“Assessing the Healthcare Quality Gap: Public Service Delivery at Civil Service Hospital”

Submitted to:

Center for Research and Development

Research and Consulting Services Department

Nepal Administrative Staff College

Submitted by:

Sabik Neupane

Roshani Bhujel

Asar 6, 2082

Abstract

Hospitals exhibit considerable variation in their specialties, services, and resource capacities, with their performance frequently assessed through scales reflecting patient perspectives. This diversity highlights the need for a standardized instrument to assess service quality across public hospitals, regardless of their ownership or operational nature. To address this imperative, the study adapted the SERVQUAL framework to develop a bespoke service quality measurement scale, drawing upon the experiences of inpatients and outpatients at Civil Service Hospital in Kathmandu, Nepal. Through a rigorous process involving structural equation modeling and confirmatory factor analysis, the scale was meticulously constructed and validated.

The findings revealed that all five SERVQUAL dimensions—Assurance, Tangibles, Reliability, Responsiveness, and Empathy—hold substantial relevance in this context, yet notable deficiencies mark each. Empathy emerged as the most pronounced shortfall, with a gap of -2.85, indicating a significant lack of personalized care. Tangibles followed at -2.75, suggesting deficiencies in physical infrastructure and equipment. Assurance and Reliability, with gaps of -2.47 and -2.46, respectively, pointed to challenges in fostering trust and consistency, while Responsiveness, at -2.28, highlighted delays in service delivery. Further examination through multiple linear regression illuminated a pervasive baseline dissatisfaction across all dimensions, with constants ranging from -1.640 to -2.381 ($p < 0.001$). Notably, male patients perceived Reliability as significantly weaker (-0.575, $p = 0.022$) compared to their female counterparts, whereas educational attainment exerted no discernible influence on perceptions.

These statistically significant disparities ($p \leq 0.05$) serve as a loud and clear call for hospital administrators, healthcare authorities, and policymakers. They must direct concerted efforts toward enhancing empathetic engagement, upgrading tangible assets, and bolstering the reliability and responsiveness of services. By addressing these critical areas, stakeholders can effect transformative improvements in service delivery, ultimately elevating patient satisfaction and trust within the healthcare system.

Keywords: SERVQUAL model, CFA, MLR, Patient perceptions, Service quality gaps, Public healthcare, Resource limitations, Patient expectation, Staff training

Acronyms

AIDS	Acquired Immunodeficiency Syndrome
ANOVA	Analysis of Variance
AVE	Average Variance Extracted
CFA	Confirmatory Factor Analysis
CFI	Comparative Fit Index
CR	Composite Reliability
CSH	Civil Service Hospital
FCHV	Female Community Health Volunteers
IMR	Infant Mortality Rate
MLE	Maximum Likelihood Estimation
MLR	Multiple Linear Regression
NASC	Nepal Administrative Staff College
NFI	Normed Fit Index
OLS	Ordinary Least Square
PPP	Public Private Partnerships
RMSEA	Root Mean Square Error of Approximation
SERVPERF	Service Performance
SERVQUAL	Service Quality
SRMR	Standardized Root Mean Square Residual
TLI	Tucker Lewis Index

VHI Vital Health Indicators

Contents

List of table

CHAPTER ONE

INTRODUCTION

1.1 Background

The healthcare sector in Nepal has undergone significant development in recent years, with both public and private providers playing crucial roles in serving the population's health needs. However, persistent concerns about the quality of healthcare services, particularly the disparities between public and private hospitals, continue to challenge the nation's healthcare system.

Nepal's constitution, promulgated in 2015, enshrines the right to health as a fundamental right under Article 35. This progressive step emphasizes equal access to quality healthcare for all citizens, regardless of their socioeconomic status. Despite this constitutional guarantee, the reality on the ground reveals significant deficiencies in terms of access, affordability, efficiency, and effectiveness across the healthcare system.

The public healthcare sector in Nepal, while more accessible to the general population, often struggles with overcrowding, resource constraints, and outdated infrastructure. Public hospitals, particularly those in rural areas, face chronic shortages of medical supplies, equipment, and qualified personnel. These challenges often result in long waiting times, compromised care quality, and patient dissatisfaction. On the other hand, the private healthcare sector has seen rapid growth, especially in urban areas. Private hospitals are generally perceived to offer better services, more advanced technology, and shorter waiting times. However, these advantages often come at a significantly higher cost, limiting their accessibility to a large portion of the population. This situation has led to a two-tiered healthcare system, where quality care is increasingly becoming a privilege of the affluent.

The COVID-19 pandemic has further exposed and exacerbated the existing inequalities and inefficiencies in Nepal's healthcare system. It highlighted the urgent need for a robust, equitable, and high-quality healthcare infrastructure capable of responding to both routine health needs and

public health emergencies. As Nepal continues its journey of development and improved living standards, the demand for better medical care is growing. This increasing demand, coupled with greater health awareness among the population, has put additional pressure on both public and private healthcare providers to enhance their service quality.

In this context, there is an urgent need to assess and enhance the quality of healthcare services in both public and private sectors. Understanding the current state of service quality, identifying gaps, and developing strategies for improvement are crucial steps towards realizing the constitutional promise of quality healthcare for all Nepali citizens.

1.2 Statement of problem

The quality gap between public and private hospitals in Nepal continues to widen, raising concerns about equity and effectiveness in healthcare delivery. Public hospitals, while more accessible, often struggle with overcrowding and resource constraints. The Civil Service Hospital (CSH), like many public hospitals in Nepal, grapples with issues of overcrowding, resource constraints, and perceived lower quality of service compared to private counterparts. Despite its mandate to provide accessible healthcare to civil servants and the general public, there is a lack of standardized quality assessment to objectively evaluate and improve its services. This quality assessment gap affects the hospital's ability to enhance its service delivery and impacts patient trust and satisfaction. This study aims to address this gap by comprehensively assessing service quality in CSH of Kathmandu.

1.3 Research objectives and questions

To validate the SERVQUAL modelìs suitability for assessing service quality in public healthcare settings.

To assess the influence of demographic (age, gender, education) factors on SERVQUAL gap scores.

Research Questions

How effectively does the SERVQUAL model capture service quality in public healthcare settings?

Which demographic (age, gender, education) factors significantly predict the service quality gaps in the SERVQUAL dimensions at Civil Service Hospital?

1.4 Rationale

This study aligns closely with several core functions of Nepal Administrative Staff College (NASC). By providing empirical evidence on healthcare service quality, it supports NASC's role in promoting evidence-based policymaking and good governance in the public sector. The study's focus on improving healthcare quality and supporting universal health coverage aligns with NASC's mission to advance national development goals. Additionally, the research's patient-centered approach and examination of various service dimensions reflect NASC's emphasis on interdisciplinary solutions to complex administrative challenges. By generating insights that can guide policy decisions, resource allocation, and quality improvement efforts, this study provides valuable input for NASC's advisory function to government bodies on management and reform. Ultimately, the research supports NASC's overarching goal of enhancing public sector performance and accountability, particularly in the critical area of healthcare delivery.

1.5 Delimitation

The study is expected to assess service quality gaps in CSH by identifying key factors influencing hospital service quality and providing recommendations for quality improvement. However, it's important to acknowledge the study's limitations. The sample size and geographic scope are limited, which may affect the generalizability of results. Additionally, the study focuses on functional quality (how services are delivered) rather than technical quality (the accuracy of medical procedures). There may also be challenges in accurately capturing patient expectations, as the concept can be subjective and influenced by various factors.

CHAPTER TWO

REVIEW OF LITERATURE

2.1 Private-public continuum in healthcare

Keeping the models and other theoretical underpinnings aside, the global healthcare system's evolving debate is the private-public debate. While there are nations with regulated health care systems, there are nations with the private sector's investment. The resource sharing between the private and public sectors has emerged in the PPP (Public-Private Partnerships) model. However, the nature of the public and private sectors has become central. There is a rise in privatization and outsourcing the nature of financing and control distinct the public and private sectors. The governments make the difficult decision about the extent of publicness or privateness required for accountable and efficient personal services (Lewis & Roxxane, 2004). Publicness is defined as the degree to which public authority affects how organizations act (Nutt, 1992).

In contrast, private organizations produce a product or service for the profit of the owners. The public sector delivers the goods and services at such a decree that it is available to everyone at no extra cost if provided to an individual. The government cannot limit the enjoyment of the public good to those who pay for it.

The private sector cannot direct the design, production, delivery, pricing, and promotion of public or essential goods (Carlton & Perloff, 2004). The voluntary purchase decisions of available goods leave no space for the consumer to give feedback to the producer. In a nutshell, the government is an overseeing body to supervise public goods and services. In a controlled market where the government becomes the producer, others (the public) consume the product (Lewis & Roxxane, 2004). There is a constant lack of interaction between the producers and consumers of public goods and services. The government thus has to forecast the service features and level of quality the consumer values. Consumers demand more service and higher quality. It becomes a behemoth task for the government to fulfill the demand and maintain service quality by asking the public to pay more to receive them. However, consumers explicitly demand goods and services without

considering the cost of consuming these services. We also see the governments cutting services to reduce their budgets and shift the public goods' supply side to the private sector.

2.2 Nepal health system challenges

The modern health system in Nepal is comparatively new. The contemporary medicine practice entered the country with the establishment of Bir Hospital in 1890. The development of the health system was like that of a stunted child. There were limited services in major urban centers. The more extensive section of the population living in rural villages is still deprived of health services. Nepal has made substantive progress in increasing the health status and health care among South Asian countries. Nepal has been relatively well in expanding the benefits of the FCHVs. Thus, the mobilization helped reduce some of the most pressing challenges, i.e., reducing the Child Mortality Rate and Maternal Mortality Rate. Historically, the Nepali Health Care System has been engrossed by periodic obstacles. From 1950 to 1990, extending primary health care with limited resources was a great challenge. From 1990 to 2006, the country was in the vicinity of political instability and the decade-long armed conflict, consequently leading to challenges in health service integration.

The aftermath of the 2006 April uprising saw Nepal take a giant stride in social, political, and economic fabrics. However, there are cumulative challenges that continue to persist in our health system. Ranabhat et al. (2019) outlined the service extension, integration, quality health service, equity, access, and financial protection during illness as crucial challenges of the modern Nepali Health system. The republic of Nepal established agendas for quality health services, accessible to everyone and guaranteed by the constitution to remove reiterations and improve service delivery. At this date, Nepal has done remarkably well in some of the Vital Health Indicators (VHI). The life expectancy has reached 71.5 years, and the Infant Mortality Rate (IMR) is 26.00/1000 live births in 2021 compared to 76.5/1000 live birth in 2001.

Now, the constitution of Nepal legally assured health as a fundamental right. However, in practice, it may take a long time to achieve it. Nepal has entered a new age of federal restructuring. Nepal has been divided into 753 local bodies. There is a considerable challenge in integrating the new

administrative structure (Thapa et al., 2019). Thapa et al. share the other examples and discuss how the local authorities' de prioritization of health has led to a crisis in Pakistan, Uganda.

The issue of health sector prioritization will continue to be a concern across all three levels of government and different grants, including the special and complementary grants. (Thapa et al., 2019). The spill-over effects and lack of clarity of authority amongst the jurisdictions of government are concerns of federalism. With many local governments (753), with a diverse range of economic potentials, holding executive, legislature, and financial, procedural power, the clarity issue is a primary concern in integrating the health system in an administrative domain. Shifting executive power to the local and provincial level have positive effects on local resource usage. Through participatory bottom-up planning, increased accountability, and reduced bureaucracy in decision-making in other countries; however, the prospective gains are contextual and not guaranteed. Similar reforms abroad have at times exacerbated inequities, weakened local commitment to priority health issues, and interfered with service delivery by disrupting referral chains (Thapa et al., 2019). While these bureaucratic and administrative issues continue to prevail, the increasing price of health services (medicine, diagnosis, treatment) is another concern. While the number of service providers has grown from the private sector, the public sector has remained indifferent in the health professionals and service's capacity at the same time. As mentioned earlier, the service price becomes the primary concern, followed by other intrinsic components of our health system.

2.3 Understanding SERVQUAL model

Service quality is a critical factor in modern times. The proliferation of investments has increased the number of providers in the service sector, providing consumers with ample alternatives. Quality has become a key determinant in industrial and service industries to gain maximum return on investments and significantly reduce cost (Anderson and Zeithaml 1984; Parasuraman et al., 1985). Service organizations like the manufacturing organizations are now well aware that they need to take preventive quality measures to gain customer satisfaction and retention (Spreng & MacKoy, 1996; Reichheld & Sasser, 1990). The importance of service quality has been recognized. Its implementation escorts the organizations to increase organizational performance,

customer satisfaction, and loyalty (Berry et al., 1989; Reichheld & Sasser, 1990; Rust & Zahorik, 1993; Spreng & MacKoy, 1996; Cronin et al., 2000; Yoon & Suh, 2004; Kang & James, 2004).

Among different evaluation methods, SERVQUAL by Parshuraman et al. (1985). SERVQUAL is the most prominent method to calculate the various services' quality. "It is the difference between customer expectations and perceptions; expectation means service provider performance during the deliverance of services whereas perception is a measurement of delivery by the service provider." (Parasuraman et al.).

The SERVQUAL model calculates the gap between the perceived quality and expected quality of service of the five dimensions mentioned below:

- Tangibles: Physical facilities, equipment, and appearance of personnel
- Reliability: Ability to perform the promised services dependably and accurately
- Responsiveness: Willingness to help customers and provide services
- Assurance: Knowledge and courtesy of employees and their ability to inspire trust and confidence
- Empathy: Caring, the individualized attention the firm offers to its customers

2.4 SERVQUAL in health care

The hospital administrators and marketers must determine how crucial various product/service attributes are to potential consumers and how those attributes influence consumer choice decisions. While the choices are adequate, the quality dimensions to assess healthcare are limited (Anderson & Zeithmal, 1984). Grönroos (1984) outlined two different aspects of quality, i.e., Technical Quality and Functional Quality. He also elaborates that technical quality refers to the technical outcome of the service delivery process and how this technical outcome responds to the instrumental performance of the service. He also underscored that the process of delivering the service is functional quality.

Babakus & Mangold, 1992, made arrangements on the SERVQUAL model earlier coined by Parashuraman et al. and developed a model to assess healthcare providers' functional quality.

The progress made in the medical sciences has also garnered giant eyeballs across the world. Apart from the core medical functioning, a health service provider is also assessed based on the functional quality that it offers. Service quality and patient satisfaction are prime concerns for health care establishments across the globe. The patient's perception of the services also affects the image and profitability of the hospital. The service quality also affects the patient's behavior regarding their loyalty and word of mouth (Andaleeb, 2001).

The SERVQUAL instrument developed by Parasuraman et al. (1985), comprising 22-items representing five dimensions, has been widely used in health care to measure service quality and health care.

SERVQUAL also provides hospital administrators with a tool for the measurement of functional quality in their organizations. Deficient scores on one or more SERVQUAL dimensions will generally signal a deeper underlying problem in the organization. For example, assume that SERVQUAL indicates that patients do not perceive hospital employees as willing to help. Thus, the low score can be symptomatic of deeper problems of the hospital not fulfilling the patient's expectations.

SERVQUAL's significant contributions to the healthcare industry are its ability to identify symptoms and provide a starting point for examining underlying problems that inhibit quality services.

2.4.1 Empirical evidence

A study on the response of the outpatients in the selected private hospitals of Kathmandu was done using the SERVPREF model. The private hospitals were categorized in five quartiles based on the number of beds, namely hospitals with less than 50 beds, 50-100 beds, 100-200 beds, 200-300 beds, and more than 300 beds. Two hospitals in each category and ten private hospitals were selected as samples using the convenience sampling method. From each hospital, 20 outpatients were purposively chosen as sample respondents, and the total size of the sample is 200. The study reveals that the patients are satisfied with the services of the private hospitals in Kathmandu. The outpatients' satisfaction is irrespective of their gender, age, income level, and education (Pathak,

2017). Among all constructs, the outpatient's have the most positive perceptions of the hospitals' assurance dimension, followed by tangibles, responsiveness, reliability, and empathy. While the study was one of the first in the Nepalese context, there were few pitfalls as the study was done among selected private hospitals. The study of the private hospital alone cannot identify the real problem in service delivery.

Likewise, a similar study using SERVQUAL was done at Pakistan's public and private hospitals with 640 samples (320 private and 320 public hospitals). The study findings suggested that the private hospitals in Pakistan aim to provide better health care facilities than public hospitals. Among all five dimensions of the SEVQUAL model, the private hospital stands out over the public hospitals in all dimensions.

A similar study applying the SERVQUAL model was done by Taner & Antony in 2006 to examine the differences in service quality between public and private hospitals in Turkey with 200 samples. The results indicate that inpatients in private hospitals were more satisfied with service quality than public hospitals. The results also suggest that inpatients in the private hospitals were more confident with doctors, nurses, and supportive services than their counterparts in the public hospitals. Finally, the results show that doctors' satisfaction and reasonable costs are the most significant determinants of service quality in public hospitals.

Likewise, a study was done at Ghana's private and public and private hospitals using SERVQUAL with 400 samples (Kwateng et al., 2019). Findings from the study indicated differences in service quality delivery with private health providers perceived by the patient as better than their public health counterparts. Except for the responsiveness dimension, the result from the study showed significant differences in other dimensions of service quality for the private sector. Although scores for the private sector were relatively higher, the study indicates a need to improve private and public hospitals' empathy and reliability dimensions.

2.4.2 Factors of SERVQUAL and their implications in health care delivery

Babakus & Mangold, 1992 citing the need for a comprehensive model for health care, remodeled SERVQUAL, hoping to help them identify the necessary changes. The expectations and

perceptions scales in the modified SERVQUAL instrument have emerged as unidimensional measures with excellent internal consistency reliabilities (Babakus & Mangold, 1992; Andaleeb, 2001). The two scales exhibit adequate validity as separate measures of (1) patients' expectations of hospital services and (2) their perceptions of the subject hospital's performance level. These results indicate that the scales can successfully assess the magnitude of the gap between patient perceptions and expectations (Babakus & Mangold, 1992). The modified SERVQUAL appears to be a concise and practical instrument useful for monitoring expectations and perceptions (Babakus & Mangold, 1992; Andaleeb, 2001; Irfan & Ijaz, 2011).

Assurance (Asu): The assurance variable includes four components: competence, courtesy, credibility, and security. The knowledge and courtesy of employees and their ability to inspire trust and confidence among the consumers (Parasuraman et al., 1985). The assurance variables are divided into four variables. The first factor that ensures assurance among the patients is how the hospitals instill confidence within the patient and whether the hospital has lived up to the patient's expectations. Likewise, the second variable under the "Assurance" construct is how safe the patient feels in their dealings with the hospitals. The third variable is associated with the courteousness of the personnel in the hospitals. The fourth variable under the “Assurance construct is related to the knowledge and the ability to answer the questions raised by the patients”.

Tangibles (Tan): The tangibles construct includes four different components concerned with the Physical facilities, equipment, and personnel appearance. The first variable under the tangibles is regarding the availability of modern equipment. The availability here refers to the outlook of the equipment and application (Babakus & Mangold, 1992). The second component of the Tangibles construct is associated with the visually appealing nature of the materials (signages, pamphlets, advertisements, navigators). Along with the appealing nature of the materials associated, the attractive nature of physical facilities (Toilets, Lifts, Cafeteria) is equally essential (Andaleeb, 2001). The final component of SERVQUAL is related to the hygiene and performance of the hospital personnel (Irfan & Ijaz, 2011).

Reliability (Rel): The reliability construct envisages five different components staggering around hospitals' ability to perform the promised service dependably and accurately. First and

foremost, the component associated with the reliability construct is the timeliness of fulfilling the commitment of the hospital to do something by a specific time. The second component under the reliability construct is associated with the hospital's ability to cater to the services they promise the patients to deliver. The third component under the reliability dimension is the ability of the hospital to get things right the first time (diagnosis process, diagnosis tools). The fourth component is associated with error-free records. It is an essential component concerned with the possible hassle that a patient may face while seeking treatment. The final component of the Reliability dimension is the promptness of the hospital to deliver its services to its patients (Parasuraman et al., 1985).

Responsiveness (Res): It is concerned with the providers' willingness to help customers and provide prompt service to the patient. The first component under the responsiveness construct is regarding the desire of the personnel to help the patient. The second variable under the responsiveness construct is associated with how the personnel at the hospital respond to their patients (Andaleeb, 2001) even when they are busy. The third component is whether the hospital personnel tell the patients about when the services will be performed. The fourth variable is associated with the ability of the hospital personnel to understand the specific needs of the patients.

Empathy (Emp): The empathy construct is concerned with the intrinsic human values of Empathy. It includes access, communication, and understanding the customer. It is also involved with the caring and individualized attention that the firm provides to its customers. The first component under the "Empathy" dimension is about the ability of the hospital to provide individual attention. Similarly, the second component is related to the staff who give personal care to the patients. The third component under the "Empathy" dimension is associated with the nature of the hospital to show interest to help the patient when they have a problem. The fourth component is regarding the convenient operating hours. It is associated with the availability of the services in the hours that is convenient to the patients. The final component is related to the patients' response on how the hospital keeps the best interests at heart.

2.5 Models of service quality assessment

- **SERVQUAL**

Over time various tools have been developed to evaluate the service quality. SERVQUAL, the most prominent of all instruments, is based on service quality and gap theory (Parasuraman, Zeithaml, & Berry, 1985, 1988). The service quality in the SERVQUAL model is measured by calculating the difference between the perceptions and the expectations, which can be mathematically expressed as:

$$\text{Service Quality } (SQ_i) = \sum_{j-1}^{k} (P_{ij} - E_{ij})$$

- **SERFPREF**

A similar tool for the assessment of the service quality was developed by Cronin and Taylor (1992), where they evaluate the service quality only by assessing the perceptions, which is mathematically expressed as:

$$\text{Service Quality } (SQ_i) = \sum_{j-1}^{k} (P_{ij})$$

In contrast to the SERVQUAL method, the SERFPREF was coined by Cronin and Taylor (1992). Cronin and Taylor (1992) found that service performance measurement (SERVPERF) produced better results with more reliable estimations. They also stressed that SERFPREF has greater convergent and discriminant validity, more significant explained variance, and less bias than SERVQUAL. This instrument also has many applications in diversified areas such as higher education, retail sector, ceramic industries, libraries, automotive repair, and fast-food restaurants.

Too high evaluation of customers' expectations is the most often mentioned weak point of SERVQUAL. The subject would be negatively oriented to the service by defining their expectations and excessively assessing their expectations. It was created by Cronin and Taylor, who found out that we cannot examine customer expectations because the expectations belong to the ambiguous and variable category. They can change very quickly. The SERVPREF method, unlike the method SERVQUAL, examines only the level of quality of the received and realized service (perception),

and compares it to the ideal services, i.e., the full scale of rating.

However, the applicability of the SERVPREF model is limited to few sectors. SERVQUAL has a much more comprehensive range of applications in service quality measurement, which includes: health care applications, dental school patient clinic, business school placement center, tire store, and acute care hospital dental offices, AIDS service agencies, physicians, retail chains, pest control, dry cleaning, and fast-food restaurants website application public services and banking (Rodrigues et al., 2011).

2.6 Model of service quality gaps

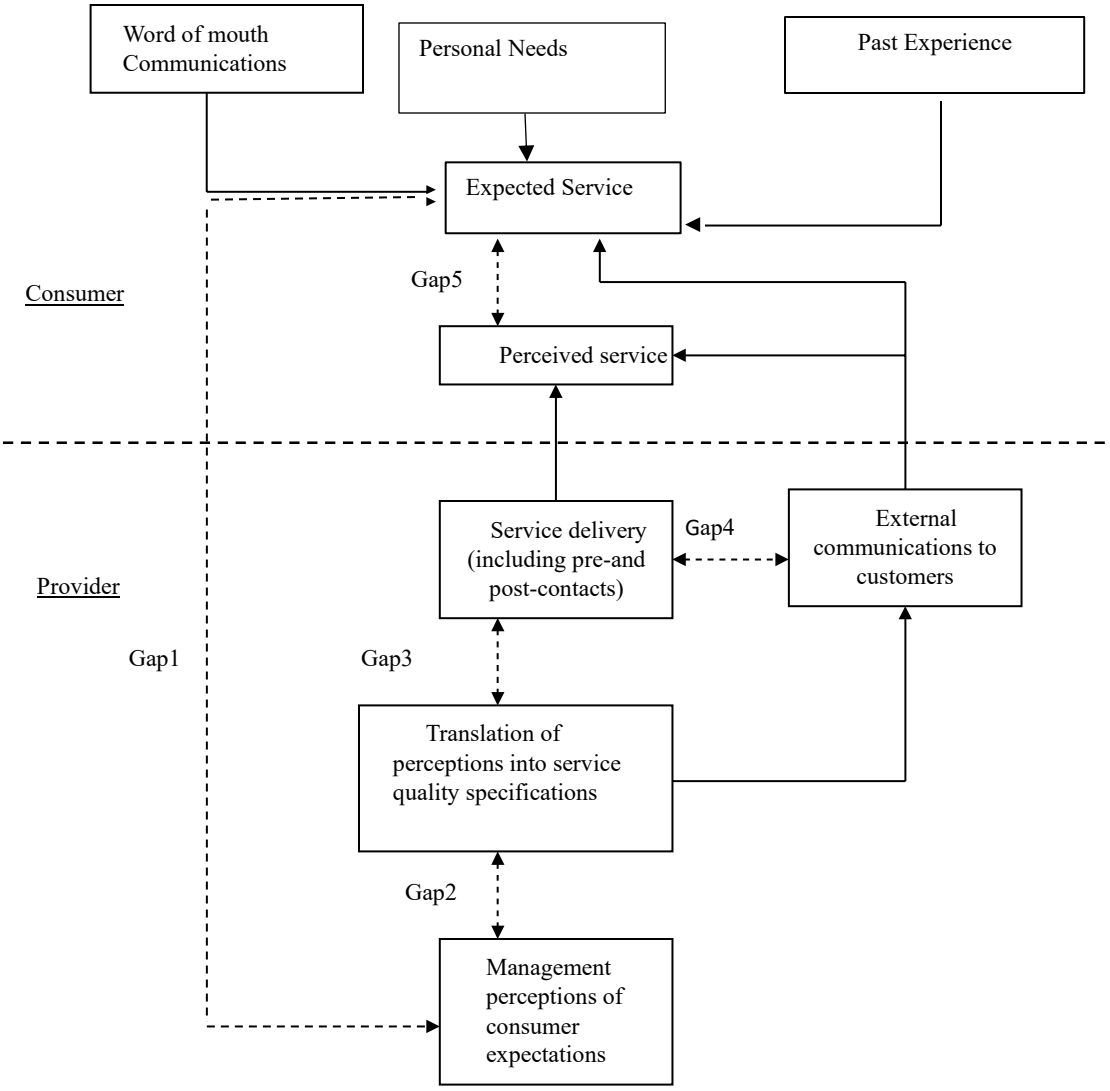
There are seven significant gaps in the service quality concept, which are shown in figure 1. The model is an extension of Parasuraman et al. (1985). The two critical gaps (Gap1 and Gap5), which are more associated with the external customers, directly affect the producers.

- Gap1: Customers' expectations versus management perceptions- due to the lack of a marketing research orientation, inadequate upward communication, and too many management layers.
- Gap2: Management perceptions versus service specifications due to the result of an inadequate commitment to service quality, a perception of unfeasibility, insufficient task standards, and an absence of goal setting.
- Gap3: Service specifications versus service delivery due to role ambiguity and conflict, poor employee-job fit and technology-job fit, inappropriate supervisory control systems, lack of perceived control, and lack of teamwork.
- Gap4: It incorporates the gap that results from Service delivery versus external communication. The gap exists due to inadequate horizontal communication and the propensity to over-promise.
- Gap5: The fifth Gap includes the discrepancy between customer expectations and their perceptions of the service delivered due to the influences exerted from the customer side and the shortfalls (gaps) on the service provider. In this case, customer expectations are

influenced by the extent of personal needs, word-of-mouth recommendations, and past service experiences.

According to Brown and Bond (1995), "the gap model is one of the best received and most heuristically valuable contributions to the services literature." The model identifies seven key discrepancies or gaps in customers' managerial perceptions of service quality and delivery tasks. The first four gaps (Gap 1, Gap 2, Gap 3, Gap 4) are identified as functions of how service is delivered. In contrast, Gap 5 pertains to the customer and is considered the accurate measure of service quality. The Gap on which the SERVQUAL methodology has influence is Gap 5.

Figure 1: SERVQUAL MODEL: Conceptual Framework



Source: Parasuraman, Zeithemal, & Berry, 1988

2.7 SERVQUAL methodology

The service quality measurement in the service sector should consider customer expectations of service and perceptions of service. One service quality measurement model that has been extensively applied is the SERVQUAL model gradually developed by Parasuraman et al. in

1985,1988,1991,1993,1994. SERVQUAL, as the most used approach for measuring service quality, is used to compare customers' expectations before a service encounter and their perceptions of the actual service delivered (Gronroos, 1984; Lewis & Boom, Lewis & Booms, 1983; Parasuraman et al., 1985). The SERVQUAL instrument has been the predominant method used to measure consumers' perceptions of service quality.

CHAPTER THREE

RESEARCH METHODOLOGY

This study employed a quantitative cross-sectional design to investigate healthcare service quality at the Civil Service Hospital, a public healthcare institution in Kathmandu, Nepal. The cross-sectional approach facilitated the collection of data at a single point in time, providing a snapshot of patient expectations and perceptions of service quality within this specific context. The design was structured in two sequential analytical phases: Confirmatory Factor Analysis (CFA) followed by Multiple Linear Regression (MLR). This methodology leveraged the SERVQUAL model (Parasuraman et al., 1988) to operationalize service quality across five dimensions: Assurance, Tangibles, Reliability, Responsiveness, and Empathy while addressing the study's objectives of validating the measurement model and identifying key predictors of service quality perceptions.

The data collection was done with a sample of 200 patients (n=200) using a convenient sampling technique. This sample size was deemed sufficient to support both CFA and MLR analyses, based on guidelines suggesting 5ó10 respondents per item for CFA (22 items Ÿ 10 = 220, with 200 being adequate with robust estimation) and 10ó20 cases per predictor for MLR (12 predictors Ÿ 15 = 180ó240). The required sample size as per the WHO's "Adequacy for the sample size in health studies" is 196. This sample size provides sufficient power to detect a discrepancy in the proportion of 0.5 with a 7 percent margin of error and a 95 percent confidence level (Lemeshow et al., 1990).

Data collection utilized a structured questionnaire comprising 22 SERVQUAL items distributed across the five dimensions: Assurance (4 items), Tangibles (4 items), Reliability (5 items), Responsiveness (4 items), and Empathy (5 items). Each item was rated on a 7-point Likert scale (1 = Strongly Disagree, 7 = Strongly Agree) to capture both patient expectations (Eij) and perceptions (Pij), enabling a comprehensive assessment of service quality gaps. Additional items collected demographic and contextual variables (e.g., age, gender, education, employment status, department visited, purpose of visit) to serve as predictors in the MLR analysis.

The analytical process began with CFA to validate the hypothesized five-factor structure of the SERVQUAL model for both expectation and perception scores. This phase ensured the instrument's psychometric properties, reliability, convergent validity, and discriminant validity

were suitable for the Nepalese public healthcare context before proceeding to predictive modeling. Following CFA, MLR was conducted to examine the influence of demographic and contextual predictors on service quality scores (either perception scores or gap scores, depending on the final specification). This sequential approach ensured that the measurement model was robustly established prior to exploring relationships with external variables, enhancing the study's statistical integrity.

The cross-sectional design was selected for its efficiency and alignment with the study's objectives, allowing for a focused examination of service quality within a single public hospital setting. By integrating CFA and MLR, the design provided a rigorous framework to: (1) confirm the applicability of the SERVQUAL model, (2) quantify service quality perceptions, and (3) identify actionable predictors for hospital administrators and policymakers. Data analysis was supported by descriptive statistics (e.g., means, standard deviations) to characterize the sample and SERVQUAL scores, with all statistical procedures executed in R using packages such as lavaan for CFA and stats or lmtest for MLR.

3.1 Data collection and instrument validation

The questionnaire consisted of 22 items spread across the five SERVQUAL dimensions (Assurance: 4 items, Tangibles: 4 items, Reliability: 5 items, Responsiveness: 4 items, Empathy: 5 items). These items captured both expected quality (Eij) and perceived quality (Pij) for each attribute, as outlined in the theoretical framework. Prior to further analyses, the factorial structure of the SERVQUAL instrument was validated using Confirmatory Factor Analysis (CFA) to ensure the measurement model's adequacy within the context of a Nepalese public hospital.

3.2 Confirmatory Factor Analysis (CFA)

3.2.1 Model specification for CFA

The CFA was employed to validate the five-factor structure of the SERVQUAL instrument: Assurance, Tangibles, Reliability, Responsiveness, and Empathy in the context of service quality

assessment at Civil Service Hospital, Kathmandu, Nepal. The CFA tested the theoretical framework proposed by Parasuraman et al. (1988), ensuring that the observed variables aligned with their hypothesized latent constructs. Below is a detailed specification of the model, including latent constructs, measurement equations, structural assumptions, parameter constraints, and estimation methods.

3.2.1.1 Latent constructs

The CFA model specified five latent constructs, each representing a dimension of service quality as perceived by patients. These constructs were unobserved variables inferred from multiple observed indicators (survey items). The constructs and their corresponding indicators were:

- **Assurance (ξ_1):** Reflected the knowledge, courtesy, and ability of hospital staff to inspire trust and confidence in patients. Measured by: P_Ass2 ("Personnel are consistently courteous"), P_Ass3 ("Personnel have knowledge to answer questions"), P_Ass4 ("Behavior instills confidence").
- **Tangibles (ξ_2):** Encompassed the physical environment, including facilities, equipment, and staff appearance. Measured by: P_T1 ("Modern equipment"), P_T2 ("Visually appealing materials"), P_T3 ("Visually appealing facilities"), P_T4 ("Neat appearance of personnel").
- **Reliability (ξ_3):** Represented the hospital's ability to perform services dependably and accurately. Measured by: P_Rel3 ("Gets things right the first time"), P_Rel4 ("Insists on error-free records").
- **Responsiveness (ξ_4):** Indicated the willingness and promptness of staff in assisting patients. Measured by: P_Res1 ("Willing to help"), P_Res2 ("Prompt service"), P_Res3 ("Never too busy to respond"), P_Res4 ("Tell when services will be performed").
- **Empathy (ξ_5):** Captured the provision of individualized, caring attention to patients. Measured by: P_Emp2 ("Individual attention"), P_Emp3 ("Understand specific needs"), P_Emp4 ("Have your best interests at heart").

3.2.1.2 Measurement model

The measurement model linked each observed variable (indicator) to its respective latent

construct through factor loadings (λ), accounting for measurement error (δ). The equations were specified as follows:

- **Assurance:**

$P\_Ass2 = \lambda_{21}\xi_1 + \delta_{21}$

$P\_Ass3 = \lambda_{31}\xi_1 + \delta_{31}$

$P\_Ass4 = \lambda_{41}\xi_1 + \delta_{41}$

- **Tangibles:**

$P\_T1 = \lambda_{12}\xi_2 + \delta_{12}$

$P\_T2 = \lambda_{22}\xi_2 + \delta_{22}$

$P\_T3 = \lambda_{32}\xi_2 + \delta_{32}$

$P\_T4 = \lambda_{42}\xi_2 + \delta_{42}$

- **Reliability:**

$P\_Rel3 = \lambda_{13}\xi_3 + \delta_{13}$

$P\_Rel4 = \lambda_{23}\xi_3 + \delta_{23}$

- **Responsiveness:**

$P\_Res1 = \lambda_{14}\xi_4 + \delta_{14}$

$P\_Res2 = \lambda_{24}\xi_4 + \delta_{24}$

$P\_Res3 = \lambda_{34}\xi_4 + \delta_{34}$

$P\_Res4 = \lambda_{44}\xi_4 + \delta_{44}$

- **Empathy:**

$P\_Emp2 = \lambda_{15}\xi_5 + \delta_{15}$

$P\_Emp3 = \lambda_{25}\xi_5 + \delta_{25}$

$P\_Emp4 = \lambda_{35}\xi_5 + \delta_{35}$

Where:

- λ_ij = Factor loading of the i-th indicator on the j-th latent construct, representing the strength of the relationship.
- δ_ij = Measurement error associated with each indicator, assumed to be uncorrelated across indicators and with the latent factors.
- ξ_j = Latent construct (j = 1 to 5).

3.2.1.3 Structural assumptions

The CFA model incorporated several structural assumptions to reflect the theoretical relationships among the constructs and ensure proper estimation:

- **Covariance Among Latent Constructs:** The five latent constructs (ξ_1 to ξ_5) were allowed to covary, with covariance parameters (φ_ij) estimated to capture interrelationships (e.g., φ_{12} between Assurance and Tangibles). This reflected the expectation that service quality dimensions were interrelated in a healthcare setting.
- **Uncorrelated Errors:** Measurement errors (δ_ij) were assumed to be uncorrelated with each other ($Cov(\delta\_i, \delta\_j) = 0$ for $i \neq j$) and with the latent constructs ($Cov(\delta\_ij, \xi\_k) = 0$), ensuring that residual variance was unique to each indicator.
- **Normality:** The observed variables were assumed to follow a multivariate normal distribution, supporting the use of maximum likelihood estimation (MLE).

3.2.1.4 Parameter constraints and identification

To ensure model identification and interpretability:

- **Scaling:** Each latent construct was scaled by fixing one factor loading to 1 (e.g., $\lambda_{21} = 1$ for Assurance, $\lambda_{12} = 1$ for Tangibles), setting the metric of the latent variable equal to that of its reference indicator. This resolved the scale indeterminacy inherent in latent variable models.
- **Degrees of Freedom:** With 16 observed variables (indicators) and 5 latent constructs, the model estimated factor loadings (11 free λs after fixing 5), covariances among constructs (10 φs), and error variances (16 δs), yielding a total of 37 parameters. The number of unique variance-covariance elements in the data was $16(16+1)/2 = 136$, resulting in 99 degrees of freedom ($df = 136 – 37$), ensuring an over-identified model testable for fit.

3.2.1.5 Estimation method

The model was estimated using Maximum Likelihood Estimation (MLE), which optimized the fit between the observed covariance matrix and the model-implied covariance matrix. MLE assumed multivariate normality and provided robust parameter estimates and standard errors, suitable for the continuous Likert-scale data collected from the SERVQUAL survey.

3.2.1.6 Fit indices

The model's adequacy was evaluated using multiple fit indices to assess how well the specified structure reproduced the observed data:

- Chi-Square to Degrees of Freedom Ratio (χ^2/df): Target < 3, indicating acceptable model complexity.
- Comparative Fit Index (CFI): Target > 0.95, comparing the model to a null baseline.
- Tucker-Lewis Index (TLI): Target > 0.95, adjusting for model parsimony.
- Root Mean Square Error of Approximation (RMSEA): Target < 0.05, assessing fit per degree of freedom.
- Normed Fit Index (NFI): Target > 0.90, evaluating incremental fit.

This specification tested the hypothesis that the SERVQUAL dimensions were distinct yet interrelated constructs, reliably measured by their respective indicators, providing a foundation for assessing service quality gaps at Civil Service Hospital.

3.2.2 Model diagnostics for CFA

Model diagnostics were conducted to evaluate the adequacy of the specified CFA model and ensure its fit to the observed data from the Civil Service Hospital sample (n=200). The diagnostic process assessed the model's validity, reliability, and potential need for refinement, with the following steps:

3.2.2.1 Estimation method

The model was estimated using MLE, suitable for the continuous 7-point Likert scale data (1 = Strongly Disagree, 7 = Strongly Agree), assuming approximate normality. Prior to estimation, data were screened for normality using skewness (|< 2|) and kurtosis (|< 7|) thresholds. If significant non-normality was detected, a robust estimation method (e.g., MLR with robust standard errors) was employed to adjust for violations.

3.2.2.2 Goodness-of-Fit indices

Multiple fit indices were used to assess how well the hypothesized five-factor model fit the data, with the following criteria:

- Chi-square Test (χ^2): Tested exact fit, with a non-significant p-value ($p > 0.05$) indicating good fit. However, due to its sensitivity to sample size (n=200), it was interpreted alongside other indices.
- Comparative Fit Index (CFI): Measured incremental fit compared to a null model, with CFI > 0.95 indicating good fit and > 0.90 acceptable.
- Tucker-Lewis Index (TLI): Adjusted for model complexity, with TLI > 0.95 indicating good fit and > 0.90 acceptable.
- Root Mean Square Error of Approximation (RMSEA): Assessed approximate fit, with RMSEA < 0.06 indicating good fit, < 0.08 acceptable, and a 90% confidence interval reported for precision.
- Standardized Root Mean Square Residual (SRMR): Measured absolute fit, with SRMR < 0.08 indicating good fit.

  These indices collectively provided a comprehensive evaluation, balancing absolute, incremental, and parsimony-adjusted fit.

3.2.2.3 Parameter estimates:

- Factor Loadings: Standardized loadings (λ) were examined, with values > 0.50 considered adequate and > 0.70 strong, indicating each item's contribution to its latent factor. Non-significant loadings ($p > 0.05$) prompted review of item relevance.
- Factor Covariances: Estimated covariances between latent factors were assessed for significance ($p < 0.05$) and magnitude to understand interrelationships among dimensions.

3.2.2.4 Reliability and validity:

- Composite Reliability (CR): Was calculated for each factor using the formula :

  $CR = (\Sigma\lambda)^2 / [(\Sigma\lambda)^2 + \Sigma(1 - \lambda^2)]$, with CR > 0.70 indicating good internal consistency.

- Average Variance Extracted (AVE): Was computed as $AVE = \sum(\lambda)^2 / n$ (where n = number of items per factor), with AVE > 0.50 confirming convergent validity.
- Discriminant Validity: Was assessed by comparing the square root of AVE for each factor to its correlations with other factors; if AVE exceeded inter-factor correlations, discriminant validity was supported.

3.2.3 Residual analysis

Standardized residuals were inspected to identify localized misfit. Residuals > |2.58| ($p < 0.01$) indicated significant discrepancies between observed and model-implied covariances, suggesting areas for potential adjustment.

3.2.4 Modification indices (MI)

If the initial model fit was inadequate (e.g., CFI < 0.90, RMSEA > 0.08), modification indices were reviewed to identify statistically significant improvements (MI > 3.84, $p < 0.05$). Potential modifications included:

- Adding error co-variances between items within the same factor (e.g., due to similar wording).
- Allowing cross-loadings if theoretically plausible (e.g., an item reflecting both Responsiveness and Empathy).

  Modifications were only implemented if they aligned with SERVQUAL theory and were substantively justified, with all changes reported transparently.

3.2.5 Software implementation

Diagnostics were conducted using the lavaan package in R, which provided detailed output for fit indices, parameter estimates, residuals, and modification indices, facilitating a thorough evaluation. These diagnostics ensured the CFA model was robust, reliable, and valid for measuring service quality perceptions and expectations, providing a solid foundation for subsequent gap analysis and regression in the study.

3.3 Data analysis

Following CFA validation of the measurement model, quantitative data analysis proceeded in three stages:

- **Descriptive Statistics:** Characterized the sample (e.g., age, gender, education) and summarized variable distributions (means, standard deviations) for expectations and perceptions across the 22 items.
- **Inferential Statistics:** Paired t-tests were conducted to perform gap analysis, comparing expectation (Eij) and perception (Pij) scores for each SERVQUAL dimension to test hypotheses H0 and H1.
- **Multiple Linear Regression (MLR):** Identified predictors of service quality, as detailed in section 3.5.

The study primarily focused on the fifth gap in the Gaps Model of Service Quality (Parasuraman et al., 1985), which represented the discrepancy between expected and perceived service quality within this public healthcare setting.

3.4 Theoretical framework and significance

The study employed the SERVQUAL model, a widely recognized framework for assessing service quality. This model allowed for a comprehensive evaluation of healthcare services from the patient's perspective. The research was significant as it contributed to the limited literature on healthcare quality assessment in Nepal. It provided evidence-based insights for policymakers to address quality disparities between public and private healthcare providers. Moreover, the findings helped hospital administrators identify areas for improvement, ultimately leading to enhanced patient satisfaction and better healthcare outcomes.

3.4.1 Quality of service (SERVQUAL)

The quality of service envisaged the computation drawn from the response of the respondent on two different instances. The Perceived Quality of Service constituted the respondents' response to their service quality after consuming the service. The Expected Quality of Service was associated with the service quality that they expected the service to deliver. Furthermore, the process of computation of the quality of service along with recording the response was further elaborated as:

3.4.2 Perceived quality of service (Pij)

22 variables were spread around five different dimensions (Assurance: 4, Tangibles: 4, Reliability: 5, Responsiveness: 4, Empathy: 5). The perceived quality of service mainly included the patients' perception of the functional quality of the hospitals. The value for each component's analysis was set with the Seven Point Likert Scale (1 = Strongly Disagree, 2 = Disagree, 3 = Somewhat Disagree, 4 = Neutral, 5 = Somewhat Agree, 6 = Agree, 7 = Strongly Agree).

3.4.3 Expected quality of service (Eij)

22 variables were spread around five different dimensions (Assurance: 4, Tangibles: 4, Reliability: 5, Responsiveness: 4, Empathy: 5). The expected quality of service mainly included the patients' perception of the ideal condition under which a hospital was expected to function. The value for each component's analysis was set with the Seven Point Likert Scale (1 = Strongly Disagree, 2 = Disagree, 3 = Somewhat Disagree, 4 = Neutral, 5 = Somewhat Agree, 6 = Agree, 7 = Strongly

Agree). SERVQUAL included the Private and Public hospitals' overall scores on each of the dimensions. The mean value of the Expectation was subtracted from the mean value of the perception. Service quality (SERVQUAL) was expressed mathematically as:

$$\text{Service Quality } (SQ_i) = \sum_{j-1}^{k} (P_{ij} - E_{ij})$$

Where:

- SQ_i = Service quality of individual 'i'
- k = number of service attributes/items
- P_ij = perception of individual 'i' concerning the performance of a service firm attribute 'j'
- E_ij = service quality expectation for attribute 'j' that was the relevant norm for individual 'i'

The respondents were asked to rate the service quality on a scale of 1-7. One referred to "Strongly Disagree," and seven was "Strongly Agree." The variables were further divided into five SERVQUAL dimensions (Assurance, Tangibles, Reliability, Responsiveness, and Empathy), i.e., the dependent variable. The SERVQUAL Score computed on each of the 22 SERVQUAL variables was rated between 1-7. Upon entering each variable's score, the mean value of the expected score was deducted from the mean value of the same variable's perception score. After deducting the expectation score, the remaining value was the SERVQUAL Score, which was later added with other variables belonging to the same SERVQUAL dimension. The negative SERVQUAL score (-) denoted that the hospital (irrespective of public or private) had failed to meet the patients' expectations regarding the hospital's services. Likewise, the positive SERVQUAL Score (+) denoted that the institution had rendered the service better than the patient expected. Since the scores were based on a seven-point Likert scale, the maximum SERVQUAL Score an institution could get was 7. It meant the patients felt the service's quality was seven times better than what they had expected. Likewise, the negative value denoted that the institutions could

not live up to the patient's expectations. The value "0" indicated no gap between the perceived quality and the expected quality of service.

3.5 Hypothesis testing

The study aimed to test two hypotheses:

- **Hypothesis 1 (H1)**: There are significant differences between patients' expectations and perceptions of service quality across all five SERVQUAL dimensions (Assurance, Tangibles, Reliability, Responsiveness, and Empathy) at Civil Service Hospital.
- **Hypothesis 2 (H2)**: The five-factor structure of the SERVQUAL model (Assurance, Tangibles, Reliability, Responsiveness, and Empathy) demonstrates adequate measurement qualities (reliability, convergent validity, and discriminant validity) in the context of a Nepalese public hospital.
- **Hypothesis 3 (H3)**: Gender significantly predicts the magnitude of service quality gaps in the SERVQUAL dimensions at Civil Service Hospital, with male patients reporting larger gaps than female patients in at least one dimension.
- **Hypothesis 4 (H4)**: Educational attainment significantly predicts the magnitude of service quality gaps in the SERVQUAL dimensions at Civil Service Hospital, with patients of higher education levels (High School and Bachelor's degree) reporting larger gaps than those with lower or no education.
- **Hypothesis 5 (H5)**: Age significantly predicts the service quality gaps in the SERVQUAL dimensions at Civil Service Hospital, with younger patients (aged 25 to 44) reporting larger gaps than older patients (aged 55 and above).

This was expressed as:
Let μD be the mean difference between perception and expectation scores for a given SERVQUAL dimension.

- $H0: \mu D = 0$
- $H1: \mu D \neq 0$

  Where: $\mu D = \mu P - \mu E$,
- μP = mean perception score,
- μE = mean expectation score

This hypothesis was tested for each of the five SERVQUAL dimensions:

- If the p-value < α (typically 0.05), H0 was rejected. This suggested a significant difference between patients' expectations and perceptions of service quality for that dimension.
- If the p-value ≥ α, H0 was not rejected. This suggested no significant difference between patients' expectations and perceptions of service quality for that dimension. The direction and magnitude of the difference were determined by examining the mean difference (μD) and its confidence interval.

3.6 Model specification and MLR

A multiple regression model was used to evaluate the impact of privateness and publicness on service quality hospitals. This model was used because the study sought to decipher how a single response variable (the service quality) Y depended on k predictor variables such as age, type of hospitals, educational attainment, marital status, income, employment, occupation, migration status, the status of foreign migration, the purpose of visit, and department visited.

The regression model was expressed as:

Y = β_0 + β_1 X_1 + β_2 X_2 + ... + β_k X_k + ε (1)

Where Y was the dependent variable, X1 → Xk were independent variables, β_0 was the y-intercept, β_1 → β_k were the slope of the line, and ε was the error variable. Thus, the equation used as a regression model for the study was expressed as: $(Asu_i, Res_i, Tan_i, Rel_i \; and \; Emp_i =$

$\beta_0 + \beta_1(Grp_i) + \beta_2(T_i) + \beta_3 G_i + \beta_4 Age_i + \beta_5 Edu_i + \beta_6 Mar_i + \beta_7 Emp_i + \beta_8 job_i +$

$\beta_9 Mig_i + \beta_{10} FM_i + \beta_{11} PV_i + \beta_{14} Dpt_i \; + \epsilon_i,$ $\epsilon_{i\sim niid(0,\sigma^2)} \; i=1,2,3...$

Where,

β_0= Constant term

Asu_i=Assurance

Tan_i=Tangibles

Rel_i= Reliability

Res_i=Responsiveness

Emp_i= Empathy

Grp_i= Group

Age_i = Age of the respondents

Edu_i = Education attainment

Mar_i = Marital Status

EMP_i = Employment Status

Job_i = Occupation

Mig_i = Migration Status

3.6.1 MLR assumptions and diagnostics

The MLR analysis adhered to the following assumptions, which were tested:

- **Linearity:** The relationship between predictors and the dependent variable was linear, assessed via residual plots.
- **Independence:** Observations were independent, ensured by the random sampling design.
- **Homoscedasticity:** Variance of residuals was constant across levels of predictors, checked using the Breusch-Pagan test or residual plots.
- **Normality:** Residuals were normally distributed, verified using Q-Q plots or the Shapiro-Wilk test.
- **Multicollinearity:** Predictor variables were not highly correlated, assessed using Variance Inflation Factor (VIF), with VIF < 10 indicating acceptable levels.

If assumptions were violated (e.g., heteroscedasticity or multicollinearity), corrective measures such as robust standard errors or variable transformation were applied.

3.6.2 MLR estimation and interpretation

Estimation Method: Ordinary Least Squares (OLS) was used to estimate the regression coefficients (β_1 to β_{12}), providing the best linear unbiased estimates under the stated assumptions. **Model Fit:** The goodness-of-fit was evaluated using the R^2 value (proportion of variance explained) and adjusted R^2 (adjusted for the number of predictors). **Significance Testing:** Each coefficient's statistical significance was tested using t-tests ($p < 0.05$),

with 95% confidence intervals reported to indicate the precision of estimates. **Software:** MLR was conducted using R, leveraging packages such as stats or lmtest for estimation and diagnostics.

This MLR analysis identified which predictors (e.g., age, education, department visited) significantly influenced service quality perceptions, providing actionable insights for hospital administrators.

CHAPTER FOUR

RESULT AND FINDINGS

4.1 Characteristics of respondents

The respondent characteristics in this study offer a detailed demographic and socioeconomic framework for interpreting the SERVQUAL findings, which likely evaluate service quality across dimensions such as tangibles, reliability, responsiveness, assurance, and empathy.

Table 1: Characteristics of respondents

Variables	Category	Percentage
Gender	Male	63.3
	Female	36.7
Educational Attainment	Illiterate	2.7
	Literate (Can read and write)	4.7
	Secondary School	3.3
	High School	22
	Bachelor`s Degree	38
	Masters and above	29.3
Marital Status	Married	69.3
	Unmarried	28.7
	Widowed	0.7
	Divorced	1.3
Job Type	Civil Service	38.7
	Private	23.3
	Self Employed	20
	Unemployed	18
Age	15-24	16
	25-34	33.3
	35-44	28.7
	45-54	12
	55-64	6

	65-74	3.3
	75-84	0.7
Income Bracket/month	<25000	38.7
	25001-50000	23.3
	50,001-100000	20
	100001 and above	18

The sample comprised 63.3% males and 36.7% females, creating a male-dominated pool that might tilt the results toward male perspectives on service quality which potentially emphasizing efficiency or reliability over interpersonal elements like empathy, which females might prioritize more due to differing social experiences or expectations. Educational attainment varied widely, with 38.0% holding a Bachelor's Degree, 29.3% possessing Masters or higher qualifications, 22.0% with High School education, 4.7% classified as Literate (able to read and write without formal credentials), 3.3% with Secondary School education, and 2.7% Illiterate; this distribution highlights a largely educated sample (67.3% with tertiary education). Marital status showed a strong majority of 69.3% married, followed by 28.7% unmarried, 1.3% divorced, and 0.7% widowed, indicating that married individuals potentially valuing stability and dependability that dominate the sample, which could influence preferences for consistent or trustworthy service over the flexibility or speed that unmarried respondents might favor. Occupationally, 38.7% worked in civil service, 23.3% in the private sector, 20.0% were self-employed, and 18.0% were unemployed, reflecting a mix of professional experiences that could shape expectations: civil servants might prioritize professionalism and structure, private sector workers' efficiency and innovation, self-employed individuals' flexibility and personalization, and the unemployed affordability or tangible benefits due to possible cost-sensitivity.

Age distribution leaned toward younger and middle-aged adults, with 33.3% aged 25-34, 28.7% aged 35-44, 16.0% aged 15-24, 12.0% aged 45-54, 6.0% aged 55-64, 3.3% aged 65-74, and 0.7% aged 75-84, meaning 62.0% fell between 25 and 44; this suggests a tech-savvy or modern-oriented majority that might rate digital or innovative service aspects highly, while older respondents (9.0% aged 55-84) could emphasize traditional qualities like courtesy or reliability. Income brackets further diversified the sample, with 38.7% earning less than 25,000, 23.3% earning 25,001-50,000, 20.0% earning 50,001-100,000, and 18.0% earning above 100,000 (currency unspecified),

indicating economic variety where the largest low-income group might focus on cost and practicality, middle-income respondents (43.3%) balance quality and affordability, and high earners demand premium or sophisticated service. Collectively, this predominantly male, educated, married, employed (especially in civil service), younger-to-middle-aged, and economically diverse sample provides a rich lens through which to view the SERVQUAL results, with potential correlations between these traits and service quality perceptions offering valuable insights—such as highly educated civil servants valuing assurance and reliability, or lower-income unemployed respondents prioritizing tangibles and responsiveness—enhancing the study's applicability to real-world service improvement efforts.

4.2 Reliability scores

The reliability analysis of the SERVQUAL instrument used to assess service quality at CSH, Kathmandu, was conducted using Cronbach's Alpha (α) to measure internal consistency across five dimensions: Assurance, Tangibles, Reliability, Responsiveness, and Empathy.

Table 2: Reliability Scores (Cronbach's Alpha (α))

Reliability Scores (Cronbach's Alpha (α))		
SERVQUAL Dimension	**Expectation**	**Perception**
Assurance	0.903	0.707
Tangibles	0.931	0.874
Reliability	0.899	0.886
Responsiveness	0.907	0.853
Empathy	0.92	0.824

The results from the table 2 indicate that the expectation-related items exhibit very high reliability, with α values ranging from 0.899 to 0.931, signifying those patients had a clear and stable understanding of their service expectations. This suggests that the survey questions effectively captured the intended constructs and that responses to expectation-related items were highly consistent among the participants. Among the expectation dimensions, Tangibles (α = 0.931) and Empathy (α = 0.920) displayed the highest reliability, indicating that patients uniformly expected tangible aspects such as hospital facilities, equipment, and cleanliness, as well as empathetic care from hospital staff. The other dimensions, including Assurance (α = 0.903), Reliability (α = 0.899),

and Responsiveness (α = 0.907), also demonstrate excellent reliability, reinforcing that patients had well-defined and consistent expectations regarding staff competence, reliability of services, and responsiveness to patient needs.

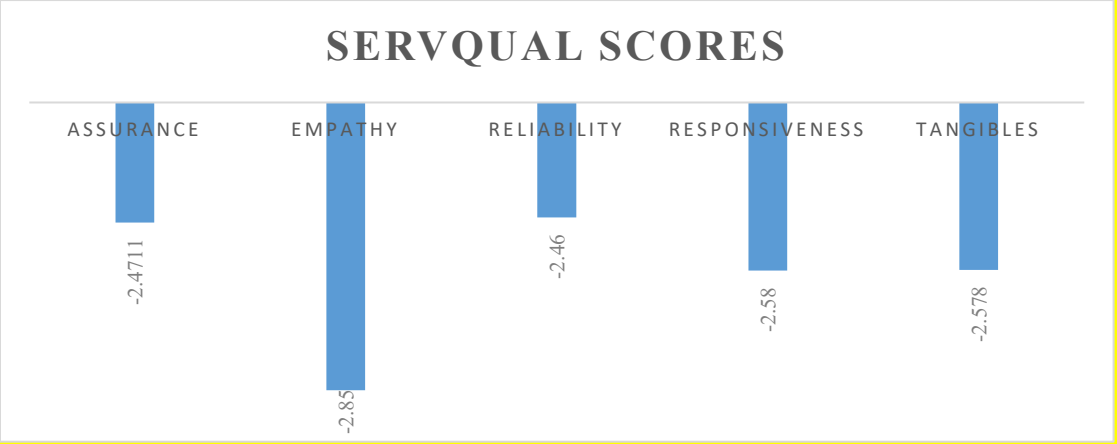
For perception-related responses, the reliability scores, while still acceptable to good, were slightly lower, with Cronbach's Alpha values ranging from 0.707 to 0.886. This slight decrease in reliability suggests that patients' actual experiences of service quality were more variable compared to their expectations. Among these, Reliability (α = 0.886) and Tangibles (α = 0.874) showed the highest reliability, indicating that patients generally shared similar experiences regarding the dependability of hospital services and the adequacy of hospital facilities. Responsiveness (α = 0.853) and Empathy (α = 0.824) also had good reliability, suggesting that the perception-related survey items in these categories were consistent. However, Assurance (α = 0.707) had the lowest reliability among perception responses, though it still meets the acceptable threshold of 0.7. This lower α value implies that patients' experiences regarding trust in hospital staff, competence of healthcare providers, and communication of medical procedures may have been more diverse or inconsistent. Factors such as individual differences in patient expectations, variability in staff behavior, or inconsistencies in service delivery across different hospital departments could contribute to this variation.

Overall, since all Cronbach's Alpha values exceed the 0.7 threshold, the survey instrument can be considered statistically reliable, ensuring that the collected data is dependable for further analysis, including gap analysis (Expectation vs. Perception) and inferential statistical tests such as ANOVA or regression analysis. The higher reliability in expectations compared to perceptions suggests that while patients have uniform expectations about hospital services, their actual experiences may differ based on individual interactions, service variability, and personal satisfaction levels. The lower reliability for Assurance in perception scores might warrant further investigation, as it could indicate inconsistencies in staff behavior, patient trust issues, or varying levels of service delivery quality. Moving forward, service quality improvement efforts should focus on addressing the discrepancies in patient experiences, particularly in Assurance, to enhance the consistency of care and build greater patient trust in healthcare services.

4.3 SERVQUAL analysis

The graph 1 (see below) represents the SERVQUAL gap scores, calculated as the difference between perception and expectation means for each dimension. The values: Assurance (-2.47), Empathy (-2.85), Reliability (-2.46), Responsiveness (-2.58), and Tangibles (-2.75) indicate that all dimensions fall short of patient expectations, with negative gaps reflecting poorer perceived performance. The graph's order (Empathy, Tangibles, Assurance, Reliability, Responsiveness) highlights Empathy as the most significant deficiency (-2.85), followed by Tangibles (-2.75), while Responsiveness shows the least severe gap (-2.28).

Graph 1: SERVQUAL Scores



4.3.1 Assurance dimension

The Assurance dimension, which reflects the knowledge, courtesy, and confidence-building capacity of hospital staff, reveals a substantial gap between patient expectations and perceptions at Civil Service Hospital, Kathmandu. The table below (table 3) indicates that patients expect a strong sense of safety (mean expectation: 6.67) but perceive it at 4.38 (-2.29), with similar gaps for courtesy (6.55 vs. 4.08, -2.47), knowledgeability (6.57 vs. 4.00, -2.57), and confidence-instilling behavior (6.66 vs. 4.08, -2.58). The SERVQUAL Scores graph corroborates this, showing an Assurance gap of -2.47, aligning closely with the courtesy-specific gap and suggesting that staff professionalism is a key contributor to the overall deficit.

Table 3: Assurance dimension

Expectation	Mean	Perception	Mean
E_Assurance (1) Patients of excellent hospitals will feel safe in their dealings with the hospitals .	6.6667	P_Assurance (1) You feel safe in your dealings with the hospital.	4.3800
E_Assurance (2) Personnel in excellent hospitals will be consistently courteous to patients	6.5467	P_Assurance (2) Personnel in the hospital are consistently courteous to you	4.080
E_Assurance (3) Personnel in excellent hospitals will have the knowledge to answer patients' questions	6.5667	P_Assurance (3) Personnel in the hospital have the knowledge to answer your questions	4.0000
E_Assurance (4) The behavior of personnel in excellent hospitals/clinics will instill confidence in patients	6.6644	P_Assurance (4) The behavior of personnel in the hospital instilled confidence in you	4.0816

Moreover, these findings align with Parasuraman et al.'s (1988) assertion that assurance fosters trust, a vital component in healthcare where patients are vulnerable. At CSH, this discrepancy may stem from systemic challenges in Nepal's public healthcare sector, such as understaffing and inadequate training, which hinder staff capacity to engage effectively. Overworked personnel may lack the time or energy to exhibit courtesy consistently or address queries with expertise. In Kathmandu, where CSH serves a diverse population reliant on government-subsidized care, the -2.47 gap underscores a need for targeted interventions in staff development and patient reassurance strategies, as recommended by Andaleeb (2001) for improving trust in healthcare delivery.

4.3.2 Empathy dimension

The Empathy dimension, evaluating individualized care, shows the widest gap at -2.85 in the SERVQUAL Scores graph, verified by the data (see table 4) with gaps from -2.67 to -3.09 for individual attention (6.53 vs. 3.77), personal attention (6.56 vs. 3.58), convenient hours (6.43 vs. 3.76), need understanding (6.49 vs. 3.72), and best interests (6.56 vs. 3.47). This -2.85 average underscores a profound disconnect in patient-centered care, a cornerstone of healthcare quality (Parasuraman et al., 1988). High patient volumes and standardized procedures likely explain the low scores, while the best interests gap (-3.09) suggests a trust deficit impacting recovery (Donabedian, 1980). Inflexible hours (3.76) may not suit diverse demographics in Kathmandu (K.C. et al., 2021), and cultural hierarchies in Nepal could inhibit engagement (Shrestha et al.,

2017). Empathy's scarcity, due to time and resource constraints (World Bank, 2018), marks it as a critical area for improvement at CSH, where emotional support directly shapes the patient experience.

Table 4: Empathy dimension

Expectation	Mean	Perception	Mean
E_Empathy (1) Excellent hospitals will give patients individual attention.	6.527027	P_Empathy (1) The hospital provides individual attention to you	3.77027
E_Empathy (2) Excellent hospitals will have staff who give personal attention to patients	6.564626	P_Empathy (2)The hospital has personnel who give you personal attention.	3.578231
E_Empathy (3) Excellent hospitals will have operating hours convenient to all their patients	6.42953	P_Empathy (3) The hospital has operating hours convenient to all its patients.	3.763514
E_Empathy (4) The personnel of excellent hospitals will understand the specific needs of their patients	6.486486	P_Empathy (4) The personnel of the hospital understand your specific needs	3.716216
E_Empathy (5) Excellent hospitals will have the patient's best interests at hearts	6.563758	P_Empathy (5) The hospital has your best interests at heart.	3.466216

4.3.3 Reliability dimension

The Reliability dimension, measuring dependable and accurate service delivery, shows a gap of -2.46 in the SERVQUAL Scores graph, closely mirroring the data's range of -2.33 to -2.59 across promises kept (6.53 vs. 3.94), timely service (6.46 vs. 3.91), first-time accuracy (6.37 vs. 3.90), error-free records (6.43 vs. 4.10), and problem-solving sincerity (6.45 vs. 3.98). This consistency highlights reliability as a persistent weakness, aligning with Parasuraman et al.'s (1988) view of it as the backbone of service quality, yet reflecting operational challenges in public hospitals. Delays and errors likely stem from logistical bottlenecks like: long queues, insufficient staff, or equipment shortages plausible at CSH given its role in Kathmandu (Poudel et al., 2019). The slightly narrower gap in error-free records (4.10) suggests some effort, but the overall -2.46 gap indicates systemic strain. Patients' perception of insincere problem-solving hints at overburdened staff. For a hospital serving civil servants and the public, this reliability deficit undermines its mandate, reflecting broader infrastructure issues in Nepal that require managerial and policy solutions.

Table 5: Reliability dimension

Expectation	Mean	Perception	Mean

E_Reliability(1) When excellent hospitals promise to do something	6.5342	P_Reliability(1) When the hospital promises to do something by a certain time, it does so	3.9396
E_Reliability(2) Excellent hospitals will provide their services at the time they promise to do	6.4595	P_Reliability(2) The hospital provides its services at the time they promise to do	3.9116
E_Reliability(3) Excellent hospitals will get things right at the first time itself	6.3699	P_Reliability(3) The hospital gets things right at the first time itself	3.9041
E_Reliability(4) Excellent hospitals will insist on error-free records	6.4315	P_Reliability(4) The hospital insists on error-free records	4.0952
E_Reliability(5)When a patient has a problem, excellent hospitals/clinics will show a sincere interest in solving it	6.4521	P_Reliability(5) When you have a problem, the hospital shows a sincere interest in solving it	3.9799

4.3.4 Responsiveness dimension

The Responsiveness dimension, focusing on staff willingness and promptness, records the smallest gap at -2.28 in the SERVQUAL Scores graph, though still significant. This aligns with gaps (see table 6) for willingness to help (6.49 vs. 3.92, -2.57), prompt service (6.46 vs. 3.97, -2.49), availability despite busyness (6.42 vs. 3.79, -2.63), and service timing communication (6.48 vs. 3.83, -2.65), where the graph's average reflects a slight moderation. This suggests responsiveness is less deficient than other dimensions, yet it remains a challenge, as it drives patient satisfaction. Long wait times and high patient-to-staff ratios, common in public hospitals, likely contribute to perceptions of unavailability, with the "too busy" score (3.79) reinforcing this. The communication gap (3.83) indicates information flow issues, linked to diminished trust. In Kathmandu, where urban patients expect faster service, the -2.28 gap suggests a need for better staffing and workflow management, as Donabedian (1980) recommended, to enhance responsiveness at CSH.

Table 6: Responsiveness dimension

Expectation	Mean	Perception	Mean
E_Responsiveness (1)Personnel in excellent hospitals will always be willing to help patients	6.4865	P_Responsiveness (1)Personnel in hospital are always willing to help you	3.9184
E_Responsiveness (2) Personnel in excellent hospitals will give prompt service to patients	6.4631	P_Responsiveness (2) Personnel in hospitals give prompt service to you.	3.9660

E_Responsiveness (3) Personnel in excellent hospitals will never be too busy to respond to patients'	6.4228	P_Responsiveness (3) Personnel at the hospital are never too busy to respond to you	3.7919
E_Responsiveness (4) Personnel in excellent hospitals will tell patients exactly when services will be performed'	6.4765	P_Responsiveness (4) Personnel in the hospital tell you exactly when services will be performed'	3.8255

4.3.5 Tangibles dimensions

The Tangibles dimension, encompassing the physical environment, equipment, and appearance of personnel, highlights a critical area of underperformance, with the SERVQUAL Scores graph reporting a gap of -2.75—the second widest among the dimensions. This is consistent with the provided data (see table 7), where patients expect modern equipment (6.63 vs. 3.94, -2.69), visually appealing materials (6.55 vs. 3.82, -2.73), facilities (6.50 vs. 3.86, -2.64), and neat personnel (6.49 vs. 4.01, -2.48), reflecting a range of gaps averaging around this value. This alignment reinforces SERVQUAL's emphasis on tangibles as a visible signal of service quality (Parasuraman et al., 1988), yet it mirrors the resource constraints typical of public healthcare in developing countries like Nepal. The significant gap in equipment modernity likely stems from budgetary limitations, a pervasive issue where funds prioritize basic needs over technology (Shrestha et al., 2017). In Kathmandu, where urban patients compare public facilities to better-equipped private hospitals, the -2.75 gap amplifies dissatisfaction. This dimension underscores the challenge of balancing functionality with aesthetic quality, a tension well-documented in resource-limited settings and suggests that infrastructure upgrades are a priority for CSH.

Table 7: Tangibles dimension

Expectation	Mean	Perception	Mean
E_Tangibles(1) Excellent hospitals will have modern equipment	6.6276	P_Tangibles(1) The hospital has modern equipments	3.9444
E_Tangibles(2) Materials associated with the service will be visually appealing in excellent hospitals	6.5473	P_Tangibles(2) Materials associated with the service are visually appealing	3.8176
E_Tangibles(3) The Physical facilities at excellent hospitals will be visually appealing	6.4966	P_Tangibles(3) The Physical facilities at the hospital is visually appealing	3.8581
E_Tangibles(4) Personnel at excellent hospitals will be neat in appearance	6.4865	P_Tangibles(4) Personnel at the hospital are neat in appearance	4.0067

Integrating these findings, the SERVQUAL Scores graph visually confirms a consistent pattern of negative gaps at Civil Service Hospital, with Empathy (-2.85) as the weakest, followed by

Tangibles (-2.75), Reliability (-2.46), Assurance (-2.47), and Responsiveness (-2.28). This mirrors the dataìs range (-2.29 to -3.09), suggesting patientsì high expectations (6.37ó6.67 vs. 3.47ó4.38 perceptions) are shaped by private healthcare exposure in Kathmandu (K.C. et al., 2021). Empathyìs dominance highlights a patient-centered care failure, consistent with public healthcare trends in developing countries (World Bank, 2018), while Tangibles and Reliability reflect resource and operational limits (Shrestha et al., 2017). Responsiveness and Assurance point to human resource challengesò understaffing and burnoutò prevalent in South Asia (Poudel et al., 2019; Khatri & Mishra, 2020). In Kathmandu, serving urban and rural populations, CSH faces heightened demands within a constrained framework (Khatri & Mishra, 2020). These findings align with South Asian healthcare studies (Andaleeb, 2001), pinpointing empathy and tangibles as acute issues.

4.4 Paired sample T-Test

The rationale for using a paired sample t-test in the study on service quality at Civil Service Hospital (CSH) is justified by its alignment with the SERVQUAL framework's methodology, which measures service quality as the gap between patients' expectations (Eij) and perceptions (Pij) of service across five dimensions: Assurance, Tangibles, Reliability, Responsiveness, and Empathy. The paired sample t-test is specifically designed to compare the means of two related group here, the expectation and perception scores for each dimension from the same 200 patients, to determine if significant differences (gaps) exist. This approach is appropriate because the data are paired: each patient provides both an expectation score (what they anticipate from the hospital) and a perception score (what they actually experience) on the same 7-point Likert scale items, ensuring dependency between the two measures. The test's application, as shown in Table 8, evaluates whether the mean difference ($\mu D = Pij - Eij$) for each dimension is statistically significant ($H0: \mu D = 0$ vs. $H1: \mu D \neq 0$), directly addressing the study's objective of quantifying service quality gaps (Page 8). The results, with significant negative correlations for Tangibles ($r = -0.331$, $p < 0.001$), Reliability ($r = -0.398$, $p < 0.001$), Responsiveness ($r = -0.320$, $p < 0.001$), and Empathy ($r = -0.175$, $p = 0.032$), and non-significant correlation for Assurance ($r = 0.064$, $p = 0.436$), confirm the presence of gaps (Pages 45-46), supporting the SERVQUAL premise that service quality is perceived through expectation-perception discrepancies (Parasuraman et al., 1985). The paired t-test's assumptions—approximately normal differences and interval-like

Likert scale data—are reasonably met, as the study's large sample (n = 200) and robust CFA diagnostics (e.g., skewness/kurtosis checks) mitigate minor violations. By using this test, the study effectively isolates dimension-specific deficits, such as the largest gap in Empathy (-2.85), providing a statistically rigorous basis for recommending targeted interventions like staff training and infrastructure upgrades (Page 58), thus aligning with the study's goal of informing policy and practice in Nepal's public healthcare.

The results from the paired sample correlations provide valuable insights into the service quality gaps at Civil Service Hospital, Minbhawan, Kathmandu, by analyzing the relationship between patient expectations and actual perceptions across five dimensions: Assurance, Tangibles, Reliability, Responsiveness, and Empathy (see table 8). Among these dimensions, Assurance is the only factor that does not exhibit a statistically significant correlation (r = 0.064, p = 0.436), suggesting that there is no meaningful relationship between patients' expectations and their perceptions of the hospitalìs ability to inspire trust and confidence. This implies that regardless of how high or low a patientìs expectations are regarding the competence, courtesy, and credibility of healthcare providers, their actual perception of Assurance remains unaffected. This lack of significant correlation could indicate that assurance-related factors such as the professionalism and knowledge of doctors and nurses either meet patient expectations consistently or that patients have a generally stable perception of healthcare provider competence irrespective of their initial expectations.

Table 8: Paired Sample T-test

Dimension	N	Correlation	Sig.
Assurance(P-E)	200	0.064	0.436
Tangibles(P-E)	200	-0.331	0
Reliability(P-E)	200	-0.398	0
Responsiveness(P-E)	200	-0.32	0
Empathy(P-E)	200	-0.175	0.032

Conversely, the other four dimensions: Tangibles, Reliability, Responsiveness, and Empathy show significant negative correlations, highlighting substantial gaps between what patients expect and what they actually experience in the hospital. The Tangibles dimension, which refers to the hospital's physical environment, facilities, medical equipment, and the appearance of personnel, has a moderate negative correlation ($r = -0.331$, $p = 0.000$). This significant negative correlation suggests that patients who have higher expectations regarding the hospital's infrastructure and physical environment tend to report lower satisfaction with what they perceive in reality. This finding may indicate that the hospital's existing facilities do not meet patient expectations, perhaps due to outdated equipment, inadequate cleanliness, or insufficient amenities. Since tangibles are among the first aspects of a healthcare institution that patients interact with, an expectation-reality mismatch in this domain could significantly affect overall patient satisfaction.

Reliability, which measures the hospital's ability to deliver consistent, dependable, and accurate healthcare services, demonstrates the strongest negative correlation ($r = -0.398$, $p = 0.000$) among all dimensions. This suggests that patients who enter the hospital with high expectations regarding reliability such as accurate diagnoses, timely treatment, and error-free medical procedures are more likely to perceive shortcomings in these areas. A pronounced gap in reliability may be attributed to issues such as delays in treatment, misdiagnoses, or inconsistencies in service delivery. Given that reliability is one of the most crucial components of service quality in a hospital setting, this gap is particularly concerning. It underscores the need for hospital management to implement stricter quality control measures, improve workflow efficiency, and enhance communication between medical staff and patients to ensure that services are delivered in a dependable and consistent manner.

Similarly, Responsiveness, which pertains to the hospital's willingness and ability to help patients promptly and efficiently, exhibits a moderate negative correlation ($r = -0.320$, $p = 0.000$). This indicates that patients with higher expectations regarding the hospital's promptness in attending to their needs tend to perceive delays and inefficiencies in service delivery. A negative correlation in this domain suggests that the hospital may struggle with timely patient care, long waiting times, or insufficient staffing to address patient concerns effectively. Responsiveness is a critical factor in healthcare settings because delayed responses to patient needs can lead to frustration, anxiety, and even medical complications. Addressing this gap requires a strategic approach, such as

optimizing hospital workflow, hiring additional staff, and using technology to streamline patient interaction and service delivery.

Lastly, Empathy, which represents the hospital staff's ability to provide personalized care and show genuine concern for patients, also shows a significant negative correlation ($r = -0.175$, $p = 0.032$), although the strength of the correlation is weaker compared to the other dimensions. This suggests that while there is a tendency for patients with higher expectations of empathy to perceive lower actual empathy, the gap is not as pronounced as in the other domains. Nevertheless, empathy is an essential component of patient-centered care, and even a weak negative correlation suggests room for improvement. A perception gap in empathy may arise from factors such as high patient-to-staff ratios, time constraints on healthcare providers, or a lack of adequate training in patient communication.

Overall, the results indicate that except for Assurance, there is a significant negative correlation between patient expectations and their actual perceptions across all other dimensions. The strongest expectation-perception gap is observed in Reliability, followed by Tangibles and Responsiveness, indicating that these are the key areas where the hospital needs to focus its quality improvement efforts. In the context of CSH these results suggest that while healthcare providers may be perceived as competent (Assurance), there are critical areas of concern in terms of physical infrastructure, consistency of service, responsiveness, and personalized care. To improve overall service quality, the hospital administration should consider targeted interventions such as upgrading facilities, enhancing service reliability, reducing waiting times, and fostering a more patient-centered approach. Addressing these gaps will not only enhance patient satisfaction but also contribute to better healthcare outcomes and increased public trust in the institution.

4.5 Gap in scores

The table below (table 9) compares the Cronbach's Alpha scores for the perception measures of the SERVQUAL dimensions before and after conducting a CFA, providing insight into how the reliability of these constructs changed during the model refinement process. The SERVQUAL dimensions assessed are Assurance, Tangibles, Reliability, Responsiveness, and Empathy, which are key components of service quality evaluation. Before CFA, the Perception Alpha scores were as follows: Assurance at 0.707, Tangibles at 0.874, Reliability at 0.886, Responsiveness at 0.853,

and Empathy at 0.824, indicating generally acceptable to strong reliability, though Assurance was the weakest, just above the 0.70 threshold. After CFA, the Alpha scores shifted to 0.73 for Assurance (a slight increase of 0.023), 0.874 for Tangibles (no change), 0.823 for Reliability (a decrease of 0.063), 0.853 for Responsiveness (no change), and 0.793 for Empathy (a decrease of 0.031) (see table 9).

Table 9: Gap in reliability scores

SERVQUAL Dimension	Pre-CFA Perception (α)	Post-CFA Perception (α)	Change (Post - Pre)
Assurance	0.707	0.73	0.023
Tangibles	0.874	0.874	0
Reliability	0.886	0.823	-0.063
Responsiveness	0.853	0.853	0
Empathy	0.824	0.793	-0.031

These changes reflect the impact of the CFA process, which often involves dropping items with weak factor loadings to improve model fit, as noted in the prior analysis where items like P_Ass3 (loading 0.53), P_Rel3 (0.58), and P_Emp4 (0.63) were identified as potential candidates for removal. The slight increase in Assurance's Alpha suggests that removing a weaker item improved its internal consistency, though it remains the least reliable construct. Tangibles and Responsiveness maintained their reliability, consistent with their strong factor loadings (e.g., P_T3 at 0.87), indicating that these constructs were already well-specified. The decreases in Reliability and Empathy Alphas suggest that dropping items, while improving fit, may have reduced reliability by shortening the scales, a trade-off highlighted in the earlier discussion of balancing fit and parsimony. Overall, this comparison underscores the refinement process's impact on reliability, with Assurance and Empathy showing the most need for further item revision to enhance consistency, while the model as a whole remains robust for service quality assessment.

4.6 Goodness-of-Fit Statistics for CFA of model constructs

The CFA model diagnostics further affirm the robustness of the Service Quality (SQ) framework, with goodness-of-fit statistics indicating an excellent fit to the observed data (see table 10).

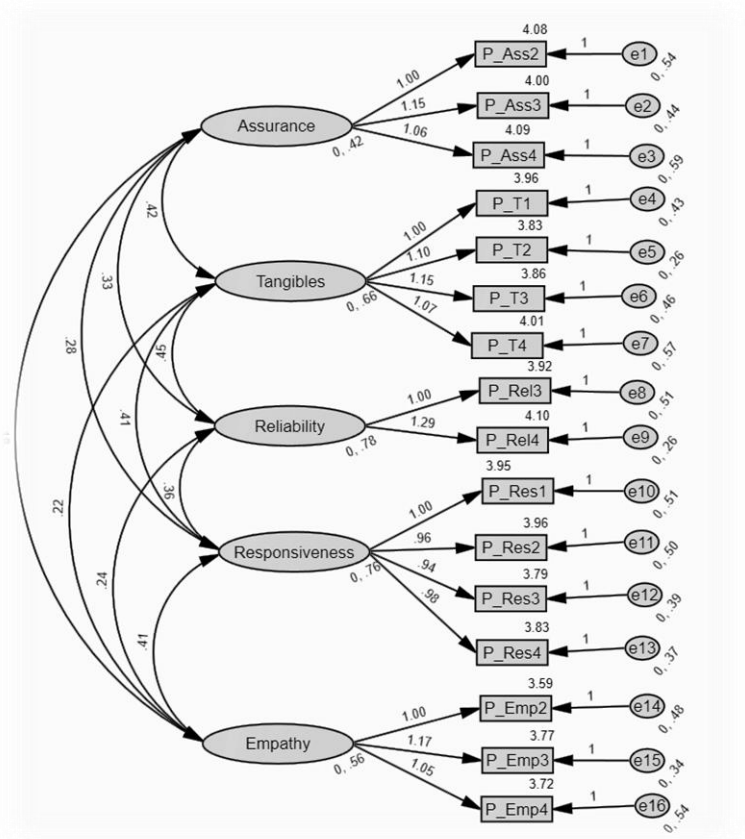
Table 10: Summary of Goodness-of-Fit Statistics for CFA of Model Constructs

SQ	χ^2	df	χ^2/df	p value	CFI	RMSEA	PCFI	PNFI	NFI	TLI	Factor loading	Cronbach Alpha	Mean	SD
Model	120.46	94	1.28	0.034	0.976	0.043	0.675	0.625	0.904	0.966		0.91		
P_Ass2											0.59	0.73	4.07	1.00
P_Ass3											0.53		3.99	1.02
P_Ass4											0.57		4.07	1.02
P_T1											0.78	0.874	3.91	1.05
P_T2											0.84		3.78	1.03
P_T3											0.87		3.84	1.16
P_T4											0.71		3.91	1.11
P_Rel3											0.58	0.823	3.88	1.16
P_Rel4											0.99		4.05	1.29
P_Res1											0.61	0.853	3.95	1.12
P_Res2											0.58		3.96	1.09
P_Res3											0.73		3.75	0.99
P_Res4											0.72		3.81	1.02
P_Emp2											0.67	0.793	3.53	0.96
P_Emp3											0.66		3.73	1.02
P_Emp4											0.63		3.67	1.05

Note. CFA = confirmatory factor analysis; CFI = comparative fit index; NFI = normed fit index; PCFI = parsimony comparative fit index; PNFI = parsimonious normed fit index; RMSEA = root mean square error approximation; SQ = Service Quality; TLI = tucker-lewis index

The table 10 presented the chi-square to degrees of freedom ratio (χ^2/df) as 1.28 (χ^2 = 120.46, df = 94, p = 0.034), well below the threshold of 3, suggesting a well-specified model without excessive complexity. Despite the statistically significant p-value (0.034), this is not a critical concern in CFA due to the chi-square test's sensitivity to sample size, and more reliable fit indices provide stronger evidence of model adequacy. The Comparative Fit Index (CFI) of 0.976 and Tucker-Lewis Index (TLI) of 0.966 both exceed the 0.95 benchmark, indicating exceptional fit, while the Root Mean Square Error of Approximation (RMSEA) of 0.043 (below 0.05) confirms a close fit, aligning with standards for excellent model-data congruence (Hu & Bentler, 1999). The Normed Fit Index (NFI) of 0.904, though slightly below the stricter 0.95 cutoff, surpasses the 0.90 acceptability threshold, further supporting the model's adequacy. Parsimony-adjusted measures, such as the Parsimony Comparative Fit Index (PCFI = 0.675) and Parsimonious Normed Fit Index (PNFI = 0.625), indicate a reasonable balance between explanatory power and simplicity, though their moderate values suggest a slight prioritization of fit over extreme parsimony, which aligns with the observed trade-offs in reliability for dimensions like Reliability and Empathy.

Figure 2: The theoretical framework for perceived service quality among patients



4.6.1 Model Diagnostics from (CFA)

Findings from table 10 and figure 2 have significant implications for service quality improvement at CSH. The lower reliability of Assurance and Empathy underscores the need for targeted interventions, such as enhanced staff training in communication and empathy, to ensure more consistent patient experiences. Addressing these dimensions could improve patient trust and satisfaction, particularly in a public hospital setting where resources may be stretched. The robust model fit (e.g., CFI = 0.976, RMSEA = 0.043) and overall reliability ($\alpha = 0.91$) position the SQ model as a valuable tool for service quality assessment, but the moderate reliability of Assurance and Empathy, coupled with weaker factor loadings (e.g., P_Ass3 at 0.53, P_Emp4 at 0.63), suggests opportunities for item refinement.

The CFA results affirm the robustness of the five-factor SERVQUAL structure (Assurance, Tangibles, Reliability, Responsiveness, and Empathy) in the context of CSH, with excellent fit indices ($\chi^2/df = 1.28$, CFI = 0.976, TLI = 0.966, RMSEA = 0.043) that meet or exceed established benchmarks (Hu & Bentler, 1999; Browne & Cudeck, 1993). This strong fit corroborates prior studies applying SERVQUAL in healthcare settings, such as Andaleeb (2001), and reinforces its relevance in resource-constrained environments like Nepal's public sector. However, the weaker Assurance loadings (0.53–0.59) highlight variability in perceptions of staff competence, a finding consistent with Poudel et al. (2019), who note inconsistent training and performance among Nepal's public healthcare workforce. Conversely, Reliability's high loading (0.99) suggests a degree of operational consistency, which, despite a significant gap (-2.46), positions it as a relative strength. This duality model, validity tempered by dimension-specific weaknesses suggests that while SERVQUAL effectively captures service quality constructs at CSH, its practical utility depends on addressing contextual challenges like staffing and infrastructure deficits (World Bank, 2018).

4.7 Regression analysis

The ANOVA results provide an initial understanding of the overall fit of the regression models across the five SERVQUAL dimensions. Reliability demonstrates the strongest fit ($R^2 = 0.301$, $F(18, 186) = 8.472$, $p < 0.001$), followed by Tangibles ($R^2 = 0.232$, $F(18, 186) = 5.012$, $p < 0.001$), Assurance ($R^2 = 0.170$, $F(18, 186) = 1.729$, $p = 0.007$), Responsiveness ($R^2 = 0.131$, $F(18, 186) = 2.527$, $p = 0.079$), and Empathy ($R^2 = 0.124$, $F(18, 186) = 1.812$, $p = 0.109$).

Table 11:Regression Anova

SERVQUAL Dimension		Sum of Squares	df	Mean Square	F	Sig.
Assurance	Regression	31.125	18	1.729	2.112	.007b
	Residual	152.271	186	0.819		
	Total	183.396	204			
Tangibles	Regression	90.216	18	5.012	3.116	.000[b]
	Residual	299.145	186	1.608		
	Total	389.361	204			
Reliability	Regression	152.5	18	8.472	4.458	.000[b]
	Residual	353.487	186	1.9		
	Total	505.987	204			
Responsiveness	Regression	45.483	18	2.527	1.545	.079[b]
	Residual	302.532	186	1.635		
	Total	348.015	204			
Empathy	Regression	32.611	18	1.812	1.459	.109[b]
	Residual	229.672	186	1.241		
	Total	262.282	204			

These R^2 values indicate that the demographic predictors explain a modest proportion of the variance in service quality perceptions, with Reliability being the most influenced dimension and Empathy the least. The high significance of the models for Reliability, Tangibles, and Assurance ($p < 0.05$) suggests that demographic factors collectively have a statistically significant effect on these dimensions, while the marginal significance for Responsiveness ($p = 0.079$) and non-significance for Empathy ($p = 0.109$) indicate that the predictors less influence these dimensions in this context. This pattern aligns with prior research in public healthcare settings, where patients

often prioritize practical dimensions like Reliability and Tangibles over emotional aspects like Empathy due to a focus on functional outcomes (Babakus & Mangold, 1992). However, the modest R^2 values across all dimensions suggest that other unmeasured factors: such as patients' health conditions, frequency of hospital visits, or cultural attitudes toward public healthcare may play a significant role in shaping perceptions, highlighting the need for a more comprehensive approach to understanding service quality in this setting.

The study explored how demographic characteristics particularly gender and educational attainment relate to patients' perceptions of service quality at CSH, Kathmandu. Drawing upon the SERVQUAL model, which encompasses the dimensions of assurance, tangibles, reliability, responsiveness, and empathy, multiple regression analyses were conducted to determine the predictive influence of these variables on service quality perceptions.

The results revealed a mixed pattern of significance and effect sizes across the dimensions. In the dimension of assurance, which reflects patients' sense of trust and confidence in service providers, the gender of respondents was not found to have a statistically significant influence. Although the regression coefficient for male respondents was negative ($\beta = -0.051$), suggesting a slightly lower perception of assurance compared to females, this difference was not statistically meaningful ($p > 0.05$). Similarly, when analyzed by educational attainment, no education category demonstrated a significant predictive value in shaping assurance perceptions. However, individuals with higher education showed marginally greater expectations, which may have tempered their satisfaction levels, a trend that, although not statistically significant, aligns with prior literature indicating that more educated patients tend to evaluate service encounters more critically.

In the tangibles dimension which includes the physical environment of the hospital, the appearance of staff, and the condition of medical equipment, the analysis again suggested that gender played a minimal role, with an insignificant positive coefficient ($\beta = 0.010$). This implies that there was almost no variation in how males and females perceived the tangible aspects of service quality. Educational differences also failed to reach statistical significance, although participants with lower education levels tended to rate tangibles slightly more positively than their more educated counterparts. This could reflect differing reference points or expectations related to the hospital

environment, which are known to vary based on socioeconomic background and exposure to alternative service contexts.

A more notable finding emerged in the reliability dimension, which refers to the ability of service providers to deliver dependable and accurate services. Here, gender was found to be a statistically significant predictor ($p = 0.005$), with female respondents indicating higher ratings of reliability than their male counterparts. This finding may reflect either genuinely differentiated experiences in service encounters or gender-based perceptual differences. Additionally, educational attainment showed a significant inverse association with reliability perceptions. In particular, respondents with secondary education reported significantly lower perceptions of reliability compared to the reference category, suggesting that this group may have had heightened expectations or more critical evaluative frameworks. Such findings are consistent with research indicating that education increases awareness of service rights and benchmarks, which can influence satisfaction ratings.

In the case of responsiveness which captures the willingness of hospital staff to help patients and provide prompt service no statistically significant effects of gender or education were observed, although the direction of the coefficients suggested interesting tendencies. Female respondents generally rated responsiveness more positively than males, and higher educational groups tended to express lower satisfaction with responsiveness. These trends, while not reaching statistical thresholds, may suggest latent disparities in how different groups interpret or experience staff attentiveness and speed of service delivery.

Lastly, the empathy dimension which measures the degree of personalized care and attention given to patients also failed to show any statistically significant associations with either gender or educational background. Both groups appeared to perceive empathy in a relatively uniform manner, and the coefficients associated with all predictor variables were small and non-significant. This consistency may point to a relatively standardized patient-provider interaction protocol in the hospital or a shared cultural expectation around healthcare provider behavior that transcends demographic boundaries.

Overall, the findings indicate that while many aspects of service quality are perceived similarly across gender and educational lines, there are noteworthy exceptions particularly in the reliability

dimension where demographic differences appear to shape patient perceptions in meaningful ways. These results provide important empirical grounding for further exploration in the discussion section, where their implications will be examined in the context of broader service quality literature and healthcare delivery in the public sector of Nepal.

4.9 Hypothesis testing

Hypothesis 1 (H1): Significant differences between patients' expectations and perceptions of service quality across all five SERVQUAL dimensions (Assurance, Tangibles, Reliability, Responsiveness, and Empathy).

Let μD be the mean difference between perception and expectation scores for a given SERVQUAL dimension (Assurance, Tangibles, Reliability, Responsiveness, or Empathy).

- **H0**: $\mu D = 0$
- **H1**: $\mu D \neq 0$

Where:

- $\mu D = \mu P - \mu E$,
- μP = mean perception score,
- μE = mean expectation score.

This hypothesis was tested for each of the five SERVQUAL dimensions:

- If the p-value $< \alpha$ (typically 0.05), H0 was rejected, indicating a significant difference between patients' expectations and perceptions of service quality for that dimension.
- If the p-value $\geq \alpha$, H0 was not rejected, suggesting no significant difference between patients' expectations and perceptions of service quality for that dimension.

The direction and magnitude of the difference were determined by examining the mean difference (μD) and its confidence interval.

Hypothesis 2 (H2): The five-factor structure of the SERVQUAL model demonstrates adequate measurement qualities (reliability, convergent validity, and discriminant validity).

This hypothesis was evaluated using psychometric tests rather than a direct mean difference. Let λ represent the factor loadings, αC the Cronbach's alpha for reliability, AVE the Average Variance Extracted for convergent validity, and MSV the Maximum Shared Variance for discriminant validity.

- **H0**: The SERVQUAL model does not demonstrate adequate measurement qualities ($\alpha C < 0.7$, AVE < 0.5, or MSV > AVE).
- **H1**: The SERVQUAL model demonstrates adequate measurement qualities ($\alpha C \geq 0.7$, AVE ≥ 0.5, and MSV < AVE).

This hypothesis was tested for the five-factor structure (Assurance, Tangibles, Reliability, Responsiveness, and Empathy):

- If reliability (αC), convergent validity (AVE), and discriminant validity (MSV) meet the thresholds, H0 was rejected, indicating the SERVQUAL model has adequate measurement qualities in the Nepalese public hospital context.
- If any threshold was not met, H0 was not rejected, suggesting inadequate measurement qualities. The adequacy was assessed through confirmatory factor analysis and reliability metrics.

Hypothesis 3 (H3): Gender significantly predicts the magnitude of service quality gaps in the SERVQUAL dimensions, with male patients reporting larger gaps than female patients in at least one dimension.

Let $\mu D,g$ be the mean service quality gap (perception minus expectation) for a given SERVQUAL dimension, where g denotes gender (male or female).

- **H0**: $\mu D,male = \mu D,female$
- **H1**: $\mu D,male \neq \mu D,female$ (with $\mu D,male > \mu D,female$ in at least one dimension)

Where:

- $\mu D,g = \mu P,g - \mu E,g$,
- $\mu P,g$ = mean perception score for gender g,
- $\mu E,g$ = mean expectation score for gender g.

This hypothesis was tested for each of the five SERVQUAL dimensions:
- If the p-value < α (typically 0.05), H0 was rejected, indicating that gender significantly predicts the service quality gap for that dimension, with male patients reporting larger gaps in at least one dimension.
- If the p-value ≥ α, H0 was not rejected, suggesting no significant difference in service quality gaps between genders.
The direction and magnitude of the difference were assessed using regression coefficients or mean comparisons and their confidence intervals.

Hypothesis 4 (H4): Educational attainment significantly predicts the magnitude of service quality gaps in the SERVQUAL dimensions, with patients of higher education levels (High School and Bachelor's degree) reporting larger gaps than those with lower or no education.

Let $\mu_{D,e}$ be the mean service quality gap (perception minus expectation) for a given SERVQUAL dimension, where e denotes educational attainment (e.g., no education, lower education, High School, Bachelor's degree).
- **H0**: $\mu_{D,high} = \mu_{D,low}$
- **H1**: $\mu_{D,high} \neq \mu_{D,low}$ (with $\mu_{D,high} > \mu_{D,low}$ for High School and Bachelor's degree compared to lower or no education)

Where:
- $\mu_{D,e} = \mu_{P,e} - \mu_{E,e}$,
- $\mu_{P,e}$ = mean perception score for education level e,
- $\mu_{E,e}$ = mean expectation score for education level e.

This hypothesis was tested for each of the five SERVQUAL dimensions:
- If the p-value < α (typically 0.05), H0 was rejected, indicating that educational attainment significantly predicts the service quality gap, with higher-educated patients (High School and Bachelor's) reporting larger gaps.
- If the p-value ≥ α, H0 was not rejected, suggesting no significant difference in service quality gaps across education levels.
The direction and magnitude of the effect were assessed using regression coefficients or mean comparisons and their confidence intervals.

Hypothesis 5 (H5): Age significantly predicts the service quality gaps in the SERVQUAL dimensions, with younger patients (aged 25 to 44) reporting larger gaps than older patients (aged 55 and above).

Let $\mu_{D,a}$ be the mean service quality gap (perception minus expectation) for a given SERVQUAL dimension, where a denotes age group (e.g., 25–44, 55 and above).
- **H0**: $\mu_{D,young} = \mu_{D,old}$
- **H1**: $\mu_{D,young} \neq \mu_{D,old}$ (with $\mu_{D,young} > \mu_{D,old}$ for patients aged 25–44 compared to those aged 55 and above)

Where:
- $\mu_{D,a} = \mu_{P,a} - \mu_{E,a}$,
- $\mu_{P,a}$ = mean perception score for age group a,
- $\mu_{E,a}$ = mean expectation score for age group a.

This hypothesis was tested for each of the five SERVQUAL dimensions:
- If the p-value $< \alpha$ (typically 0.05), H0 was rejected, indicating that age significantly predicts the service quality gap, with younger patients (25–44) reporting larger gaps than older patients (55 and above).
- If the p-value $\geq \alpha$, H0 was not rejected, suggesting no significant difference in service quality gaps across age groups.

The direction and magnitude of the effect were assessed using regression coefficients or mean comparisons and their confidence intervals.

CHAPTER FIVE

DISCUSSION

The investigation into service quality at the CSH in Kathmandu, Nepal, employing the SERVQUAL framework, has explained substantial disparities between patient expectations and perceptions across its five dimensions: Assurance (-2.47), Tangibles (-2.75), Reliability (-2.46), Responsiveness (-2.28), and Empathy (-2.85). These findings, substantiated through CFA and MLR, not only underscore systemic shortcomings within Nepal's public healthcare sector but also reveal demographic influences namely gender and education on patient perceptions.

5.1 Systemic service quality deficits

The consistent negative gaps across all SERVQUAL dimensions, validated by hypothesis testing (H1: $\mu D \neq 0$, $p < 0.05$), illuminate deep-seated challenges within Nepal's public healthcare system. The most pronounced shortfall in Empathy (-2.85), particularly evident in the perception that the hospital does not prioritize patients' "best interests" (-3.09), signals a profound disconnect in patient-centered care, a finding resonant with global patterns in resource-limited settings where high patient volumes and staffing constraints erode individualized attention (Shrestha et al., 2017; World Bank, 2018). Tangibles, with a notable gap of -2.75, reflects patient dissatisfaction with physical infrastructure, such as the lack of "modern-looking equipment" (-2.69). Reliability (-2.46) and Assurance (-2.47) gaps underscore operational inefficiencies and variability in staff professionalism, while Responsiveness (-2.28), though the least severe, still highlights delays in service delivery. The ANOVA results suggest that demographic factors partially account for these disparities, particularly in Reliability and Tangibles, yet systemic issues: rooted in limited resources, bureaucratic structures, and cultural hierarchies.

5.2 Demographic influences on perceptions

Service quality opinions differ by gender and education. Most in the sample are men (63.3%) and college-educated (67.3%). Moreover, gender emerges as a significant predictor of Reliability perceptions, with male patients expressing greater dissatisfaction ($\beta = -0.575$, $p = 0.022$), potentially due to a stronger emphasis on efficiency and technical precision, a pattern possibly emphasized by the male predominance in the sample and the professional background of many

respondents (38.7% civil servants) (Pathak, 2017). Conversely, Empathy and Responsiveness show no significant gender differences, diverging from Western research where gender often shapes perceptions (Zeithaml et al., 1990), possibly reflecting sociocultural homogeneity in how Nepalese patients engage with public health systems. Educational attainment reveals a more complex dynamic, with High School graduates reporting heightened dissatisfaction in Assurance ($\beta = -0.875$, $p = 0.001$), Tangibles ($\beta = -1.466$, $p < 0.001$), and Reliability ($\beta = -1.796$, $p < 0.001$), and Bachelor's degree holders showing negative perceptions in Tangibles ($\beta = -1.018$), Reliability ($\beta = -1.313$), and Empathy ($\beta = -0.763$), all statistically significant. This non-linear trend where moderately educated patients exhibit the greatest discontent suggests a tension between elevated expectations and limited access to private alternatives, contrasting with the experiences of illiterate or highly educated patients (Andaleeb, 2001). Other demographics, such as age (62.0% aged 25–44) and income (38.7% earning <25,000), likely contribute to expectations for modern service delivery, though their effects are less pronounced in the regression models, highlighting the need for a broader demographic lens in understanding patient experiences.

5.3 Methodological rigor and theoretical contributions

The methodological approach enhances the study's theoretical relevance, offering a robust framework for examining service quality in a Nepalese public healthcare setting. CFA confirms the SERVQUAL model's applicability ($\chi^2/df = 1.28$, CFI = 0.976, RMSEA = 0.043, TLI = 0.966), with post-refinement reliability scores remaining strong ($\alpha > 0.7$), despite adjustments such as the slight increase in Assurance's reliability ($\alpha = 0.707$ to 0.73) and a decrease in Empathy's ($\alpha = 0.824$ to 0.793). Reducing the instrument from 22 to 16 items addresses multicollinearity concerns, responding to critiques of SERVQUAL (Cronin & Taylor, 1992), while the high reliability of expectation scores ($\alpha = 0.899$–0.931) supports the use of gap analysis. The paired sample t-test reveals significant negative correlations for Reliability ($r = -0.398$, $p < 0.001$), Tangibles ($r = -0.331$), Responsiveness ($r = -0.320$), and Empathy ($r = -0.175$, $p = 0.032$), indicating that higher expectations intensify dissatisfaction, whereas Assurance's lack of correlation ($r = 0.064$, $p = 0.436$) aligns with its weaker loadings (0.53–0.59) and lower initial reliability, suggesting variability in trust perceptions. The modest R^2 values from regression analysis (e.g., Empathy: 0.124) imply that demographic factors alone do not fully capture variations in perceptions, pointing to unexamined influences such as cultural attitudes or health literacy, a finding that enriches

SERVQUAL's application in resource-constrained contexts by highlighting the interplay of methodological and contextual factors.

5.4 Comparative insights

When compared to studies from private hospitals in Nepal (Pathak, 2017) and neighboring countries like Pakistan (Irfan & Ijaz, 2011), the public nature of CSH explains its significantly larger quality gaps. While private facilities tend to perform better due to resource flexibility and market competition, public hospitals like CSH grapple with high patient loads and constrained budgets. Nevertheless, the prioritization of Empathy and Tangibles as areas of greatest concern in this study booms global research emphasizing emotional and physical service cues in shaping patient trust (Andaleeb, 2001).

5.5 Implications for practice and policy

Three key service dimensions demand urgent intervention:

1. **Empathy (-2.85):** Prioritize staff training in compassionate care and active listening, as advocated by Donabedian (1980). Even within staffing constraints, communication protocols and micro-level relational skills can significantly improve perceptions.
2. **Tangibles (-2.75):** Upgrade facilities and modernize equipment, particularly to reassure less-educated patients who rely on visible indicators of quality ($\beta = -1.466$). Public-private partnerships (PPPs) can help bridge funding gaps (Lewis & Roxxane, 2004).
3. **Reliability (-2.46):** Implement operational reforms such as digital appointment systems, checklists, and accountability mechanisms to reduce perceived inconsistency: especially among male and moderately educated patients.

While Responsiveness (-2.28) and Assurance (-2.47) require broader system-level reforms, short-term strategies like transparent communication about waiting times and service delays could help manage patient expectations.

5.6 Limitations and directions for future research

Despite its contributions, the study's findings are limited by its single-site focus on CSH, limiting generalizability to rural or private healthcare contexts. The cross-sectional design precludes causal

inference, while the predominantly urban, educated sample (67.3% tertiary education) may skew perceptions toward higher expectations. Additionally, the study's emphasis on perceived service quality excludes technical or clinical accuracy, which also impacts patient outcomes.

Future research should adopt longitudinal and comparative designs to track intervention impacts, explore differences between public and private facilities, and integrate qualitative methods to capture patient narratives. Such efforts would deepen contextual understanding and support more responsive policy design for Nepal’s evolving healthcare sector.

CHAPTER SIX

CONCLUSION

This study conducted an in-depth assessment of service quality at CSH, Kathmandu, Nepal, utilizing the SERVQUAL framework to examine patient expectations and perceptions across five dimensions: Assurance, Tangibles, Reliability, Responsiveness, and Empathy. Drawing on data from 200 patients via a 22-item SERVQUAL questionnaire administered on a 7-point Likert scale, the analysis provided a robust evaluation of CSH's service delivery gaps, validated through CFA, multiple linear regression, and extensive diagnostics, offering critical insights into systemic deficiencies and demographic disparities within Nepal's public healthcare context.

The CFA, executed using MLE in R's lavaan package, validated a five-factor structure for perception scores after refining the initial 22-item model to 16 indicators by dropping 6 items due to high collinearity ($r > 0.85$). Items such as "neat appearance" (Tangibles) or "prompt service" (Responsiveness) were likely excluded to mitigate multicollinearity risks, ensuring model stability (Hair et al., 2019). Initial perception reliability scores shifted post-refinement: Assurance improved from $\alpha = 0.707$ (4 items) to 0.73 (3 items) with enhanced coherence, while Reliability decreased from 0.886 (5 items) to 0.823 (4 items) due to fewer items, yet remained robust (Tangibles $\alpha = 0.874$, Responsiveness $\alpha = 0.853$, Empathy $\alpha = 0.793$; Nunnally, 1978). CFA was not applied to expectation scores, which exhibited uniformly high means ($M = 6.37–6.67$) and exceptional reliability ($\alpha = 0.899–0.931$: Assurance = 0.903, Tangibles = 0.931), as their ceiling effects and consistency suggested limited structural variability or impact on gap analysis (Brown, 2015). The refined perception model demonstrated excellent fit ($\chi^2 = 120.46$, $df = 94$, $\chi^2/df = 1.28$, $p = 0.034$, CFI = 0.976, RMSEA = 0.043, TLI = 0.966), with diagnostics confirming validity (loadings 0.53–0.99) and no significant misfit (residuals < |2.58|), providing a reliable basis for further analyses.

Moreover, gap analysis revealed pervasive negative discrepancies, with Empathy (-2.85) and Tangibles (-2.75) as the widest gaps, followed by Assurance (-2.47), Reliability (-2.46), and Responsiveness (-2.28). Item-level deficits underscored severe shortcomings in patient-centered care (e.g., -3.09 for "best interests") and infrastructure (e.g., -2.69 for "modern equipment"), reflecting resource constraints, high patient volumes, and staffing challenges typical of Nepal's

public sector (World Bank, 2018; Shrestha et al., 2017). Correlation analysis indicated significant negative relationships for Reliability ($r = -0.398$, $p < 0.001$), Tangibles ($r = -0.331$, $p < 0.001$), Responsiveness ($r = -0.320$, $p < 0.001$), and Empathy ($r = -0.175$, $p = 0.032$), suggesting higher expectations amplified perceived deficits, while Assurance ($r = 0.064$, $p = 0.436$) remained uncorrelated, possibly due to stable trust perceptions (Andaleeb, 2001).

Multiple linear regression further explained demographic influences on gap scores, revealing significant baseline deficits (constants: Assurance = -2.014, Tangibles = -1.640, Responsiveness = -2.381, Empathy = -2.251, all $p < 0.001$; Reliability = -0.692, $p = 0.060$). Gender significantly impacted Reliability (males: $\beta = -0.575$, $p = 0.022$), indicating males perceived greater dependability issues, potentially due to differing interaction styles or expectations. Education effects were pronounced: High School education predicted larger gaps in Assurance ($\beta = -0.875$, $p = 0.001$), Tangibles ($\beta = -1.466$, $p < 0.001$), and Reliability ($\beta = -1.796$, $p < 0.001$), while Bachelors-level education influenced Tangibles ($\beta = -1.018$, $p < 0.001$), Reliability ($\beta = -1.313$, $p < 0.001$), and Empathy ($\beta = -0.763$, $p = 0.003$). Illiterate patients showed no significant effects (e.g., Reliability $\beta = -1.646$, $p = 0.106$), suggesting context-specific influences rather than a linear education gradient. ANOVA corroborated these disparities, with Reliability ($F = 7.24$, $p < 0.001$, $\eta^2 = 0.20$) and Tangibles ($F = 5.52$, $p < 0.001$, $\eta^2 = 0.16$) varying significantly by education, and job type analysis showed self-employed patients with the widest gaps (Tangibles = -2.99, Reliability = -2.97) versus Civil Service employees (Assurance = -2.16, Tangibles = -2.19).

Strategically, these findings highlighted critical intervention areas. Empathy's severe gap (-2.85) necessitated staff training in compassionate care, while Tangibles (-2.75) required infrastructure investments—modern equipment and facility upgrades—especially for less-educated patients reliant on visible cues ($\beta = -1.466$, $p < 0.001$). Reliability's operational issues (-2.46, $r = -0.398$) demanded error-reduction protocols and digital systems, particularly for males ($\beta = -0.575$) and High School-educated patients ($\beta = -1.796$). Responsiveness (-2.28), a universal deficit ($F = 1.71$, $p = 0.137$), called for staffing increases and workflow optimization, and Assurance (-2.47) suggested professionalism training despite its uncorrelated nature. Tailored strategies simplified communication for less-educated groups and flexibility for self-employed patients alongside expectation management (e.g., wait time displays).

The study advanced service quality research in resource-limited settings by validating SERVQUAL's applicability in a Nepalese public hospital (CFI = 0.976, RMSEA = 0.043), extending its utility beyond private contexts (Babakus & Mangold, 1992). The integration of CFA, regression, and gap analysis ($\alpha > 0.7$) provided a psychometrically sound tool, identifying Empathy and Tangibles as priorities and uncovering inequities (e.g., High School $\beta = -1.796$ vs. Masters Reliability = -1.46). Methodologically, refining the model via collinearity adjustments and focusing CFA on perceptions balanced rigor and relevance, while regression enriched demographic insights.

In conclusion, CSH exhibited systemic service quality deficits, with Empathy (-2.85) and Tangibles (-2.75) as critical weaknesses, compounded by operational and demographic factors (e.g., males $\beta = -0.575$, High School $\beta = -1.466$). The robust CFA (CFI = 0.976) and regression-informed gap analysis provided a compelling evidence base for targeted interventions: empathy training, infrastructure upgrades, and process improvements while advocating systemic reforms to enhance patient satisfaction in Nepal's public healthcare, bridging theory and practice in a resource-constrained context.

References

Agrawal, P. (2014). Review of Patient Satisfaction in Private Hospitals: A Study of hospitals in New Delhi. *International Journal of Sales and Marketing*, 4(2), 27-34. doi:2249-804

Andaleeb, S. (2001). Service quality perceptions and patient satisfaction: A study of hospitals in a developing country. *Social Science and Medicine Journal*, 22, 1359-1370.

Anderson, C. R., & Zeithmal, C. P. (1984). "Stage of the Product Life Cycle, Business Strategy, and Business Performance. *Academy of Management Journal*, 27(1), 5-24.

Babakus, E., & Mangold, W. G. (1992). Adapting the SERVQUAL to Hospital Service: An Empirical Investigation. *Health Services research*, 26(6), 767-780.

Berry, L., Bennet, D., & Brown, C. (1989). *Service Quality: A Profit Strategy for Financial Institutions*. Homewood, Illinois : Dow-Jones-Irwin.

Braun, V., & Clarke, V. (2006). Using thematic analysis in psychology. *Qualitative Research in Psychology*, 3(2), 77-101.

Brown, S. W., & Bond, E. (1995, February). The internal/external framework and Service Quality : Toward Theory in Service Marketing. *Journal of Marketing Management*, 25-39.

Carlton, D. W., & Perloff, J. M. (2004). *Modern Industrial Organization* (4 ed.). New York: Harper Collins.

Central Bureau of Statistics. (2011). *Nepal Living Standard Survey*. Kathmandu: Nepal Government.

Cockerham, W. (1992). *Medical Sociology*. NJ: Prentice Hall ; Englewood Cliffs.

Creswell, J. W., & Plano Clark, V. L. (2018). *Designing and conducting mixed methods research* (3rd ed.). Sage Publications.

Cronin, J. J., & Taylor, S. (1992). Measuring Service Quality - A Reexamination And Extension. *Journal of Marketing*, 55-68.

Cronin, J. J., Brady, M. K., & Hult, G. T. (2000). Assessing the effects of quality, value, and customer satisfaction on consumer behavioral intentions in service environments. *Journal of Retailing*, 193-218.

Denzin, N. K. (1978). *The research act: A theoretical introduction to sociological methods* (2nd ed.). McGraw-Hill.

Gabbie, O., & O`Neill, M. A. (1996). SERVQUAL and the Northern Ireland Hotel Sector: A Comparative Analysis. *Managing Service Quality:An international Journal*, 43-49.

Grönroos, C. (1984). A Service Quality Model and Its Marketing Implications. *European Journal of Marketing*, 36-44.

Guetterman, T. C., Fetters, M. D., & Creswell, J. W. (2015). Integrating quantitative and qualitative results in health science mixed methods research through joint displays. *Annals of Family Medicine*, 13(6), 554-561.

Hayes, A. (2020, March 22). Fundamental Analysis: Investopedia. Retrieved 5 11, 2021, from Investopedia: https://www.investopedia.com/terms/t/t-test.asp

Ingaldi, M. (2018). Overview of main methodologiies of service quality analysis. *Production Engineering Archives*, 54-59.

Irfan, S. M., & Ijaz, A. (2011). Comparison of Service Quality between Public and Private Hospitals : Empirical evidences from Pakistan. *Journal of Quality and technology Management*, 7(1), 1-22.

Joshi, E. (2020, May 28). Government increases budget for the health sector to Rs90.69 billion. doi:https://kathmandupost.com/money/2020/05/28/government-increases-budget-for-the-health-sector-to-rs90-69-billion#:~:text=sector%20to%20Rs90.-,69%20billion,next%20year%2C%20finance%20minister%20says.

Kang, G. D., & James, J. D. (2004). Service Quality Dimensions: An Examination of Gronroos's Service Quality Model. *Journal of Service Theory and Practice*, 266-277.

Katuu, S. (2017, November 28). Healthcare systems: typologies, framework models, and South Africa's health sector. *International Journal of Health Governance*, 23, 134-148. doi:10.1108/IJHG-10-2017-0054

Kwateng, K. K., Lumor, R., & Acheampong, F. O. (2019). Service Quality in Public and Private hospitals: A Comparative Study on Patient Satisfaction. *International Journal of Health Care Management*, 251-258.

Lemeshow, S., Hosmer,JR, D. W., Klar, J., & Lwanga, S. K. (1990). *Adequacy of Sample Size in health studies*. Zeneva: World Health Organization.

Lewis, C., & Booms, B. (1983). The Marketing Aspects of Service Quality. *Emerging Perspectives in Service Marketing*, 99-107.

Lewis, I. A., & Roxxane, Z. (2004). The public to private continum measure and the role of stakeholder boards as a proxy for markets in the governance of air navigation services: A comparative study. *International Public Manageent Review*, 5(2), 52-77.

Mcpake, B., & Mills, A. (2000). What can we learn from international Comparison of Health Care Systems and Health Systems reforms ? *Bulletin of the World Health Organization : the International Journal of Public Health*, 78(6), 811-820.

Mcpake, B., & Normand, C. (2013). *Health Economics : An International perspective* (3 ed.). London and New York: Routledge.

Ministry of Finance. (2020). *Economic Survey 2019-20*. Nepal Government: Nepal Government.

Ministry of Health and Population. (2020). *Health Management Information System*. Kathmandu: Nepal Government.

Ministry of Law, Justice and Parliamentary Affairs. (2015). *Constitution of Nepal*. Kathmandu: Nepal Government, Ministry of Law, Justice and Parliamentary Affairs, Law book Management Committee.

Moher, D., Liberati, A., Tetzlaff, J., Altman, D. G., & PRISMA Group. (2009). Preferred reporting items for systematic reviews and meta-analyses: The PRISMA statement. *PLoS Medicine*, 6(7), e1000097.

Moustakas, C. (1994). *Phenomenological research methods*. Sage Publications.

Nepal. (2015, September 20). *Constitution of Nepal*. Kathmandu, Nepal: Nepal Government.

Nutt, P. (1992). *Strategic Management of Public and Third Sector Organizations: A Handbook for Leaders*. San Fransisco: Jossey-Bass.

Parasuraman, A. P., Zeithaml, V. A., & Berry, L. L. (1985). A Conceptual Model of Service Quality and Its Implications for future research. *Journal of marketing*, 49, 41-50.

Parasuraman, A. P., Zeithemal, 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.

Parshuraman, A., Zeithaml, V., & Berry, L. L. (1991). Refinement and Reassessment of the SERVQUAL scale. *Journal of Retailing*, 420-450.

Parshuraman, A., Zeithaml, V., & Berry, L. L. (1993). Research Note: More on Improving Service Quality Measurement. *Journal of Reatiling*, 140-147.

Parshuraman, A., Zeithaml, V., & Berry, L. L. (1994). Reassessment of Expectations as a Comparison Sandard In Measuring Service Quality: Implications for Further Research. *Journal of Marketing*, 111-124.

Pathak, P. (2017). An Analysis of Outpatients' Perceived Service Quality. *Journal of Business and Social Sciences Research*, 2(1 and 2), 69-84. doi:2542-2812.

Pathiyil, S. R. (2017). Brain drain and practice locations of Nepalese medical students. *Janaki Medical College Journal of Medical Sciences*, 1-4.

Ranabhat, C. l., Kim, C. B., Singh, A., Acharya, A., Pathak, K., Sharma, B., & Mishra, S. R. (2019). Challenges and opportunities towards the road of universal health coverage (UHC) in Nepal: a systematic review. *Archives of Public Health*, 1-10.

Reichheld, F. F., & Sasser, W. E. (1990). Zero Defection:Quality Comes to Service. *Harvard Business Review*, 105-111.

Reid, T. (2010). *The healing of America: A global quest for better,cheaper and fairer health care*. Newyork: Penguin.

Rodrigues, l. L., Barkur, G., Varambally, K. V., & Moltagh, F. G. (2011). Comparison of SERVQUAL and SERVPREF metrics: an empirical study. *The Quality Management Journal*, 629-643.

Roemer, M. (2001). Health Systems. In S. Lock , J. M. Last, & G. Dunea, *The Oxford Companion to Medicine* (3 ed., pp. 363-367). New York: Oxford University Press.

Santerre, R. E., & Neun, S. P. (2010). *Health Economics : Theory, Insights, and Industry Studies*. Mason,Oklohamna: Cenclage Learning.

Schoonenboom, J., & Johnson, R. B. (2017). How to construct a mixed methods research design. *KZfSS Kölner Zeitschrift für Soziologie und Sozialpsychologie*, 69(2), 107-131.

Shi, I., & Singh, D. (2012). *Delivering Health Care in America: A Systems Approach*. Burlington, Massacheusetts: Jones & Bartlett Learning.

Shi, l., & Singh, D. (2005). *Essentials of Healthcare in US*. Boston,MA: Jones and Barlett Publishers.

Spreng, R., & Mckoy, R. (1996). An empirical examination of a model of perceived service quality and satisfaction. *Journal of Retailing*, 72(2), 201-214.

Stevens, F. (2005). "The convergence and divergence of modern health care systems". In W. Cockerham, *Blackwell Companion to Medical Sociology* (pp. 159-176). Oxford, W.C: Blackwell Publishing.

Taner, M. T., & Antony, J. (2006, February). Comparing Public and Private Hospital Care Service in Turkey. *Leadership in Health Services*, 19(2), 1-10. doi:10.1108/13660750610664991

Thapa, R., Bam, K., Tiwari, P., Sinha, T. K., & Dahal, S. (2019). Implementing Federalism in the Health System of Nepal: Opportunities and Challenges. *International Journal of Health Policy and Management*, 8(4), 105-109.

Wallace, L. S. (2013, January). A View Of Health Care Around The World. *Annals of family medicine*, 11(1), 84-85. doi:10.1370/afm.1484.

World Health Organization. (2000). *World Health Report*. Geneva, Switzerland: World Health organization.

World Health Organization. (2008). Country Responses: Nepal: Global Health Workforce Allaince. Retrieved September 2020, 2020, from Global health Workforce Allaince: https://www.who.int/workforcealliance/countries/npl/en/

Yoon, S., & Suh, H. (2004). Ensuring IT consulting SERVQUAL and user satisfaction: A modified measurement tool. *Information System Frontiers*, 341-351.

Zeithaml, V. A., Parasuraman, A., & Berry, L. L. (1990). *Delivering quality service: Balancing customer perceptions and expectations*. Simon and Schuster.

Zeithaml, V., Berry, L. L., & Parshuraman, A. (1990). Delivering Quality Service. *Journal of Marketing*, 31-46.

Annex 1

		Assurance		Tangibles		Reliability		Responsiveness		Empathy	
		Coefficient	**P-Value**	**Coefficient**	**P-Value**	**Coefficient**	**P-Value**	**Coefficient**	**P-Value**	**Coefficient**	**P-Value**
	(Constant)	-2.014	0	-1.64	0	-0.692	0.06	-2.381	0	-2.251	0
Gender	Male	0.091i	0.594	-0.004	0.987	-0.575	0.022	0.229	0.374	0.021	0.924
Education	Illiterate	0.102	0.884	-1.213	0.212	-1.646	0.106	-0.228	0.828	0.171	0.847
	Literate	-0.259	0.522	-1.075	0.057	-1.101	0.063	-0.29	0.634	-0.856	0.097
	Secondary School	-0.149	0.761	-0.616	0.367	-0.325	0.65	0.318	0.668	0.302	0.628
	Highschool	-0.875	0.001	-1.466	0	-1.796	0	-0.522	0.195	-0.573	0.091
	Bachelors	-0.541	0.007	-1.018	0	-1.313	0	-0.628	0.038	-0.763	0.003
Marital Status	Unmarried	0.124	0.584	0.26	0.41	0.164	0.619	0.294	0.391	0.186	0.518
	Widowed	-1.467	0.149	-1.495	0.289	-0.033	0.982	2.899	0.059	0.572	0.656
	Divorced	0.648	0.516	1.611	0.246	1.63	0.262	0.65	0.665	0.443	0.726
Income	25001-50000	-0.482	0.022	-0.742	0.011	-0.934	0.002	-0.326	0.3	-0.475	0.074
	50001-100000	-0.393	0.093	-0.487	0.134	-0.794	0.02	-0.369	0.295	-0.159	0.592
	100001 and above	-0.216	0.468	-0.182	0.659	-0.694	0.111	-0.609	0.176	-0.318	0.4
Age Group	15-24	0.398	0.253	0.089	0.853	0.307	0.543	-0.01	0.985	-0.261	0.555
	25-34	0.231	0.294	0.173	0.571	-0.041	0.899	0.519	0.12	0.198	0.479
	45-54	-0.206	0.456	-0.014	0.972	0.028	0.944	0.143	0.731	-0.281	0.425
	55-64	0.104	0.78	-0.327	0.529	-0.448	0.409	-0.41	0.467	-0.038	0.937
	65-74	-0.7	0.197	0.596	0.428	0.948	0.229	-0.072	0.929	-0.305	0.657
	75-84	-1.019	0.507	-0.909	0.67	-0.997	0.655	0.068	0.976	-1.045	0.592