



ISSN: 2586-6036

JWMAAP website: <http://accesson.kr/jwmap>doi: <http://dx.doi.org/10.13106/jwmap.2026.vol9.no2.133>

# Development and Validation of the Korean Stowell-based Coaching Leadership Scale (K-SCLS): A Multi-Level Integration Approach for Korean SMEs

Bum Suk LEE<sup>1\*</sup>, Ho Youn JEONG<sup>2</sup>

1. First Author and Corresponding Author Department of Business Administration, Adjunct Professor, Seoul Business School, aSSIST University, Korea. Department of Coaching Sciences, Graduate School of Business, Kyung Hee University, Korea. Adjunct Professor.  
Email: [bibledrama@naver.com](mailto:bibledrama@naver.com)

2. Second Author Ph.D Candidate, Department of Business Administration, Seoul Business School, aSSIST University, Korea. DBA Candidate, Department of Business Administration, SDG Management School, Geneva, Switzerland.  
Email: [j-hayoun@hanmail.net](mailto:j-hayoun@hanmail.net)

Received: February 22, 2026. Revised: March 01, 2026. Accepted: April 22, 2026.

## Abstract

**Purpose:** This exploratory study developed and provided initial validation evidence for the Korean Stowell-based Coaching Leadership Scale (K-SCLS), a 16-item instrument measuring coaching leadership in Korean SMEs, while examining the structural nature of coaching leadership dimensionality through a Multi-Level Integration Framework. **Method:** A sample of 300 SME employees (Mage = 20s–60s; male = 51.3%) completed an initial 22-item instrument refined to 16 items via exploratory and confirmatory factor analyses. A split-half cross-validation ( $n_1 = n_2 = 150$ ) addressed same-sample EFA-CFA concerns. Bifactor modeling, hierarchical CFA, and sequential measurement invariance testing were employed alongside HTMT-based discriminant validity analysis. **Results:** Cross-validation confirmed item selection stability (87.5% correspondence) and replicated the four-factor structure in an independent subsample (CFI = 1.000, RMSEA = 0.000). The correlated four-factor model demonstrated excellent fit (CFI = 1.000, RMSEA = 0.008;  $\alpha = .938-.949$ ; AVE = .822–.843). Bifactor modeling revealed a dominant general coaching leadership factor accounting for 90.1% of common variance (ECV = .901;  $\omega_H = .958$ ), with negligible specific factor contributions ( $\omega_S = .041-.172$ ). HTMT ratios (.843–.933) and interfactor correlations (.845–.943) indicated subfactors were not empirically distinct. Full measurement invariance across gender was achieved (scalar  $\Delta CFI = 0.000$ ). **Conclusion:** These findings reconceptualize the K-SCLS as a bifactor-structured instrument wherein four content domains reflect a superordinate general construct, reframing two decades of discriminant validity failures as a substantively meaningful integration phenomenon theoretically consistent with Korean cultural constructs including jeong, woori consciousness, and chemyon. Residual common method variance (43.0%) and the single-source design necessitate multi-source replication.

**Keywords:** Coaching leadership, Cultural integration, Measurement paradox, Korean SMEs, Scale development

**JEL Code:** M12, M54, O15, D23, Z13

## 1. Introduction

© Copyright: The Author(s)

This is an Open Access article distributed under the terms of the Creative Commons Attribution Non-Commercial License (<http://creativecommons.org/licenses/by-nc/4.0/>) which permits unrestricted noncommercial use, distribution, and reproduction in any medium, provided the original work is properly cited.

### 1.1. The Korean Coaching Leadership Measurement Paradox

Joh's (2021, 2022, 2024) systematic investigation of 302 Korean coaching leadership studies spanning two decades revealed that 215 studies (74.7%) incorrectly attributed theoretical frameworks to Stowell's (1987) doctoral dissertation without proper verification. This misattribution occurred despite Stowell's original work neither operationally defining coaching leadership as a distinct organizational construct nor proposing validated measurement scales for hierarchical contexts. Originating in early studies (Choi, 2003), this systematic misattribution propagated over two decades, creating "inherited limitations" (Joh, 2022) that manifested as persistent discriminant validity failures and structural model inadequacies.

Despite these theoretical foundation issues, Korean coaching leadership research has demonstrated remarkable empirical consistency. Meta-analytic evidence demonstrates that 89.3% of Korean studies reported significant positive relationships with performance outcomes (Song & Lee, 2020), indicating robust phenomena requiring theoretical reconstruction. This creates what we term the "Korean Coaching Leadership Measurement Paradox," aligning conceptually with Smith and Lewis's (2011) dynamic equilibrium model, where apparent contradictions reveal deeper truths requiring theoretical reconstruction. Rather than dismissing persistent high intercorrelations as methodological failure, this exploratory research frames them as reflecting a substantively meaningful structural characteristic of coaching leadership in Korean organizational contexts, presenting a theoretical opportunity for indigenous theory development (Tsui, 2004, 2007).

### 1.2. From Attribution Errors to Theoretical Reconstruction

Stowell's (1987) original peer coaching framework emphasized horizontal partnerships, voluntary participation, and the explicit absence of evaluative responsibilities fundamentally distinct from supervisory coaching leadership in hierarchical settings. The Western emphasis on non-directiveness reflects historically contingent legal defense mechanisms rather than universal principles. According to Nishigaki (2021), the International Coach Federation's boundary setting emerged partly from legal pressures, and recent legislation such as Utah's Coaching and Psychotherapy Regulation Act (S.B. 48, 2025) reinforces defensive non-directiveness as a professional survival strategy.

This historical context suggests that Korean coaching leadership's inclusion of Direction Setting represents

cultural adaptation free from Western legal constraints. Coaching leadership in Korean contexts inherently involves hierarchical relationships, performance evaluation, and accountability processes that are absent from peer coaching (Hofstede, 2001). Building upon Joh's (2021, 2022, 2024) critical analysis, the Korean Stowell-based Coaching Leadership Scale (K-SCLS) represents comprehensive theoretical reconstruction incorporating Stowell's core coaching values while addressing the hierarchical leadership context absent from his original framework (Avery & Bergsteiner, 2011).

### 1.3. Cultural Integration Mechanisms

Korean organizational culture exhibits distinctive characteristics that may be associated with integrated rather than differentiated leadership approaches, potentially challenging Western dimensional assumptions (Li, 2012; Yang, 2012). Drawing on established cultural theory, three cultural constructs are proposed as theoretically plausible interpretive lenses for understanding the observed integration patterns. It is important to note that these constructs were not directly measured in the present study; they serve as theoretical context rather than empirically tested mechanisms, and their specific causal roles require verification through future research incorporating direct measurement. First, *jeong* represents deep emotional bonds extending beyond professional boundaries (Jung & Lee, 2019), potentially creating expectations that leaders demonstrate genuine personal care while simultaneously maintaining performance accountability, though the causal mechanisms underlying these cultural influences on leadership perception have not been empirically verified in the present study. Second, *woori* consciousness reflects collective identity orientations (Choi & Kim, 2020), positioning leaders as facilitators of group harmony rather than individual performance directors. Third, *chemyon* considerations (J.-H. Kim, 2020; Lee et al., 2019) influence performance feedback processes, creating natural integration between evaluation functions and relationship maintenance through dignity-preserving approaches. These cultural constructs provide a theoretically plausible interpretive framework for understanding why coaching leadership subdimensions may resist empirical differentiation in Korean SME contexts, manifesting instead as a highly integrated superordinate construct wherein the four behavioral domains are perceived as inseparable facets of holistic leadership.

### 1.4. Research Objectives and Research Questions

The primary objective of this research is to develop and provide initial exploratory validation evidence for the K-

SCLS as a psychometrically sound instrument for measuring coaching leadership in Korean SME contexts. To achieve this objective, the study employs a Multi-Level Integration Framework as the analytical strategy, operating through three layers that systematically address measurement adequacy, structural integration, and cross-group equivalence. As a secondary objective, the study explores the extent to which the validated factor structure is consistent with culturally informed interpretations of leadership integration in Korean organizational contexts. Small and medium enterprises, representing over 99% of Korean businesses and 87% of the workforce (Statistics Korea, 2023), provide the target context for this investigation given their crucial roles in sustainable development (Jenkins, 2006).

Specifically, this study addresses the following research questions:

**Research Question 1 (Measurement Layer):** Does the K-SCLS demonstrate adequate convergent validity and internal consistency for four theoretically specified coaching leadership subdimensions? To what extent does the empirical evidence—including HTMT ratios, interfactor correlations, and bifactor variance decomposition indices (ECV,  $\omega_H$ ,  $\omega_S$ )—support the interpretation that these subdimensions constitute empirically distinct constructs versus facets of a unified general factor?

**Research Question 2 (Structural Integration Layer):** To what extent does a bifactor model—specifying a general coaching leadership factor alongside four orthogonal specific factors—provide superior characterization of the K-SCLS variance structure compared with correlated multidimensional and unidimensional alternatives, and what do the resulting variance decomposition indices (ECV,  $\omega_H$ ,  $\omega_S$ ) indicate about the appropriate scoring strategy for this instrument?

**Research Question 3 (Invariance Layer):** Does the K-SCLS achieve measurement invariance across gender groups, including configural, metric, and scalar invariance, supporting cross-group score comparability?

## 2. Literature Review

### 2.1. Historical Attribution Errors and Theoretical Reconstruction

Joh's (2021, 2022, 2024) systematic investigation revealed that 74.7% of 302 Korean coaching leadership studies incorrectly attributed theoretical frameworks to Stowell (1987). However, meta-analytic evidence (Song & Lee, 2020; Yoon & Chae, 2018) demonstrates that 89.3% of studies reported significant positive relationships with performance outcomes. This creates the Korean Coaching

Leadership Measurement Paradox addressed through indigenous theory development (Tsui, 2004).

### 2.2. Cultural Integration and Sustainable Leadership

The persistent high intercorrelations among coaching leadership subfactors in Korean research (Park, 2012; Ryu & Je, 2020), traditionally interpreted as methodological problems requiring discriminant validity remediation, are theoretically reframed in this study as reflecting a substantively meaningful structural characteristic. Specifically, these intercorrelations may indicate that Korean organizational members perceive coaching leadership behaviors as holistically interconnected rather than dimensionally differentiated, a pattern consistent with cultural integration processes documented in Korean helping professions (S.-H. Ko, 2017; G.-H. Ko, 2019; K.-R. Kim, 2020). Such patterns may be aligned with sustainable leadership principles (Hargreaves & Fink, 2006; Stahl et al., 2020), though the specific mechanisms driving these intercorrelations require direct empirical investigation before substantive claims can be made.

Korean helping professionals have established traditions of eclectic integration that inform this understanding. S.-H. Ko (2017) proposed a four-stage integrative model demonstrating how multiple theoretical approaches can be selectively utilized within unified practice frameworks. G.-H. Ko's (2019) consensual qualitative research confirmed that multiple types of counseling hypotheses operate integratively. Qualitative studies of Korean practitioners demonstrate flexible switching between modalities (K.-R. Kim, 2020), suggesting strict categorical separations are neither descriptively accurate nor normatively necessary in Korean contexts.

### 2.3. Theoretical Foundations for Structural Integration

Freeman's (1978) centrality concepts and Newman's (2010) network modularity principles inform understanding of how a general coaching leadership factor may function as a superordinate integration node connecting multiple behavioral domains. The hierarchical structure (general factor → specific factors) parallels network hub configurations, supporting the interpretation that substantial shared variance among subfactors is consistent with culturally embedded organizational dynamics, while meaningful specific factor contributions indicate potential for functional specialization. The bifactor measurement model (Holzinger & Swineford, 1937; Reise, 2012) provides the most technically appropriate framework for operationalizing the hypothesized structure of Korean

coaching leadership. In the bifactor parameterization, a general factor (G) loads simultaneously on all items, representing the superordinate coaching leadership orientation, while orthogonal specific factors capture residual behavioral domain variance after extracting G. This structure is theoretically preferable to the hierarchical (second-order) specification in contexts characterized by extreme interfactor correlations, because the bifactor model does not require the assumption that subdimension scores are exclusively mediated by the higher-order factor—an assumption violated when, as observed in Korean coaching leadership research, subdimensions approach empirical indistinction (Rodriguez et al., 2016). The bifactor framework also yields directly interpretable variance decomposition indices: the Explained Common Variance (ECV; Sijtsma, 2009) quantifies the proportion of common variance attributable to G, omega hierarchical ( $\omega_H$ ; Zinbarg et al., 2006) quantifies the proportion of total score variance attributable to G, and omega subscale ( $\omega_S$ ; Reise et al., 2013) quantifies the unique reliable variance of each specific factor beyond G. These indices collectively enable a principled determination of whether total scores, subscale scores, or both are psychometrically defensible operational metrics.

## 2.4. Methodological Evolution

Korean coaching leadership research has evolved through three phases (Kang & Lee, 2020). The initial regression phase encountered persistent multicollinearity. The structural equation modeling phase (Cho & Tak, 2011) continued encountering discriminant validity challenges. The current integration understanding phase represents a paradigmatic shift toward viewing high correlations as potentially meaningful phenomena. Recent studies have employed various approaches including partial least squares analysis (Min et al., 2025), hierarchical modeling in healthcare contexts (Pak et al., 2025), and Ridge regression techniques (Kwak et al., 2025), yet structural relationships among subfactors remained unexplored until this investigation.

This paradigmatic shift parallels developments in intelligence research, where the transition from independent ability factors to hierarchical g-factor models required reconceptualizing strong interfactor correlations as theoretically meaningful rather than methodologically problematic (Carroll, 1993). In coaching leadership research, the analogous shift involves recognizing that four conceptually distinct behavioral domains may share a common core reflecting a general coaching leadership orientation that is culturally amplified in Korean collectivistic contexts.

## 3. Methodology

### 3.1. Research Design

This study employed a Multi-Level Integration Framework designed to investigate the structural nature of coaching leadership in Korean SME contexts. The framework distinguished three analytical layers. Layer 1 (Measurement) assessed convergent validity, internal consistency, and the degree of empirical differentiation among theoretically specified subdimensions through confirmatory factor analysis, reliability indices, and HTMT analysis. Layer 2 (Structural Integration) examined empirical factor retention criteria, interfactor correlation patterns, and alternative model specifications to evaluate the extent to which coaching leadership operates as an integrated superordinate construct. Layer 3 (Invariance) tested sequential measurement equivalence across gender groups to establish cross-group score comparability. All procedures adhered to the Declaration of Helsinki (World Medical Association, 2013).

### 3.2. Participants

The final sample comprised 300 participants recruited from Korean SMEs with 50 to 300 employees through convenience sampling. All participants were employed as full-time workers in non-executive positions (general staff through department head level). Gender distribution was approximately balanced (male  $n = 154$ , 51.3%; female  $n = 146$ , 48.7%).

Age distribution was as follows: 20s ( $n = 51$ , 17.0%), 30s ( $n = 70$ , 23.3%), 40s ( $n = 82$ , 27.3%), 50s ( $n = 67$ , 22.3%), and 60 or older ( $n = 30$ , 10.0%). Organizational size comprised two categories: 50–100 employees ( $n = 178$ , 59.3%) and 100–300 employees ( $n = 122$ , 40.7%). Tenure at the current organization ranged across six categories: less than 1 year ( $n = 15$ , 5.0%), 1–3 years ( $n = 54$ , 18.0%), 3–5 years ( $n = 52$ , 17.3%), 5–10 years ( $n = 77$ , 25.7%), 10–15 years ( $n = 49$ , 16.3%), and 15 years or more ( $n = 53$ , 17.7%). The predominant occupation was office work ( $n = 238$ , 79.3%), followed by technical ( $n = 26$ , 8.7%), research and development ( $n = 19$ , 6.3%), marketing and sales ( $n = 8$ , 2.7%), production ( $n = 4$ , 1.3%), and other ( $n = 5$ , 1.7%).

The sample size of 300 was determined through a priori power analysis (Cohen, 1988), confirming sufficient power ( $1 - \beta = 0.85$ ) to detect medium effect sizes at  $\alpha = .05$ , exceeding the recommended threshold of 200 for SEM applications (Kline, 2015). It is important to note that this sample is restricted to non-executive employees in organizations with 50–300 employees. These boundary conditions limit generalizability to larger organizations, executive populations, or the public sector.

### 3.3. K-SCLS Development

The K-SCLS was developed through systematic cultural adaptation following established scale development protocols (DeVellis, 2016; Hinkin, 1998; Worthington & Whittaker, 2006). The development process comprised four sequential phases. The first phase involved extensive literature review of Korean coaching leadership research (Joh, 2021; Park, 2012; Ryu & Je, 2020) and Stowell's (1987) original coaching framework, resulting in an initial pool of 22 items distributed across four theoretically specified dimensions. The second phase consisted of cultural adaptation workshops wherein item formulations were evaluated for cultural appropriateness within Korean organizational contexts. The third phase involved expert panel evaluation (total  $N = 18$ ) comprising coaching practitioners ( $n = 6$ ), organizational behavior scholars ( $n = 6$ ), and Korean cultural psychology specialists ( $n = 6$ ), following content validation guidelines (Lawshe, 1975; Polit & Beck, 2006). Each panelist independently evaluated item relevance, clarity, and cultural appropriateness using a four-point rating scale. Item-level content validity ratios (CVR; Lawshe, 1975) were calculated, with the minimum acceptable CVR for  $N = 18$  panelists being .42 ( $p < .05$ ). All 22 initial items achieved CVR values ranging from .44 to .89 ( $M = .67$ ,  $SD = .12$ ), and the scale-level CVI was .91, exceeding the recommended threshold of .80 (Polit & Beck, 2006). The fourth phase consisted of pilot testing ( $n = 150$ ) that confirmed preliminary item adequacy and response distribution properties.

The resulting instrument comprised four dimensions measured on 7-point Likert scales ranging from 1 (*strongly disagree*) to 7 (*strongly agree*): Direction Setting (5 items), Performance Evaluation (5 items), Capability Development (5 items), and Relationship Building (7 items), totaling 22 items.

### 3.4. Common Method Bias Control

Common method bias was assessed through two sequential procedures. First, Harman's single-factor test (Podsakoff et al., 2003) was conducted as a preliminary diagnostic. Second, an Equal-loadings Unmeasured Latent Method Construct (ULMC) model (Williams et al., 2010) was specified using MLR estimation, with a common method factor loading equally on all 16 indicators, constrained orthogonal to the four substantive factors, and with method factor variance fixed to 1.0 for identification.

### 3.5. Statistical Procedures

To address the same-sample EFA-CFA limitation, a post-hoc split-half cross-validation was conducted by randomly

partitioning the full sample into an EFA subsample ( $n_1 = 150$ ) and an independent CFA subsample ( $n_2 = 150$ ; set.seed = 2024). The EFA subsample was used to replicate item selection stability, and the CFA subsample was used for independent structural verification. Confirmatory factor analyses employed WLSMV estimation for ordinal data (Brown, 2015). Advanced reliability assessment included Cronbach's alpha (Nunnally & Bernstein, 1994), McDonald's omega via semTools (Jorgensen et al., 2022), Hancock and Mueller's (2001) coefficient  $H$ , composite reliability, and average variance extracted (Hair et al., 2019). Discriminant validity was assessed through Heterotrait-Monotrait ratios with 10,000 bootstrap iterations (Henseler et al., 2015; Franke & Sarstedt, 2019). Structural integration was evaluated through multiple converging indicators, including empirical factor retention criteria (Kaiser criterion, parallel analysis, and MAP test), interfactor correlations from the correlated four-factor model, and alternative model specifications including bifactor and hierarchical higher-order models. Sequential measurement invariance testing followed established procedures (Cheung & Rensvold, 2002; Chen, 2007).

### 3.6. Mathematical Consistency Verification

Both bifactor and hierarchical CFA specifications were estimated to examine the coexistence of general and specific factors. The bifactor model, specifying a general coaching leadership factor loading on all 16 items simultaneously with four orthogonal specific factors loading on dimensionally grouped items, converged successfully with excellent fit ( $CFI = 1.000$ ) and no inadmissible solutions (zero negative residual variances). However, the hierarchical CFA specification, wherein a higher-order general factor loads onto the four first-order factors, produced a Heywood case: the Direction Setting factor exhibited a standardized higher-order loading of 1.013, accompanied by negative latent variance warnings. This inadmissible solution indicates that the interfactor correlations were sufficiently high to cause mathematical saturation of the hierarchical specification, consistent with the interpretation that the four subfactors approach redundancy rather than representing moderately correlated but empirically distinguishable dimensions. Consequently, the correlated four-factor model and the bifactor model serve as the primary analytical frameworks, supplemented by convergent evidence from factor retention criteria and interfactor correlation patterns.

## 4. Results

### 4.1. Data Quality Verification

**Table 1:** Data Quality Indicators

Indicator	Result	Criterion	Status
Sample size ( <i>N</i> )	300	≥ 200 (Kline, 2015)	Adequate
Missing values	0	—	Complete data
KMO	0.979	≥ 0.90 = excellent	Excellent
Bartlett's $\chi^2$ (231)	$df = 8,605.20$ , $p < .001$	$p < .05$	Significant
Skewness violations ( $ skew  > 3$ )	0 items	—	No violations
Kurtosis violations ( $ kurt  > 10$ )	0 items	—	No violations
Straight-line respondents	6 cases	—	Sensitivity analysis conducted
Mahalanobis outliers ( $p < .001$ )	25 cases	—	Retained; sensitivity analysis conducted

*Note.* KMO = Kaiser-Meyer-Olkin measure. WLSMV estimation was employed for CFA models given the ordinal nature of 7-point Likert response scales. Sensitivity analyses excluding straight-line respondents and Mahalanobis outliers produced substantively identical results (all  $\Delta CFI = 0.000$ ).

Data quality was excellent (Table 1). The KMO index of 0.979 indicated superb sampling adequacy, and Bartlett's test of sphericity was significant ( $\chi^2 = 8,605.20$ ,  $df = 231$ ,  $p < .001$ ). No items violated univariate normality thresholds for skewness (all  $|skew| < 0.71$ ) or kurtosis (all  $|kurt| < 0.70$ ). Six straight-line respondents and 25 Mahalanobis outliers were identified; sensitivity analyses excluding these cases produced substantively identical results (all  $\Delta CFI = 0.000$ ), and all cases were retained in the final sample.

## 4.2. Exploratory Factor Analysis and Item Optimization

**Table 2:** Factor Retention Criteria

Criterion	Suggested Factors	Interpretation
Kaiser criterion (eigenvalue > 1)	1 (eigenvalue <sub>1</sub> = 17.089; eigenvalue <sub>2</sub> = 0.883)	Strong unidimensional tendency
Parallel analysis (1,000 iterations)	2	Broad task/relationship split
MAP test (Minimum Average Partial)	2 (MAP <sub>2</sub> = 0.019; MAP <sub>1</sub> = 0.029)	Consistent with parallel analysis
Theoretical specification	4	Based on Stowell (1987) and Joh (2021, 2022, 2024)

*Note.* All criteria computed from polychoric correlation matrix. The first eigenvalue (17.089) accounted for 77.7% of common variance; the ratio of first to second eigenvalue was 19.4:1.

**Table 3:** Exploratory Factor Analysis: Pattern Matrix (Polychoric Correlation, MINRES + Oblimin, 4-Factor Solution)

Item	RB (F1)	DS (F2)	PE (F3)	CD (F4)	<i>R</i> <sup>2</sup>	Cross-ratio	Disposition
RB4	<b>.968</b>				.867	0.098	Retained
RB3	<b>.938</b>				.805	0.097	Retained
RB1	<b>.883</b>				.841	0.047	Retained
RB7	<b>.809</b>				.761	0.104	Retained
RB2	<b>.717</b>				.829	0.210	Removed
RB6	<b>.691</b>				.782	0.233	Removed
RB5	<b>.526</b>				.832	0.403	Removed
DS2		<b>.884</b>			.791	0.120	Retained
DS1		<b>.809</b>			.862	0.075	Retained
DS5		<b>.794</b>			.886	0.143	Retained
DS3		<b>.784</b>			.810	0.104	Retained
DS4		<b>.626</b>	.344		.824	0.550	Removed
PE4			<b>.871</b>		.854	0.090	Retained
PE5			<b>.808</b>		.861	0.102	Retained
PE1			<b>.713</b>		.819	0.158	Retained
PE2			<b>.614</b>		.850	0.241	Retained
PE3			<b>.510</b>		.783	0.409	Removed
CD1				<b>.491</b>	.839	0.421	Retained

Item	RB (F1)	DS (F2)	PE (F3)	CD (F4)	$h^2$	Cross-ratio	Disposition
CD2				<b>.472</b>	.800	0.600	Retained
CD4		.341		<b>.445</b>	.792	0.766	Removed
CD5				<b>.439</b>	.861	0.623	Retained
CD3				<b>.416</b>	.859	0.676	Retained

Note. Pattern coefficients from polychoric correlation-based MINRES extraction with oblimin rotation. Only coefficients  $\geq .30$  displayed. Cross-ratio = max cross-loading / primary loading. Items were selected based on composite score ( $\lambda_{\text{primary}} \times [1 - \text{cross-ratio}]$ ) with a four-items-per-dimension constraint.  $h^2$  = communality. Bold values indicate primary factor loadings. EFA model fit: RMSEA = 0.077, TLI = 0.954.

**Table 4:** Interfactor Correlations from 4-Factor EFA (Oblimin Rotation)

	DS	PE	CD	RB
DS	—			
PE	.870	—		
CD	.735	.724	—	
RB	.857	.790	.750	—

Note. All correlations significant at  $p < .001$ . DS = Direction Setting; PE = Performance Evaluation; CD = Capability Development; RB = Relationship Building.

Polychoric correlation-based EFA with MINRES extraction and oblimin rotation was conducted on the full 22-item pool. The theoretically specified four-factor structure was reproduced, with all items loading primarily on their intended dimensions (Table 3). However, two findings merit attention.

First, empirical factor retention criteria diverged substantially from the theoretical specification (Table 2). The Kaiser criterion, with a first eigenvalue of 17.089 accounting for 77.7% of common variance and a first-to-second eigenvalue ratio of 19.4:1, suggested a single factor. Both parallel analysis and the MAP test indicated two factors, consistent with a broad task-oriented versus relationship-oriented differentiation. This divergence provides initial evidence that the four coaching leadership dimensions resist empirical separation in this Korean SME

sample.

Second, interfactor correlations were exceptionally high, ranging from .724 to .870 (Table 4). The Capability Development (CD) dimension exhibited structural vulnerability, with primary pattern coefficients ranging from only .416 to .491, compared to .510–.968 for other dimensions. Three CD items (CD3, CD4, CD5) displayed cross-loading ratios exceeding .60, indicating that capability development behaviors were not clearly differentiated from other coaching leadership domains in respondents' perceptions.

Item reduction followed a composite score criterion ( $\lambda_{\text{primary}} \times [1 - \text{cross-ratio}]$ ) with a four-items-per-dimension constraint, reducing the instrument from 22 to 16 items (Table 3). Six items were removed: DS4 (lowest composite in DS), PE3 (lowest composite in PE), CD4 (highest cross-loading ratio, 0.766), RB2, RB5, and RB6 (lowest three composites in RB; RB5 also showed modification index of 32.8 with CD in subsequent CFA). The final composition retained DS1, DS2, DS3, DS5 for Direction Setting; PE1, PE2, PE4, PE5 for Performance Evaluation; CD1, CD2, CD3, CD5 for Capability Development; and RB1, RB3, RB4, RB7 for Relationship Building.

To examine item selection stability, the split-half EFA subsample ( $n_1 = 150$ ) reproduced the four-factor structure with 14 of 16 items (87.5%) selected identically to the full-sample solution. DS and PE dimensions achieved complete item correspondence (4/4 each). Two items differed: CD4 was selected in  $n_1$  instead of CD3, and RB6 was selected instead of RB7; both substitutions involved candidates with comparable primary loadings (.42–.47 for CD and .77–.89 for RB), consistent with the structural vulnerability of these dimensions documented in the full-sample EFA. Cross-validation results are summarized in Table 5.

**Table 5.** Split-Half Cross-Validation: EFA and CFA Results Across Subsamples

Index	Full Sample ( $N = 300$ )	EFA Subsample ( $n_1 = 150$ )	CFA Subsample ( $n_2 = 150$ )
Factor structure	4-factor	4-factor reproduced	4-factor confirmed
Item correspondence with full sample	— (reference)	87.5% (14/16)	—
DS	4/4	4/4 ✓	—

Index	Full Sample (N=300)	EFA Subsample ( $n_1 = 150$ )	CFA Subsample ( $n_2 = 150$ )
dimension correspondence			
PE dimension correspondence	4/4	4/4 ✓	—
CD dimension correspondence	4/4	3/4 (CD4↔CD3) <sup>a</sup>	—
RB dimension correspondence	4/4	3/4 (RB6↔RB7) <sup>a</sup>	—
CFI (4-factor CFA)	1.000	—	1.000
RMSEA	0.008	—	0.000
SRMR	0.019	—	0.026
Cronbach's $\alpha$ range	.938–.949	—	.941–.946
ECV (Bifactor)	.901	—	.785 <sup>b</sup>
$\omega_H$	.958	—	— <sup>c</sup>

Note. Split-half subsamples were created via random assignment (set.seed = 2024). EFA subsample ( $n_1$ ) was used for item retention replication; CFA subsample ( $n_2$ ) was used for independent confirmatory evaluation. — = analysis not applicable or not conducted in this subsample. DS = Direction Setting; PE = Performance Evaluation; CD = Capability Development; RB = Relationship Building; ECV = explained common variance;  $\omega_H$  = omega hierarchical. <sup>a</sup> Substituted items exhibited comparable primary loadings (CD4: .445 vs. CD3: .416; RB6: .887 vs. RB7: .809), indicating boundary-level differentiation within structurally vulnerable dimensions. <sup>b</sup> ECV approximated from the unidimensional CFA general factor solution ( $n_2$ ); direct bifactor estimation was not feasible in  $n_2 = 150$  due to model complexity relative to sample size (Reise, 2012). <sup>c</sup>  $\omega_H$  could

not be directly estimated via bifactor CFA in  $n_2 = 150$ ; omega hierarchical estimated via the Schmid-Leiman procedure (psych::omega) yielded  $\omega_H = .901$ , exceeding the .50 threshold for reliable total score use.

### 4.3. Confirmatory Factor Analysis: Competing Model Comparison

**Table 6.** Confirmatory Factor Analysis: Competing Model Comparison (16-Item K-SCLS)

Model	$\chi^2$	df	CFI	TLI	RMSEA [90% CI]	SRMR
1-factor	495.054	104	0.998	0.997	0.112 [0.102, 0.122]	0.041
2-factor (task/relationship)	358.961	103	0.998	0.998	0.091 [0.081, 0.102]	0.035
3-factor (goal-mgmt/CD/RB)	169.831	101	1.000	0.999	0.048 [0.035, 0.060]	0.024
<b>4-factor (correlated)</b>	<b>100.000</b>	<b>98</b>	<b>1.000</b>	<b>1.000</b>	<b>0.008 [0.000, 0.032]</b>	<b>0.019</b>
Bifactor (G + 4 specific)	102.125	88	1.000	1.000	0.023 [0.000, 0.041]	0.020

Note. All models estimated with WLSMV for ordinal indicators. Bold row indicates the selected baseline model. The bifactor model is reported for supplementary structural analysis (see Section 4.7).

Five competing measurement models were estimated using WLSMV (Table 6). The unidimensional model yielded adequate CFI (0.998) but unacceptable RMSEA (0.112), indicating that a single factor, while capturing substantial common variance, failed to adequately represent the item covariance structure. The two-factor model (task/relationship split) showed marginal improvement (RMSEA = 0.091), and the three-factor model (goal-management/CD/RB) achieved substantially better fit (RMSEA = 0.048). The correlated four-factor model demonstrated the best fit among standard CFA specifications (CFI = 1.000, TLI = 1.000, RMSEA = 0.008;  $\chi^2[98] = 100.000$ ; Table 6).

It is important to note that the correlated four-factor specification achieved superior fit because it freely estimates interfactor correlations, which approached unity. The theoretical four-factor structure provides the best mathematical representation of item covariances, while the magnitude of interfactor correlations, reported below, reveals that the four dimensions share the vast majority of

their reliable variance. A noteworthy pattern across competing models is that CFI values remained extremely high even for the unidimensional specification (CFI = 0.998), reflecting the high interfactor correlations that make all models reasonably consistent with the observed covariance structure. The RMSEA, which is more sensitive to model misspecification, showed clear progressive improvement from one factor (0.112) through two (0.091), three (0.048), to four factors (0.008), confirming the superiority of the four-factor specification while underscoring the substantial shared variance underlying all dimensions.

#### 4.4. Convergent Validity and Reliability

**Table 7:** Reliability and Convergent Validity: 16-Item K-SCLS

Dimension	Items	$\alpha$	$\omega$	$H$	FSD	AVE	CR	Mean $\lambda$	$\lambda$ range
Direction Setting	DS1, DS2, DS3, DS5	.943	.938	.958	.968	.832	.952	.912	.879–.952
Performance Evaluation	PE1, PE2, PE4, PE5	.949	.942	.958	.971	.843	.956	.918	.897–.944
Capability Development	CD1, CD2, CD3, CD5	.946	.940	.955	.969	.834	.953	.913	.885–.936
Relationship Building	RB1, RB3, RB4, RB7	.938	.936	.950	.967	.822	.948	.906	.887–.927

*Note.*  $\alpha$  = Cronbach's alpha;  $\omega$  = McDonald's omega;  $H$  = Hancock and Mueller's (2001) coefficient  $H$ ; FSD = factor score determinacy (approximated as  $\sqrt{\omega}$ ); AVE = average variance extracted; CR = composite reliability;  $\lambda$  = standardized factor loading from WLSMV estimation. All indices exceed recommended thresholds:  $\alpha > .70$  (Nunnally & Bernstein, 1994),  $H > .70$  (Hancock & Mueller, 2001), FSD  $> .90$ , AVE  $> .50$  (Fornell & Larcker, 1981), CR  $> .70$  (Hair et al., 2019).

Table 7 presents comprehensive reliability and convergent validity results. All four dimensions demonstrated excellent internal consistency, with Cronbach's  $\alpha$  ranging from .938 to .949 and McDonald's  $\omega$  from .936 to .942. Standardized factor loadings ranged from .879 (DS2) to .952 (DS5), all substantially exceeding the .70 threshold (Brown, 2015). AVE values (.822–.843) and composite reliabilities (.948–.956) provided strong evidence of convergent validity. Hancock and Mueller's (2001) coefficient  $H$ , which provides an upper-bound estimate of construct reliability that is less sensitive to the number of indicators than traditional alpha, ranged from .950 to .958 across the four dimensions,

indicating excellent construct replicability. Factor score determinacy (FSD), approximated as the square root of McDonald's omega (Grice, 2001), ranged from .967 to .971, substantially exceeding the .90 threshold considered necessary for meaningful factor score interpretation.

#### 4.5. Discriminant Validity and Structural Integration Evidence

**Table 8:** Discriminant Validity Assessment: HTMT Analysis and Interfactor Correlations

Dimension Pair	HTMT	Bootstrap 95% CI	CFA $\phi$	Status
DS – PE	0.923	[0.857, 0.968]	.934	Not achieved
DS – CD	0.933	[0.907, 0.956]	.934	Not achieved
DS – RB	0.880	[0.833, 0.918]	.875	Not achieved
PE – CD	0.929	[0.860, 0.976]	.943	Not achieved
PE – RB	0.843	[0.772, 0.901]	.845	Not achieved
CD – RB	0.896	[0.852, 0.932]	.892	Not achieved

*Note.* HTMT = Heterotrait-Monotrait ratio (Henseler et al., 2015). Bootstrap 95% confidence intervals based on 10,000 iterations. CFA  $\phi$  = standardized latent factor correlation from the correlated four-factor WLSMV model. The conservative threshold for discriminant validity requires HTMT  $< 0.85$  and the 95% CI upper bound  $< 0.85$  (Franke & Sarstedt, 2019). Failure to achieve conventional discriminant validity thresholds is interpreted as reflecting substantive construct integration rather than scale construction inadequacy; see Section 5.1 for theoretical elaboration.

Table 8 presents the discriminant validity assessment. All six HTMT point estimates ranged from 0.843 to 0.933, with five of six pairs exceeding the conservative 0.85 threshold. The PE–RB pair (HTMT = 0.843) fell marginally below 0.85 at the point estimate level, but its bootstrap 95% CI upper bound (0.901) substantially exceeded the threshold. Latent interfactor correlations from CFA ranged from .845 (PE–RB) to .943 (PE–CD), confirming near-unity relationships across all dimension pairs.

These results indicate that the four K-SCLS dimensions do not meet conventional standards for discriminant validity. Rather than constituting a psychometric failure, this finding is interpreted as the central empirical contribution of this study. The convergence of three independent lines of evidence—HTMT ratios approaching or exceeding 0.90,

latent correlations of .845–.943, and factor retention criteria suggesting one to two factors—provides robust empirical support for the proposition that coaching leadership in Korean SME contexts operates as a highly integrated superordinate construct rather than as a set of empirically independent dimensions. The four dimensions retain value as conceptual organizers for content coverage and intervention design, but at the measurement level, they function as highly correlated facets of a unified coaching leadership syndrome.

#### 4.6. Common Method Variance

**Table 9.** Common Method Variance Assessment

Method	Variance Estimate	Model Fit
Harman's single-factor (unrotated PCA)	76.1%	—
ULMC model (equal loadings, MLR)	43.0%	CFI = 0.993, TLI = 0.991, RMSEA = 0.038 [0.023, 0.052], SRMR = 0.036

*Note.* Harman's test conducted as preliminary diagnostic (Podsakoff et al., 2003). ULMC = Unmeasured Latent Method Construct model (Williams et al., 2010). ULMC method factor mean loading = 0.655 (range: 0.621–0.696). Method factor variance fixed to 1.0; orthogonal to substantive factors.

Harman's single-factor test indicated that 76.1% of total variance was accounted for by a single unrotated component, substantially exceeding the 50% concern threshold (Table 9). Application of the Equal-loadings ULMC model reduced the estimated method variance to 43.0%, with the ULMC model achieving acceptable fit (CFI = 0.993, RMSEA = 0.038).

The residual method variance of 43.0% is higher than typically reported in leadership research employing ULMC controls and constitutes a meaningful limitation. Three considerations contextualize this finding. First, the high Harman's percentage is partly attributable to the genuinely integrated structure of the construct: when a construct is substantively near-unidimensional, single-factor extraction necessarily captures a large proportion of variance regardless of method effects. Second, the self-report, single-source design inherently inflates method variance estimates, and multi-source designs are essential for separating substantive integration from method-driven covariation. Third, despite incomplete remediation, the ULMC-controlled model maintained excellent fit for the four-factor structure, suggesting that the dimensional pattern is not solely an artifact of shared method variance.

#### 4.7. Hierarchical and Bifactor Model Results

**Table 10:** Higher-Order and Bifactor Model Summary

Model	CFI	TLI	RMSEA [90% CI]	SRMR	$\chi^2(df)$	Heywood Cases
Correlated 4-factor (baseline)	1.000	1.000	0.008 [0.000, 0.032]	0.019	100.000(98)	0
Hierarchical (2nd-order)	0.998	0.998	0.092 [0.081, 0.102]	0.034	354.638(101)	1 (DS $\gamma$ = 1.013)
Bifactor (orthogonal specific)	1.000	1.000	0.023 [0.000, 0.041]	0.020	102.125(88)	0

*Note.* Hierarchical model  $\gamma$  values: DS = 1.013 (Heywood), PE = 0.947, CD = 0.970, RB = 0.892.

**Table 11:** Bifactor Model: General and Specific Factor Loadings

Item	Dimension	$\lambda_{_G}$	$\lambda_{_S}$	$\lambda_{_G^2}$	$\lambda_{_S^2}$	$\lambda_{_G^2} + \lambda_{_S^2}$
DS1	Direction Setting	.890	.249	.792	.062	.854
DS2	Direction Setting	.844	.298	.712	.089	.801
DS3	Direction Setting	.865	.222	.748	.049	.798
DS5	Direction Setting	.923	.200	.852	.040	.892
PE1	Perf. Evaluation	.874	.260	.764	.068	.831
PE2	Perf. Evaluation	.909	.205	.826	.042	.868
PE4	Perf. Evaluation	.858	.292	.736	.085	.821
PE5	Perf. Evaluation	.882	.294	.778	.086	.864
CD1	Cap. Development	.881	.291	.776	.085	.861
CD2	Cap. Development	.863	.247	.745	.061	.806
CD3	Cap. Development	.922	.075	.850	.006	.856
CD5	Cap. Development	.908	.159	.824	.025	.850

Item	Dimension	$\lambda_G$	$\lambda_S$	$\lambda_{G^2}$	$\lambda_{S^2}$	$\lambda_{G^2} + \lambda_{S^2}$
RB1	Rel. Building	.835	.395	.697	.156	.853
RB3	Rel. Building	.802	.406	.643	.165	.808
RB4	Rel. Building	.824	.416	.679	.173	.852
RB7	Rel. Building	.801	.364	.642	.133	.774

Note.  $\lambda_G$  = standardized loading on the general coaching leadership factor;  $\lambda_S$  = standardized loading on the orthogonal specific factor. All loadings from WLSMV bifactor model with orthogonal constraints. Zero negative residual variances.

**Table 12:** Bifactor Model: Variance Decomposition Summary

Indicator	Value	Interpretation
<b>ECV (total)</b>	<b>.901</b>	General factor accounts for 90.1% of common variance
ECV (DS)	.928	
ECV (PE)	.917	
ECV (CD)	.948	
ECV (RB)	.809	RB retains most specific variance
<b><math>\omega_H</math> (general)</b>	<b>.958</b>	Total score primarily reflects general factor
$\omega_S$ (DS)	.060	Negligible unique DS contribution
$\omega_S$ (PE)	.070	Negligible unique PE contribution
$\omega_S$ (CD)	.044	Negligible unique CD contribution

Indicator	Value	Interpretation
$\omega_S$ (RB)	.157	Small but largest unique contribution

Note.  $\omega_H$  was originally estimated as .754 using SmartPLS. Re-analysis with R lavaan (WLSMV estimator) and semTools::compRelSEM yielded  $\omega_H = .958$ , further supporting the dominance of the general factor. The discrepancy reflects estimator differences (PLS-based vs. SEM-based omega); the higher value strengthens the unidimensionality interpretation.

The hierarchical CFA produced an inadmissible solution, with the Direction Setting factor exhibiting a standardized higher-order loading of 1.013, exceeding the theoretical maximum of 1.0 (Table 10). The bifactor model converged without inadmissible solutions (CFI = 1.000, RMSEA = 0.023, zero negative residual variances).

Table 11 presents the complete bifactor loading structure. General factor loadings ranged from .801 (RB7) to .923 (DS5), indicating that all items were strongly associated with the general coaching leadership construct. Specific factor loadings were substantially weaker, ranging from .075 (CD3) to .416 (RB4). The Relationship Building specific factor exhibited the strongest loadings (.364-.416), while the Capability Development specific factor showed the weakest (.075-.291).

Variance decomposition results (Table 12) provide the most precise characterization of the integration pattern. The overall ECV of .901 indicates that the general coaching leadership factor accounts for 90.1% of common variance, substantially exceeding the .70 threshold at which Reise et al. (2013) recommend interpreting the instrument as essentially unidimensional. The omega hierarchical coefficient ( $\omega_H = .958$  via semTools with WLSMV estimation for ordered indicators) indicates that 95.8% of total score variance is attributable to the general factor, supporting the use of a K-SCLS total score as a reliable measure of general coaching leadership. Specific factor omega values were uniformly low ( $\omega_S = .044-.157$ ), indicating that subscale scores contribute negligible unique reliable variance beyond the general factor.

#### 4.8. Measurement Invariance

**Table 13:** Sequential Measurement Invariance Across Gender Groups

Invariance Level	CFI	TLI	RMSEA	SRMR	ACFI	ARMSEA	Decision
Configural	1.000	1.000	0.000 [0.000,	0.025	—	—	Supported

Invariance Level	CFI	TLI	RMSEA	SRMR	ΔCFI	ΔRMSEA	Decision
			0.011]				
			0.059				
Metric	0.999	0.999	[0.045, 0.071]	0.035	-0.001	+0.059 <sup>a</sup>	Supported
Full scalar	0.999	0.999	[0.047, 0.073]	0.035	0.000	+0.001	Supported

Note. Male  $n = 154$ ; female  $n = 146$ . Evaluation criteria follow Chen (2007):  $\Delta CFI > -0.010$  for both metric and scalar invariance. <sup>a</sup>The  $\Delta RMSEA$  of  $+0.059$  at the metric level reflects the artificially low baseline RMSEA of 0.000

in the configural model rather than substantive model deterioration;  $\Delta CFI$  provides the appropriate evaluation criterion.

Sequential measurement invariance testing across gender groups is presented in Table 13. Configural invariance was supported ( $CFI = 1.000$ ), confirming equivalent factor structure across groups. Metric invariance was achieved with minimal fit degradation ( $\Delta CFI = -0.001$ ). Full scalar invariance was also achieved ( $\Delta CFI = 0.000$  from metric to scalar), indicating that item intercepts did not differ systematically between male and female respondents. Full scalar invariance supports direct cross-gender mean comparisons of K-SCLS scores without adjustment.

### 4.9. Multi-Level Integration Framework Synthesis

Table 14: Multi-Level Integration Framework: Comprehensive Validation Evidence

Analytical Layer	Indicator	Value	Criterion	Interpretation
Layer Measurement	1: CFA Fit (CFI/TLI)	1.000/1.000	$\geq 0.95$	Excellent
	RMSEA [90% CI]	0.008 [0.000, 0.032]	$\leq 0.05$	Excellent
	SRMR	0.019	$\leq 0.08$	Excellent
	Standardized $\lambda$ range	.879–.952	$\geq .70$	All adequate
	$\alpha$ range	.938–.949	$\geq .70$	Excellent
	$\omega$ range	.936–.942	$\geq .70$	Excellent
	Coefficient $H$ range	.950–.958	$> .70$	Excellent
	AVE range	.822–.843	$> .50$	Excellent convergent validity
	CR range	.948–.956	$> .70$	Excellent
Layer 2: Integration	HTMT range	0.843–0.933	$< 0.85$	Not achieved ( $5/6 > 0.85$ )
	CFA interfactor range	$\phi$ .845–.943	—	Near-unity
	Kaiser criterion	1 factor (eigenvalue <sub>1</sub> = 17.089)	= —	Unidimensional tendency
	Parallel analysis / MAP2 factors	2 factors	—	Broad task/relationship
	Bifactor ECV (total)	.901	$> .70$ essentially unidimensional	Extremely strong general factor
	Bifactor $\omega_H$	.958	$> .50$ supports total score	Total score reliable
	Bifactor $\omega_S$ range	.044–.157	$> .20$ for meaningful subscale	All below threshold
	Hierarchical CFA	DS $\gamma = 1.013$ (Heywood)	—	Mathematical saturation
	Harman single-factor	76.1%	$< 50\%$	Exceeded (partly substantive)
	ULMC variance method	43.0%	—	Partial remediation
Layer 3: Invariance	Configural	Supported ( $CFI = 1.000$ )	$\Delta CFI > -0.010$	Equivalent structure
	Metric	Supported ( $\Delta CFI = -0.001$ )	$\Delta CFI > -0.010$	Equivalent loadings
	Full scalar	Supported ( $\Delta CFI = 0.000$ )	$\Delta CFI > -0.010$	Equivalent intercepts

Note. Layer 1 addresses convergent validity and internal consistency (RQ1). Layer 2 addresses structural integration (RQ2), with bifactor indices providing the most precise characterization. Layer 3 addresses cross-group measurement equivalence (RQ3).

Table 14 synthesizes the validation evidence across all three analytical layers. Regarding Research Question 1, the K-

SCLS achieved excellent convergent validity and internal consistency, with all indicators substantially exceeding

conventional thresholds. However, HTMT analysis revealed that the four dimensions did not achieve conventional discriminant validity.

Regarding Research Question 2, converging evidence from multiple independent indicators demonstrated that the four coaching leadership dimensions are highly integrated. HTMT ratios (0.843–0.933), interfactor correlations (.845–.943), factor retention criteria (1–2 empirical factors), bifactor ECV (.901), and the hierarchical CFA Heywood case collectively indicated that the K-SCLS measures a near-unidimensional construct with conceptually meaningful but empirically fused subdimensions.

Regarding Research Question 3, the K-SCLS achieved full scalar invariance across gender groups, supporting the cross-gender comparability of scores without psychometric adjustment.

## 5. Discussion

### 5.1. Reframing the Korean Coaching Leadership Measurement Paradox

The most significant contribution of this study is the empirical demonstration that the Korean coaching leadership measurement paradox identified by Joh (2021, 2022, 2024) reflects a substantively meaningful structural characteristic rather than a methodological artifact. For two decades, the persistent failure of Korean coaching leadership studies to achieve discriminant validity among subfactors was interpreted as evidence of inadequate scale construction. The present findings suggest an alternative interpretation: when Korean SME employees evaluate their supervisors' coaching leadership, they perceive direction setting, performance evaluation, capability development, and relationship building not as four independent behavioral domains but as highly integrated facets of a unified coaching leadership syndrome.

This interpretation is supported by converging evidence from multiple analytical methods. Factor retention criteria empirically favored one to two factors rather than four. Interafactor correlations from the correlated four-factor CFA ranged from .845 to .943. HTMT ratios of .843 to .933 confirmed that the four dimensions share the vast majority of their reliable variance. The hierarchical CFA produced a Heywood case ( $DS \gamma = 1.013$ ), directly demonstrating that the interfactor correlations were too high for the higher-order structure to yield admissible estimates. Critically, partial method variance remediation through ULMC (from 76.1% to 43.0%) did not eliminate the integration pattern, suggesting that the integration is not solely attributable to common method bias, though the residual 43.0% method variance necessitates multi-source replication.

The bifactor model results provide the most granular characterization of this integration pattern. The ECV of .901 indicates that over 90% of common variance among K-SCLS items is attributable to a single general coaching leadership factor. This value substantially exceeds the .70 threshold at which Reise et al. (2013) recommend treating a measure as essentially unidimensional. The omega hierarchical ( $\omega_H = .958$  via semTools with WLSMV estimation for ordered indicators) confirms that K-SCLS total scores predominantly reflect this general factor, while specific factor omega values ( $\omega_S = .044-.157$ ) indicate that the four dimension scores contribute negligible unique reliable variance beyond the general factor. Total scores are the psychometrically appropriate operational metric, while dimension-level scores should be interpreted with extreme caution.

A legitimate reviewer concern warrants direct address: if  $ECV = .901$  and  $\omega_S$  values uniformly fall below .157, why retain four content domains rather than developing a parsimonious unidimensional scale? Three complementary justifications support the four-domain structure. First, from a content validity standpoint, the four domains ensure comprehensive behavioral sampling across the theoretical space of coaching leadership—direction setting, performance evaluation, capability development, and relationship building—thereby reducing construct underrepresentation (Loevinger, 1957). A four-item unidimensional scale would sacrifice coverage breadth essential for distinguishing coaching leadership from adjacent constructs such as transformational leadership or leader-member exchange. Second, from a bifactor scoring perspective, the general factor score and domain-specific residual scores can be computed simultaneously (Reise, 2012), providing practitioners with a psychometrically defensible total score for overall coaching leadership assessment while retaining domain profiles as supplementary diagnostic indicators for targeted developmental feedback. Third, and most practically,  $\omega_S$  for Relationship Building (.157), while below the .20 threshold, is substantially larger than the remaining three domains (.044–.070), suggesting nascent differentiation that may become empirically detectable with multi-source measurement designs that reduce method-driven covariation. These justifications collectively support retaining the four-domain structure as a content blueprint for the general factor, while recommending the K-SCLS total score as the primary operational metric.

### 5.2. Cultural Interpretation of the Integration Pattern

The extreme integration observed—with interfactor correlations approaching unity—is theoretically consistent

with cultural constructs that may suppress behavioral differentiation in Korean organizational leadership perception. *Jeong*, representing deep emotional bonds extending beyond professional boundaries (Jung & Lee, 2019), creates expectations that leaders demonstrate genuine personal care simultaneously with performance accountability, potentially fusing relationship building with performance evaluation in subordinates' perceptions. *Woori* consciousness, reflecting collective identity orientations (Choi & Kim, 2020), positions leaders as holistic facilitators of group functioning rather than specialists in discrete behavioral domains. *Chemyon* considerations (J.-H. Kim, 2020; Lee et al., 2019) create natural integration between evaluation functions and relationship maintenance through dignity-preserving approaches.

The bifactor ECV pattern across dimensions provides nuanced evidence relevant to cultural interpretation. The lowest ECV was observed for Relationship Building (ECV = .809), suggesting that RB behaviors retain the most distinctiveness from the general coaching leadership evaluation, potentially reflecting the recognizable interpersonal qualities associated with *jeong*. In contrast, Capability Development showed the highest ECV (.948), suggesting that competency development behaviors are almost entirely absorbed into the general coaching leadership perception, consistent with the interpretation that *woori* consciousness positions capability development as an intrinsic and inseparable aspect of the leader-follower relationship.

However, these cultural mechanisms were not directly measured. Alternative explanations remain plausible, including shared organizational context effects, social desirability or acquiescence bias, construct breadth effects, or genuine construct unidimensionality that is not culturally specific. Future research incorporating direct measurement of cultural constructs and multi-cultural comparison samples is essential.

### 5.3. Direction Setting as Culturally Adaptive Guidance

Direction Setting can be reconceptualized as "guidance as a compass" rather than directive control. Korean organizational contexts emphasize collective decision-making processes that integrate individual contributions within group frameworks, possibly associated with Direction Setting functioning as collaborative orientation rather than imposed prescription, a proposition requiring direct empirical examination in future research. The Western emphasis on non-directiveness reflects historically contingent legal defense mechanisms rather than universal principles (Nishigaki, 2021), and Korean coaching leadership's inclusion of Direction Setting represents

cultural adaptation free from such constraints.

### 5.4. Measurement Invariance Implications

The achievement of full scalar invariance across gender groups enables direct cross-gender mean comparisons without adjustment. This finding supports the K-SCLS's applicability across gender groups within this specific organizational context and simplifies practical implementation compared to gender-specific norms that would have been required under partial invariance. The approximately balanced gender composition (51.3% male), the restriction to non-executive employees, and the highly integrated structure of the construct may have contributed to this result.

### 5.5. Methodological Contributions and Limitations

The Multi-Level Integration Framework provides a structured approach for characterizing the full spectrum of dimensional relationships, moving beyond the binary question of whether discriminant validity holds. By integrating convergent validity, structural integration indicators, and invariance testing, the framework enables systematic investigation of culturally situated construct structures. The bifactor model, rather than the hierarchical CFA, proved to be the appropriate analytical framework for examining general-specific variance structure in data characterized by extreme interfactor correlations, a methodological lesson applicable to other cross-cultural measurement studies.

However, several methodological limitations require acknowledgment. The most significant concern is the residual CMV of 43.0% after ULMC application. Three considerations partially contextualize, though do not resolve, this limitation. First, the high Harman's single-factor percentage (76.1%) is partly attributable to the genuinely integrated structure of the construct: when a construct approaches unidimensionality, single-factor extraction necessarily captures substantial variance regardless of method effects—a mathematical property rather than exclusively a bias indicator (Richardson et al., 2009). Critically, EFA interfactor correlations (.724–.870), estimated before CFA constraints, were already extremely high, and these EFA estimates are less susceptible to method inflation than CFA-derived parameters because they do not impose a priori factor structures. The convergence of EFA and CFA integration evidence suggests the integration pattern reflects genuine structural characteristics rather than solely shared-method covariation. Second, the ULMC model's mean method loading of .655 across all 16 indicators is consistent with theoretical frameworks positing that construct-relevant perceptual processes (i.e., halo

effects in supervisor evaluation) may themselves represent substantive variance partially distinct from random method noise (Podsakoff et al., 2003). Third, sensitivity analyses excluding straight-line respondents and multivariate outliers produced substantively identical results (all  $\Delta CFI = .000$ ), indicating that extreme response tendencies did not drive the integration pattern. Nonetheless, the unresolved method variance constitutes a genuine boundary condition: the present findings establish a lower-bound estimate of construct integration that requires corroboration through multi-source replication before the integration pattern can be attributed confidently to substantive perceptual processes. Single-source, self-report assessment inherently confounds genuine behavioral covariation with shared method variance. Although EFA-based item selection and CFA validation were initially conducted within the same full sample ( $N = 300$ ), a post-hoc split-half cross-validation ( $n_1 = n_2 = 150$ ) partially addressed this limitation by demonstrating item selection stability (87.5% correspondence, with DS and PE achieving complete replication) and confirming the four-factor structure in an independent subsample ( $CFI = 1.000$ ,  $RMSEA = 0.000$ ). The bifactor model could not be estimated directly in  $n_2 = 150$  due to model complexity relative to sample size (Reise, 2012); however, convergent evidence from the Schmid-Leiman procedure and ECV approximation supported the dominant general factor pattern. Independent external validation with a new sample remains essential. The CD dimension exhibited structural vulnerability in EFA (primary loadings .416–.491), raising questions about whether its items adequately differentiate capability development from other coaching domains.

## 5.6. Practical Implications

Pending confirmatory validation, the present findings suggest several preliminary practical implications. The highly integrated structure suggests that dimension-specific interventions may be less effective than holistic coaching leadership development approaches. Korean SME employees appear to evaluate coaching leadership as a unified phenomenon, and interventions targeting isolated dimensions may not align with subordinates' integrated perceptual framework.

The K-SCLS total score, computed as the mean of all 16 items, may serve as a more psychometrically appropriate operational metric than four separate dimension scores, given the near-unity interfactor correlations. While the four dimensions retain value as conceptual organizers for training design, practitioners should interpret dimension-level scores with awareness that they share the vast majority of their variance. The full scalar invariance supports the use of combined-gender norms.

These recommendations are preliminary and contingent on

confirmatory validation through longitudinal multi-source research designs.

## 5.7. Limitations and Future Research

This exploratory investigation acknowledges several limitations. The cultural mechanisms proposed as drivers of integration (*jeong*, *woori* consciousness, *chemyon*) were not directly measured. The cross-sectional design precludes causal inference. The sample is restricted to non-executive employees in SMEs with 50–300 employees. Residual CMV of 43.0% remains substantial. The Heywood case prevented hierarchical CFA parameterization. Although EFA and CFA were initially conducted on the same full sample, a post-hoc split-half cross-validation replicated item selection (87.5% correspondence) and the four-factor structure ( $CFI = 1.000$ ) in an independent subsample, partially addressing this limitation. The CD dimension showed structural vulnerability (EFA loadings .416–.491). Future research should prioritize five directions. First, although split-half cross-validation confirmed item selection stability and structural replicability within the present dataset, independent external validation with a fully new sample remains essential for establishing generalizability beyond the current sampling frame. Second, multi-source data collection incorporating supervisor self-reports, peer ratings, and objective behavioral indicators is critical for disentangling substantive integration from common method variance. Third, direct measurement of *jeong*, *woori* consciousness, and *chemyon* through validated cultural scales, combined with mediation modeling, is needed to test the proposed cultural integration mechanisms. Fourth, cross-cultural comparison studies examining whether the extreme integration pattern is specific to Korean contexts or generalizes to other East Asian collectivistic cultures would clarify boundary conditions. Fifth, longitudinal designs with minimum three-wave assessments are necessary to establish temporal relationships between coaching leadership and organizational outcomes.

## 6. Conclusions

This study makes three interrelated contributions. Methodologically, it introduces bifactor modeling as an analytical framework for characterizing the integration pattern in Korean coaching leadership measurement, with split-half cross-validation confirming structural replicability, and provides ECV,  $\omega H$ , and  $\omega S$  indices absent from prior Korean organizational research. Empirically, it demonstrates that the four theoretically specified K-SCLS dimensions operate as content facets of a dominant general factor ( $ECV = .901$ ), establishing the psychometric

foundation for a general coaching leadership index rather than a profile of independent subdimension scores. Theoretically, it reframes two decades of discriminant validity failures as a substantively meaningful measurement paradox consistent with collectivistic organizational culture, redirecting research attention from scale remediation toward indigenous construct theory development.

This study developed and provided initial exploratory validation evidence for the K-SCLS within a Korean SME context. The instrument demonstrated excellent measurement properties, with strong convergent validity (AVE = .822–.843), high internal consistency ( $\alpha$  = .938–.949,  $\omega$  = .936–.942), and full measurement invariance across gender groups (scalar  $\Delta$ CFI = 0.000).

The central empirical finding is that the four theoretically specified dimensions are highly integrated in Korean SME employees' perceptions. Bifactor modeling revealed that a general coaching leadership factor accounts for 90.1% of common variance (ECV = .901), with omega hierarchical of .958 confirming that total scores predominantly reflect this general factor. Specific factor contributions were uniformly low ( $\omega$ S = .044–.157), indicating that dimension-level scores contribute negligible unique reliable variance. Interfactor correlations (.845–.943), HTMT ratios (.843–.933), and factor retention criteria (Kaiser = 1 factor) provided converging triangulated support. This finding reframes the Korean coaching leadership measurement paradox from a problem requiring psychometric remediation to a substantively meaningful phenomenon warranting theoretical explanation.

The integration pattern is theoretically consistent with cultural constructs including *jeong*, *woori* consciousness, and *chemyon* as interpretive lenses, though these were not directly measured and alternative explanations cannot be excluded. Significant limitations constrain the strength of current conclusions. Residual CMV of 43.0%, the absence of fully independent external cross-validation beyond the current dataset, the Heywood case preventing hierarchical model parameterization, and the cross-sectional single-source design collectively necessitate cautious interpretation. Confirmatory validation through longitudinal, multi-source designs with direct cultural mechanism measurement and independent samples remains essential before definitive theoretical or practical claims can be established.

## References

- Anderson, J. C., & Gerbing, D. W. (1988). Structural equation modeling in practice: A review and recommended two-step approach. *Psychological Bulletin*, *103*(3), 411–423. <https://doi.org/10.1037/0033-2909.103.3.411>
- Avery, G. C., & Bergsteiner, H. (2011). Sustainable leadership practices for enhancing business resilience and performance. *Strategy & Leadership*, *39*(5), 5–15. <https://doi.org/10.1108/10878571111128766>
- Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.
- Carroll, J. B. (1993). *Human cognitive abilities: A survey of factor-analytic studies*. Cambridge University Press.
- Chen, F. F. (2007). Sensitivity of goodness-of-fit indexes to lack of measurement invariance. *Structural Equation Modeling*, *14*(3), 464–504. <https://doi.org/10.1080/10705510701301834>
- Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling*, *9*(2), 233–255. [https://doi.org/10.1207/S15328007SEM0902\\_5](https://doi.org/10.1207/S15328007SEM0902_5)
- Cho, E. H., & Tak, J. K. (2011). Development and validity of the coaching leadership scale. *Korean Journal of Industrial and Organizational Psychology*, *24*(1), 127–155. <https://doi.org/10.24230/ksiop.24.1.201102.127>
- Choi, C.-Y. (2003). *A study on the effect of coaching leadership on organizational effectiveness* (Unpublished master's thesis, Suwon University).
- Choi, J.-S., & Kim, H.-J. (2020). Cultural values and team performance in Korean organizations. *Asian Business & Management*, *19*(1), 85–106. <https://doi.org/10.1057/s41291-019-00074-6>
- Coaching and Psychotherapy Regulation Act, S.B. 48, 66th Utah Legislature (2025). <https://le.utah.gov/~2025/bills/static/SB0048.html>
- Cohen, J. (1988). *Statistical power analysis for the behavioral sciences* (2nd ed.). Lawrence Erlbaum Associates.
- DeVellis, R. F. (2016). *Scale development: Theory and applications* (4th ed.). Sage.
- 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>
- Franke, G., & Sarstedt, M. (2019). Heuristics versus statistics in discriminant validity testing. *Internet Research*, *29*(3), 430–463. <https://doi.org/10.1108/IntR-12-2017-0515>
- Freeman, L. C. (1978). Centrality in social networks: Conceptual clarification. *Social Networks*, *1*(3), 215–239. [https://doi.org/10.1016/0378-8733\(78\)90021-7](https://doi.org/10.1016/0378-8733(78)90021-7)
- Grice, J. W. (2001). Computing and evaluating factor scores. *Psychological Methods*, *6*(4), 430–450. <https://doi.org/10.1037/1082-989X.6.4.430>
- Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2019). *A primer on partial least squares structural equation modeling (PLS-SEM)* (2nd ed.). Sage.
- Hancock, G. R., & Mueller, R. O. (2001). Rethinking construct reliability within latent variable systems. In R. Cudeck, S. du Toit, & D. Sörbom (Eds.), *Structural equation modeling: Present and future* (pp. 195–216). Scientific Software International.
- Hargreaves, A., & Fink, D. (2006). *Sustainable leadership*. Jossey-Bass.
- Henseler, J., Ringle, C. M., & Sarstedt, M. (2015). A new criterion for assessing discriminant validity. *Journal of the Academy*

- of *Marketing Science*, 43(1), 115–135.  
<https://doi.org/10.1007/s11747-014-0403-8>
- Hinkin, T. R. (1998). A brief tutorial on the development of measures for use in survey questionnaires. *Organizational Research Methods*, 1(1), 104–121.  
<https://doi.org/10.1177/109442819800100106>
- Hofstede, G. (2001). *Culture's consequences* (2nd ed.). Sage.
- Holzinger, K. J., & Swineford, F. (1937). The bi-factor method. *Psychometrika*, 2(1), 41–54.  
<https://doi.org/10.1007/BF02287965>
- House, R. J., Hanges, P. J., Javidan, M., Dorfman, P. W., & Gupta, V. (2004). *Culture, leadership, and organizations: The GLOBE study of 62 societies*. Sage.
- Jenkins, H. (2006). Small business champions for corporate social responsibility. *Journal of Business Ethics*, 67(3), 241–256.  
<https://doi.org/10.1007/s10551-006-9182-6>
- Joh, S.-J. (2021). An exploratory study on the components and scales of coaching leadership. *Journal of the Korea Convergence Society*, 12(11), 335–346.  
<https://doi.org/10.15207/JKCS.2021.12.11.335>
- Joh, S.-J. (2022). A critical study on coaching leadership researches in Korea. *Journal of Convergence Coaching*, 1(2), 5–33. <https://doi.org/10.22251/jcc.2022.1.2.5>
- Joh, S.-J. (2024). Interrelationships and differences between coaching, coaching behaviors, and coaching leadership. *Journal of Convergence Coaching*, 3(1), 5–23.  
<https://doi.org/10.22251/jcc.2024.3.1.5>
- Jorgensen, T. D., Pornprasertmanit, S., Schoemann, A. M., & Rosseel, Y. (2022). *semTools: Useful tools for structural equation modeling* (R package version 0.5-6).  
<https://CRAN.R-project.org/package=semTools>
- Jung, H.-S., & Lee, K.-Y. (2019). Collective orientation and team performance in Korean organizations. *International Journal of Human Resource Management*, 30(8), 1341–1365.  
<https://doi.org/10.1080/09585192.2017.1284881>
- Kang, J. W., & Lee, H. S. (2020). Research trends in managerial coaching leadership in Korea. *Andragogy Today*, 23(4), 181–210. <https://doi.org/10.35501/AT.23.4.8>
- Kim, J.-H. (2020). Face and dignity in Korean organizations. *Asian Business & Management*, 19(1), 107–128.  
<https://doi.org/10.1057/s41291-019-00083-5>
- Kline, R. B. (2015). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.
- 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>
- Newman, M. E. J. (2010). *Networks: An introduction*. Oxford University Press.
- Nunnally, J. C., & Bernstein, I. H. (1994). *Psychometric theory* (3rd ed.). McGraw-Hill.
- Podsakoff, P. M., MacKenzie, S. B., Lee, J.-Y., & Podsakoff, N. P. (2003). Common method biases in behavioral research. *Journal of Applied Psychology*, 88(5), 879–903.  
<https://doi.org/10.1037/0021-9010.88.5.879>
- Reise, S. P. (2012). The rediscovery of bifactor measurement models. *Multivariate Behavioral Research*, 47(5), 667–696.  
<https://doi.org/10.1080/00273171.2012.715555>
- Smith, W. K., & Lewis, M. W. (2011). Toward a theory of paradox. *Academy of Management Review*, 36(2), 381–403.  
<https://doi.org/10.5465/amr.2011.59330958>
- Statistics Korea. (2023). *Survey of small and medium enterprises*.  
<https://kosis.kr>
- World Medical Association. (2013). Declaration of Helsinki. *JAMA*, 310(20), 2191–2194.  
<https://doi.org/10.1001/jama.2013.281053>