



BEYOND THE AVERAGE: A DESCRIPTIVE SEGMENT-SENSITIVITY ANALYSIS OF BANKING SERVICE-QUALITY GAPS

Snežana Ristevska-Jovanovska 

Ss. Cyril and Methodius University, Faculty of Economics, Skopje, North Macedonia
<https://orcid.org/0000-0003-1019-0218>

ARTICLE INFO



 Open access

JEL Category:
G21, M31

Keywords:

SERVQUAL
banking service quality
customer segmentation
subgroup analysis
service-quality gaps
North Macedonia
retail banking

ABSTRACT

A sector-wide average can correctly identify a service-quality weakness and still lead managers to the wrong intervention. This study presents a secondary comparative analysis of service-quality subgroup means derived from a 2025 survey of 196 retail-bank customers in North Macedonia. The study applies a descriptive diagnostic scheme to compare age, gender, and bank-size subgroup profiles using three descriptive indicators: segment differentials, cross-profile ranges, and qualitative rank stability. The aim is to distinguish structural service-quality weaknesses from those showing greater descriptive variation across subgroup profiles. In this study, segment sensitivity refers exclusively to observed descriptive differences between subgroup means and does not imply a statistically tested effect or statistically significant difference between groups. Reliability remains a major deficit across customer groups, while tangibility is consistently the relative strength. Observed gender differentials are modest but consistent, with the largest female-male widening observed in empathy (0.19). Observed subgroup differences were larger for bank-size profiles than for gender profiles. Responsiveness shows a 0.49 gap differential between large and small/medium banks, representing the largest observed subgroup difference in the sample. Among the two adequately represented age groups, reliability differs by 0.29 and responsiveness by 0.19. The findings show that the largest aggregate deficit is not necessarily the dimension displaying the greatest observed descriptive variation across subgroup profiles and suggest that structural weaknesses may require broader managerial attention, whereas weaknesses displaying greater descriptive subgroup variation may benefit from more targeted diagnosis and intervention.

Address of the author:
Snežana Ristevska - Jovanovska
snezanarj@eccf.ukim.edu.mk

Received: 23.08.2026
Revised: 22.09.2026
Accepted: 07.10.2026
Available online: 10.10.2026



1 INTRODUCTION

An average can identify the banking sector's weakest service dimension and still misdirect the next managerial investment. It shows where service quality falls short overall, but it says much less about whether the same weakness is equally severe across customer groups and institutional settings. In banking, that distinction matters because a weakness that appears broadly stable across subgroup profiles may warrant system-wide attention, whereas one displaying greater observed variation across profiles may warrant more targeted diagnosis. Service-quality research traditionally compares what customers expect with what they perceive they receive. Grönroos (1984) distinguished technical from functional quality, while Parasuraman et al. (1985, 1988) formalised the expectation-perception gap and the five SERVQUAL dimensions of tangibility, reliability, responsiveness, assurance, and empathy. In retail banking, these dimensions capture both operational execution and the interpersonal experience through which customers evaluate accuracy, speed, confidence, access, and individual attention.

The aggregate SERVQUAL analysis based on this dataset was reported in an earlier study by Ristevska-Jovanovska (2026). That study examined the magnitude and ranking of expectation-perception gaps at the item and dimension level and identified reliability as the largest overall deficit and tangibility as the smallest. Although both studies use the same survey dataset and SERVQUAL instrument, their analytical focus differs. The earlier study concentrated on aggregate item-level and dimension-level gap magnitudes, whereas the present study examines subgroup profiles, comparative variability, cross-profile ranges, and qualitative rank stability. The subgroup results reported here are derived from the same original dataset, but they are reorganised and analysed to address a distinct research question focused on descriptive subgroup variation, cross-profile ranges, and rank stability rather than on the magnitude and ranking of overall service-quality deficits. Building on that empirical baseline, the present paper addresses a different question: are the identified service-quality weaknesses structurally stable, or does their observed severity

vary across customer and institutional profiles? Rather than reproducing the earlier item-level or aggregate gap analysis, this study uses the reported subgroup means as inputs for a secondary descriptive comparison of subgroup variation. Its analytical unit is therefore not the absolute SERVQUAL gap itself, but the observed variation, range, and rank stability of those gaps across subgroup profiles. This progression turns an aggregate diagnosis into a more targeted descriptive question about where a system-wide response may be appropriate and where more segment-specific investigation may be warranted.

This distinction matters because two dimensions can have similar overall gaps but display very different degrees of observed variation across customer and institutional subgroup profiles. A dimension that remains weak across gender, age, and bank-size profiles may be interpreted as a structural managerial priority. A dimension whose observed gap varies more substantially across subgroup profiles may instead warrant more targeted diagnosis. The difference between the "largest gap" and the "largest observed differential" is therefore analytically important.

The contribution of this paper is the proposal of a descriptive diagnostic scheme for organising observed subgroup differences in SERVQUAL gaps according to severity, cross-profile variability, and qualitative rank stability. The scheme is based on three descriptive indicators. First, segment differentials describe the magnitude and direction of observed differences between comparable subgroup means. Second, cross-profile ranges identify which SERVQUAL dimensions display greater descriptive variation across the examined subgroup profiles. Third, qualitative rank-stability assessment indicates whether dimensions maintain a similar relative position across profiles. These indicators are intended to summarise observed patterns in subgroup means and should not be interpreted as statistically tested effects or evidence of statistically significant differences. Rather than introducing a new statistical methodology, the proposed scheme reorganises existing descriptive evidence into a structured diagnostic perspective that supports interpretation of service-quality weaknesses and managerial prioritisation.

The study addresses four research questions:

- RQ1:** Which SERVQUAL dimensions show the greatest and smallest observed descriptive variation across gender profiles?
- RQ2:** Which dimensions show the greatest observed variation between the two adequately represented age groups?
- RQ3:** Which SERVQUAL dimensions show the greatest and smallest observed descriptive variation across bank-size profiles?
- RQ4:** Which service-quality weaknesses appear relatively stable across the examined subgroup profiles, and which display greater descriptive variation that may warrant segment-specific managerial investigation?

2 LITERATURE REVIEW

SERVQUAL conceptualises service quality as the difference between perceived performance and expected performance (Parasuraman et al., 1988). Negative gaps indicate underdelivery. The model has remained influential because it converts a broad perception into identifiable operational dimensions, although its dimensional stability and the role of expectations have been debated (Asubonteng et al., 1996; Cronin & Taylor, 1992; Ladhari, 2009; Newman, 2001).

For managerial purposes, gap magnitude is only one layer of information. An overall mean can reveal the strongest average deficit, while descriptive comparison of subgroup means can indicate whether that deficit appears relatively stable or varies across customer and institutional profiles. This is consistent with broader service-quality approaches that emphasise context, interaction, and hierarchical evaluation rather than a single undifferentiated score (Brady & Cronin, 2001). The present study therefore treats observed subgroup variation not as evidence of causation or statistical significance, but as a descriptive diagnostic signal within the service-quality profile.

2.1 Customer Heterogeneity in Banking

Banking service quality combines highly standardised processes with heterogeneous customer experience. Banking-specific instruments have repeatedly identified transaction

accuracy, fulfilment of promises, employee behaviour, communication, security, access, and personal attention as central elements (Aldlaigan & Buttle, 2002; Avkiran, 1994; Bahia & Nantel, 2000; Karatepe et al., 2005). Digitalisation has broadened rather than eliminated these concerns. Contemporary research continues to emphasise reliability, responsiveness, security, ease of use, and integrated-channel experience as determinants of customer outcomes (Ayinaddis et al., 2023; Sharma et al., 2024; Sreejesh, 2024; Yesmin et al., 2023).

Customer heterogeneity is relevant because expectations and evaluations may vary with experience, interaction needs, digital familiarity, perceived risk, and the institutional setting in which service is delivered. However, observed subgroup differences should not automatically be interpreted as demographic or institutional effects. In the present study, such differences are treated as descriptive signals indicating where further respondent-level explanation and statistical testing may be warranted, rather than as evidence of causal relationships or statistically significant differences between groups.

Recent banking research provides direct support for examining demographic heterogeneity rather than assuming uniform service evaluations. Joshi and Dabas (2022) found that age moderated the effects of responsiveness and competence on online-banking satisfaction, while gender moderated the effect of competence. Le et al. (2023) similarly reported differences in digital-banking service-quality perceptions by age and gender. These findings support further examination of whether observed service-quality patterns remain broadly stable or vary across customer profiles.

2.2 Banking service quality in North Macedonia

Evidence from North Macedonia confirms the strategic importance of service quality. Yzeiri Baftijari et al. (2020) reported high expectations and highlighted reliability in bank-service evaluation. Naumovska-Saveska et al. (2021) linked service quality and service characteristics to customer satisfaction and loyalty. The present analysis adds a different layer: rather than asking only which dimension performs worst, it asks

which dimensions remain stable and which change most across available customer and institutional profiles. Despite the extensive application of SERVQUAL in banking research, most studies focus primarily on the magnitude of service-quality gaps, customer satisfaction outcomes, or relationships between service quality and behavioural intentions. Comparatively less attention has been paid to the descriptive variability of SERVQUAL gaps across customer and institutional subgroup profiles. As a result, existing research provides limited guidance on whether observed service-quality weaknesses should be interpreted as broadly distributed structural problems or as context-sensitive priorities. The present study addresses this gap through a secondary descriptive comparison of subgroup variability, cross-profile ranges, and rank stability.

3 RESEARCH METHODOLOGY

The present secondary analysis is based on the 2025 cross-sectional survey dataset of 196 retail-bank customers in North Macedonia reported in Ristevska-Jovanovska (2026). The original survey instrument contained 20 paired SERVQUAL expectation and perception statements, four for each of the five standard SERVQUAL dimensions, measured on a five-point Likert scale. SERVQUAL gap scores were defined as perception minus expectation ($P - E$), with more negative values indicating larger service-quality shortfalls. The present study adopts this existing measurement framework and focuses on descriptive comparisons of the resulting subgroup mean gap scores rather than on additional measurement validation. Survey design and respondent classification. The present study is based on the same dataset and SERVQUAL survey instrument reported in Ristevska-Jovanovska (2026). The original study employed a quantitative cross-sectional design and a non-probability electronic survey of retail-bank customers in North Macedonia. Data were collected in December 2025 through online distribution via email and social media channels. Respondents identified the bank they used most frequently, which was treated as their principal bank for classification purposes. The present paper does not introduce a new sample or a new measurement instrument; instead, it addresses a different research question

through secondary descriptive analysis of subgroup means derived from the original dataset. Further details regarding questionnaire development, administration, and respondent recruitment are provided in the original study. The sample comprised 102 men and 94 women. The age distribution was 128 respondents aged 20–45, 65 aged 46–65, and three over 65. Because the oldest category contains only three observations, it is excluded from the comparative age-group calculations. For bank size, 112 respondents identified a large bank as their principal bank and 84 identified a small or medium-sized bank. Bank-size categories follow the classification reported in Ristevska-Jovanovska (2026), based on the relative share of banking-system assets in North Macedonia, where the five largest banks were grouped as large banks and the remaining institutions as small/medium-sized banks.

Secondary analytical design - The present paper is a secondary descriptive analysis derived from the same research project as the earlier aggregate SERVQUAL study (Ristevska-Jovanovska, 2026). The earlier paper provides the empirical baseline by examining the magnitude and ranking of expectation-perception gaps at the item and dimension level. The present study extends that work by deriving comparative descriptive indicators from subgroup means to examine observed variation, cross-profile ranges, and qualitative rank stability. In this study, references to segment sensitivity denote descriptive variation in observed subgroup means and do not represent statistically tested effects or statistically significant differences between groups. The two studies therefore share the dataset and measurement instrument but differ in research questions, analytical unit, derived outputs, and managerial interpretation. The original study reported the overall SERVQUAL gaps at item and dimension level, whereas the present study focuses on descriptive comparison of subgroup profiles, observed variability, and qualitative rank stability across segmentation dimensions. No new respondents are introduced, and the secondary transformations are not presented as independent inferential evidence. Instead, they are used to extract an additional descriptive layer of managerial meaning from the previously established gap structure. Three derived

descriptive indicators are used. A segment differential is defined as the absolute difference between the two reported subgroup mean gap scores for a dimension. Direction is interpreted separately to identify which subgroup reports the wider observed negative gap. A cross-profile range is defined as the difference between the least negative and most negative observed subgroup mean gap for a dimension across six profiles derived from three segmentation dimensions: gender, age, and bank size. Because these profiles are not mutually exclusive, the range is interpreted solely as a descriptive measure of variation among observed subgroup means and not as a statistical measure of heterogeneity or a tested effect. Rank stability is assessed qualitatively by ordering the five SERVQUAL dimensions within each subgroup profile according to the magnitude of the observed gap and examining the consistency of their relative positions across profiles.

Table 1. *Analytical groups used in the secondary analysis*

Segmentation variable	Groups retained	N
Gender	Male / Female	102 / 94
Age	20-45 / 46-65	128 / 65
Bank size	Large / Small-medium	112 / 84

Note: The over-65 group (N = 3) is not used in comparative age group calculations

Source: (Author's original research, 2025)

Construction of subgroup means. The subgroup means used in the present analysis were derived from the same original SERVQUAL dataset reported in Ristevska-Jovanovska (2026). The dataset contained complete responses for all SERVQUAL items included in the analysis; therefore, no imputation or missing data procedures were required. For each respondent, SERVQUAL gap scores were calculated as perception minus expectation ($P - E$), and dimension-level gaps were obtained by averaging the four equally weighted items belonging to each SERVQUAL dimension. The present study uses

these subgroups mean gap scores as the basis for newly derived descriptive comparisons of subgroup differentials, cross-profile ranges, and qualitative rank stability. It does not introduce new respondents or additional respondent-level inferential testing.

Scale reliability and dimensional structure. The present study adopts the SERVQUAL measurement structure previously used and assessed in Ristevska-Jovanovska (2026). No additional reliability analysis, factor analysis, validity assessment, or other psychometric testing was conducted for the present secondary analysis. The standard five-dimensional SERVQUAL structure and the corresponding dimension-level gap scores are therefore treated as part of the established measurement framework inherited from the original study. The analytical contribution of the present paper begins with the descriptive comparison of subgroup mean gap scores through segment differentials, cross-profile ranges, and qualitative rank stability. Readers are referred to Ristevska-Jovanovska (2026) for details concerning questionnaire development, measurement procedures, and the scale assessment conducted in the original study.

4 RESULTS

Gender profiles: same hierarchy, different observed intensity – Observed subgroup means indicate modest descriptive variation in service-quality gaps between male and female respondents. Female respondents report more negative mean gaps in all five dimensions. The largest observed female-male differential is found in empathy (0.19), followed by reliability (0.17), assurance (0.15), responsiveness (0.13), and tangibility (0.09). These differences are descriptive and should not be interpreted as statistically significant gender effects. Reliability, despite having the largest absolute deficit, does not show the largest observed gender differential. Empathy displays the largest descriptive gender differential, whereas reliability remains the dominant observed weakness for both groups.

Table 2. *Gender gap differentials*

Dimension	Male gap	Female gap	Absolute differential	Wider observed gap
Tangibility	-1.04	-1.13	0.09	Female
Reliability	-1.62	-1.79	0.17	Female

Dimension	Male gap	Female gap	Absolute differential	Wider observed gap
Responsiveness	-1.55	-1.68	0.13	Female
Assurance	-1.48	-1.63	0.15	Female
Empathy	-1.43	-1.62	0.19	Female

Source: Calculated by the author from original subgroup means (2025).

Age profiles: observed differences between the two adequately represented age groups. Among the two adequately represented age groups, reliability shows the largest observed descriptive differential (0.29): the mean gap is -1.79 among respondents aged 20–45 and -1.50 among those aged 46–65. Responsiveness follows with a differential of 0.19 (-1.68 versus -1.49). Tangibility differs by 0.10. Empathy also differs by 0.10, but

in the opposite direction: respondents aged 46–65 report the wider observed empathy gap (-1.60 versus -1.50). Assurance differs by 0.10. These differences are descriptive and do not establish statistically significant age-group effects. The observed variation is greater for reliability and responsiveness than for the remaining dimensions and may warrant further respondent-level statistical investigation.

Table 3. Age-group gap differentials

Dimension	20-45 gap	46-65 gap	Absolute differential	Wider observed gap
Tangibility	-1.12	-1.02	0.10	20-45
Reliability	-1.79	-1.50	0.29	20-45
Responsiveness	-1.68	-1.49	0.19	20-45
Assurance	-1.60	-1.50	0.10	20-45
Empathy	-1.50	-1.60	0.10	46-65

Source: Calculated by the author from original subgroup means (2025).

Bank-size profiles: responsiveness shows the largest observed differential. The largest observed descriptive subgroup differentials in the present sample are associated with the bank-size comparison. Responsiveness differs by 0.49 between large-bank customers (-1.82) and small/medium-bank customers (-1.33), representing the largest observed two-group differential across the three segmentation dimensions examined. Reliability differs by 0.25, with the wider observed gap among small/medium-bank customers (-1.85 versus -1.60). Empathy differs by 0.22 and assurance by 0.14, again with wider observed gaps among

small/medium-bank customers. Tangibility is virtually invariant, differing by only 0.01. These differences are descriptive and do not establish statistically significant bank-size effects. The observed pattern nevertheless changes the managerial interpretation of the original gap hierarchy: reliability remains a broad weakness, while responsiveness displays the greatest descriptive variation across the two bank-size profiles. Large-bank customers show a wider observed responsiveness gap, whereas customers of small and medium-sized banks show wider observed gaps in reliability, empathy, and assurance.

Table 4. Bank-size gap differentials

Dimension	Large-bank gap	Small/medium gap	Absolute differential	Wider observed gap
Tangibility	-1.09	-1.08	0.01	Large (negligible difference)
Reliability	-1.60	-1.85	0.25	Small/medium
Responsiveness	-1.82	-1.33	0.49	Large
Assurance	-1.49	-1.63	0.14	Small/medium
Empathy	-1.43	-1.65	0.22	Small/medium

Source: Calculated by the author from original subgroup means (2025).

Cross-profile range: structural versus descriptively variable dimensions - Comparing the six subgroup profiles across the three segmentation dimensions provides a second descriptive view of observed variation. Responsiveness has the widest observed cross-profile range (0.49), from -1.82 among large-bank customers to -1.33 among small/medium-bank customers. Reliability has the second-widest observed range (0.35), followed by empathy (0.22), assurance (0.15), and tangibility (0.11). These ranges are descriptive summaries of

subgroup mean variation and should not be interpreted as statistical effect measures or evidence of statistically significant differences. Tangibility is therefore not only the relative strength of the service-quality profile but also the most descriptively stable dimension across the examined subgroup profiles. Responsiveness, by contrast, combines a comparatively large deficit with the widest observed descriptive variation across profiles.

Table 5. Cross-profile descriptive variation range

Dimension	Most negative gap	Least negative gap	Range	Descriptive variation interpretation
Tangibility	-1.13	-1.02	0.11	Very stable
Reliability	-1.85	-1.50	0.35	Moderate variation
Responsiveness	-1.82	-1.33	0.49	Highest observed
Assurance	-1.63	-1.48	0.15	Relatively stable
Empathy	-1.65	-1.43	0.22	Moderate variation

Note. Range is calculated across six subgroup profiles representing three segmentation dimensions (gender, age, and bank size). Because profiles are not mutually exclusive, the range should be interpreted as a descriptive measure of variation among subgroup means.

Source: Author's calculations

Rank ordering provides a third descriptive perspective on variation across service-quality subgroup profiles. Table 6 summarises the relative position of each SERVQUAL dimension within the six subgroup profiles. Tangibility occupies the fifth position in every profile, indicating strong descriptive rank stability as the

least negative dimension. Reliability ranks first in four of the six profiles and remains among the most negative dimensions throughout. Responsiveness and empathy display greater observed rank variability across profiles, indicating that their relative positions are less stable across the examined customer and institutional profiles.

Table 6. Qualitative rank stability across subgroup profiles

Dimension	Male	Female	Age 20-45	Age 46-65	Large bank	Small/medium
Tangibility	5	5	5	5	5	5
Reliability	1	1	1	2.5	2	1
Responsiveness	2	2	2	4	1	4
Assurance	3	3	3	2.5	3	3
Empathy	4	4	4	1	4	2

Note: 1 = largest observed deficit; 5 = smallest observed deficit. Tied values receive the average of the corresponding rank positions. Rankings are based on the observed subgroup mean gap scores and are descriptive only.

Source: Author's calculations

The managerial classifications presented in Table 7 are descriptive and are based on the combined consideration of three elements: absolute severity of the observed gap, cross-profile variation, and qualitative rank stability.

Dimensions with consistently large negative gaps across most subgroup profiles are interpreted as structural priorities. Dimensions displaying comparatively greater observed cross-profile variation and changes in relative rank are

interpreted as priorities that may warrant more targeted segment-level investigation. Dimensions with consistently smaller negative gaps and stable rankings are interpreted as relative strengths.

These classifications are intended solely as a descriptive diagnostic aid and do not represent statistically derived categories or statistically significant group differences.

Table 7. Descriptive service-quality variation matrix

Dimension	Absolute severity	Observed descriptive variation	Managerial classification
Tangibility	Lowest deficit	Very low	Maintain; broad relative strength
Reliability	Very high	Moderate	Broad structural priority
Responsiveness	High	Very high	Broad priority with targeted follow-up
Assurance	High	Low	Broad improvement priority
Empathy	High	Moderate	Interpersonal priority for targeted follow-up

Note: Classifications are based on descriptive evaluation of observed gap severity, cross-profile variation, and qualitative rank stability. They are intended as managerial diagnostic categories only and do not represent statistically derived classifications, statistically significant group differences, or causal effects.

Source: Author's synthesis of gap magnitude, differential, range, and rank patterns.

5 DISCUSSION

The main contribution of this analysis is the distinction between severity and observed descriptive variation across subgroup profiles. A dimension can show a severe service-quality deficit while varying relatively little across profiles, or it can display greater descriptive variation depending on the customer or institutional profile examined. This distinction is not visible in a single aggregate SERVQUAL ranking and should be interpreted as descriptive rather than as evidence of statistically significant subgroup effects.

Reliability illustrates a descriptively stable pattern of high severity. It remains one of the largest observed deficits across nearly all profiles and is the dominant weakness for both genders, respondents aged 20–45, and customers of small and medium-sized banks. Its cross-profile range of 0.35 indicates some descriptive variation, but the consistency of its high deficit ranking suggests that reliability may reasonably be treated as a broad service-system priority rather than as a weakness confined to a particular subgroup.

This interpretation is consistent with North Macedonian evidence highlighting reliability (Yzeiri Baftijari et al., 2020) and with banking-specific quality frameworks that emphasise transaction accuracy and fulfilment of promises (Aldlaigan & Buttle, 2002; Bahia & Nantel, 2000; Karatepe et al., 2005).

Responsiveness illustrates a different descriptive pattern. Its cross-profile range of 0.49 is the largest observed in the study, driven primarily by the contrast between large- and small/medium-bank subgroup means. It therefore combines a comparatively large deficit with the greatest observed descriptive variation across the examined profiles. The practical implication is that responsiveness should not be interpreted solely through the sector-wide average. Possible explanations for the observed bank-size difference may include factors such as workload, customer volume, process complexity, communication practices, or access to employees; however, these factors were not measured in the present study, and no causal explanation can be inferred from the observed subgroup means. Such mechanisms should be examined directly in future respondent-level research.

Empathy occupies an intermediate descriptive position. Its overall deficit is substantial, its cross-profile range is 0.22, and it shows the largest observed gender differential (0.19). It is also the widest observed gap among respondents aged 46–65. These patterns indicate greater descriptive variation across some subgroup profiles and may justify more targeted investigation of interpersonal aspects of service quality. They do not establish statistically significant gender or age effects, nor do they imply that these characteristics cause differences in expectations or evaluations. Rather,

they indicate where further research could examine individual attention, explanation, problem ownership, and understanding of customer circumstances.

Tangibility is analytically important because it displays comparatively little variation across the examined subgroup profiles. Its observed cross-profile range is only 0.11, and its bank-size differential is 0.01. It remains the least negative dimension throughout the six profiles. This descriptive stability suggests that tangibility represents relative strength within the sample examined. The observed pattern is consistent with a possible contrast between relatively stable physical service attributes and more variable operational and interpersonal dimensions; however, this interpretation remains descriptive and should be examined further using respondent-level analysis.

The findings suggest a two-level approach to service-quality management. Weaknesses that remain comparatively severe across subgroup profiles may warrant broader process-level attention, whereas weaknesses displaying greater observed descriptive variation may warrant more targeted diagnosis. The largest aggregate gap and the largest observed subgroup differential therefore provide different managerial signals. Treating these descriptive signals as equivalent may obscure potentially important differences in how service-quality weaknesses are distributed across the examined profiles.

This distinction between broadly stable weaknesses and greater observed subgroup variation also helps position the findings within recent digital-banking research. Studies that formally test moderation indicate that demographic characteristics can alter the strength of service-quality relationships rather than simply shift an overall mean (Joshi & Dabas, 2022), while evidence from digital banking has also documented differences in perceived service quality across demographic groups (Le et al., 2023). The present study contributes a complementary descriptive perspective: before formal moderation or other inferential models are estimated, observed subgroup differentials and rank-stability patterns can help identify where further investigation of heterogeneity may be

warranted and where broadly shared service-quality weaknesses appear to remain dominant.

6 MANAGERIAL IMPLICATIONS

First, banks should add a descriptive subgroup-variation layer to conventional service-quality dashboards. Alongside the overall gap for each SERVQUAL dimension, management could track the range and direction of observed subgroup differentials. A dimension with a large overall deficit and relatively low descriptive variation may warrant broad corrective attention, whereas a dimension with a larger observed subgroup differential may warrant more targeted investigation of possible segment- or institution-related factors.

Second, reliability may reasonably be treated as a broad structural priority within the observed sample. Banks should monitor error rates, first-time-right completion, missed deadlines, repeated corrections, and unresolved problems across channels and customer groups. Because the reliability deficit remains comparatively large and descriptively stable across most examined subgroup profiles, improvement efforts may appropriately focus on core systems and processes rather than on isolated customer segments.

Third, responsiveness may warrant more targeted managerial investigation. The observed 0.49 bank-size differential is the largest subgroup difference identified in the present sample and may be obscured when only aggregate sector means are considered. For large banks, the wider observed responsiveness gap suggests that factors such as response times, employee availability, queue and workload patterns, status communication, escalation routes, and ownership of unresolved requests may be useful areas for further examination. However, these factors were not measured in the present study and should be treated as possible explanatory areas rather than established causes of the observed difference. Small and medium-sized banks may likewise use the observed pattern as a basis for examining how responsiveness and reliability are balanced in their service processes.

Fourth, empathy may be monitored as an interpersonal dimension showing comparatively greater descriptive variation across some subgroup profiles. The observed gender

differential and the wider gap among respondents aged 46–65 suggest that individual attention and understanding of customer circumstances may warrant more targeted qualitative follow-up. These patterns should not be interpreted as evidence that gender or age causes the observed differences. Interviews, complaint coding, and service-recovery data could help explore possible explanations that cannot be established from the descriptive subgroup means alone.

Finally, tangibility may reasonably be maintained as a relative strength rather than treated as the primary improvement priority. Its consistently smaller gap and comparatively stable descriptive pattern suggest that, within the observed sample, managerial attention may be more usefully directed toward dependable execution, responsive support, and individualised communication, while maintaining existing standards of physical and visible service quality.

7 LIMITATIONS AND FUTURE RESEARCH

This study is a secondary descriptive analysis, and its contribution lies in the transformation and comparison of existing subgroup means rather than in new data collection. The present study uses the same dataset and SERVQUAL measurement instrument as the earlier aggregate study (Ristevska-Jovanovska, 2026) but addresses a different research question and analytical focus. Specifically, it examines observed subgroup variation, cross-profile ranges, and qualitative rank stability rather than reproducing the earlier item-level and aggregate gap analysis. This relationship and analytical distinction are disclosed explicitly throughout the manuscript.

The present manuscript focuses on derived descriptive indicators of subgroup variation, cross-profile ranges, and qualitative rank stability rather than reproducing the earlier item-level and aggregate analysis. The non-probability sample cannot be generalised statistically to all bank customers in North Macedonia. The analysis does not include respondent-level inferential tests, confidence intervals, or effect-size estimates. The reported differentials and ranges are descriptive indicators of observed variation among subgroup means and should not be interpreted as statistical

effect measures or evidence of statistically significant differences between groups. These limitations define the contribution appropriately: the study offers a structured secondary descriptive diagnostic scheme whose observed patterns should be examined further using respondent-level statistical analysis.

The over-65 group contains only three respondents and was therefore excluded from the main comparative calculations. Future research should use stratified recruitment to achieve adequate age representation. Bank-size categories are also broad and should not be interpreted as rankings of individual banks.

Most importantly, future studies should test the descriptive patterns identified here using respondent-level data. Multivariate models could examine whether age, gender, bank size, channel use, financial literacy, and digital literacy are associated with meaningful variation in SERVQUAL gaps. Interaction terms could test whether institutional characteristics moderate demographic patterns. Repeated measurements could also examine whether broadly stable weaknesses and those displaying greater observed subgroup variation respond differently to managerial interventions.

8 CONCLUSION

Service-quality averages reveal where banks underdeliver. Descriptive comparison of subgroup profiles adds a second layer by showing whether observed service-quality gaps remain relatively stable or vary across customer and institutional profiles. The distinction matters because severity and descriptive subgroup variation are not the same phenomenon.

In the present analysis, reliability emerges as the clearest broadly stable weakness: it remains severe across most customer and institutional profiles. Responsiveness displays the greatest observed descriptive variation, with a cross-profile range of 0.49 driven primarily by the bank-size contrast. Empathy shows moderate observed variation and the largest gender differential, while tangibility is consistently the most stable relative strength. These patterns are descriptive and should not be interpreted as statistically significant subgroup effects.

The managerial lesson is therefore more precise than a simple ranking of SERVQUAL gaps. Banks may distinguish weaknesses that appear broadly stable across subgroup profiles from those displaying greater observed variation and therefore warranting more targeted investigation. Reliability may justify broad process improvement, while responsiveness may warrant both general attention and further institution-specific diagnosis. Empathy deserves closer segment-level investigation, whereas tangibility may reasonably be maintained as a relative strength.

The study therefore extends the earlier aggregate SERVQUAL diagnosis by introducing a descriptive comparison of observed subgroup variability and qualitative rank stability across customer and institutional profiles. The average customer remains useful for describing the sector; for managing service quality, however, an additional question is whether an observed weakness appears broadly stable across profiles or displays greater descriptive variation in particular customer or institutional contexts. This distinction can help inform whether broader system-level improvement or more targeted follow-up investigation may be appropriate.

REFERENCES

- Aldlaigan, A. H., & Buttle, F. A. (2002). SYSTRA-SQ: A new measure of bank service quality. *International Journal of Service Industry Management*, 13(4), 362-381. <https://doi.org/10.1108/09564230210445041>
- Asubonteng, P., McCleary, K. J., & Swan, J. E. (1996). SERVQUAL revisited: A critical review of service quality. *Journal of Services Marketing*, 10(6), 62-81. <https://doi.org/10.1108/08876049610148602>
- Avkiran, N. K. (1994). Developing an instrument to measure customer service quality in branch banking. *International Journal of Bank Marketing*, 12(6), 10-18. <https://doi.org/10.1108/02652329410063223>
- Ayinaddis, S. G., Taye, B. A., & Yirsaw, B. G. (2023). Examining the effect of electronic banking service quality on customer satisfaction and loyalty: An implication for technological innovation. *Journal of Innovation and Entrepreneurship*, 12, Article 22. <https://doi.org/10.1186/s13731-023-00287-y>
- Bahia, K., & Nantel, J. (2000). A reliable and valid measurement scale for the perceived service quality of banks. *International Journal of Bank Marketing*, 18(2), 84-91. <https://doi.org/10.1108/02652320010322994>
- Brady, M. K., & Cronin, J. J., Jr. (2001). Some new thoughts on conceptualizing perceived service quality: A hierarchical approach. *Journal of Marketing*, 65(3), 34-49. <https://doi.org/10.1509/jmkg.65.3.34.18334>
- 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>
- Grönroos, C. (1984). A service quality model and its marketing implications. *European Journal of Marketing*, 18(4), 36-44. <https://doi.org/10.1108/EUM0000000004784>
- Joshi, M., & Dabas, R. (2022). Exploring the nexus among internet banking service quality, customer satisfaction, age, and gender: A developing country perspective during COVID-19. *Indian Journal of Marketing*, 52(9), 8-25. <https://doi.org/10.17010/ijom/2022/v52/i9/171982>
- Karatepe, O. M., Yavas, U., & Babakus, E. (2005). Measuring service quality of banks: Scale development and validation. *Journal of Retailing and Consumer Services*, 12(5), 373-383. <https://doi.org/10.1016/j.jretconser.2005.01.001>
- Ladhari, R. (2009). A review of twenty years of SERVQUAL research. *International Journal of Quality and Service Sciences*, 1(2), 172-198. <https://doi.org/10.1108/17566690910971445>
- Le, N. P., Ninh, N. T. H., & Hung, N. M. (2023). Factors influencing customers' perception of digital banking service quality in Vietnam. *International Journal of Professional Business Review*, 8(7), e02861. <https://doi.org/10.26668/businessreview/2023.v8i7.2861>
- Naumovska-Saveska, M., Tomovska Misoska, A., Efremov, K., & Petrovska, I. (2021). The impact of service quality and service characteristics on customer satisfaction in the North Macedonian banking sector. *Bankarstvo*, 50(2), 34-48. <https://doi.org/10.5937/bankarstvo2102034N>

- Newman, K. (2001). Interrogating SERVQUAL: A critical assessment of service quality measurement in a high street retail bank. *International Journal of Bank Marketing*, 19(3), 126-139. <https://doi.org/10.1108/02652320110388559>
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1985). A conceptual model of service quality and its implications for future research. *Journal of Marketing*, 49(4), 41-50. <https://doi.org/10.1177/002224298504900403>
- 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.
- Ristevska-Jovanovska, S. (2026). When banks underdeliver: A SERVQUAL gap analysis of North Macedonia's banking sector. *MEST Journal*, 15(1), 1–13. https://mest.meste.org/MEST_Najava/XXIX_Ristevska_Jovanovska.pdf
- Sharma, V., Jangir, K., Gupta, M., & Rupeika-Apoga, R. (2024). Does service quality matter in FinTech payment services? An integrated SERVQUAL and TAM approach. *International Journal of Information Management Data Insights*, 4(2), Article 100252. <https://doi.org/10.1016/j.jjime.2024.100252>
- Sreejesh, S. (2024). Integrated banking channel service quality (IBCSQ): Role of IBCSQ for building consumers' relationship quality and brand equity. *Journal of Retailing and Consumer Services*, 76, Article 103616. <https://doi.org/10.1016/j.jretconser.2023.103616>
- Yesmin, M. N., Hoque, S., Hossain, M. A., Jahan, N., Fang, Y., Wu, R., & Alam, M. J. (2023). SERVQUAL to determine relationship quality and behavioral intentions: An SEM approach in retail banking service. *Sustainability*, 15(8), 6536. <https://doi.org/10.3390/su15086536>
- Yzeiri Baftijari, A., Ismaili, R., & Aziri, B. (2020). Client expectations and preferences for service quality dimensions in bank services in Republic of North Macedonia. *Economic Vision*, 7(13-14), 13-21. <https://www.ceeol.com/search/article-detail?id=939543>