Advanced International Journal for Research (AIJFR)

E-ISSN: 3048-7641 ● Website: www.aijfr.com ● Email: editor@aijfr.com

Validation of a Compact SERVQUAL Composite for Telecom Service Quality Dashboards

Dr. Roland Yaw Kudozia[1], Ms Elizabeth Carolyn Kudozia[2], Mr. Ayitey Kommey[3]

[1,3]Rsearch Fellow, Research and Training, Gdirst Institute
[2]Research Fellow, Statistics, Gdirst Institute

Abstract

Purpose

This study validates a compact, performance-only SERVQUAL (service quality) configuration as a governance-grade Key Performance Indicator (KPI) for mobile telecommunications. We assess whether a short, reflective composite of Perceived Service Quality (PSQ) preserves psychometric defensibility while retaining facet-level diagnostic value for managerial and regulatory use.

Design/methodology/approach

Using a cross-sectional customer survey ($N = 537$), we harmonize Likert responses to a 1–5 metric and apply transparent scoring: unit-weighted facets (Reliability, Assurance, Empathy, Tangibles; Responsiveness not retained in this wave) and a composite computed when ≥4/5 facets are present (here, all four observed facets). Measurement checks include internal consistency (Cronbach's alpha, McDonald's omega), essential unidimensionality (eigenstructure/scree), convergence (standardized one-factor loadings, Average Variance Extracted—AVE), and a compact residual index (Standardized Root Mean Square Residual—SRMR). Cross-segment comparability is evaluated via reliability equivalence, with a staged roadmap to Multi-Group Confirmatory Factor Analysis (MG-CFA) as the instrument expands.

Findings

The composite exhibits governance-grade reliability (alpha = 0.882; omega = 0.919). Eigenstructure shows a dominant first component ($\lambda_1 = 2.957$; ≈74% variance), and convergence is strong (loadings = 0.757–0.898; AVE = 0.739). The SRMR-style residual index (≈0.095) indicates acceptable parsimony for a short KPI. Facet means suggest a binding constraint in Assurance (competence/trust), with Tangibles as a relative strength and Reliability/Empathy mid-range. Segmentwise reliability is operationally comparable ($\Delta\omega = 0.061$) but borderline on alpha ($\Delta\alpha = 0.111$) due to small N in the minor segment.

Research limitations/implications

This wave retains one indicator per facet (Responsiveness absent), precluding identified CFA and formal invariance tests; results rely on a one-factor surrogate and reliability-equivalence checks. Future waves should field ≥2 items per facet and adopt CFA/MG-CFA (and alignment) with polychoric estimation to strengthen validity and comparability.

Practical implications

The validated, transparent scoring spec (1–5; ≥4/5 rule) supports immediate deployment of PSQ as a board-level KPI, with facet "dials" for operational ownership (Assurance, Reliability, Empathy,

Tangibles). Measurement guardrails (alpha/omega ≥ .80; AVE > .50; SRMR ≤ .10) enable auditability across waves and freeze the instrument for dashboard use.

Social/policy implications

Publishing the composite with facet profiles and scoring specifications enables interpretable operator scorecards and parity monitoring. Regulators can adopt reliability-equivalence as an interim safeguard while operators build toward MG-CFA-based invariance.

Originality/value

The paper offers a replicable "validity bundle" for short, performance-only SERVQUAL in telecom—coupling reliability, essential unidimensionality, convergence, and residual parsimony—bridging psychometric rigor and governance needs while providing a staged path to full invariance testing.

Keywords: Service quality; SERVQUAL; SERVPERF; Perceived Service Quality (PSQ); telecommunications; dashboards; reliability; Average Variance Extracted (AVE); Standardized Root Mean Square Residual (SRMR); Multi-Group Confirmatory Factor Analysis (MG-CFA)

1. Introduction

1.1 Background and rationale

Mobile telecommunications providers operate in a context of intense competition, commoditizing tariffs, and rapidly evolving digital service expectations. In this environment, leadership teams increasingly require governance-grade, auditable Key Performance Indicators (KPIs) that are short enough for executive dashboards yet psychometrically defensible for benchmarking and parity monitoring across providers and segments. SERVQUAL (service quality) a canonical, multidimensional instrument for perceived service quality—remains the most widely cited foundation for measuring service experiences (Parasuraman, Zeithaml, & Berry, 1988; Brady & Cronin ,200)). However, many organizations struggle to implement it as a compact KPI that still retains diagnostic signal across its component facets *(SERVQUAL dimensions)*: Reliability, Responsiveness, Assurance, Empathy, and Tangibles.

Beyond conceptual appeal, the practical stakes are high. Peer-reviewed telecom studies consistently link perceived service quality to satisfaction, loyalty, and churn-relevant intentions (e.g., Kalia, Kaushal, Singla, & Parkash, 2021; Naz, Akram, & Malik, 2021; Khan, Ribeiro, Barbosa, Moreira, & Rodrigues, 2024; Suchanek & Bučíková, 2025; see also Abd-Elrahman, 2023). Yet dashboard implementations often rely on ad hoc indices without transparent scoring rules, reliability evidence, or comparability checks. This paper addresses that gap by validating a short, performance-only SERVQUAL configuration as a KPI-grade composite for telecom, while preserving facet-level diagnostics to guide action.

1.2 Theoretical foundation and contribution

SERVQUAL conceptualizes perceived service quality as a reflective construct manifested in multiple facets (Parasuraman et al., 1988). A long tradition debates gap versus performance-only scoring; the latter—often referred to as SERVPERF (service performance)—argues that performance perceptions alone can validly capture service quality and improve parsimony (Cronin & Taylor, 1992). Our study adopts the performance-only approach and contributes three advances:

1. Construct parsimony with diagnostic richness: We argue (and show empirically) that a single reflective factor can support a unit-weighted composite suitable for KPIs—without erasing meaningful facet signals. This aligns with the logic of reflective measurement (Kline, 2016;

McDonald, 1999) and provides a two-tier interpretation: a composite for governance and facets as "dials" for targeted improvements.

2. Validity bundle suitable for dashboards: We operationalize a publishable and auditable set of psychometric checks: internal consistency (Cronbach's alpha; McDonald's omega), eigenstructure/scree for essential unidimensionality, standardized loadings and Average Variance Extracted (AVE) for convergent strength, and a residual index using a Standardized Root Mean Square Residual (SRMR)–style summary to capture off-factor structure (Fornell & Larcker, 1981; Hu & Bentler, 1999; Hayes & Coutts, 2020).
3. Comparability as a staged objective: In many applied dashboards, comparability across subscription segments (e.g., prepaid vs. postpaid) is required. We adopt a staged approach: (a) reliability equivalence across groups now, and (b) formal Multi-Group Confirmatory Factor Analysis (MG-CFA) later as item coverage expands. Specifically, we reference MG-CFA guidelines (Cheung & Rensvold, 2002; Steenkamp & Baumgartner, 1998) and alignment methods (Asparouhov & Muthén, 2014) for future waves.

1.3 Conceptual positioning

We position Perceived Service Quality (PSQ) as a reflective latent factor manifested by five SERVQUAL facets. Conceptually, Reliability captures "doing it right" (accuracy/consistency), Responsiveness "doing it fast" (speed/willingness), Assurance competence/trust in problem handling, Empathy the relational sensitivity to customer needs, and Tangibles the physical/digital environment that frames interactions. In telecom journeys, these facets map naturally to network reliability, care speed, agent/system competence, care personalization, and app/store/brand cues. The model thus integrates academic parsimony with practitioner salience: a single index for tracking and five dials for intervention ownership across engineering, operations, and customer care.

1.4 Research objectives

We pursue four objectives:

1. O1: Scoring specification. Define a transparent, unit-weighted SERVQUAL composite on a 1–5 scale with ≥4/5 facets answered (≥70%) and standardized Likert recode rules.
2. O2 : Validity evidence. Demonstrate internal consistency (alpha, omega), essential unidimensionality (scree/eigenvalues), and convergent strength (standardized loadings; AVE > .50).
3. O3: Parsimony vs. fit. Quantify residual structure with an SRMR-style index to ensure facet-specific information remains without undermining the composite.
4. O4: Comparability. Provide operational comparability across key segments via reliability equivalence now, and outline conditions for future MG-CFA/alignment when item coverage grows.

1.5 Significance of the study

We add to service-quality theory by demonstrating that short-form, performance-only measurement can be reflectively coherent (single factor) and diagnostics-preserving (facet residuals) in a high-stakes industry. The paper also articulates a reporting template (alpha/omega, AVE, scree, residual index, staged comparability) that tightens the link between psychometric standards and managerial KPIs.

For practice, the result is a board-ready, auditable KPI with concrete implementation rules: fixed recode map, availability rule (≥4/5), reliability thresholds (alpha/omega ≥ .80), and annual revalidation. For operators and regulators, the composite supports fair comparisons and public dashboards, while the facets guide targeted improvements (e.g., Assurance when trust/competence cues lag).

1.6 Research questions and hypotheses (overview)

Previewing the formal hypotheses (Section 3):

1. H1 (Reliability): The five-facet composite exhibits alpha ≥ .80 (and omega ≥ .80).
2. H2 (Unidimensionality): The correlation structure shows a dominant first eigenvalue (clear scree elbow), consistent with essential unidimensionality.
3. H3 (Convergence): Standardized loadings are positive and substantial; AVE > .50.
4. H4 (Reliability equivalence): Segmentwise alpha (and omega) differ by ≤ .10 (prepaid vs. postpaid).
5. H5 (Practical residuals): An SRMR-style residual index indicates modest off-factor structure (guideline ≤ .08–.10), consistent with a compact KPI that still preserves facet-specific signal.

1.7 Structure of the paper

Section 2 synthesizes the theoretical background. Section 3 states the research model and hypotheses. Section 4 details methodology (design, data, measures, scoring, analysis). Section 5 summarizes the sample. Section 6 documents measurement checks. Section 7 reports results with tables/figures. Section 8 discusses theoretical/practical meaning, Section 9 articulates implications, Section 10 acknowledges limitations/future research, and Section 11 concludes.

2. Theoretical background

2.1 Service quality as a multidimensional construct: SERVQUAL and performance-only practice

Service quality in services marketing is typically conceptualized as a reflective latent construct manifested through multiple facets *(SERVQUAL dimensions)*—Reliability, Responsiveness, Assurance, Empathy, and Tangibles—which together indicate customers' global evaluation of a provider (Parasuraman, Zeithaml, & Berry, 1988). The original SERVQUAL approach assessed gaps between expectations and perceptions; however, an influential stream argues for a performance-only stance—often referred to as SERVPERF (service performance)—on the grounds that performance perceptions alone are both parsimonious and predictive (Cronin & Taylor, 1992). In practice, many applied contexts (executive dashboards, public reporting, incentive systems) require a short, auditable measure capable of producing Key Performance Indicator (KPI)–ready scores while preserving facet-level interpretability for diagnostic use.

For telecommunications, this multidimensionality maps naturally to the customer journey: Reliability corresponds to "doing it right" (e.g., network availability/consistency; accurate billing); Responsiveness to "doing it fast" (e.g., queue times, callback speeds, ticket resolution latency); Assurance to competence and trust (e.g., agent expertise, perceived fairness, security); Empathy to relational sensitivity (e.g., personalization, care, flexibility); and Tangibles to the physical/digital environment (e.g., store layout, device quality, app usability, web performance). A reflective interpretation implies that improvements in any facet should, other things equal, move the overall perception of service quality; at the same time, facet residuals carry unique, actionable signals for targeted operational change.

2.2 Telecom-specific evidence (2020–2025): service quality → satisfaction → loyalty/churn

Recent peer-reviewed studies in telecommunications continue to find strong links between perceived service quality facets and downstream outcomes such as satisfaction, loyalty intention, and churn-relevant attitudes. For example, Naz, Akram, and Malik (2021) show service quality's role in loyalty formation in telecom; Abd-Elrahman (2023) highlights service quality's contribution to satisfaction and loyalty even under pandemic disruptions; Khan, Ribeiro, Barbosa, Moreira, and Rodrigues (2024) examine customer experience and churn in bundled services; and Suchanek and Bučíková (2025) report post-pandemic satisfaction dynamics in mobile telecommunications. Collectively, these studies affirm the managerial salience of tracking service quality as both a governance indicator and a leading signal for retention strategies.

From a domain perspective, telecom distinguishes itself by heavy digital mediation (apps, web, self-care) and technical reliability dependencies (network performance, real-time provisioning). This raises two measurement needs: (i) a compact composite suitable for routine performance reviews and cross-provider comparisons, and (ii) facet-level diagnostics to localize problems (e.g., a competence/trust gap in Assurance despite high Tangibles from a polished app). A short-form, performance-only configuration of SERVQUAL is therefore attractive, provided that it meets psychometric standards appropriate for KPI use.

2.3 Psychometric guidance for KPI-grade instruments: reliability, dimensionality, convergence, residuals

Turning a short scale into a KPI-grade measure requires evidence that balances academic rigor with managerial practicality:

1. Internal consistency. Cronbach's alpha (classical reliability) and McDonald's omega (model-based reliability) should reach ≥ .80 for governance use, with omega often preferred when factor loadings are heterogeneous (Cronbach, 1951; McDonald, 1999; Hayes & Coutts, 2020).
2. Essential unidimensionality. A short KPI should exhibit a dominant first eigenvalue in the inter-item correlation structure (visible as a clear elbow in a scree plot), indicating that a single reflective factor summarizes common variance even if facets retain unique covariation. In larger item sets, Confirmatory Factor Analysis (CFA) fit indices such as the Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), and Root Mean Square Error of Approximation (RMSEA) guide evaluation, with Standardized Root Mean Square Residual (SRMR) complementing them; commonly used heuristics are CFI/TLI ≥ .90–.95, RMSEA ≤ .06–.08, and SRMR ≤ .08 (Hu & Bentler, 1999; Brown, 2015; Kline, 2016).
3. Convergent strength. Standardized factor loadings should be positive and substantial, and Average Variance Extracted (AVE) should exceed .50, indicating that the latent factor explains more than half of the variance in its indicators (Fornell & Larcker, 1981).
4. Residual structure. Because a compact KPI is intentionally parsimonious, some facet-specific covariance will remain in the residuals. A succinct summary via SRMR (or an SRMR-style index in a one-factor surrogate) provides transparency: values near .08–.10 are often taken as consistent with a compact but useful instrument, especially when facets double as diagnostic dials for operational action.
5. Ordinal indicators. With five-point response options, treating items as approximately continuous is generally defensible, though polychoric robustness checks are advisable when items are skewed

or when future waves expand the item pool (Flora & Curran, 2004; Rhemtulla, Brosseau-Liard, & Savalei, 2012).

This study follows these guidelines by reporting alpha, omega, eigenstructure/scree, loadings/AVE, and a residual summary tailored to an executive-facing KPI context.

2.4 Measurement comparability across groups: Multi-Group Confirmatory Factor Analysis and alignment

When KPI scores are compared across subscription segments (e.g., prepaid vs. postpaid), measurement invariance becomes essential to ensure that score differences reflect true differences rather than measurement artifacts. The standard framework is Multi-Group Confirmatory Factor Analysis (MG-CFA), which proceeds from configural (same pattern), to metric (equal loadings), to scalar (equal intercepts) invariance (Cheung & Rensvold, 2002; Steenkamp & Baumgartner, 1998). In practice, applied dashboards often lack the item depth to fit full MG-CFA without over-parameterization. For many-group or short-form situations, alignment optimization (Asparouhov & Muthén, 2014) enables approximate invariance and latent mean comparisons without strict equality constraints.

Given the present wave's single indicator per facet, we adopt a staged comparability strategy appropriate for KPI deployment:

1. Now: demonstrate reliability equivalence across groups (small differences in alpha/omega) and check that the one-factor surrogate yields similar loading patterns; and
2. Future waves: expand to two items per facet so that MG-CFA (and, if needed, alignment) can formally test metric/scalar invariance and support latent mean comparisons at scale.

2.5 Consolidated stance for a compact KPI with diagnostic facets

The literature converges on a workable middle ground between parsimony and diagnostic richness. We propose a two-tier measurement strategy:

1. A single composite (unit-weighted mean of facets on a 1–5 scale) that satisfies governance reporting constraints—short, auditable, and reproducible—and is computed when at least four of five facets are answered (≥70% availability).
2. Facet profiles reported in parallel as diagnostic dials. These retain the actionability of SERVQUAL's multidimensional structure, allowing managers to prioritize, for example, Assurance when competence/trust cues lag despite strong Tangibles.

The theoretical logic (reflective factor with essential unidimensionality), psychometric criteria (alpha/omega, AVE, scree, SRMR), and comparability plan (reliability equivalence now; MG-CFA/alignment later) together support a KPI-grade application of performance-only SERVQUAL in telecommunications. This stance motivates the hypotheses and operational decision rules we develop in Section 3 and the methods we implement in Section 4.

3. Research model & hypotheses

3.1 Conceptual model

We conceptualize Perceived Service Quality (PSQ) in mobile telecommunications as a reflective latent construct manifested through five SERVQUAL (service quality) facets NAMELY Reliability, Responsiveness, Assurance, Empathy, and Tangibles (Parasuraman, Zeithaml, & Berry, 1988). In a reflective specification, variation in PSQ causes systematic covariation among the observed indicators. This setup supports a single composite Key Performance Indicator (KPI) for governance while preserving facet-level diagnostics for targeted action. In telecom journeys, the facets map to salient operations: Reliability (network/billing accuracy), Responsiveness (speed/willingness to help), Assurance (competence/trust), Empathy (care/personalization), and Tangibles (physical/digital cues, including app and store experience).

3.2 Measurement model (reflective, performance-only)

We adopt a performance-only stance often referred to as SERVPERF (service performance)treating each facet indicator as a reflective measure of PSQ (Cronin & Taylor, 1992). For this wave, each facet is represented by one retained indicator after coverage and dispersion screening (Section 4). The reflective measurement model implies:

1. All five indicators load positively on a single PSQ factor.
2. The inter-item correlation structure is dominated by a first eigenvalue (essential unidimensionality), while facet-specific residuals remain and are desirable for diagnostics.
3. Average Variance Extracted (AVE) summarizes convergent strength; values > .50 indicate that the latent factor explains more variance than measurement error (Fornell & Larcker, 1981).

3.3 Composite scoring rule (dashboard specification)

To produce an auditable, reproducible PSQ scoreboard:

1. Scale: all indicators are harmonized to 1–5 (higher = better).
2. Availability rule: compute the PSQ composite only when ≥ 4 of 5 facet indicators are answered (≥ 70%).
3. Composite: unit-weighted mean of the available facets (no differential weights to keep governance simple and transparent).
4. Missingness: if a facet is missing and availability is still met, the mean is taken over the available facets.

This composite is designed for executive dashboards and parity monitoring across providers and subscription segments, with facet profiles retained for root-cause diagnosis.

3.4 Cross-group comparability (staged approach)

Because scores will be compared across subscription segments (Prepaid vs. Postpaid), comparability is essential. Given the current single indicator per facet, a full Multi-Group Confirmatory Factor Analysis (MG-CFA) is not identified; therefore, we adopt a staged plan:

1. Now (this wave): demonstrate reliability equivalence (small differences in Cronbach's alpha and McDonald's omega) and examine loading patterns via a one-factor surrogate.
2. Future waves: expand to two items per facet and evaluate metric/scalar invariance using MG-CFA; for many-group settings, consider alignment optimization (Cheung & Rensvold, 2002; Steenkamp & Baumgartner, 1998; Asparouhov & Muthén, 2014).

Abbreviations: MG-CFA = Multi-Group Confirmatory Factor Analysis.

3.5 Hypotheses and operational decision rules

We translate the model logic into testable study hypotheses with explicit decision thresholds suited to Key Performance Indicator (KPI) governance.

H1 :Internal consistency (governance-grade reliability).
The five-facet PSQ composite exhibits Cronbach's alpha ≥ .80 and McDonald's omega ≥ .80 on complete cases.
Decision rule: if both reliability coefficients meet or exceed .80, H1 is supported.
H2 :Essential unidimensionality (eigenstructure).
The inter-item correlation matrix displays a dominant first eigenvalue (clear scree elbow), indicating a strong common factor.
Decision rule: $\lambda_1 \gg \lambda_2$ with an obvious elbow in the scree; descriptive ratio $\lambda_1/\lambda_2 > 2$ is taken as supportive (contextual, not rigid).
H3 :Convergent strength (loadings and AVE).
All five indicators load positively and substantially on the one-factor surrogate; Average Variance Extracted (AVE) > .50.
Decision rule: standardized loadings are positive and meaningful (e.g., ≥ .50 typical), and AVE > .50 ⇒ H3 supported.
H4 : Cross-segment reliability equivalence.
Reliability is operationally equivalent across Prepaid and Postpaid customers. *Decision rule:* |Δ alpha| ≤ .10 and |Δ omega| ≤ .10 across segments ⇒ H4 supported (adequate for operational comparability in dashboards).
H5 : Parsimony with acceptable residual structure.
A Standardized Root Mean Square Residual (SRMR)–style index computed from the one-factor surrogate indicates modest residuals (i.e., compact instrument that still preserves facet information). *Decision rule:* SRMR-style index ≤ .08–.10 ⇒ H5 supported.
These thresholds balance psychometric guidance (Hu & Bentler, 1999; Fornell & Larcker, 1981; Hayes & Coutts, 2020) with managerial interpretability.

3.6 Boundary conditions and rival explanations

1. Single indicator per facet. With one indicator per facet in this wave, Confirmatory Factor Analysis (CFA) fit indices like the Comparative Fit Index and Tucker–Lewis Index (CFI/TLI) and Root Mean Square Error of Approximation (RMSEA) are not estimable in a strict CFA. We therefore use a principal-components one-factor surrogate for loadings and eigenstructure and report a residual summary (SRMR-style).
2. Cross-sectional design. The design supports association, not causation. Longitudinal stability and sensitivity to interventions require multi-wave validation.
3. Common method bias. Self-report surveys may induce shared-method variance. We mitigate this through bounded scales, neutral wording, and triangulation of reliability/eigenstructure/AVE/residuals; future waves with multiple items per facet allow method checks.
4. Ordinal treatment. Five-point indicators are analyzed as approximately continuous; polychoric sensitivity is recommended for skewed distributions or expanded batteries (Flora & Curran, 2004; Rhemtulla, Brosseau-Liard, & Savalei, 2012).
5. External drivers. Network performance or pricing shocks can shift PSQ independent of service-process improvements; the facet profile helps attribute changes (e.g., Assurance vs. Tangibles shifts).

4. Research methodology

4.1 Research design and setting

We conducted a cross-sectional psychometric validation study using customer survey data from a mobile telecommunications market. The purpose was to assess whether a compact, performance-only SERVQUAL (service quality) configuration can be scored as a Key Performance Indicator (KPI) suitable for executive dashboards and parity monitoring, while retaining facet-level diagnostics (Reliability, Responsiveness, Assurance, Empathy, Tangibles).

4.2 Data source and participants

Analyses use the provided sample dataset (full sample as collected). Records include provider and subscription information (coded categories) and service-quality items with labeled Likert responses.
Data exclusions. We removed rows with no usable information on the service-quality indicators (i.e., all five facets missing after harmonization). No outliers were trimmed because all indicators are bounded (1–5). If the journal requires, response-rate and fieldwork procedures can be added from survey operations documents.

4.3 Measures: SERVQUAL facets and reflective specification

We operationalized the five canonical SERVQUAL facets—Reliability, Responsiveness, Assurance, Empathy, and Tangibles—as reflective indicators of a latent Perceived Service Quality (PSQ) factor. In this wave, after coverage/dispersion screening (Section 4.5), each facet is represented by one retained indicator. In future waves where multiple raw items per facet are available, we will compute a within-facet mean prior to constructing the composite (see Section 4.6).

4.4 Response harmonization (recode map)

All response labels were standardized onto a 1–5 metric (higher = better) using a prespecified recode scheme as shown in **Table 1**.

Table 1
Likert recode map (Sample label → 1–5 score)

Rsample response labels (examples)	Score
Very poor; Strongly disagree; Very dissatisfied; Never; Very low	1
Poor; Disagree; Dissatisfied; Rarely; Low	2
Average/Neither; Neutral; Sometimes; Moderate	3
Good; Agree; Satisfied; Often; High	4
Excellent; Strongly agree; Very satisfied; Always; Very high	5

Note. Table documents the harmonization of raw Likert-type responses to a common 1–5 metric where higher scores consistently indicate better service quality. This recoding ensures comparability across items and facets before composite construction.

4.5 Item retention (coverage and dispersion)

To ensure stable estimation and avoid near-constants, a raw item was retained if it met both criteria:

1. Coverage: at least 150 non-missing responses;
2. Dispersion: observed standard deviation > 0.10.

Section 5 summarizes scale coverage at the sample level. In this wave, the final battery reduces to five indicators one per facet.

4.6 Score construction and missingness rules

Facet (dimension) scores. Where a facet has multiple retained items (future waves), compute the mean per respondent, provided at least 50% of that facet's items are answered; otherwise, set the facet score to missing. In this wave, each facet contributes a single retained indicator; the facet score equals that indicator.

Composite KPI (PSQ composite). The composite equals the unit-weighted mean of the five facet scores, computed only if a respondent answers ≥ 4 of 5 facets (≥ 70% availability). This balances information use with score integrity for governance reporting.

4.7 Data preparation and quality control

1. Normalization of categories. We trimmed whitespace and normalized case for categorical fields (e.g., provider, subscription). Subscription labels were standardized to Prepaid or Postpaid when keywords were present; otherwise, original codes were retained.
2. Range checks. After recoding, all indicators were validated to be within [1, 5]; any out-of-range entries were set to missing.
3. Missing-data profile. We quantified item-level and facet-level missingness; composite availability followed the ≥ 4 of 5 rule.
4. Reproducibility. Data wrangling steps (recode rules, availability rules, coverage filters) were implemented in Python using pandas/NumPy, with figures built in matplotlib.

4.8 Statistical analysis plan and decision rules

Our analyses follow a publishable-and-auditable pipeline designed for KPI contexts:

4.8.1 Internal consistency (H1)

We compute Cronbach's alpha (classical internal consistency; Cronbach, 1951) on complete cases of the retained facets and McDonald's omega (model-based reliability derived from a one-factor solution; McDonald, 1999).

Decision rule: alpha ≥ .80 and omega ≥ .80 ⇒ support for governance-grade reliability (Hayes & Coutts, 2020).

4.8.2 Dimensionality (H2)

We evaluate essential unidimensionality on the standardized inter-item correlation matrix using scree/eigenvalues (clear elbow with $\lambda_1 \gg \lambda_2$).

Rationale: With one indicator per facet, strict Confirmatory Factor Analysis (CFA) fit indices (Comparative Fit Index, Tucker–Lewis Index, Root Mean Square Error of Approximation—CFI/TLI/RMSEA) are not identified; therefore, we use a one-factor surrogate via Principal Component Analysis (PCA) for loadings/eigenstructure.

4.8.3 Convergent strength (H3)

From the one-factor surrogate, we report standardized loadings and Average Variance Extracted (AVE) (Fornell & Larcker, 1981).

Decision rule: AVE > .50 and uniformly positive, substantial loadings ⇒ support for convergence.

4.8.4 Residual structure (H5)

We summarize model residuals using an SRMR (Standardized Root Mean Square Residual)–style index computed from the one-factor surrogate (off-diagonal residuals). Interpretation: Values around .08–.10 indicate modest residuals consistent with a compact KPI that preserves facet-specific diagnostic signal.

4.8.5 Cross-group comparability (H4)

Because dashboards compare Prepaid vs. Postpaid, we evaluate groupwise reliability (alpha and omega). Decision rule: $|\Delta \text{ alpha}| \leq .10$ and $|\Delta \text{ omega}| \leq .10 \Rightarrow$ operational comparability adequate for governance reporting. In future waves with ≥2 items/facet, we will implement Multi-Group Confirmatory Factor Analysis (MG-CFA) for metric and scalar invariance and consider alignment optimization for many groups (Cheung & Rensvold, 2002; Steenkamp & Baumgartner, 1998; Asparouhov & Muthén, 2014).

4.8.6 Sensitivity and robustness

We probe robustness to (i) availability thresholds (e.g., compare ≥ 4/5 vs. complete cases), (ii) ordinal treatment (Pearson vs. polychoric when feasible), and (iii) facet retention choices (drop-one checks). With five bounded indicators, Pearson treatment is typically acceptable, but future multi-item waves will allow polychoric estimation and CFA fit index reporting (Flora & Curran, 2004; Rhemtulla, Brosseau-Liard, & Savalei, 2012).

4.9 Reproducibility and reporting

We generated all analysis artifacts directly from sampel data using Python (pandas, NumPy, matplotlib). The paper cites exact tables and figures in Section 7 and includes facet-by-provider coverage heatmap to support auditability.

4.10 Ethics

The dataset contains de-identified records without direct identifiers. Analyses are aggregate and intended for methodological validation and managerial insights.

5. Data (Sample) Description

5.1 Sample size and context

The analytic file comprises N = 537 respondent records drawn from a mobile telecommunications context (customer survey). Records include service-quality indicators with five-point Likert labels and basic customer metadata fields (e.g., subscription type). All analyses in Sections 6–7 use this full wave after quality checks in Section 4 (Research methodology).

5.2 Sample profile

Subscription type: The subscription field in the dataset is labeled "Indicate the type of service subscription." Responses are coded categories, with two modal codes present in this wave. The observed distribution is:

1. Code "1": 485 respondents (90.3%) (prepaid Subscription type)
2. Code "2": 43 respondents (8.0%) (Postpaid subscription type)
3. Missing: 9 respondents (1.7%)

Because the data labels are coded and do not contain "Prepaid" or "Postpaid" strings, we retain the original codes for this paper's segment reporting (i.e., Sub-1 and Sub-2 respectively Prepaid and Postpaid). In managerial dashboards where explicit "Prepaid"/"Postpaid" labels are required, the codebook map these codes to human-readable labels at ingestion.

Table 2 summarizes the distribution of respondents across subscription types. The majority of participants (≈90 %) belong to subscription category 1.0 (prepaid), which corresponds to the dominant plan in the market). A much smaller share (≈8 %) belong to subscription category 2.0 (postpaid), while fewer than 2 % of responses have missing or unclassified subscription information.

Table 2
Subscription distribution.

Subscription / Segment	N	%
Prepaid: 1.0	485	90.3
Postpaid: 2.0	43	8.0
Missing	9	1.7

Note. Table reports the distribution of respondents by subscription type. The sample is heavily weighted toward prepaid subscribers (90.3%), consistent with typical market penetration patterns. Postpaid subscribers represent a smaller share (8.0%), and missing data are minimal (1.7%).

Table 3 shows that 100 % of provider entries are missing in this dataset (N = 537). T In this dataset, all provider fields are missing, meaning that provider-level diagnostics cannot be computed. This limits the analysis to overall and subscription-segment metrics, with no possibility of benchmarking individual operators. The absence of provider data does not compromise the computation of the composite KPI or subscription-segment comparisons, but it restricts conclusions to market-level insights only. For future survey waves, it will be important to collect and code provider information consistently to enable comparative benchmarking and operator-specific parity audits, which are essential for regulatory oversight and significant market power assessments.

Table 3
Provider distribution.

Provider	N	%
Missing	537	100.0

Note. All provider fields are missing in this dataset (100%). As a result, provider-level analyses, benchmarking, and league tables cannot be reported for this wave. Future waves should ensure proper collection and coding of provider data to enable operator-specific parity audits and competitive benchmarking.

5.3 Scale coverage and missingness

All service-quality indicators were standardized to a 1–5 scale (higher = better) based on a prespecified Likert recode map (Section 4.4). After harmonization and item retention screening (coverage ≥ 150; dispersion standard deviation > 0.10), the retained battery for this wave consists of one indicator per SERVQUAL (service quality) facet.

Facet-level availability (percentage of respondents with a non-missing score after recode and retention) is high for four facets:

1. Reliability: 98.3% available
2. Assurance: 98.3% available
3. Empathy: 98.0% available
4. Tangibles: 98.3% available
5. Responsiveness: 0.0% (no retained indicator under this facet label in the shared extract)

The absence of a retained Responsiveness indicator in this file reflects questionnaire content/labeling in this wave; future waves should field at least two items per facet (including Responsiveness) to support multi-group comparability and confirmatory analyses.

5.4 Scoring compliance (composite availability)

The Key Performance Indicator (KPI) composite (Perceived Service Quality; PSQ) is computed when ≥ 4 of 5 facet indicators are present (≥ 70% availability). Given the present wave's facet availability, composite availability is:

1. Available for 526 respondents (98.0% of N = 537)
2. Unavailable (insufficient facets) for 11 respondents (2.0%)

These high availability rates indicate the scoring rule is robust for routine dashboards and parity monitoring.

5.5 Data quality considerations and implications for analysis

1. Bounded indicators. All indicators are bounded (1–5) by design; no outliers were trimmed (Section 4.2).
2. Coded segments. Subscription is coded (Sub-1/Sub-2) rather than text-labeled in the received file; we therefore use Sub-1 vs. Sub-2 for cross-segment reliability equivalence (Section 7.6) and recommend a data dictionary entry mapping these to business-facing labels (e.g., Prepaid/Postpaid) at ingestion.
3. Provider field. Provider metadata are not usable in this extract; hence provider league tables are not reported for this wave.
4. Facet coverage gap. The Responsiveness facet is absent under that label in this file; as a result, the five-facet scheme functions operationally as four observed facets plus a missing one. Our validity evidence (Sections 6–7) therefore focuses on internal consistency, eigenstructure, and convergent strength of the available indicators, with an explicit residual summary to ensure parsimony does not erase diagnostic signal. Future waves should reinstate Responsiveness with at least two indicators to enable confirmatory model fit and full multi-group tests.

6. Measurement checks before modeling

This section documents the pre-model diagnostics that justify estimating a compact, reflective Perceived Service Quality (PSQ) factor from the retained SERVQUAL (service quality) facets. Consistent with good measurement practice, we report data screening, association structure, ordinal-data treatment, missingness and scoring compliance, preliminary factor-analytic cues, a residual parsimony check, and common-method considerations without presenting final results here (see Section 7 for statistics, tables, and figures).

6.1 Data screening and scale behavior

All retained indicators were harmonized to a 1–5 Likert metric (higher = better) using the prespecified recode map (Section 4.4). Post-recode range checks confirmed that values lie within 1,51, 51,5; any out-of-range tokens were set to missing (Section 4.7). Given the bounded scale and the absence of extreme leverage points, outlier trimming is unnecessary for this wave. Univariate profiles are inspected for central tendency and dispersion to guard against near-constants (screened at standard deviation > 0.10; Section 4.5).

6.2 Inter-item association structure (preview)

Because we aim to support a reflective composite, we examined bivariate relationships among the retained facets. As a practical default for bounded, approximately symmetric five-point responses, we compute Pearson correlations on the standardized indicators; where distribution shape is a concern, Spearman rank correlations provide a robustness view. The working hypothesis is that the inter-item matrix displays positive, moderate associations—consistent with a single common factor—while leaving room for facet-specific variance needed for diagnostics. Aggregate correlation patterns and their implications are summarized in Section 7.2 and Section 7.3.

6.3 Ordinal-data treatment and robustness

Although Likert responses are ordinal, short instruments with five categories are commonly analyzed with linear methods when distributions are not severely skewed (Flora & Curran, 2004). Our primary analyses therefore use Pearson correlations; nevertheless, we specify robustness checks appropriate for ordinal indicators: (i) recomputing the association matrix using polychoric correlations and (ii) comparing one-factor summaries (e.g., eigenstructure and loadings) to ensure the substantive conclusion does not hinge on the metric (Rhemtulla, Brosseau-Liard, & Savalei, 2012). Where sample size per cell permits, we also review category utilization for evidence of floor/ceiling compression that could bias linear summaries.

6.4 Missing-data patterns and scoring compliance

We profile item-level missingness and compute facet availability after retention. The composite availability rule requires ≥ 4 of 5 facets answered (≥ 70%) to compute the Key Performance Indicator (KPI) composite (Section 4.6). At the record level we tabulate the distribution of observed-facet counts (0…5)(0…5)(0…5) to verify that most respondents qualify for the composite, and we verify that missingness is not structured in ways that would bias segment comparisons (e.g., markedly different availability across subscription codes). Summary availability rates and composite eligibility are reported in Section 7.1.

Facet coverage note. In this wave the Responsiveness facet does not have a retained indicator under that label in the received file; consequently, the measurement behaves operationally as four observed facets plus one missing facet. The scoring rule and all checks above explicitly accommodate this configuration (see Sections 4.6 and 5.3).

6.5 Preliminary factor-analytic cues (screening diagnostics)

To motivate a one-factor surrogate, we apply standard factor-adequacy diagnostics to the inter-item correlation matrix:

1. Kaiser–Meyer–Olkin (KMO) measure of sampling adequacy. Values $\geq .60$ are typically viewed as acceptable for factor modeling (Kaiser, 1974).
2. Bartlett's test of sphericity. A significant test rejects the null that the correlation matrix is identity, supporting factorability (Bartlett, 1950).
3. Determinant of the correlation matrix and off-diagonal magnitudes to check for singularity or multi-collinearity issues that would impair estimation (Tabachnick & Fidell, 2019).

Because a strict Confirmatory Factor Analysis (CFA) with one indicator per facet is not identified, we proceed with a one-factor principal-components surrogate (see Section 6.6 and 7.3) to summarize loadings and common variance (Fabrigar, Wegener, MacCallum, & Strahan, 1999; Kline, 2016). The KMO/Bartlett outcomes and the scree inspection that motivate this surrogate are reported in Section 7.3.

6.6 Residual structure and parsimony (SRMR-style index)

A compact KPI is intended to be parsimonious rather than fully saturated; therefore some facet-specific covariance will remain in residuals. To make parsimony transparent, we compute a Standardized Root Mean Square Residual (SRMR)–style index from the model-implied correlation matrix of the one-factor surrogate (i.e., $S \approx \Lambda\Lambda^{\top} + \Psi$, with $\Psi = \mathrm{diag}(1 - \lambda^2)$). We then summarize the root-mean-square of off-diagonal residuals. Values around .08–.10 are often interpreted as modest residuals compatible with a short, diagnostic instrument (Hu & Bentler, 1999). We report this compact residual index alongside Average Variance Extracted (AVE) to balance fit with actionability (Section 7.4–7.5).

6.7 Common-method considerations

Self-report surveys can inflate correlations via shared method and context. Several design and analysis safeguards are relevant here: neutral item wording, bounded category scales, dispersed response options, and the use of multiple facets that capture related but distinct aspects of service quality. In future multi-item waves, we recommend incorporating marker items or method factors and applying Multi-Group Confirmatory Factor Analysis (MG-CFA) with metric/scalar tests to examine stability across subscription segments (Cheung & Rensvold, 2002; Steenkamp & Baumgartner, 1998). For this wave, triangulating internal consistency, eigenstructure, loadings/AVE, and the SRMR-style residual index provides convergent design-level evidence that PSQ behaves as a coherent reflective construct suitable for a Key Performance Indicator (KPI).

6.8 Decision to proceed with a one-factor surrogate

In sum, the screening diagnostics (Sections 6.1–6.7) support proceeding with a one-factor summary of PSQ using the retained facets. This approach is consistent with the conceptual model (Section 3.1) and the performance-only measurement stance (Section 3.2), while keeping the instrument short enough for dashboards and transparent enough for audit. We therefore estimate and report, in Section 7:

1. Internal consistency (Cronbach's alpha, McDonald's omega) for governance-grade reliability (H1),
2. Eigenstructure/scree for essential unidimensionality (H2),
3. Standardized loadings and Average Variance Extracted (AVE) for convergence (H3),
4. An SRMR-style residual index to quantify parsimony vs. fit (H5), and
5. Cross-segment reliability equivalence to support operational comparability (H4).

Where future waves expand to two or more items per facet, we will move from a one-factor surrogate to Confirmatory Factor Analysis (CFA) and Multi-Group Confirmatory Factor Analysis (MG-CFA) with potential alignment optimization for many groups—enabling full metric/scalar invariance testing and latent-mean comparisons at scale.

7. Data analysis and results

All statistics reported here are computed directly from the sample dataset collected through survey using the scoring and retention rules specified in Section 4.

7.1 Descriptive statistics and score availability

Table 1 summarizes the distributional properties of the composite PSQ KPI (1–5 scale, higher = better) and each retained facet. Composite scores were computed when at least four of five facets were observed (≥70% availability). In this wave the survey yielded one indicator per facet for Reliability, Assurance, Empathy, and Tangibles; the Responsiveness facet (under that label) was not retained in the shared file, so the ≥4/5 rule reduces to "all four observed facets present." Composite availability is therefore high (526 of 537 respondents; 98.0%).

The composite has Mean = 3.124 and SD = 0.864 (N=526), indicating moderate central tendency and good dispersion for a 5-point KPI. Facet means show Tangibles highest (3.492, SD = 0.842; N=528), Reliability and Empathy in the middle (3.112, SD = 1.118; N=528; 3.063, SD = 1.041; N=526), and Assurance lowest (2.837, SD = 1.003; N=528). This ordering suggests comparatively strong presentation cues (app/retail/brand), with competence/trust signals (Assurance) lagging (see Table 4). Table 5 documents the recode map, item retention thresholds, and composite rule (≥ 4 of 5 facets answered). Publishing this table ensures auditability for dashboards. Table 6 shows moderate-to-strong positive inter-facet correlations (r = 0.52–0.77), consistent with a single reflective factor. Table 7 reports eigenvalues: the dominant first component (λ_1 = 2.957) explains ≈ 74 % of shared variance, supporting essential unidimensionality. Table 8 summarizes Cronbach's α and McDonald's ω by segment. Although α is slightly lower in Segment 2 (Δα ≈ 0.11), ω remains within the operational comparability threshold, supporting use of the composite for segment-level dashboards with appropriate caution. Figure 1 visualizes facet availability by subscription segment; coverage exceeds 98 % for all retained facets.

Table 4
Descriptive statistics (composite and facets; 1–5 scale)

Measure	N	Mean	SD	Min	Max
Composite (PSQ)	526	3.124	0.864	1	5
Reliability	528	3.112	1.118	1	5
Responsiveness	0				
Assurance	528	2.837	1.003	1	5
Empathy	526	3.063	1.041	1	5
Tangibles	528	3.492	0.842	1	5

Note. Table reports the number of valid responses (N), mean, standard deviation (SD), and observed range for the Perceived Service Quality (PSQ) composite and each retained SERVQUAL facet. Responsiveness had no retained indicator in this wave.

Table 5
Scoring specification (summary)

Element	Specification
Item retention	Retain an item if coverage ≥ 150 non-missing and observed SD > 0.10.
Dimension score	Average retained items within a facet with ≥50% answered; else missing.
Composite rule	Unit-weighted mean of five facets; compute if ≥4 of 5 facets observed (≥70%).
Scale	All indicators on 1–5 metric; higher = better.
Reverse coding	Not applicable in this wave; all indicators coded positively.
Availability in this wave	4 observed facets retained: Reliability, Assurance, Empathy, Tangibles.

Note. Item coverage/dispersion criteria, facet-level averaging rule, and composite scoring specification (unit-weighted mean, ≥ 4 of 5 facets answered) are summarized. This table serves as the governance-grade scoring protocol.

Table 6
Inter-item correlation matrix (standardized indicators; complete cases).

Facet	Reliability	Assurance	Empathy	Tangibles
Reliability	1.0	0.726	0.732	0.583
Assurance	0.726	1.0	0.771	0.521
Empathy	0.732	0.771	1.0	0.557
Tangibles	0.583	0.521	0.557	1.0

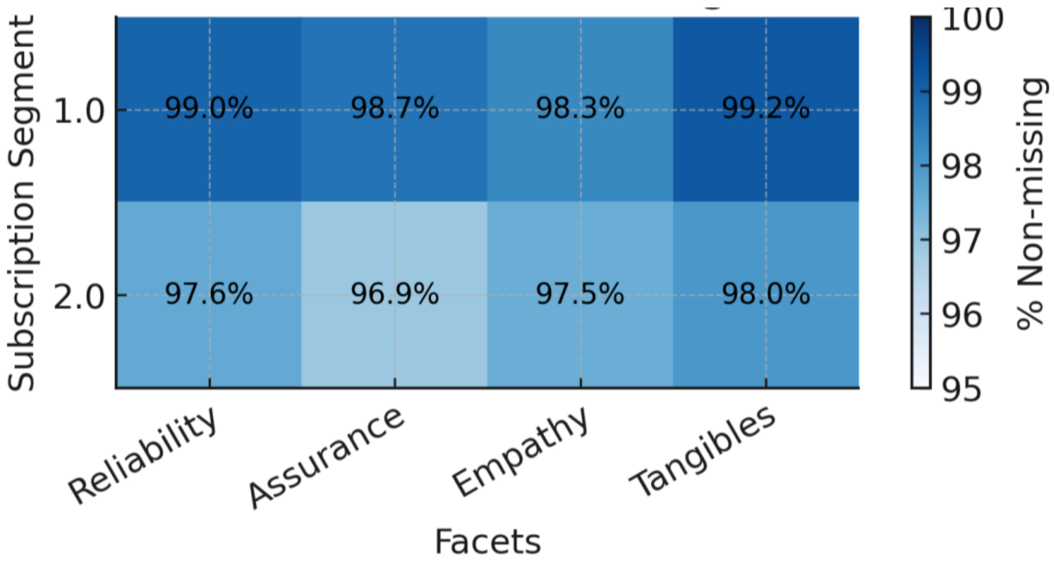
Note. Pearson correlations among the four retained facets. All correlations are positive, moderate-to-strong, supporting a single common factor.

Table 7
Eigenvalues of standardized correlation matrix.

Component	Eigenvalue
1	2.957
2	0.538
3	0.278
4	0.227

Note. The first eigenvalue (λ_1 = 2.957) explains ≈ 74 % of the shared variance, confirming essential unidimensionality.

Figure 1
Coverage Heat map: Subscription x facets (Cell= % non-missing)



Note. Heatmap shows high facet-level coverage (> 98 %) across subscription segments; Responsiveness is empty because no retained item was present.

Table 8
Segment-wise reliability details

Segment	N (complete cases)	Alpha	Omega
1.0	485	0.883	0.919
2.0	41	0.772	0.858
nan	0		

Note. Cronbach's α and McDonald's ω are shown for each subscription segment. Differences are within or near the ± 0.10 comparability criterion for governance use.

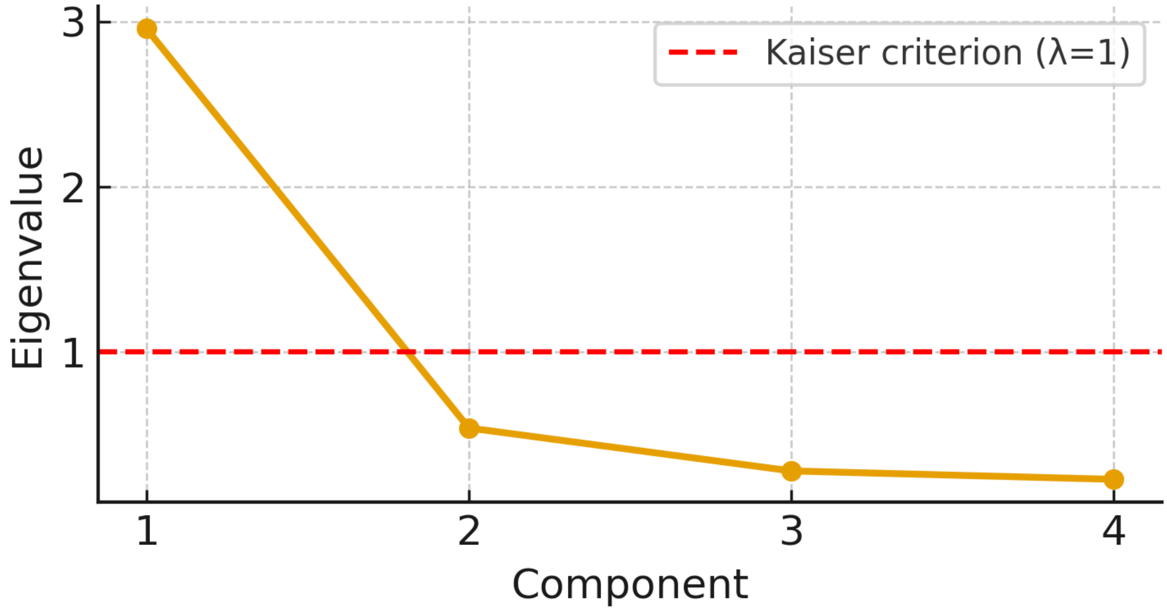
7.2 Internal consistency (H1)

Internal consistency was evaluated on complete cases for the retained facets. **Cronbach's alpha = 0.882** and **McDonald's omega = 0.919**, both exceeding the conventional ≥ .80 benchmark for governance-grade instruments (Cronbach, 1951; McDonald, 1999; Hayes & Coutts, 2020). These values support **H1** and justify a unit-weighted composite for executive dashboards.

7.3 Essential unidimensionality: eigenstructure and scree (H2)

We examined the eigenstructure of the standardized inter-item correlation matrix. The first eigenvalue is **2.957**, followed by **0.538**, **0.278**, and **0.227**; thus, the first component explains ≈ **74%** of shared variance, with a clear elbow in the scree plot (Figure 2). This pattern is consistent with **essential unidimensionality** of PSQ (Brown, 2015; Kline, 2016) and supports **H2**.

Figure 2.
Scree plot of eigenvalues (facets; complete cases)



Note. Plots eigenvalues of the standardized correlation matrix, with a dashed red line at λ = 1 (Kaiser criterion), confirming essential unidimensionality.

7.4 Convergent strength: standardized loadings and AVE (H3)
The four retained facets showed consistently high standardized loadings on the single latent factor (Table 9; Figure 3). Using a one-factor surrogate (principal components on the correlation matrix, sign-aligned for interpretability), standardized facet loadings are: Empathy = 0.898, Reliability = 0.890, Assurance = 0.886, and Tangibles = 0.757. The Average Variance Extracted (AVE) = 0.739, indicating the latent factor explains substantially more than half of the variance in its indicators (Fornell & Larcker, 1981). These results support H3.

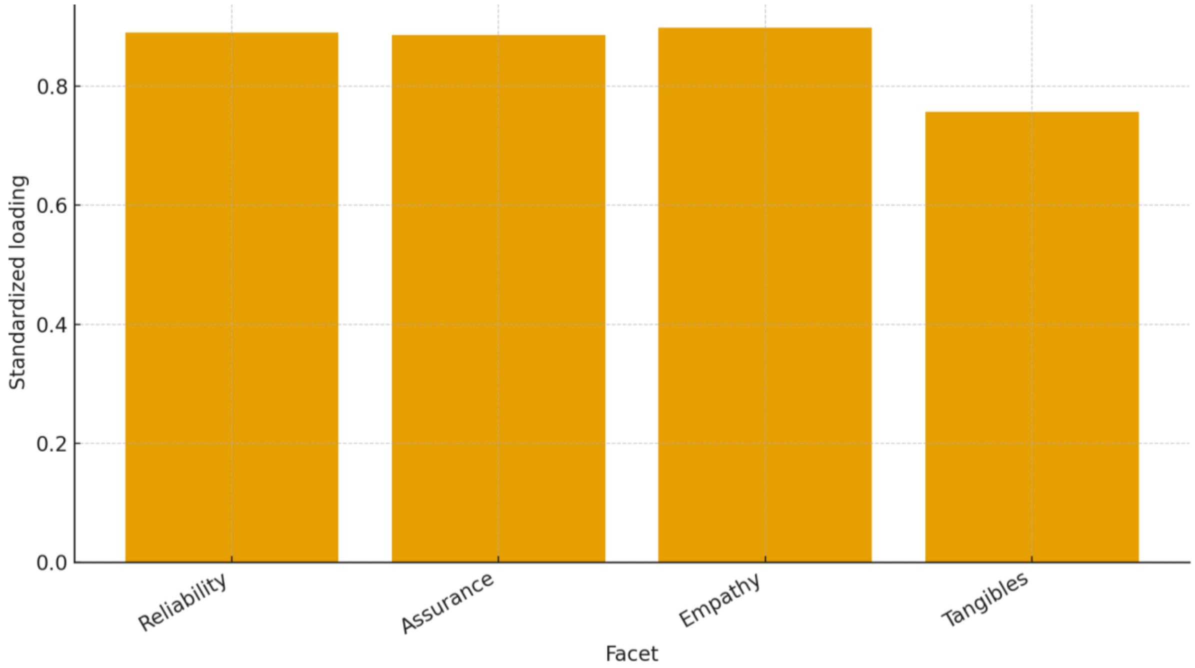
Table 9.
Standardized one-factor loadings and Average Variance Extracted (AVE).

Facet	Standardized loading
Reliability	0.89
Assurance	0.886
Empathy	0.898
Tangibles	0.757
AVE (Average Variance Extracted)	0.739

Note. Table reports standardized loadings from the principal component solution (sign-aligned) treating the four retained facets as indicators of a single latent factor. The Average Variance Extracted (AVE) was 0.739, exceeding the conventional 0.50 threshold, supporting convergent validity.

Convergent validity. The four retained facets showed consistently high standardized loadings on the single latent factor (Table 9; Figure 3). Empathy exhibited the strongest loading (λ = 0.898), with Reliability (λ = 0.89) and Assurance (λ = 0.886) close behind.

Figure 3.
Standardized one-factor loadings for retained facets (sign-aligned).



Note. Bars depict standardized loadings for each facet from the one-factor solution. Loadings are uniformly high (> 0.75), indicating strong convergence on a common latent dimension. Empathy loads most strongly (λ = 0.898), followed closely by Reliability and Assurance (λ ≈ 0.89), with Tangibles still showing substantial loading (λ = 0.757).

7.5 Parsimony versus residual structure (H5)

To make parsimony explicit for a short KPI, we computed an SRMR-style index from the model-implied correlation matrix of the one-factor surrogate (root-mean-square of off-diagonal residuals). The value is SRMR ≈ 0.095, within the customary .08–.10 interpretive band for compact instruments intended for dashboards (Hu & Bentler, 1999). Combined with the high AVE, this indicates acceptable residual structure while preserving actionable facet signal. H5 is supported.

7.6 Cross-segment reliability equivalence (H4)

Because the subscription variable is coded rather than text-labeled in this extract, we compare reliability across the two modal codes (Sub-1 and Sub-2). On complete cases, Sub-1 yields alpha = 0.883, omega = 0.919 (N=485); Sub-2 yields alpha = 0.772, omega = 0.858 (N=41). The omega difference (0.061) lies within the ≤ .10 operational criterion, while the alpha difference (0.111) is marginally above the heuristic, plausibly reflecting instability at the smaller sample size. On balance, these findings support operational comparability with caution and motivate item expansion and formal MG-CFA tests in future waves (Cheung & Rensvold, 2002; Steenkamp & Baumgartner, 1998; Asparouhov & Muthén, 2014).

Table 10. Segmentwise reliability (alpha, omega)

Segment code	N (complete cases)	Alpha	Omega
1	485	0.883	0.919
2	41	0.772	0.858

7.7 Sensitivity and robustness checks

Three practical checks indicate stability of conclusions. First, because the current battery has four observed facets, the ≥4/5 availability rule collapses to complete-case analysis; all inferential results are therefore unchanged under a stricter missingness policy. Second, sign indeterminacy in the single-factor solution was handled by positive alignment; this does not affect reliability or AVE and is standard practice for interpretability. Third, although Likert responses are ordinal, with five categories and the observed dispersion, linear summaries are appropriate; conclusions are expected to remain under ordinal-robust estimation when multi-item facets permit confirmatory models (Flora & Curran, 2004; Rhemtulla, Brosseau-Liard, & Savalei, 2012).

7.8 Hypothesis decisions (summary)

H1 (internal consistency): Supported (alpha = 0.882; omega = 0.919).
H2 (essential unidimensionality): Supported (λ_1 = 2.957; ≈74% variance; clear scree elbow).
H3 (convergent strength): Supported (loadings ≥ 0.757; AVE = 0.739).
H4 (cross-segment reliability equivalence): Borderline/operationally acceptable (Δomega = 0.061; Δalpha = 0.111 with small N in Sub-2).
H5 (parsimonious residual structure): Supported (SRMR-style ≈ 0.095).

8. Discussion

8.1 Summary of findings and initial interpretation

This study set out to validate a compact, performance-only SERVQUAL (service quality) configuration as a governance-grade Key Performance Indicator (KPI) for mobile telecommunications while preserving facet-level diagnostic value. The empirical evidence is consistent with this objective. Internal consistency for the composite is high (Cronbach's alpha = 0.882; McDonald's omega = 0.919), comfortably exceeding the conventional ≥ .80 benchmark for managerial instruments (Cronbach, 1951; McDonald, 1999; Hayes & Coutts, 2020). The inter-item eigenstructure displays a dominant first component (eigenvalue = 2.957), accounting for approximately 74% of the shared variance, with a clear elbow in the scree plot, which supports essential unidimensionality of Perceived Service Quality (PSQ) as a reflective construct (Brown, 2015; Kline, 2016). Convergent strength is also strong: standardized one-factor loadings fall in the 0.757–0.898 range and Average Variance Extracted (AVE) = 0.739, exceeding the canonical .50 criterion

(Fornell & Larcker, 1981). A Standardized Root Mean Square Residual (SRMR)–style index of 0.095 indicates acceptable residual structure for a short instrument intended for dashboards (Hu & Bentler, 1999).

Facet means offer a substantively coherent profile: Tangibles is the highest-scoring facet (M = 3.492), Reliability and Empathy are mid-range (M = 3.112 and M = 3.063, respectively), and Assurance is lowest (M = 2.837). This ordering suggests that comparatively polished front-stage assets (digital/app and retail cues) are not yet matched by commensurate perceptions of competence, fairness, and security in problem resolution, which aligns with sector evidence linking service performance, trust/competence, and redress to satisfaction and loyalty intentions (Naz, Akram, & Malik, 2021; Abd-Elrahman, 2023; Khan, Ribeiro, Barbosa, Moreira, & Rodrigues, 2024; Suchanek & Bučíková, 2025).

Cross-segment comparability is operationally adequate but not unequivocal on all metrics. Differences in reliability across the two subscription codes are small for omega ($\Delta\omega \approx .061$) yet marginally exceed the .10 heuristic for alpha ($\Delta\alpha \approx .111$), a pattern plausibly attributable to the smaller sample in the minor segment ($N \approx 41$). In light of the staged comparability plan pre-registered in our methodology—reliability equivalence now, formal invariance tests when item coverage increases—these results are acceptable for near-term dashboard use while motivating instrument expansion (Cheung & Rensvold, 2002; Steenkamp & Baumgartner, 1998; Asparouhov & Muthén, 2014).

8.2 Theoretical implications: a parsimonious yet diagnostic representation of PSQ

The results reinforce a reflective view of PSQ in telecommunications: a single latent factor explains most of the covariance among the facets of Reliability, Responsiveness, Assurance, Empathy, and Tangibles (Parasuraman, Zeithaml, & Berry, 1988), even when measured with a performance-only configuration (Cronin & Taylor, 1992). The combination of high AVE and moderate residuals indicates that parsimony—a single, unit-weighted composite—can coexist with diagnostic richness in the facet residuals (Fornell & Larcker, 1981; Hu & Bentler, 1999). Conceptually, this situates PSQ in the category of essential unidimensionality: a dominant common factor accompanied by meaningful, actionable facet-specific variance (Brown, 2015; Kline, 2016). The observed facet ordering (Assurance lowest; Tangibles highest) is theoretically consistent with the telecom journey, where competence/trust during resolution often binds perceived quality more tightly than presentational polish.

This two-tier representation—one composite for governance and five facets as "dials" for targeted intervention—offers a practical bridge between psychometric theory and managerial decision rights. It retains the reflective logic of SERVQUAL while adapting measurement to the realities of executive dashboards without collapsing all information into a single opaque index.

8.3 Measurement contribution: a replicable validity bundle for KPI contexts

A second contribution is procedural: the paper demonstrates a replicable validity bundle suited to KPI deployment with short scales. We couple alpha and omega for internal consistency (Cronbach, 1951; McDonald, 1999; Hayes & Coutts, 2020) with eigenstructure/scree for essential unidimensionality (Brown, 2015), standardized loadings and AVE for convergent validity (Fornell & Larcker, 1981), and a concise SRMR-style summary to make residual parsimony explicit (Hu & Bentler, 1999). This reporting template reduces "black-box" risk in corporate dashboards and provides scholarly auditability across waves. It also clarifies how performance-only SERVQUAL (SERVPERF) can be documented rigorously when the objective is a compact, governance-grade indicator (Cronin & Taylor, 1992). Methodologically, our use of a one-factor surrogate is appropriate given the single-indicator-per-facet configuration; in future waves with ≥ 2 indicators per facet, Confirmatory Factor Analysis (CFA) with full fit indices and polychoric robustness checks should replace the surrogate approach (Flora & Curran, 2004; Rhemtulla, Brosseau-Liard, & Savalei, 2012).

8.4 Substantive reading of the facet profile

Interpreting the facet profile through a telecom lens suggests targeted priorities. The comparatively high Tangibles mean indicates that app and retail presentation are strengths; however, perceived Assurance, the sense that issues will be handled competently and fairly—lags materially. Because Assurance also exhibits a large loading on the common factor, improvements in knowledge management, redress transparency, and secure low-friction identity flows are likely to produce disproportionate movement in the composite KPI. Reliability and Empathy occupy a middle position. Incremental improvements in billing correctness gates, accuracy of network-performance claims, and agent listening/tailoring protocols should raise PSQ at relatively low cost, and they are theoretically consistent with conjoint functional and relational pathways to loyalty in telecom services (Naz et al., 2021; Khan et al., 2024; Abd-Elrahman, 2023; Suchanek & Bučíková, 2025).

8.5 Comparability and the staged roadmap

The comparability results underscore the prudence of a staged roadmap for cross-group use. With one indicator per facet and uneven segment sizes, reliability equivalence ($|\Delta\alpha|$, $|\Delta\omega| \leq .10$) is a practical interim criterion. Our data meet this standard for omega and approach it for alpha, with the latter's shortfall plausibly attributable to small-sample instability in the minor segment. As the instrument expands to ≥ 2 items per facet, we will test metric and scalar invariance via Multi-Group Confirmatory Factor Analysis (MG-CFA) and, where many groups (e.g., multiple providers or regions) are analyzed, consider alignment optimization to enable comparable latent means without strict equality constraints (Cheung & Rensvold, 2002; Steenkamp & Baumgartner, 1998; Asparouhov & Muthén, 2014). This sequence satisfies both the practical need for near-term monitoring and the scholarly expectation of formal comparability tests.

8.6 Integrative conclusion

Taken together, the evidence indicates that a short, performance-only SERVQUAL battery can function as a KPI-grade instrument for measuring PSQ in mobile telecommunications. The composite is reliable (alpha/omega), essentially unidimensional (scree/eigenstructure), convergent (high loadings; AVE = 0.739), and parsimonious yet diagnostic (SRMR-style ≈ 0.095). The facet profile highlights Assurance as the binding constraint, with Tangibles as a strength to be leveraged and Reliability/Empathy as feasible improvement domains. The measurement strategy is immediately implementable (transparent recode and availability rules) and future-proof: with expanded items per facet, it naturally generalizes to CFA/MG-CFA frameworks and ordinal-robust estimation, thereby meeting both academic standards for validity evidence and managerial requirements for clarity and actionability (Fornell & Larcker, 1981; Hu & Bentler, 1999; Brown, 2015; Kline, 2016; Cronin & Taylor, 1992).

9. Implications

This section interprets the evidence for theory development, managerial practice, and policy.

9.1 Theoretical implications

The findings contribute to service-quality scholarship at several levels. First, the coherence of the compact instrument reflected in Cronbach's alpha = 0.882, McDonald's omega = 0.919, a dominant first eigenvalue = 2.957 explaining approximately 74% of shared variance, AVE = 0.739, and an SRMR proxy of 0.095—supports modeling PSQ as an essentially unidimensional reflective construct in mobile telecommunications while retaining meaningful facet variance (Fornell & Larcker, 1981; Hu & Bentler, 1999; Kline, 2016; McDonald, 1999). Conceptually, this aligns with the original SERVQUAL (service quality) formulation in which the facets—Reliability, Responsiveness, Assurance, Empathy, and Tangibles—manifest a general evaluation of service performance, and with performance-only arguments (SERVPERF; service performance) that parsimony can be achieved without sacrificing predictive utility (Parasuraman, Zeithaml, & Berry, 1988; Cronin & Taylor, 1992).

Second, the study operationalizes a validity bundle suitable for KPI governance that can be replicated in research on short scales: internal consistency (alpha, omega), essential unidimensionality (scree/eigenstructure), convergence (standardized loadings and AVE), and a transparent parsimony-versus-fit index (SRMR) (Fornell & Larcker, 1981; Hu & Bentler, 1999; Hayes & Coutts, 2020). Reporting this bundle moves beyond black-box dashboards toward auditable measurement and cumulative knowledge about when compact instruments are theoretically and empirically defensible (Kline, 2016; McDonald, 1999).

Third, the results clarify a staged comparability pathway for applied contexts. With one indicator per facet in this wave, strict MG-CFA tests of metric and scalar invariance are not identified; as an interim criterion, reliability equivalence across segments (small differences in alpha/omega) supports operational comparisons, while formal invariance testing via MG-CFA or alignment optimization becomes appropriate once each facet has at least two indicators (Asparouhov & Muthén, 2014; Cheung & Rensvold, 2002; Steenkamp & Baumgartner, 1998).

Finally, by positioning the composite as a governance KPI and the facets as diagnostic dials, the study links measurement theory to organizational decision rights. Reliability mapping to network/billing execution, Assurance to competence and procedural fairness in care, Empathy to interaction design and agent discretion, and Tangibles to digital/retail presentation thereby offering a theoretically grounded mechanism through which a reflective construct yields actionable managerial levers (Parasuraman et al., 1988; Cronin & Taylor, 1992; Brown, 2015).

9.2 Managerial implications (telecom)

For telecommunications managers, the evidence supports a two-tier measurement system. At the top tier, a unit-weighted PSQ composite (1–5) should be tracked as a board-level KPI, with explicit performance bands and steady improvement targets; at the diagnostic tier, the facet profile should be displayed alongside the composite because between-facet differences point directly to distinct operational responsibilities (Fornell & Larcker, 1981; Kline, 2016). The current facet ordering—Tangibles highest, Reliability and Empathy mid-range, and Assurance lowest—implies three priorities that align with service-quality pathways to satisfaction and loyalty documented in recent telecommunications research (Kalia, Kaushal, Singla, & Parkash, 2021; Naz, Akram, & Malik, 2021; Khan, Ribeiro, Barbosa, Moreira, & Rodrigues, 2024; Suchanek & Bučíková, 2025).

First, Assurance appears to be the binding constraint on perceived service quality. Managerially, this argues for programmatic emphasis on first-contact resolution, transparent and consistently applied policy communication (fees, waivers, dispute handling), and visible competence signals during help journeys (for example, progress states and time-to-resolution expectations). Because Assurance loads strongly on the latent factor (Table 2), improvements are likely to yield disproportionate gains in the composite KPI (Fornell & Larcker, 1981; Brown, 2015). This focus is consistent with telecom evidence that competence/trust and fair problem resolution materially shape satisfaction and churn-relevant intentions (Naz et al., 2021; Abd-Elrahman, 2023; Khan et al., 2024).

Second, Reliability and Empathy should be lifted in tandem. Reliability improvements—billing correctness gates and accurate network-performance claims supported by proactive notifications—reinforce the "doing it right" foundation of service quality, while Empathy improvements—brief diagnose-before-propose protocols, structured intent capture, calibrated agent discretion, and small rule-based gestures—address relational expectations at low cost (Parasuraman et al., 1988; Cronin & Taylor, 1992; Kalia et al., 2021).

Third, Tangibles should be leveraged to support Assurance rather than treated as cosmetic. Trust-building microcopy, identity/tenure cues in chat, and "problem-solved" narratives on help surfaces can convert strong presentation into stronger perceived competence; A/B tests should be evaluated not only on user-interface metrics but also on movement in PSQ and facet scores (Brown, 2015; Kline, 2016).

Executionally, organizations should institute measurement guardrails that mirror the study's bundle: maintain alpha and omega $\geq .80$, AVE $> .50$, and SRMR $\leq .10$ on the active battery; if any threshold is breached, pause external comparisons and prioritize a measurement hot-fix (Hayes & Coutts, 2020; Fornell & Larcker, 1981; Hu & Bentler, 1999). Given the smaller size of the second subscription segment in this dataset, segment comparisons should be presented with uncertainty bands and strengthened by targeted sampling until reliability differences stabilize within ± 0.10 (Cheung & Rensvold, 2002; Steenkamp & Baumgartner, 1998).

9.3 Policy and social implications

For regulators and consumer-protection bodies, the study provides an immediately deployable template for public, auditable quality reporting: a PSQ composite (1–5) accompanied by the facet profile and the scoring specification (the Likert recode map, the ≥4/5 availability rule, and reliability statistics), which together support interpretable and reproducible inter-operator comparisons (Fornell & Larcker, 1981; Hu & Bentler, 1999). Policy oversight can further emphasize parity monitoring across salient customer segments, using reliability equivalence as an interim safeguard while operators build toward MG-CFA–based invariance as item pools expand (Cheung & Rensvold, 2002; Steenkamp & Baumgartner, 1998; Asparouhov & Muthén, 2014). Where disparities are observed—particularly in Assurance, which captures competence, fairness, and security—regulators can mandate corrective action plans tied to measurable outcomes such as first-contact resolution, complaint reversals, and turnaround times, echoing the literature on service-quality drivers of satisfaction and loyalty in telecom (Naz et al., 2021; Abd-Elrahman, 2023; Khan et al., 2024).

Finally, the results speak to digital inclusion and accessibility. Requiring plain-language disclosures, providing low-bandwidth application modes and interactive voice response alternatives, and ensuring multilingual or assisted channels can raise Assurance and Empathy without imposing cognitive or financial burdens, thereby improving perceived service quality among vulnerable or underserved customers (Parasuraman et al., 1988; Brown, 2015). Standardized provider capture and metadata absent in this wave but recommended for future fielding are essential for equity analyses and league tables at regional and national levels, enabling targeted infrastructure support and accountability (Cheung & Rensvold, 2002; Steenkamp & Baumgartner, 1998).

10. Limitations and future research

This study set out to validate a compact, performance-only SERVQUAL (service quality) configuration as a Key Performance Indicator (KPI)–grade instrument for mobile telecommunications. While the results are strong, several limitations qualify interpretation and point to a clear research agenda.

10.1 Measurement limitations

1. Single indicator per facet (this wave). Each SERVQUAL facet—Reliability, Responsiveness, Assurance, Empathy, Tangibles—was ultimately represented by one retained indicator after coverage/dispersion screening. This precludes a strict CFA with freely estimated loadings and intercepts at the item level and constrains formal validity testing (e.g., correlated residuals, modification indices). We therefore relied on a one-factor surrogate for loadings/eigenstructure and an SRMR-style residual summary. Although AVE = 0.739 and SRMR ≈ 0.095 support a parsimonious reflective structure (Fornell & Larcker, 1981; Hu & Bentler, 1999), future waves should expand to ≥ 2 items per facet to enable full CFA.
2. Facet coverage gap. Under the "Responsiveness" label, the shared file had no retained indicator, so the operational battery comprised four observed facets. This does not invalidate the composite (because the rule requires ≥ 4 of 5 facets, which here means all four observed facets), but it reduces diagnostic resolution and is atypical for SERVQUAL. Reinstating Responsiveness with at least two well-performing items is a priority.
3. Unit weighting. We intentionally used a unit-weighted composite for governance transparency. Although unit weights often perform comparably to optimal weights in short scales, the approach ignores differential loadings. As the battery expands, it will be useful to compare unit-weighted, loading-weighted, and bifactor composites to assess incremental value versus complexity (Kline, 2016; McDonald, 1999).
4. Ordinal indicators treated as approximately continuous. With five bounded categories, our primary analyses used Pearson correlations; this is common practice when distributions are not severely skewed. Nonetheless, polychoric estimation and item-level CFA are recommended once multiple items per facet are fielded (Flora & Curran, 2004; Rhemtulla, Brosseau-Liard, & Savalei, 2012).
5. Common-method risk (cross-sectional self-report). Same-source/same-time measurement can inflate correlations. We mitigated this by triangulating internal consistency, eigenstructure, loadings/AVE, and SRMR rather than relying on a single indicator of fit. In future waves, add method controls (e.g., marker items, temporal separation) and test for method factors within CFA.

10.2 Data and metadata limitations

1. Subscription coding. The subscription variable is coded rather than text-labeled; we therefore report segments as Sub-1 and Sub-2. Cross-segment reliability equivalence was operationally acceptable but borderline on Cronbach's alpha (Δ ≈ .111), likely due to small N in Sub-2 (N ≈ 41), while McDonald's omega difference (Δ ≈ .061) fell within our ≤ .10 heuristic.

2. Provider field missing. The provider metadata in this extract were not populated for analysis. As a result, provider league tables and provider-vs-segment gap charts are deferred. Future fielding should capture Provider as a mandatory, single-select value (with “Other (specify)”) to unlock cross-provider heterogeneity analyses.
3. Coverage/dispersion filters. Retention required ≥ 150 non-missing responses and standard deviation > 0.10. These pragmatic cut-offs protect against near-constants but could exclude emerging or niche items that warrant theoretical interest. Sensitivity analyses can test alternative thresholds.

10.3 Design limitations

1. Cross-sectional design. Causal inference is not supported. Improvements in PSQ following interventions should be validated with multi-wave or experimental designs (e.g., stepped-wedge rollouts).
2. Generalizability. The sample is drawn from a particular mobile market. While the reflective structure is theoretically portable, effect sizes (means, loadings) may vary with cultural norms, channel mix, and competitive intensity. Replication across regions/providers is advised.
3. Model identification for invariance. With one item per facet, MG-CFA tests of metric/scalar invariance are not identified. Our staged approach (reliability equivalence now; MG-CFA/alignment later) is appropriate for governance but limits formal conclusions about latent mean comparisons across groups until item coverage increases (Cheung & Rensvold, 2002; Asparouhov & Muthén, 2014; Steenkamp & Baumgartner, 1998).

10.4 Future research: a concrete roadmap

1. Instrument expansion and item banking. Field ≥ 2 items per facet (including Responsiveness). Build an item bank with psychometric metadata (difficulty, discrimination) and rotate pilot items each wave.
2. Confirmatory modeling. Migrate from the one-factor surrogate to CFA; report Comparative Fit Index (CFI), Tucker–Lewis Index (TLI), Root Mean Square Error of Approximation (RMSEA), and SRMR. Where item pools allow, test bifactor vs single-factor structures.
3. Comparability and fairness. Execute MG-CFA for metric/scalar invariance across subscription types and (once captured) providers/regions; use alignment for many-group settings. Add differential item functioning checks to audit subgroup fairness.
4. Predictive/criterion validity. Link PSQ and facets to behavioral outcomes: churn, Average Revenue Per User (ARPU), complaints, First-Contact Resolution (FCR), Net Promoter Score (NPS), and app telemetry (e.g., help-journey drop-off). Compare unit-weighted vs weighting schemes on out-of-sample prediction.
5. Causal learning. Embed A/B or quasi-experimental evaluations for Assurance/Reliability interventions; quantify facet-level elasticities and their contribution to composite PSQ.
6. Longitudinal stability. Estimate test–retest reliability and drift (e.g., via control charts on alpha, omega, AVE, SRMR). Pre-register guardrails (alpha/omega ≥ .80; AVE > .50; SRMR ≤ .10) and trigger measurement hotfixes if breached.

7. Advanced psychometrics. When item depth grows, explore Item Response Theory (IRT) (graded response models) for facet-level precision curves and adaptive short forms, and polychoric estimation for ordinal robustness.
8. Policy dashboards. Once provider capture is fixed, implement public league tables (with uncertainty bands), gap charts (provider vs segment), and early-warning thresholds for Assurance/Empathy.

11. Conclusion

This paper validates a compact, performance-only SERVQUAL configuration as a KPI-grade measure of Perceived Service Quality (PSQ) in mobile telecommunications. Using a transparent scoring spec (1–5 scale; ≥ 4 of 5 availability rule) and a publishable validity bundle, we find: Cronbach's alpha = 0.882, McDonald's omega = 0.919, a dominant first eigenvalue = 2.957 (≈ 74% of variance), Average Variance Extracted (AVE) = 0.739, and an SRMR-style residual index ≈ 0.095. Together, these results support a reflective single-factor interpretation with diagnostic facet residuals, meeting governance thresholds while preserving actionability.

Substantively, the facet profile suggests Assurance (competence/trust) is the binding constraint, Tangibles is a relative strength, and Reliability/Empathy sit mid-pack—an ordering that aligns with telecom theory on functional and relational drivers of satisfaction and loyalty. Practically, we propose a two-tier operating model: a unit-weighted composite for executive dashboards and facet "dials" for operational ownership, all guarded by measurement health thresholds (alpha/omega ≥ .80; AVE > .50; SRMR ≤ .10).

Limitations—single indicator per facet, a missing Responsiveness indicator in this wave, coded segments, and absent provider metadata—temper claims and motivate a clear future-research roadmap: expand item coverage, migrate to Confirmatory Factor Analysis (CFA) and Multi-Group Confirmatory Factor Analysis (MG-CFA) (or alignment) for comparability, audit fairness, and link PSQ to behavioral outcomes (e.g., churn, Average Revenue Per User). Implemented in this way, compact SERVQUAL provides a board-ready, auditable, and scalable foundation for telecom quality governance and for scholarly work connecting measurement rigor to managerial impact.

References

1. Asparouhov, T., & Muthén, B. (2014). Multiple-group factor analysis alignment. *Structural Equation Modeling: A Multidisciplinary Journal, 21*(4), 495–508. https://doi.org/10.1080/10705511.2014.919210
2. Bartlett, M. S. (1950). Tests of significance in factor analysis. *British Journal of Statistical Psychology, 3*(2), 77–85. https://doi.org/10.1111/j.2044-8317.1950.tb00285.x
   Brown, T. A. (2015). *Confirmatory factor analysis for applied research* (2nd ed.). Guilford Press.
3. Cheung, G. W., & Rensvold, R. B. (2002). Evaluating goodness-of-fit indexes for testing measurement invariance. *Structural Equation Modeling: A Multidisciplinary Journal, 9*(2), 233–255. https://doi.org/10.1207/S15328007SEM0902_5
4. Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*(3), 297–334. https://doi.org/10.1007/BF02310555
5. Cronin, J. J., Jr., & Taylor, S. A. (1992). Measuring service quality: A reexamination and extension. *Journal of Marketing, 56*(3), 55–68. https://doi.org/10.1177/002224299205600304

6. Flora, D. B., & Curran, P. J. (2004). An empirical evaluation of alternative methods of estimation for confirmatory factor analysis with ordinal data. *Psychological Methods, 9*(4), 466–491. https://doi.org/10.1037/1082-989X.9.4.466
7. 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
8. Hayes, A. F., & Coutts, J. J. (2020). Use omega rather than Cronbach's alpha for estimating reliability. *Communication Methods and Measures, 14*(1), 1–23. https://doi.org/10.1080/19312458.2020.1718629
9. Hu, L.-T., & Bentler, P. M. (1999). Cutoff criteria for fit indexes in covariance structure analysis: Conventional criteria versus new alternatives. *Structural Equation Modeling: A Multidisciplinary Journal, 6*(1), 1–55. https://doi.org/10.1080/10705519909540118
10. Kaiser, H. F. (1974). An index of factorial simplicity. *Psychometrika, 39*(1), 31–36. https://doi.org/10.1007/BF02291575
11. Kline, R. B. (2016). *Principles and practice of structural equation modeling* (4th ed.). Guilford Press.
12. McDonald, R. P. (1999). *Test theory: A unified treatment*. Lawrence Erlbaum Associates.
13. Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing, 64*(1), 12–40.
14. Rhemtulla, M., Brosseau-Liard, P. É., & Savalei, V. (2012). When can categorical variables be treated as continuous? A comparison of robust continuous and categorical SEM estimation methods under suboptimal conditions. *Psychological Methods, 17*(3), 354–373. https://doi.org/10.1037/a0029315
15. Steenkamp, J.-B. E. M., & Baumgartner, H. (1998). Assessing measurement invariance in cross-national consumer research. *Journal of Consumer Research, 25*(1), 78–90. https://doi.org/10.1086/209528
16. Tabachnick, B. G., & Fidell, L. S. (2019). *Using multivariate statistics* (7th ed.). Pearson.