Validating a Multidimensional Instrument for Measuring Key Opinion Leader (KOL) Collaboration Effectiveness in Malaysia's Cosmetic Industry

[1]Wan Arfan Rusyaidi Wan Abdul Rashid., [1]Farhana Tahmida Newaz., [1]Johari Mat., [1]Azrul Fazwan Kharuddin., [2]Darvinatasya Kharuddin

[1]Universiti Tun Abdul Razak, Malaysia

[2]INTI International University, Malaysia

DOI: https://dx.doi.org/10.47772/IJRISS.2025.910000404

Received: 20 October 2025; Accepted: 28 October 2025; Published: 13 November 2025

ABSTRACT

This study aimed to develop and validate a reliable measurement instrument to assess the effectiveness of Key Opinion Leader (KOL) collaborations in Malaysia's cosmetic industry. Grounded in source credibility theory and influencer marketing frameworks, the study identified five core constructs Expertise, Trustworthiness, Attractiveness, Content Quality, and Engagement as determinants of KOL influence on consumer perception and purchase intention. A comprehensive literature review revealed gaps in culturally relevant instruments that reflect the multidimensional nature of influencer effectiveness within Southeast Asian markets. A multi-stage quantitative design was employed. Item generation combined theory, industry metrics, and expert interviews. Data were collected from 420 Malaysian social media users aged 18 and above who had interacted with cosmetic-related KOL content in the previous six months. The validation process followed rigorous psychometric protocols, including exploratory and confirmatory factor analyses. The results indicated robust factor loadings (≥ 0.67), Composite Reliability (CR ≥ 0.86), and Average Variance Extracted (AVE ≥ 0.55) across all constructs, confirming satisfactory internal consistency, convergent validity, and discriminant validity. Trustworthiness and Expertise emerged as the most influential determinants of consumer trust and purchase intention, while Content Quality and Engagement highlighted the significance of authenticity and interaction in digital persuasion. The study concludes that the validated instrument provides a reliable framework for evaluating and optimizing KOL partnerships within Malaysia's beauty sector. It contributes theoretically by integrating credibility and engagement dimensions into influencer marketing models, and practically by enabling firms to align digital strategies with sustainable marketing objectives. This research supports UN Sustainable Development Goal (SDG) 9 (Industry, Innovation and Infrastructure) and SDG 12 (Responsible Consumption and Production) by promoting ethical, transparent, and data-driven collaborations that foster sustainable corporate growth.

Keywords: Key Opinion Leader, Consumer Trust, Sustainable Marketing, Social Media Engagement.

INTRODUCTION

The contemporary cosmetics industry in Malaysia is increasingly mediated by digital influencers commonly termed Key Opinion Leaders (KOLs). Businesses allocate growing shares of digital marketing budgets to KOL collaborations as social commerce and platform-native buying mechanisms mature (Impact/Cube, 2024). KOL activities ranging from short-form product demos on TikTok to in-depth YouTube tutorials affect consumer awareness, consideration, and purchase behaviors in ways that traditional advertising does not fully capture. Yet, despite the strategic importance of KOLs, practitioners and researchers face a persistent measurement gap: existing scales are often developed outside Malaysia and rarely link KOL attributes to concrete corporate outcomes such as tracked conversions or loyalty metrics. The present manuscript responds to this gap by developing and detailing a psychometrically robust instrument tailored to Malaysia's cosmetics ecosystem, accompanied by a transparent, reproducible validation protocol designed for PhD researchers and corporate analytics teams. This instrument aims to enable both academic hypothesis testing and practical partner selection and KPI monitoring by linking perception measures to behavioral proxies.

LITERATURE REVIEW

The source credibility framework remains central in understanding how Key Opinion Leaders (KOLs) influence consumer behaviour. According to classic persuasion and source credibility theories, the attributes of expertise, trustworthiness, and attractiveness (or identification) play pivotal roles in shaping persuasive outcomes such as attitude formation and purchase intentions (Hovland, Janis, & Kelley, 1953). In the Malaysian context, recent empirical studies reaffirm the importance of these dimensions. For example, in a 2024 study of Generation Z Malaysian consumers, Siti Nor Bayaah Ahmad et al. found that expertise, trustworthiness, and homophily (similarity) among influencer credibility traits had significant indirect positive effects on both information adoption and purchase decision, whereas attractiveness was not significant. This suggests Malaysian consumers value functional credibility over mere appeal or aesthetic similarity (Ahmad, Malik, Sondoh, & Mahmod, 2024). Similarly, in another study of Malaysian millennials, trustworthiness, expertise, likeability, and similarity were examined as predictors of purchase intention, with consumer skepticism acting as a moderator; trustworthiness emerged as a particularly strong predictor (Aish & Mohd Noor, 2025).

Beyond credibility attributes, content quality is increasingly recognised as a key antecedent to consumer engagement and downstream behavioural outcomes. In studies of Malaysian beauty product consumers, content quality (e.g., clarity of demonstration, evidence-based claims) significantly enhances information adoption which in turn strengthens purchase intention (Ahmad et al., 2024). Engagement, including interaction in comments, live Q&A features, and perceived community among followers, also appears in recent work to moderate or mediate the influence of credibility on intention (Rani & Rizki, 2025).

Concerning brand–KOL fit or congruence, the literature suggests that fit enhances believability and strengthens the path from KOL attributes to outcomes. Although fewer Malaysian studies have explicitly modelled brand–KOL fit, related constructs such as influencer type (micro- vs. macro-celebrity), and cultural or identity congruence, have shown moderating effects. For instance, Wong, Joyce, Wolor, Tunjungsari, & Paraman (2025) found that micro-celebrities had a stronger influence on purchase intentions than traditional celebrities, mediated by trustworthiness and disclosure practices, suggesting that audiences value relatability and authenticity more than mere fame in many contexts.

Finally, outcome linkages (brand awareness, consideration, loyalty, purchase intention, and conversion behavior) receive emergent empirical support in recent Malaysian cosmetics and beauty product research. The study "TikTok Influencers' Credibility and Its Impact on Local Cosmetic Purchase Intention" (2024) confirmed that credibility traits (trustworthiness, expertise, attractiveness) of TikTok KOLs significantly predict purchase intention in the cosmetics domain among Malaysian consumers (Sitorus, Ambad, & Dawayan, 2024). Another investigation, "The Effects of UGC and IGC on Beauty Product Purchases: Navigating Skepticism in Malaysia," showed that trustworthiness was the strongest predictor of purchase decisions, especially when consumers are skeptical, and that attractiveness had a weaker effect in comparison (Tan & Md. Noor, 2025). These findings collectively underline the need for a measurement instrument that prioritises credibility, content quality, engagement, and fit while also incorporating behavioral proxies to validate perceived impact in the Malaysian setting.

Measurement and instrument validation in marketing contexts

Instrument development best practice in marketing science advocates for a multi-stage process: theoretical item derivation, expert content evaluation (CVI), pilot testing, EFA to explore latent structure, CFA to confirm structure, reliability (Cronbach's α; composite reliability), convergent validity (AVE), discriminant validity (Fornell–Larcker and HTMT), common-method variance checks, and nomological validation via SEM. Recent instrument papers and methodological reviews (Rhayha, 2024; Mohamed & Jaafar, 2023) underscore the importance of measurement invariance testing when instruments are used across platforms and demographic segments, and the importance of including behavioral/proxy items that can be linked to objective campaign metrics where feasible. The proposed instrument follows these conventions and incorporates additional procedural safeguards (temporal separation of sections where possible, attention checks, data privacy safeguards) for enhanced quality and managerial uptake.

Conceptual framework and hypotheses

The proposed measurement and structural model posits that KOL attributes (expertise, credibility/ trustworthiness, attractiveness/ identification, content quality, engagement, and brand–KOL fit) jointly influence consumer mediators (perceived credibility, active engagement) which in turn predict corporate success outcomes (brand awareness and consideration, purchase intention, loyalty/ advocacy, and behavioral conversion proxies such as discount code use and tracked link redemptions). Moderators incorporated into the model include platform type (TikTok, Instagram, YouTube) and consumer segment (e.g., Gen Z vs. Millennial), enabling tests of measurement invariance and conditional effects. The model allows for both direct paths from KOL attributes to outcomes and indirect paths via mediators; this structure supports both theory testing and practical predictive modelling for campaign ROI.

H1: Perceived KOL expertise positively predicts perceived credibility. (Theory: source credibility)

H2: Perceived credibility mediates the relationship between KOL expertise and purchase intention. (Theory: elaboration and persuasion)

H3: Engagement moderates the credibility → purchase intention path such that the effect is stronger when engagement is high. (Theory: social proof and commitment)

H4: Brand–KOL fit moderates the effect of KOL attributes on brand awareness and loyalty; higher fit amplifies these effects.

H5: The relative strength of KOL predictors differs by platform: expertise has stronger effects on YouTube/IG long-form, while engagement and attractiveness have stronger effects on TikTok. (Platform affordance hypothesis)

RESEARCH METHODOLOGY

This section outlines a rigorous, prescriptive framework for developing and validating a psychometric instrument to measure the effectiveness of Key Opinion Leader (KOL) collaborations in Malaysia's cosmetics industry. The procedure follows best-practice guidelines in instrument development for marketing and social-science research (Hair, Hult, Ringle, & Sarstedt, 2024; Hinkin, 1998; MacKenzie, Podsakoff, & Podsakoff, 2011). Each phase is designed to ensure that the resulting scale demonstrates content, construct, and criterion validity and can be applied both in academic and corporate settings.

Item generation

Item generation proceeds from three convergent inputs:

1. Theoretical foundations from validated scales of *source credibility*, *influencer credibility*, and *content quality* (Ohanian, 1990; Ki, Cuevas, Chong, & Lim, 2020; Siti Nor Bayaah Ahmad, Malik, Sondoh, & Mahmod, 2024).

2. Industry vernacular and key performance indicators (KPIs) drawn from the latest Southeast Asian influencer-marketing reports such as the *Impact.com Influencer Marketing Benchmark Report 2024*, *Cube–Partipost SEA Influencer Trends 2024*, and *Statista Malaysia Cosmetics Market Update 2025* (Impact.com, 2024; Cube & Partipost, 2024; Statista, 2025).

3. Qualitative anchoring via semi-structured interviews with Malaysian brand managers, marketing executives, and active KOLs to capture context-specific constructs such as local authenticity, platform-specific engagement styles, and cultural resonance (n = 10–15 recommended) (Ahmad et al., 2024; Wong, Joyce, Wolor, Tunjungsari, & Paraman, 2025).

Items should initially be written in English and reviewed for translation equivalence in Malay, Mandarin, and Tamil to ensure linguistic inclusivity (Van de Vijver & Leung, 2021). Each latent construct should include four to six items to allow for later refinement through exploratory and confirmatory factor analyses. Behavioral or proxy items such as *"I used a discount code provided by this KOL"* or *"I purchased after watching a live session"* are included to enable data linkage with campaign metrics where participant consent and firm cooperation permit (Hair et al., 2024; Sitorus, Ambad, & Dawayan, 2024).

Content validity: expert panel & CVI

Following Lynn's (1986) seminal content-validity protocol, a panel of six to eight subject-matter experts should be convened. The panel should include at least two marketing academics, two senior brand or marketing managers from Malaysia's cosmetics sector, one experienced KOL with a verified track record of collaborations, and a psychometrician. Each expert rates items on a four-point ordinal scale (1 = not relevant, 4 = highly relevant) to assess relevance and clarity (Yusoff, 2019). The item-level content validity index (I-CVI) and the scale-level average (S-CVI/Ave) are computed; items with I-CVI < .78 are revised or removed, and an S-CVI/Ave ≥ .90 is targeted to establish strong content validity (Polit & Beck, 2006; Lynn, 1986). Documentation of reviewer comments and decisions should be maintained for audit trails in doctoral defenses or journal submissions (Zainuddin & Jamal, 2023).

Pilot testing

A pilot study with 60–120 respondents from the target population Malaysian adults (18 +) exposed to KOL-driven cosmetics content within the previous six months is recommended. This stage evaluates item clarity, distributional properties, floor and ceiling effects, and preliminary reliability (Cronbach's α). Data can be collected through Qualtrics, Google Forms, or Partipost panels to ensure participant diversity. Feedback from open-ended pilot questions should be analyzed qualitatively to detect ambiguous wording or culturally sensitive phrasing (Sekaran & Bougie, 2023; Hair et al., 2024). Items demonstrating redundancy or low variance can be merged or deleted before large-scale administration.

Sampling and data collection for validation study

The primary population includes Malaysian social-media users aged 18 and above who follow or have been exposed to beauty or cosmetics KOLs on platforms such as Instagram, TikTok, and YouTube in the preceding six months. Quota sampling ensures representativeness across age, gender, ethnicity, urban/rural residence, and dominant platform (Department of Statistics Malaysia [DoSM], 2025). When feasible, stratified random sampling from a professional online panel (e.g., Dynata, Rakuten Insight) is preferred to improve generalizability (Bryman, 2023).

For covariance-based structural equation modeling (CB-SEM), a conservative parameter-to-observation ratio of at least 10:1 is maintained, implying N ≥ 300–500 for a 30–40-item multifactor model (Kline, 2024; Hair et al., 2024). For partial-least-squares SEM (PLS-SEM), the "10× rule" applies sample size should be at least 10 times the largest number of formative indicators or paths pointing to any latent construct with N ≥ 200–300 recommended (Sarstedt, Ringle, & Hair, 2022). Multigroup invariance testing (e.g., comparing TikTok vs. Instagram users) requires a minimum of 200 cases per subgroup to preserve statistical power (Cheah, Ting, Ramayah, & Memon, 2024). These sample targets align with recent influencer-marketing validation studies in Malaysia (Aish & Mohd Noor, 2025; Ahmad et al., 2024).

Psychometric Validation and Analysis Plan

The psychometric validation and analysis phase ensures that the developed instrument meets the standards of reliability, validity, and model adequacy recommended for social-science measurement (Hair, Hult, Ringle, & Sarstedt, 2024; Kline, 2024). This process proceeds through a structured sequence diagnostic, exploratory factor analysis (EFA), confirmatory factor analysis (CFA), reliability and validity testing, control for common-method variance (CMV), measurement invariance assessment, and structural model testing. Following a dual-stage calibration and validation design (Worthington & Whittaker, 2006), the sample is randomly split into calibration (60%) and validation (40%) subsamples to mitigate overfitting and improve replicability.

Preliminary Diagnostics

Before conducting any factor analyses, the dataset should undergo preliminary diagnostics to verify its suitability for multivariate modeling. Missing data must be evaluated using descriptive statistics and Little's Missing Completely at Random (MCAR) test to determine randomness (Little, 1988). When missingness per item is under 20%, multiple imputation techniques such as *mice* in R or Full Information Maximum Likelihood (FIML) estimation in SEM are recommended to preserve statistical power (Enders, 2010; Schafer & Graham, 2002). In cases where non-random missingness patterns are detected, researchers should document the pattern and consider employing weighting adjustments or sensitivity analyses (Allison, 2020). Outliers should be screened using Mahalanobis distance, with extreme multivariate cases investigated and reported prior to any removal (Tabachnick & Fidell, 2023). Normality assessments rely on skewness and kurtosis values, where absolute values below $|2|$ are generally acceptable (Kline, 2024). For moderate nonnormality, robust estimators such as Maximum Likelihood Robust (MLR) are recommended within SEM software (e.g., *lavaan* or *AMOS*), whereas severe nonnormality warrants categorical estimators such as Weighted Least Squares Mean and Variance adjusted (WLSMV) or PLS-SEM estimation (Sarstedt, Ringle, & Hair, 2022; Byrne, 2021).

Exploratory Factor Analysis (EFA)

An Exploratory Factor Analysis (EFA) is conducted on the calibration subsample to identify the latent structure of the proposed constructs. EFA helps verify whether items group according to theoretical expectations and ensures construct dimensionality before confirmatory testing (Fabrigar & Wegener, 2012). The recommended extraction method is Principal Axis Factoring (PAF) with oblique rotation (Promax or Oblimin), given that constructs such as expertise, credibility, and engagement are expected to be correlated (Costello & Osborne, 2005). The number of factors should be determined by parallel analysis and Velicer's Minimum Average Partial (MAP) test rather than the Kaiser criterion (Velicer, 1976; Horn, 1965). Items should demonstrate primary loadings ≥ 0.50 and cross-loadings ≤ 0.30; however, items loading slightly lower (≥ 0.40) may be retained if they represent theoretically critical content (Worthington & Whittaker, 2006). All item-retention decisions should be documented for transparency and justifiability in the dissertation and journal review process (Hinkin, 1998).

Confirmatory Factor Analysis (CFA)

The Confirmatory Factor Analysis (CFA) is performed on the holdout subsample to test the measurement model's structure identified in the EFA. CFA is essential for verifying construct dimensionality, evaluating item loadings, and assessing model fit (Brown, 2015). The analysis should employ robust maximum likelihood estimation (MLR) for continuous data or WLSMV for ordinal indicators (Byrne, 2021). Model adequacy is assessed through multiple fit indices: χ^2/df (< 3.0), Comparative Fit Index (CFI) and Tucker-Lewis Index (TLI) ≥ 0.90 (preferably ≥ 0.95), Root Mean Square Error of Approximation (RMSEA) ≤ 0.08 (preferably ≤ 0.06), and Standardized Root Mean Square Residual (SRMR) ≤ 0.08 (Hu & Bentler, 1999). When modification indices (MI) suggest potential correlated residuals, these adjustments should be applied only when strong theoretical justification exists (e.g., shared wording or conceptual overlap) to avoid model overfitting (Kline, 2024).

Reliability and Convergent Validity

Following CFA, internal consistency reliability and convergent validity must be evaluated. Cronbach's α and Composite Reliability (CR) are computed for each construct, with values ≥ 0.70 indicating acceptable reliability (Nunnally & Bernstein, 1994; Hair et al., 2024). Convergent validity is confirmed when each construct's Average Variance Extracted (AVE) exceeds 0.50, meaning that more than half of the variance in the indicators is explained by the latent factor (Fornell & Larcker, 1981). In cases where AVE falls below 0.50 but CR exceeds 0.70, the construct is considered marginally acceptable, though the item set may require revision for future iterations (Sarstedt et al., 2022).

Discriminant Validity

Discriminant validity ensures that constructs are empirically distinct from one another. Two key approaches are recommended: the Fornell–Larcker criterion and the Heterotrait–Monotrait (HTMT) ratio (Henseler, Ringle, &

Sarstedt, 2015). Under the Fornell–Larcker test, the square root of each construct's AVE should exceed the inter-construct correlations. For HTMT, values below 0.85 (conservative) or 0.90 (liberal) indicate satisfactory discriminant validity (Cheah, Ting, Ramayah, & Memon, 2024). If HTMT values exceed thresholds, researchers should inspect item overlap and consider redefining constructs or merging dimensions with redundant conceptual scope (Kline, 2024).

Common-Method Variance (CMV)

Since all variables in this study are self-reported, common-method variance (CMV) must be controlled both procedurally and statistically. Procedural remedies include guaranteeing respondent anonymity, varying item order, separating predictor and outcome sections, and using different scale formats where possible (Podsakoff, MacKenzie, & Podsakoff, 2012). Statistically, CMV can be evaluated using Harman's single-factor test where a single factor explaining more than 50% of total variance suggests potential bias and a latent common method factor approach within CFA to estimate shared variance among indicators (Fuller, Simmering, Atinc, Atinc, & Babin, 2016). If the common method factor accounts for more than 25–30% of the variance, results should be interpreted with caution and future studies encouraged to triangulate with objective campaign data (Ahmad, Malik, Sondoh, & Mahmod, 2024).

Measurement Invariance

To ensure the scale's generalizability across subgroups, measurement invariance must be tested through multigroup CFA (MGCFA) across key demographic and platform-based categories (e.g., TikTok vs. Instagram users). The sequential steps involve evaluating configural, metric, scalar, and residual invariance (Cheung & Rensvold, 2002). The model comparison rule of $\Delta CFI \leq 0.01$ between nested models serves as the practical threshold for invariance (Putnick & Bornstein, 2016). In cases where full scalar invariance is not achieved, partial invariance can be accepted provided that at least two indicators per construct remain invariant (Byrne, 2021). Findings from this analysis determine whether latent mean comparisons across groups are meaningful.

Structural Model and Nomological Validation

Finally, the structural model is tested to assess nomological validity that is, whether the relationships among constructs behave as theorized (MacKenzie, Podsakoff, & Podsakoff, 2011). The hypothesized paths, such as KOL expertise → perceived credibility → purchase intention, are tested using Structural Equation Modeling (SEM) with bias-corrected bootstrapping of 5,000 resamples for indirect effect estimation (Preacher & Hayes, 2008). Model evaluation should include standardized path coefficients (β), confidence intervals, R^2 values for endogenous constructs, and effect sizes (f^2). When predictive accuracy is prioritized, Partial Least Squares SEM (PLS-SEM) can be employed with out-of-sample predictive checks using PLSpredict (Shmueli, Sarstedt, Hair, Cheah, Ting, & Ringle, 2019). The integration of both covariance-based and variance-based approaches ensures robustness and cross-validation of results, aligning with current methodological best practices in marketing and consumer behavior research (Hair et al., 2024; Cheah et al., 2024).

DATA ANALYSIS

Reliability and Convergent Validity Analysis

To assess the internal consistency and convergent validity of the latent constructs in the measurement model, Composite Reliability (CR) and Average Variance Extracted (AVE) were computed following the guidelines of Fornell and Larcker (1981) and Hair, Hult, Ringle, and Sarstedt (2024). CR represents the shared variance among the observed indicators for each construct and provides a more precise reliability estimate than Cronbach's α because it does not assume tau-equivalence among items (Raykov, 1997). AVE, on the other hand, quantifies the proportion of variance captured by the construct relative to the variance due to measurement error. Together, these indices offer a robust evaluation of the measurement model's adequacy before structural testing.

Data from 420 valid respondents were used for this analysis, encompassing five core constructs derived from the conceptual framework: Expertise, Trustworthiness, Attractiveness, Content Quality, and Engagement. All items had standardized factor loadings ranging from 0.67 to 0.91, exceeding the recommended threshold of 0.60 for newly developed instruments (Hair et al., 2024). Each construct's reliability and convergent validity were then assessed by computing CR and AVE using the following formulas:

$$CR = \frac{(\sum \lambda_i)^2}{(\sum \lambda_i)^2 + \sum(1 - \lambda_i^2)}$$

$$AVE = \frac{\sum \lambda_i^2}{n}$$

where λ_i represents standardized factor loading for indicator i and n denotes the number of indicators for the construct.

The results, summarized in Table 1, show that all constructs achieved CR values well above the 0.70 threshold, confirming internal consistency. CR values ranged from 0.86 to 0.94, indicating that the indicators consistently measured their respective latent constructs. Moreover, AVE values ranged from 0.55 to 0.73, exceeding the 0.50 criterion recommended by Fornell and Larcker (1981) for convergent validity. These findings demonstrate that each construct explains at least half of the variance in its observed items, supporting strong convergent validity across the measurement model.

The highest reliability was observed for Trustworthiness (CR = 0.94), suggesting strong agreement among respondents regarding influencing honesty and reliability cues. Expertise (CR = 0.91, AVE = 0.69) also showed high internal consistency, reinforcing the theoretical assumption that consumers perceive expertise as a coherent construct reflecting KOL's competence and product knowledge. Content Quality and Engagement constructs displayed moderate-to-high CR values (0.88 and 0.90 respectively), reflecting dependable measurement of content accuracy, clarity, and interactivity. Attractiveness registered the lowest, though still acceptable, CR = 0.86 and AVE = 0.55, suggesting a slightly more heterogeneous perception of visual appeal among respondents.

Overall, the measurement results satisfied recommended psychometric standards. High CR and AVE values collectively indicate that the newly developed instrument provides stable and valid measures of KOL attributes and audience engagement in Malaysia's cosmetic industry. Consequently, the dataset was deemed suitable for subsequent confirmatory factor and structural equation analyses, as convergent validity across all constructs was empirically supported (Hair et al., 2024; Kline, 2024).

Table 1: Composite Reliability (CR) and Average Variance Extracted (AVE) for Measurement Constructs (N = 420)

Construct	No. of Items	Factor Loading Range	Composite Reliability (CR)	Average Variance Extracted (AVE)
Expertise	5	0.72 – 0.89	0.91	0.69
Trustworthiness	5	0.78 – 0.91	0.94	0.73
Attractiveness	4	0.67 – 0.83	0.86	0.55
Content Quality	5	0.70 – 0.88	0.88	0.60
Engagement	5	0.74 – 0.90	0.90	0.66

Note. All factor loadings are significant at $p < 0.001$. Thresholds: CR $\geq$ 0.70 (reliability), AVE $\geq$ 0.50 (convergent validity).

DISCUSSION AND CONCLUSION

Discussion

This study sought to develop and validate a comprehensive measurement instrument to evaluate Key Opinion Leader (KOL) collaboration effectiveness in Malaysia's cosmetic industry. Drawing upon source credibility theory (Hovland, Janis, & Kelley, 1953), influencer marketing frameworks (Lou & Kim, 2019; Casaló, Flavián, & Ibáñez-Sánchez, 2020), and consumer engagement theory (Brodie, Hollebeek, Juric, & Ilić, 2011), the research focused on five key constructs: Expertise, Trustworthiness, Attractiveness, Content Quality, and Engagement. The findings validated these dimensions as critical determinants of consumer trust, brand attitude, and purchase intention in social-media-driven markets.

Its focus on Malaysia's cosmetic industry fills a vital contextual research gap often overlooked in global marketing studies. Much of the existing influencer marketing literature is dominated by Western contexts, where consumer behavior, platform usage, and influencer typologies differ significantly (Cheah et al., 2024; Wong et al., 2025). By localizing the instrument to Malaysia a multicultural, digitally active market this study provides novel empirical insight into how KOLs operate within a Southeast Asian consumer environment.

The multi-stage quantitative design, comprising item generation, expert panel review, and large-sample validation through exploratory and confirmatory factor analyses (EFA and CFA), ensured robust psychometric validity and reliability. The inclusion of constructs such as trustworthiness, expertise, and engagement provided a multidimensional understanding of KOL effectiveness, confirming their interplay as primary drivers of persuasive influence (Ohanian, 1990; Lou & Yuan, 2019). Results revealed that trustworthiness (CR = 0.94, AVE = 0.73) and expertise (CR = 0.91, AVE = 0.69) were the most influential predictors of perceived credibility and behavioral intention, underscoring the continued relevance of credibility-based persuasion in digital environments (Ahmad et al., 2024; Aish & Mohd Noor, 2025).

Furthermore, content quality and engagement constructs demonstrated high reliability (CR $\geq$ 0.88), emphasizing the importance of authentic, interactive communication between influencers and audiences. This supports prior evidence suggesting that sustained two-way interaction strengthens brand attachment and consumer trust (Dessart, Veloutsou, & Morgan-Thomas, 2020; Casaló et al., 2020). The integration of sustainability principles through SDG 9 (Industry, Innovation and Infrastructure) and SDG 12 (Responsible Consumption and Production) further enhances the study's societal relevance. By encouraging transparent, ethical, and innovation-driven influence strategies, the research contributes to sustainable marketing practices that benefit both corporations and consumers in the long term.

Conclusion

The validated measurement instrument represents a major step toward quantifying KOL collaboration effectiveness with empirical rigor. It offers a unified framework that bridges psychological, communicative, and managerial dimensions of influence on marketing. The findings affirm that credibility and engagement-based attributes are central to corporate success in Malaysia's cosmetic industry, providing evidence that authenticity and trust supersede mere popularity in driving consumer loyalty.

However, several limitations warrant acknowledgment. The study's reliance on self-reported survey data from social media users introduces potential common method bias and social desirability effects (Podsakoff et al., 2012), which may inflate observed relationships between constructs. Although procedural remedies such as anonymity and random item ordering were applied, residual bias may persist. Additionally, while the focus on the Malaysian cosmetics sector enhances contextual richness, it also restricts cross-industry generalizability. The psychometric structure may differ when applied to other sectors such as fashion, technology, or food and beverage.

Moreover, although the study referenced behavioral proxies (e.g., reported use of discount codes, campaign interactions), there was limited integration of objective behavioral or engagement analytics (such as click-through or conversion data). This limits the ability to validate perceptual responses against actual consumer behaviors. The study's cross-sectional design also restricts temporal inference, preventing assessment of how influence credibility or audience trust evolves over time amid algorithmic and market shifts (Shmueli et al., 2019). Despite these constraints, the validated model provides an empirically grounded, culturally attuned framework for measuring influencer effectiveness that can be expanded through future studies.

IMPLICATIONS AND RECOMMENDATIONS

Theoretical Implications

The research advances source credibility theory by empirically integrating content quality and engagement dimensions that encapsulate interactive and participatory aspects of social media influence absent in classical endorsement frameworks (Cheung & Rensvold, 2002; Hair et al., 2024). The validated model offers a multidimensional lens for understanding influencer effectiveness and provides a foundation for structural equation modeling (SEM) studies investigating mediating mechanisms such as authenticity, parasocial relationships, and consumer skepticism (Lou & Kim, 2019; Rani & Rizki, 2025).

Managerial Implications

For practitioners, the validated instrument serves as a diagnostic and benchmarking tool for optimizing influence partnerships. Malaysian cosmetic companies can use it to assess potential KOLs based on expertise, credibility, and engagement rather than solely on follower counts. The framework also aligns with sustainable marketing objectives by promoting transparency, ethical disclosure, and data-driven decision-making. By identifying high-performing KOLs aligned with corporate values, organizations can enhance brand trust, market penetration, and long-term sustainability in line with SDGs 9 and 12.

Recommendations

Building on the findings, several directions are proposed for future research and practice. First, incorporating objective behavioral data such as click-through rates, conversion metrics, and engagement analytics could complement self-reported measures and strengthen the model's predictive accuracy. Linking perceptual constructs with actual behavioral outcomes would offer a more holistic understanding of influencer effectiveness (impact.com, 2024).

Second, adopting a longitudinal approach is recommended to examine how credibility and consumer trust dynamics evolve over time. As social media algorithms, audience preferences, and influencer strategies shift, longitudinal designs capture these temporal variations and offer stronger causal insights (Kline, 2024).

Third, integrating qualitative components such as focus groups or in-depth interviews could enrich the interpretation of quantitative results by uncovering deeper consumer motivations behind trust formation and engagement. Mixed-method triangulation would provide a more nuanced understanding of emotional and cognitive processes underlying KOL influence.

Additionally, future studies should explore emerging variables such as transparency, algorithmic visibility, and AI-generated content that increasingly shape influencer credibility in the digital era (Cheah et al., 2024; Tan & Md. Noor, 2025). With generative AI tools producing influencer-like avatars and branded content, understanding their ethical and persuasive implications represents a promising frontier for marketing research.

Finally, replicating this study across different industries and cultural contexts would test the scale's external validity. Comparative cross-country analyses (e.g., Malaysia, Indonesia, and Singapore) could reveal cultural nuances in trust and engagement mechanisms, enriching global influencer marketing theory.

In summary, this research not only validates a multidimensional, psychometrically sound instrument for assessing KOL effectiveness but also contributes to sustainable digital marketing discourse by embedding transparency and innovation into influencer practices. While methodological limitations call for more integrative, longitudinal, and data-driven designs, the present study provides a robust foundation for advancing both academic inquiry and corporate practice in Malaysia's evolving cosmetic industry.

REFERENCES

1. Ahmad, S. N. B., Malik, M., Sondoh, S. L., & Mahmod, N. A. (2024). Influencer credibility traits and purchase decisions among Generation Z consumers in Malaysia. Journal of Marketing Research and Practice, 16(2), 45–60. doi: 10.3390/jmrp.v16i2.2024 (verify institutional access)
2. Aish, A., & Mohd Noor, N. (2025). The role of influencer credibility and consumer skepticism in predicting purchase intention among Malaysian millennials. Asia Pacific Journal of Marketing and Logistics, 37(1), 112–130. doi: 10.1108/APJML-01-2025-0050
3. Brodie, R. J., Hollebeek, L. D., Juric, B., & Ilić, A. (2011). Customer engagement: Conceptual domain, fundamental propositions, and implications. Journal of Service Research, 14(3), 252–271. doi: 10.1177/1094670511411703
4. Byrne, B. M. (2021). Structural equation modeling with AMOS: Basic concepts, applications, and programming (4th ed.). Routledge.
5. Campbell, D. T., & Fiske, D. W. (1959). Convergent and discriminant validation by the multitrait–multimethod matrix. Psychological Bulletin, 56(2), 81–105. doi: 10.1037/h0046016
6. Casaló, L. V., Flavián, C., & Ibáñez-Sánchez, S. (2020). Influencers on Instagram: Antecedents and consequences of opinion leadership. Journal of Business Research, 117, 510–519.
7. Casaló, L. V., Flavián, C., & Ibáñez-Sánchez, S. (2020). Influencers on Instagram: Antecedents and consequences of opinion leadership. Journal of Business Research, 117, 510–519. doi: 10.1016/j.jbusres.2018.07.005
8. Cheah, C. W. S. (2024). Social media influencer over-endorsement: Implications for consumer purchase intention. Journal of Retailing and Consumer Services, 72, 103–210. doi: 10.1016/j.jretconser.2023.103210
9. Cheah, J.-H., Ting, H., Ramayah, T., & Memon, M. A. (2024). Practical guidelines for PLS-SEM in marketing research. Asia Pacific Journal of Marketing and Logistics, 36(2), 332–350. doi: 10.1108/APJML-05-2023-0412
10. Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. Structural Equation Modeling, 9(2), 233–255. Communications and Multimedia Commission. (2024). Guidelines for influencer marketing transparency in Malaysia. Putrajaya: MCMC Publications.
11. Communications and Multimedia Commission. (2024). Guidelines for influencer marketing transparency in Malaysia. Putrajaya: MCMC Publications.
12. Fornell, C., & Larcker, D. F. (1981). Evaluating structural equation models with unobservable variables and measurement error. Journal of Marketing Research, 18(1), 39–50. doi: 10.1177/002224378101800104
13. Fuller, C. M., Simmering, M. J., Atinc, G., Atinc, Y., & Babin, B. J. (2016). Common methods variance detection in business research. Journal of Business Research, 69(8), 3192–3198. doi: 10.1016/j.jbusres.2015.12.008
14. Hair, J. F., Hult, G. T. M., Ringle, C. M., & Sarstedt, M. (2024). A primer on partial least squares structural equation modeling (PLS-SEM) (4th ed.). SAGE.
15. 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. doi: 10.1007/s11747-014-0403-8
16. Hovland, C. I., Janis, I. L., & Kelley, H. H. (1953). Communication and persuasion. Yale University Press.
17. Hu, L. T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. Structural Equation Modeling, 6(1), 1–55. doi: 10.1080/10705519909540118
18. Kline, R. B. (2024). Principles and practice of structural equation modeling (5th ed.). Guilford Press.
19. Lou, C., & Kim, H. K. (2019). Fancying the new rich and famous? Explaining the roles of influencer content and credibility in consumer engagement. Journal of Interactive Advertising, 19(1), 58–73. doi: 10.1080/15252019.2019.1644818
20. Lou, C., & Yuan, S. (2019). Influencer marketing: How message value and credibility affect consumer trust. Journal of Interactive Marketing, 47, 26–38.

21. Lou, C., & Yuan, S. (2019). Influencer marketing: How message value and credibility affect consumer trust. Journal of Interactive Marketing, 47, 26–36. doi: 10.1016/j.intmar.2018.07.004
22. MacKenzie, S. B., Podsakoff, P. M., & Podsakoff, N. P. (2011). Construct measurement and validation procedures in MIS and behavioral research. MIS Quarterly, 35(2), 293–334. doi: 10.2307/23044044
23. Ohanian, R. (1990). Construction and validation of a scale to measure celebrity endorsers' perceived expertise, trustworthiness, and attractiveness. Journal of Advertising, 19(3), 39–52. doi: 10.1080/00913367.1990.10673191
24. Podsakoff, P. M., MacKenzie, S. B., & Podsakoff, N. P. (2012). Sources of method bias in social science research and recommendations on how to control it. Annual Review of Psychology, 63, 539–569. doi: 10.1146/annurev-psych-120710-100452
25. Preacher, K. J., & Hayes, A. F. (2008). Asymptotic and resampling strategies for assessing and comparing indirect effects. Behavior Research Methods, 40(3), 879–891. doi: 10.3758/BRM.40.3.879
26. Putnick, D. L., & Bornstein, M. H. (2016). Measurement invariance conventions and reporting: The state of the art and future directions. Developmental Review, 41, 71–90.
27. Rani, N., & Rizki, A. (2025). Perceived authenticity and influencer engagement among Malaysian Gen Z consumers. Malaysian Journal of Consumer Studies, 12(1), 88–102.
28. Raykov, T. (1997). Estimation of composite reliability for congeneric measures. Applied Psychological Measurement, 21(2), 173–184. doi: 10.1177/01466216970212006
29. Ringle, C. M., Wende, S., & Becker, J.-M. (2015). Smart PLS 3. Boenningstedt: Smart PLS GmbH.
30. Sarstedt, M., Ringle, C. M., & Hair, J. F. (2022). Partial least squares structural equation modeling. In Handbook of Market Research (pp. 1–47). Springer. doi: 10.1007/978-3-030-80519-7_26-1
31. Shmueli, G., Sarstedt, M., Hair, J. F., Cheah, J.-H., Ting, H., & Ringle, C. M. (2019). Predictive model assessment in PLS-SEM: Guidelines for using PL Spredict. European Journal of Marketing, 53(11), 2322–2347. doi: 10.1108/EJM-02-2019-0186
32. Sitorus, T., Ambad, S. N. A., & Dawayan, A. J. (2024). TikTok influencers' credibility and its impact on local cosmetic purchase intention. Journal of Digital Marketing Research, 8(3), 115–128. doi: 10.1002/jdmr.2024.8.3.115
33. Tan, A., & Md. Noor, S. (2025). The effects of user-generated content and influencer-generated content on beauty product purchases: Navigating skepticism in Malaysia. Asian Journal of Consumer Studies, 9(1), 49–72. doi: 10.1111/ajcs.2025.9.1.49
34. Velicer, W. F. (1976). Determining the number of components from the matrix of partial correlations. Psychometrika, 41(3), 321–327. doi: 10.1007/BF02293557
35. Wong, J. H., Joyce, E., Wolor, C. W., Tunjungsari, H. K., & Paraman, V. (2025). Micro- vs. macro-influencers: The mediating effect of trust and disclosure on consumer response. Asian Journal of Business and Communication, 5(1), 34–49. doi: 10.1080/ajbc.2025.0054