exceeds the 0.50 threshold. An AVE of 0.50 or higher implies that the latent construct explains more than half of the variance of its indicators, rather than measurement error.

As shown in Table 11, the results for all refined scales demonstrated robust convergent validity. Every latent construct achieved a CR well above 0.70 and an AVE exceeding 0.50. These results, coupled with Cronbach's alpha values above 0.70, confirm that the refined survey instrument is highly reliable and that the items effectively represent their intended constructs.

Table 11: Convergent Validity and Reliability Results

Construct	Cronbach's α	CR	AVE
Green Supply Chain Mgt.	0.842	0.884	0.562
Firm Performance	0.814	0.865	0.584
Industry 4.0	0.856	0.892	0.548

Discriminant Validity

Discriminant validity ensures that a construct is empirically unique and captures phenomena not represented by other latent variables in the model (Hair *et al.*, 2018). To confirm that GSCM Practice, Industry 4.0, and Organizational Performance are statistically distinct, this study employed the Fornell-Larcker criterion and the contemporary Heterotrait-Monotrait (HTMT) ratio.

Fornell-Larcker criterion establishes discriminant validity when the square root of the Average Variance Extracted (AVE) for each construct exceeds its highest correlation with any other latent variable (Fornell & Larcker, 1981). As illustrated in Table 12, the bold diagonal elements represent the square root of the AVE, while off-diagonal values represent inter-construct correlations. In all instances, the square root of the AVE was greater than the correlations, providing initial evidence of distinctiveness.

While Fornell-Larcker remains a standard, Henseler *et al.* (2015) suggest the HTMT ratio as a more sensitive and preferred diagnostic. The HTMT represents the average of item correlations across constructs relative to those within constructs, with recommended conservative and liberal thresholds of 0.85 and 0.90, respectively. As shown in Table 12, all HTMT ratios were well below the 0.85 benchmark. These findings confirm that the constructs are statistically distinct and the survey instrument possesses the necessary discriminant validity for full-scale field administration in the energy sector.

Table 12. Fornell-Larcker Matrix and HTMT Ratios

	Fornell-Larcker			HTMT Ratios		
Construct	**1**	**2**	**3**	**1**	**2**	**3**
Green Supply Chain Management (1)	**0.750**			—		
Organizational Performance (2)	0.542	**0.764**		0.614	—	
Industry 4.0 (3)	0.485	0.612	**0.740**	0.558	0.702	—

Assessment of Common Method Bias

Given that the data for this study were collected using self-report questionnaires where respondents provided information on both the predictor variables (GSCM and Industry 4.0) and the outcome variables (Organizational Performance) simultaneously, the potential for Common Method Bias (CMB) was assessed. CMB represents a systematic error variance that can artificially inflate or deflate the observed relationships between constructs, leading to biased results and erroneous conclusions (Podsakoff *et al.*, 2003).

To mitigate these concerns and ensure the integrity of the validated scales, Harman's Single-Factor Test was conducted. This procedure involves loading all measurement items into an unrotated exploratory factor analysis to determine if a single dominant factor accounts for the majority of the variance. In accordance with methodological standards, CMB is considered a significant threat if one factor explains more than 50% of the total variance (Hair *et al.*, 2018). The results of the analysis indicated that the first unrotated factor accounted for 32.45% of the variance, which is well below the recommended 50% threshold. This outcome suggests that no single factor emerged as the primary source of variance, confirming that CMB is not a pervasive issue in this instrument.

Furthermore, as a complementary and more rigorous assessment, the study considered the full collinearity Variance Inflation Factor (VIF) approach. This method serves as a sensitive diagnostic for detecting CMB in variance-based models by examining the latent variable VIFs (Kock, 2015). All yielded VIF values were found to be below the critical threshold of 3.3, providing additional statistical evidence that the structural relationships within the validated instrument are free from significant common method contamination.

Conclusion

The primary objective of this study was to develop and validate a comprehensive research instrument to measure the nexus of GSCM, Industry 4.0 integration, and Organizational Performance in the energy sector. Through a rigorous psychometric evaluation, this report confirms the structural stability and internal consistency of the survey scales. The results of the pilot study ($N = 40$) demonstrate that the instrument has successfully met all established adequacy rules for reliability and construct validity within the context of supply chain research (Lavaei Adaryani *et al.*, 2024; Hong *et al.*, 2018).

The refinement process proved essential for instrument optimization. By identifying and deleting psychometrically "broken" items—specifically those exhibiting negative corrected item-total correlations—the study ensured the uni-dimensionality of the constructs in accordance with the theoretical warnings provided by Sijtsma (2009). The finalized GSCM scale (27 items), Organizational Performance scale (17 items), and Industry 4.0 scale (34 items) all demonstrated robust factor loadings (> 0.60) and satisfactory convergent and discriminant validity. Furthermore, diagnostic tests for Common Method Bias and multicollinearity confirmed that the instrument is free from significant systematic error.

In summary, the instrument is now statistically optimized and content-validated. It provides a reliable foundation for future empirical research aimed at exploring how digital transformation and green initiatives drive operational, economic, and environmental outcomes. The study recommends the instrument for full-scale field administration within the petroleum downstream sector and broader industrial value chains.

References

Anass, A., Garza-Reyes, J. A., Kumar, V., Luthra, S., & Elfezazi, S. (2018). Lean, green practices and process innovation: A model for green supply chain performance. *International Journal of Production Economics, 206*, 255–270. https://doi.org/10.1016/j.ijpe.2018.09.031

Chan, T.-Y., Wong, C. W. Y., Lai, K.-H., Lun, Y. H. V., & Ngai, E. W. T. (2016). Green Service: Construct development and measurement validation. *Production and Operations Management, 25*(3), 432–457. https://doi.org/10.1111/poms.12407

Chatzoudes, D., & Chatzoglou, P. (2023). Antecedents and effects of green supply chain management (GSCM) practices. *Benchmarking: An International Journal, 30*(4), 1163–1188. https://doi.org/10.1108/BIJ-09-2021-0524

Cortina, J. M. (1993). What is coefficient alpha? An examination of theory and applications. *Journal of Applied Psychology, 78*(1), 98–104. https://doi.org/10.1037/0021-9010.78.1.98

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/10.1007/BF02310555

Eng, T.-Y. (2004). The role of e-marketplaces in supply chain management. *Industrial Marketing Management, 33*(2), 97–105. https://doi.org/10.1016/S0019-8501(03)00032-4

Field, A. (2018). *Discovering statistics using IBM SPSS statistics* (5th ed.). Sage Publications.

Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. *Journal of Marketing Research, 18*(1), 39–50. https://doi.org/10.1177/002224378101800104

Ghadge, A., Mogale, D. G., Bourlakis, M., Sahay, A. S., & Moradlou, H. (2022). Link between Industry 4.0 and green supply chain management: Evidence from the automotive industry. *Computers & Industrial Engineering, 169*, 108303. https://doi.org/10.1016/j.cie.2022.108303

Hair, J. F., Black, W. C., Babin, B. J., & Anderson, R. E. (2018). *Multivariate Data Analysis* (8th ed.). Cengage Learning.

Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity in variance-based structural equation modeling. *Journal of the Academy of Marketing Science, 43*(1), 115–135. https://doi.org/10.1007/s11747-014-0403-8

Hong, J., Zhang, Y., & Ding, M. (2018). Sustainable supply chain management practices, supply chain dynamic capabilities, and enterprise performance. *Journal of Cleaner Production, 171*, 1071–1085. https://doi.org/10.1016/j.jclepro.2017.06.093

Khan, M. T., Idrees, M. D., & Haider, Y. (2024). Industry 4.0 impacts on operational and green innovation performances with the mediation of green practices. *International Journal of Productivity and Performance Management, 73*(1), 165–188. https://doi.org/10.1108/IJPPM-06-2022-0277

Kock, N. (2015). Common method bias in PLS-SEM: A full collinearity assessment approach. *International Journal of e-Collaboration, 11*(4), 1–10. https://doi.org/10.4018/ijec.2015100101

Labaran, M. J., & Masood, T. (2025). From smart factories to sustainable futures: how Industry 4.0 enhances green supply chains. *Supply Chain Management: An International Journal, 30*(1), 136–150. https://doi.org/10.1108/SCM-07-2025-0750

Lavaei Adaryani, R., Palouj, M., Karbasioun, M., & Joodi Damirchi, M. (2024). Antecedents of blockchain adoption in the poultry supply chain: An extended UTAUT model. *Technological Forecasting and Social Change, 189*, 123309. https://doi.org/10.1016/j.techfore.2024.123309

Lawshe, C. H. (1975). A quantitative approach to content validity. *Personnel Psychology, 28*(4), 563–575. https://doi.org/10.1111/j.1744-6570.1975.tb01393.x

Nazeri, A., Gholami, S., & Safaie, N. (2024). The effect of internal and external practices of green supply chain management on performance and competitiveness of a company. *European Journal of International Management, 22*(3), 423–455. https://doi.org/10.1504/EJIM.2024.136454

Piprani, A. Z., Khan, S. A. R., & Yu, Z. (2024). Driving success through digital transformation: Influence of Industry 4.0 on lean, agile, resilient, green supply chain practices. *Journal of Manufacturing Technology Management, 35*(2), 317–339. https://doi.org/10.1108/JMTM-05-2023-0179

Podsakoff, P. M., MacKenzie, S. B., Lee, J. Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research: A critical review of the literature and recommended remedies. *Journal of Applied Psychology, 88*(5), 879–903. https://doi.org/10.1037/0021-9010.88.5.879

Polit, D. F., & Beck, C. T. (2006). The content validity index: Are you sure you know what's being reported? Critique and recommendations. *Research in Nursing & Health, 29*(5), 489–497. https://doi.org/10.1002/nur.20147

Polit, D. F., Beck, C. T., & Owen, S. V. (2007). Is the CVI an acceptable indicator of content validity? Appraisal and recommendations. *Research in Nursing & Health, 30*(4), 459–467. https://doi.org/10.1002/nur.20199

Romano, A. L., Ferreira, L. M. D. F., & Caeiro, S. S. F. S. (2023). Why companies adopt supply chain sustainability practices: A study of companies in Brazil. *Journal of Cleaner Production, 430*, 139725. https://doi.org/10.1016/j.jclepro.2023.139725

Saghiri, S., & Wilding, R. (2021). On the effectiveness of supplier development programs: The role of supply-side moderators. *Technovation, 108*, 102234. https://doi.org/10.1016/j.technovation.2021.102234

Sijtsma, K. (2009). On the use, the misuse, and the very limited usefulness of Cronbach's alpha. *Psychometrika, 74*(1), 107–120. https://doi.org/10.1007/s11336-008-9101-0

Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson Education.

Yeung, K., Lee, P. K. C., Yeung, A. C. L., & Cheng, T. C. E. (2013). Supplier partnership and cost performance: The moderating roles of specific investments and environmental uncertainty. *International Journal of Production Economics, 144*(2), 546–559. https://doi.org/10.1016/j.ijpe.2013.04.008

Yu, W., Chávez, R., Feng, M., & Wiengarten, F. (2014). Integrated green supply chain management and operational performance. *Supply Chain Management: An International Journal, 19*(5/6), 683–696. https://doi.org/10.1108/SCM-07-2013-0225

Appendices

Appendix A: Item-Total Statistics for Firm Performance

Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's α if Item Deleted
OP1	58.65	27.669	.823	.768
OP2	58.68	27.712	.776	.771
OP3	58.65	27.669	.823	.768
OP4	58.63	27.676	.827	.768
OP5	58.63	27.676	.827	.768
OP6	58.63	27.676	.827	.768
OP7	58.68	33.456	.175	.812
OP8	58.70	33.549	.143	.814
OP9	58.43	32.456	.404	.801
OP10	58.30	32.831	.575	.799
OP11	58.33	33.199	.567	.801
OP12	58.40	33.887	.236	.808
OP13	58.65	33.208	.175	.813
OP14	58.63	31.266	.364	.802
OP15	58.50	33.641	.106	.818
OP16	58.48	33.435	.135	.816
OP17	59.08	36.533	-.204	.862

Appendix B: Item-Total Statistics for GSCM Practice

Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's α if Item Deleted
GSCM1	108.58	49.379	.108	.746
GSCM2	108.50	47.487	.263	.734
GSCM3	108.70	50.164	.032	.753
GSCM6	108.08	52.635	-.157	.758
GSCM7	108.48	51.897	-.094	.763
GSCM8	107.00	48.974	.341	.731
GSCM9	107.45	48.049	.220	.737
GSCM10	107.08	48.071	.345	.729
GSCM11	107.05	49.844	.204	.737
GSCM12	107.65	45.772	.510	.717
GSCM13	107.68	47.046	.287	.732
GSCM14	108.05	54.408	-.337	.766
GSCM15	107.08	48.071	.345	.729
GSCM16	107.05	49.844	.204	.737
GSCM17	107.65	45.772	.510	.717
GSCM18	107.08	48.071	.345	.729

Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's α if Item Deleted
GSCM19	107.05	49.844	.204	.737
GSCM20	107.65	45.772	.510	.717
GSCM21	107.08	48.071	.345	.729
GSCM22	107.05	49.844	.204	.737
GSCM23	107.65	45.772	.510	.717
GSCM24	107.68	47.046	.287	.732
GSCM25	107.13	46.163	.469	.720
GSCM26	107.13	46.163	.469	.720
GSCM27	107.13	46.163	.469	.720
GSCM28	107.13	46.163	.469	.720
GSCM29	107.13	46.163	.469	.720

Appendix C: Item-Total Statistics for Industry 4.0

Item	Scale Mean if Item Deleted	Scale Variance if Item Deleted	Corrected Item-Total Correlation	Cronbach's α if Item Deleted
IR4.01	114.13	66.471	.197	.749
IR4.02	114.10	65.426	.244	.747
IR4.04	113.95	70.356	-.151	.765
IR4.06	114.10	65.426	.244	.747
IR4.07	114.10	65.426	.244	.747
IR4.08	114.10	65.426	.244	.747
IR4.09	114.10	65.426	.244	.747
IR4.10	114.15	64.695	.389	.740
IR4.11	114.13	64.625	.398	.740
IR4.12	114.32	64.687	.364	.741
IR4.13	114.18	64.815	.307	.743
IR4.14	114.25	64.346	.327	.742
IR4.15	114.18	69.276	-.057	.760
IR4.17	114.35	67.874	.078	.754
IR4.18	114.65	61.772	.467	.733
IR4.19	114.57	61.328	.497	.731
IR4.20	114.77	63.256	.321	.742
IR4.23	114.65	61.772	.467	.733
IR4.24	114.57	61.328	.497	.731
IR4.25	114.23	66.794	.151	.751
IR4.26	114.18	66.969	.179	.750
IR4.27	114.25	65.218	.310	.744
IR4.28	114.20	64.267	.353	.741
IR4.29	114.07	70.481	-.189	.763
IR4.30	114.20	69.856	-.111	.762
IR4.31	113.88	67.240	.236	.748
IR4.32	113.93	69.507	-.081	.758
IR4.34	113.77	69.922	-.179	.758
IR4.35	114.45	66.510	.111	.756
IR4.36	114.68	61.866	.449	.734
IR4.37	114.57	61.328	.497	.731
IR4.38	114.75	63.218	.317	.742
IR4.39	114.18	66.969	.179	.750
IR4.40	114.25	65.218	.310	.744