ONLINE FIRST ARTICLE
doi.org/10.36253/ijas-17401

Statistica Applicata - Italian Journal of Applied Statistics

Received: 26.03.25/ **Accepted:** 10.04.26
Published online: 16.05.26

A Statistical Framework for Measuring Contextual Familiarity in Qualitative Survey Responses

Rahim Mahmoudvand[1,2], Fatemeh Moameri[3]

[1]*Department of Statistics, Faculty of Science, Bu-Ali Sina University, Hamedan Iran*
[2]*Department of Economics and Business Sciences, University of Cagliari, Cagliari, Italy*
[3]*Department of Mathematics, Faculty of Science, Bu-Ali Sina University, Hamedan Iran*

Abstract. *Qualitative surveys aimed at evaluating various phenomena are widely conducted across diverse fields. However, many such surveys often overlook critical methodological issues that affect the reliability and validity of their results. A common challenge is respondents' limited familiarity with the subject matter, which can bias survey outcomes. This study introduces a novel approach to measure respondent familiarity in qualitative surveys by analyzing survey content to derive a familiarity index ranging from zero to one, where higher values indicate greater familiarity. Two estimators are proposed: one using a classical framework and the other a Bayesian one. The applicability and effectiveness of this approach are demonstrated through a case study assessing service quality at a public university in Iran. The results indicate a familiarity measure of approximately 0.26, revealing a significant gap in respondent familiarity and emphasizing the importance of incorporating such measures to highlight potential limitations in the validity of qualitative survey findings.*

Keywords. *Qualitative survey, Validity, Familiarity.*

1 Introduction

In this study, we focus on qualitative data collection methods, particularly semi-structured interviews, while also considering related tools such as structured interviews, focus groups, and open-ended survey questions. Throughout the paper, we use the term qualitative data collection to broadly refer to these approaches, and we specify the tool used in our empirical work where relevant. This clarification is important because the challenges and solutions we discuss apply across these qualitative methods, although our case study is based on semi-structured interview data.

Qualitative data plays a crucial role in exploring new phenomena, understanding the depth of a study, and complementing quantitative research (Karunarathna et al., 2024). However, collecting qualitative data can be time-consuming, costly, and often lacks generalizability beyond the study's specific context. Ensuring the quality of qualitative data, whether from interviews, focus groups, or open-ended survey responses, remains a fundamental challenge in research. The sources of data quality issues vary across different studies, but it is evident that many surveys are affected by such problems. Key concepts like reliability, validity, and generalizability are central to evaluating data quality in qualitative surveys (Leung, 2015). Yet, these attributes are difficult to assess in practice, as they are often challenging to measure and remain largely conceptual. Qualitative researchers have so far failed to agree on what constitutes "validity" or "quality" in their work (Garside, 2014).

Braun et al. (2021) argue against assuming a strict, clear-cut distinction between qualitative and quantitative aspects, emphasizing that the boundaries between the two are often fluid, influenced by both the nature of the data and the values of the researcher. This ambiguity underscores the importance of expertise from the outset of the study design. Without this expertise, the quality of the data can be significantly compromised. This issue is particularly relevant in academia, where students and researchers

with limited statistical knowledge may neglect to consult a statistician early in the process, seeking help only after data collection.

Another major challenge in qualitative data collection is ensuring thoughtful and accurate responses, even when respondents are appropriately selected. This is especially problematic in online settings and interviews, where respondents often rush through questions, leading to responses that suggest a lack of careful consideration. Revilla (2016) explores this issue through experiments aimed at reducing 'speeding behaviors' in web surveys. The results showed that reminders and commitment statements could slightly improve response quality for some participants. However, a large portion of respondents continued to provide low-quality answers, indicating that more substantial interventions are needed to improve the accuracy and thoughtfulness of responses.

An often underexplored but critical area of research is the impact of cognitive engagement on the validity and reliability of qualitative data. Mejia et al. (2021) highlight that in service quality assessments, measuring quality requires not only effective methodologies but also a deeper understanding of how respondents perceive and engage with survey items. Tourangeau et al. (2021) further argue that although most researchers recognize the importance of reliability, it is rarely standard practice to collect data on reliability within qualitative research. Their review stresses the need for methodologies that account for cognitive components to improve data accuracy. Similarly, Mellinger and Hanson (2020) examine issues related to validity and reliability, emphasizing the importance of proper sampling theory, item writing, and instrument adaptation. Despite existing frameworks, they argue that substantial work remains to improve the cognitive alignment between the research instrument and respondent understanding.

This paper builds upon these insights by examining the cognitive components of survey responses collected at a public university in Iran. Specifically, it introduces a novel approach to measuring respondent familiarity in qualitative surveys. This method will provide a measure alongside the analysis of the qualitative data, helping to assess the potential of the analysis and its results. In Section 2, we present the theoretical framework for introducing the measure; Section 3 outlines the design of the case study; Section 4 presents the findings from the case study; and Section 5 discusses the results and concludes the paper.

2 Theoretical framework

Service familiarity refers to the extent to which respondents in qualitative studies have been exposed to and recognize the services provided by an institution. It encompasses both direct experiences, such as using the services themselves, and indirect experiences, such as learning about the services through communication channels, word of mouth, or observation. In the context of a study, service familiarity reflects the respondents' ability to identify and acknowledge the services offered, based on their level of engagement or interaction. It does not necessarily imply a deep understanding of the services' features, quality, or alignment with the institution's mission. Instead, it serves as an indicator of respondents' baseline familiarity, shaped by their exposure to or experience with the institution. A lack of service familiarity can lead to misinformed responses and, consequently, biased evaluations. This issue is especially relevant in educational contexts, where students may lack sufficient knowledge about the university's processes or available services. As a result, their assessments may fail to accurately reflect the actual quality of the services provided.

Suppose we have conducted a qualitative study with a sample size of n and let $A_{1,j},\ldots,A_{m,j}$ represents

the answers of respondent j (for $j = 1,\ldots,n$) to the m questions about the services of the organization under study. Denote by $\theta_{1,j}$, …, $\theta_{m,j}$ the levels of service familiarity of respondent j for questions 1 to m . We posit that the $\theta_{i,j}$ values influence the interview responses. It is evident that when $\theta_{i,j}$ values are at their highest, the results are likely to be more reliable. Therefore, we suggest estimating the $\theta_{i,j}$ component and reporting the average value of this component across the study. This average can serve as an indicator of the overall service familiarity among the respondents.

In this study, we employed a straightforward approach to estimate $\theta_{i,j}$ values by evaluating responses in relation to the processes of the services provided by the organization. A panel of experts from the service provider reviewed each respondent's answers and assigned scores based on their alignment with the relevant processes. To ensure consistency and objectivity in assessing respondent' familiarity with each service, we developed a structured evaluation framework. This framework allowed the assessors to assign familiarity scores based on both the respondents' actual experience and the level of detail and accuracy in their descriptions of the service. Table 1 summarizes a typical criterion that could be used for assigning familiarity scores during the evaluation process based on a 5-scale approach.

Table 1. Criteria used by assessors to assign familiarity scores based on respondents' responses.

Response type	score
No personal experience, no knowledge or only heard from others	0
Limited use, vague or brief comment	1
Some use, basic but accurate description	2
Regular use, clear and appropriate description	3
Frequent use, detailed and accurate explanation	4

We use the notation $B_{i,j,k}$ to represents the level of service familiarity of respondent j for questions i from the point of view of expert k. Then, we employ two approaches to estimate the respondents' level of service familiarity.

2.1 Classical approach

The average of expert assessments yields a simple estimate of respondent familiarity. Table 2 illustrates the evaluation structure for respondent j as assessed by K experts. If the scores $B_{i,j,k}$ of expert k ($k = 1,\ldots,K$) range from zero to one—where zero indicates the lowest familiarity and one indicates the highest—then the overall estimate will also fall within this range. Values closer to one indicate that the survey was conducted among more aware respondents.

Table 2. Structure of data for computing the estimate of familiarity

Question	Answer	evaluation 1	evaluation 2	…	evaluation K	familiarity score
1	$A_{1,j}$	$B_{1,j,1}$	$B_{1,j,2}$	…	$B_{1,j,K}$	$\left(\frac{B_{1,j,1}+\cdots+B_{1,j,K}}{K}\right)$

2	$A_{2,j}$	$B_{2,j,1}$	$B_{2,j,2}$	…	$B_{2,j,K}$	$\left(\frac{B_{2,j,1}+\cdots+B_{2,j,K}}{K}\right)$
⋮	⋮	⋮	⋮	⋮	⋮	⋮
m	$A_{m,j}$	$B_{m,j,1}$	$B_{m,j,2}$	…	$B_{m,j,K}$	$\left(\frac{B_{m,j,1}+\cdots+B_{m,j,K}}{K}\right)$

Considering the familiarity scores in Table 2, we can compute the overall familiarity index by simple averaging as follows

$$\mathrm{FI_C} = \frac{1}{K.m.n}\sum_{i=1}^{m}\sum_{j=1}^{n}\sum_{k=1}^{K}B_{i,j,k} = \frac{1}{K.m.n}\sum_{i=1}^{m}\sum_{j=1}^{n}B_{i,j,\bullet},$$

where $B_{i,j,\bullet}$, is the sum of the scores given by all experts for question *i* and respondent j.

2.2 Bayesian approach

One practical challenge in measuring familiarity lies in the choice of scale. We recommend using simple scales such as a 5-point Likert scale. A practical option is to use a scale ranging from 0 to 4, where 0 represents the lowest level of familiarity and 4 represents the highest level. However, using a scale that includes zero can pose a problem since it is often unrealistic to assume that respondents have absolutely no awareness of the services offered by an organization. To address this issue, we suggest a Bayesian framework. In this framework, we need a prior distribution for the familiarity parameter and a likelihood function based on the observations. Here we don't want to discuss the right choices of Bayesian components as it depends on the data sets. We just give an example to show how this approach can work. Using the Bayesian framework for joint qualitative and quantitative responses is also discussed in Kang et al. (2018).

Let assume a $\mathrm{Beta}(\alpha_0, \beta_0)$ prior distribution for the familiarity parameter $\theta_{i,j}$, for any level of familiarity within the [0, 1] range. Given the 5-point Likert scale and the familiarity parameter $\theta_{i,j}$, we assume that the expert scores $B_{i,j,k}$ follow a $\mathrm{Binomial}(4, \theta_{i,j})$ distribution. Under this framework, the posterior distribution for the combined scores from K experts are given by:

$$\theta_{i,j} \left| B_{i,j,1},...,B_{i,j,K} \sim \mathrm{Beta}\left(B_{i,j,\bullet}+\alpha_0,\ 4K+\beta_0-B_{i,j,\bullet}\right).\right.$$

Using a squared-error loss function, the Bayes estimator for familiarity score is the mean of the posterior distribution, which is:

$$\hat{\theta}_{i,j} = \frac{B_{i,j,\bullet}+\alpha_0}{4K+\alpha_0+\beta_0}.$$

This estimator may have several advantages over the classical estimator given in Table 2. First, it does not produce a familiarity estimate of zero, even in cases where all experts assign a score of zero. Additionally, the structure of $\theta_{i,j}$ implies that the following inequality holds:

$$\frac{\alpha_0}{4K+\alpha_0+\beta_0} \le \hat{\theta}_{i,j} \le \frac{4K+\alpha_0}{4K+\alpha_0+\beta_0}.$$

This help avoids extreme values and provides a more realistic measure of respondent familiarity. This ensures that the evaluation reflects a more nuanced understanding of the respondent's knowledge and familiarity levels. Using this Bayesian familiarity score, we can compute the overall Bayesian familiarity index with the following formula:

$$\mathrm{FI}_B = \frac{1}{m.n}\sum_{i=1}^{m}\sum_{j=1}^{n}\hat{\theta}_{i,j}.$$

To employ this approach, we need to estimate the hyperparameters α_0, β_0. Using the previous calculations, we can derive the likelihood function as follows:

$$L(\alpha_0,\beta_0) = \prod_{i,j,k}\binom{4}{B_{i,j,k}}\frac{Be\left(B_{i,j,\bullet}+\alpha_0,\, 4K_{i,j}+\beta_0-B_{i,j,\bullet}\right)}{Be\left(\alpha_0,\,\beta_0\right)},$$

where $Be(.,.)$ is beta function and assumed that the number of evaluators varies as they may have no idea for some cases. To estimate α_0 and β_0 we can maximize this likelihood function. Although numerical methods can be employed, the VGAM package in R provides a convenient way to find the maximum likelihood estimates (MLE) of the parameters for our problem using the family function betabinomial.

The formula for $\mathrm{FI_C}$ and $\mathrm{FI_B}$ show that the following relationship holds between the two approaches:

$$\mathrm{FI}_B = \frac{K.\mathrm{FI_C}+\alpha_0}{4K+\alpha_0+\beta_0}.$$

This relationship confirms that the variance of FI_B is always less than that of $\mathrm{FI_C}$.

3 Case study

Evaluating service quality in public universities has become a critical area of research, particularly as institutions strive to enhance their offerings in response to student needs and expectations. Previous research has shown that improving service quality is a primary goal for higher education providers. However, the perspectives of students regarding service quality improvements have often been overlooked, especially in developing countries (Osman and Saputra, 2019; Bozbay *et al.*, 2020; Sohail and Hasan, 2021). Traditional survey methodologies frequently rely on student feedback to assess service quality; however, these surveys often neglect the crucial factor of respondent awareness. This oversight can significantly compromise the validity and reliability of the findings, leading to misguided conclusions about service quality.

The familiarity of respondents with the processes within a university setting is vital for any service quality survey. If students, faculty, or staff are not fully aware of or familiar with institutional processes and how different departments and services contribute to these processes, their feedback may lack depth or accuracy. Awareness involves:

- Understanding the mission of the university (e.g., providing education, fostering research, contributing to society).
- Familiarity with the processes that support these missions (e.g., how courses are managed, how student services are provided, how research is funded and supported).
- Knowledge about quality standards and what is considered good service in the context of these processes.

In the context of Iran's rapidly evolving educational landscape, with an increasing number of public universities striving to improve service quality through student feedback mechanisms, many surveys fail to consider whether respondents possess the necessary background information to make informed evaluations.

Many existing models for evaluating service quality, such as the SERVQUAL model, have been widely applied in higher education settings (Gilavand and Torabipour, 2022; Daryazadeh et al, 2022; Nooripoor and Sharifi, 2020). These models, however, often assume that respondents have a sufficient baseline understanding of the services and institutional processes being evaluated, which may not always hold true in the Iranian context. This assumption is problematic, particularly in the Iranian context, where cultural and educational factors can influence awareness levels. For example, in Iran, hierarchical structures in educational institutions and varying levels of student exposure to administrative processes can affect their understanding of service quality.

Moreover, the reliance on self-reported data from students can exacerbate the issue. Research indicates that individuals tend to provide responses based on perceptions rather than objective criteria, leading to potential bias in survey results (Tourangeau & Yan, 2007). If students are unaware of specific services, their feedback may not reflect the true quality of services, rendering the survey results less useful for institutional improvement.

This study aimed to evaluate the quality of services from the perspective of students at Bu-Ali Sina University, a public university in Iran. The university's services are organized into six main domains: the Chancellor's (Rector's) Office, Educational Affairs, Research and Technology, Administrative Affairs, Cultural and Social Affairs, and Student Affairs. Our study design ensured that all these service areas were thoroughly covered. Within the six main domains, we identified different units with which students regularly interact during their studies, and where they are capable of providing informed opinions. Table 3 summarizes the main domains of services at the university along with the most important units that were included in our evaluation. These units represent the primary points of contact between students and the institution, and together they provide a comprehensive view of the service environment assessed in this study.

Table 3. Overview of university service domains and key student-related units included in the study.

Domain	Services
Rector's office	• Communications and External Relations, • Campus Security, • Admissions and Student Records, • Quality Assurance and Academic Standards. • Disciplinary Committee
Educational Affairs	• Teaching and supervision, • Examinations, • Academic Staff and related processes, • Learning Facilities, • Student Information System.
Research and Technology Affairs	• Laboratory, • Library, • IT Center, • Research Grants.
Administrative Affairs	• Insurance,

	• Tuition accounting.
Cultural and Social Affairs	• Student Union, • Cultural and Social Activities، • Negarestan (Related web service)
Student Affairs	• Student Accommodation, • Campus Catering, • Student Financial Aid, • Campus Transport, • Student Counseling & Well-being, • University Sports Centre, • Bostan (Related web service).

Given the focus of this research, we concentrated on postgraduate students, as we believe they have a deeper understanding of the quality of services at the university compared to undergraduate students. The university has approximately 10,000 students, of whom 2,153 are at the master's level and 804 are at the Ph.D. level. These students are enrolled across 13 faculties, covering a wide range of disciplines. Of these faculties, 9 offer postgraduate programs, and these 9 faculties formed the basis for our sampling frame.

Data collection was conducted over two working weeks during the 2022–2023 academic years. Because both the faculty and level of study (MSc or Ph.D.) could contribute to heterogeneity in perceptions of service quality, we defined the combinations of these two factors (faculty × level) as strata for sampling. Within each stratum, students were randomly selected from lists provided by the university administration. Our aim was to ensure representation across the diversity of programs while controlling for potential sources of variability. In line with recommendations for qualitative research, where very large samples are typically not necessary or feasible (See for example Creswell and Poth, 2016), we determined that a total of 60 interviews would provide sufficient depth and breadth for meaningful analysis. The distribution of these 60 participants across strata was guided by proportional representation, while also ensuring at least one participant per stratum wherever feasible.

Tables 2 and 3 present the sample size and schedule of interviews, with the abbreviation F for Faculty. The faculty names include Physical Education (PE), Economics (Eco), Engineering (Eng), Veterinary Medicine (VM), Art and Architecture (Art), Agriculture (Agri), Basic Sciences (BS), Chemistry (Chem), and Humanities (Hum). Additionally, MSc and Ph.D. indicate the academic programs in which the students were enrolled. Table 4 shows the sample size for each stratum, defined by the combination of faculty and program level. Table 5 displays the schedule of interviews, where a randomized combination of faculty and level was assigned for each day. On each day, these predefined combinations determined which students were to be interviewed. In practice, if a randomly selected student from the list of candidates for a given stratum did not respond or was unavailable, the next student was randomly drawn from the same stratum list to ensure the sampling plan was followed.

Table 4. Sample size for each stratum (faculty × program level)

Level	Faculty								
	PE	Eco	Eng	VM	Art	Agri	BS	Chem	Hum
MSc	4	4	4	4	4	4	3	3	3

Ph.D.	3	3	3	3	3	3	3	3	3

Table 5. Daily randomized schedule of faculty and program level combinations for interviews

Date	F / Level	F / Level	F / Level	F / Level	F / Level
Feb 20, 2023	PE /MSc	Eco /MSc	Eng /MSc	VM /MSc	Art /MSc
Feb 21, 2023	Agri /MSc	BS /Ph.D.	Chem /Ph.D.	Hum /Ph.D.	PE /Ph.D.
Feb 22, 2023	Eco /Ph.D.	Eng /Ph.D.	VM /Ph.D.	Art /Ph.D.	Agri /Ph.D.
Feb 23, 2023	BS /MSc	Chem /MSc	Hum /MSc	PE /MSc	Eco /MSc
Feb 24, 2023	Eng /MSc	VM /MSc	Art /MSc	Agri /MSc	BS /Ph.D.
Feb 25, 2023	Chem /Ph.D.	Hum /Ph.D.	PE /Ph.D.	Eco /Ph.D.	Eng /Ph.D.
Feb 26, 2023	VM /Ph.D.	Art /Ph.D.	Agri /Ph.D.	BS /MSc	Chem /MSc
Feb 27, 2023	Hum /MSc	PE /MSc	Eco /MSc	Eng /MSc	VM /MSc
Feb 28, 2023	Art /MSc	Agri /MSc	BS /Ph.D.	Chem /Ph.D.	Hum /Ph.D.
Mar 1, 2023	PE /Ph.D.	Eco /Ph.D.	Eng /Ph.D.	VM /Ph.D.	Art /Ph.D.
Mar 2, 2023	Agri /Ph.D.	BS /MSc	Chem /MSc	Hum /MSc	PE /MSc
Mar 3, 2023	Eco /MSc	Eng /MSc	VM /MSc	Art /MSc	Agri /MSc

One notable aspect of this study is the method of sample selection, which was based on principles used in designing statistical experiments—a practice seldom applied in qualitative research. Specifically, the stratified random sampling approach helped enhance the credibility of the study by promoting diversity in the sample and satisfying assumptions of independence and homogeneity where appropriate.

The research followed a qualitative approach, with data collection carried out through semi-structured interviews. These interviews were conducted over the phone, lasting an average of 20 minutes per session to ensure that participants could express their opinions and elaborate on each topic without time pressure. However, due to the trust established during the interviews, some students engaged in extended discussions, lasting over an hour, providing detailed insights on all topics covered. The process consisted of the following stages:

1. Introduction – Brief explanation of the study purpose and assurance of confidentiality.
2. General Service awareness – Questions regarding student experiences with the identified service units (See Table 3).
3. Service evaluation – Inquiries about the perceived quality, accessibility, and responsiveness of the services.
4. Suggestions for improvement – Open-ended prompts inviting recommendations for enhancing service delivery.

To maintain the integrity of the interviews and minimize interviewer influence, we ensured semantic consistency in the core questions, while allowing participants the flexibility to elaborate on topics of interest. This consistency helped avoid any unintended biases or personal judgments being conveyed to the respondents, allowing for more authentic and reliable responses.

Additionally, where appropriate, the interviews included direct references to the specific units from which the students had received services. This approach aimed to gather targeted feedback and nuanced insights about their experiences with these units.

By encouraging participants to consider their interactions with particular units, we enhanced the depth and relevance of the data collected, aligning with the study's goal of capturing detailed, unit-specific perspectives.

To assess students' familiarity with the services provided, the authors, who are working in the university and are familiar with the services available within the university, separately evaluated the opinions and assigned scores based on students' levels of familiarity to one of the options: "not familiar at all," "somewhat familiar," "relatively familiar," "familiar," and "very familiar," corresponding to scores of 0, 1, 2, 3, and 4, respectively. These values coincide with the criteria presented in Table 1.

4 Finding

Table 6 presents the relationship between the familiarity scores assigned by two evaluators to the respondents' familiarity. The table shows the frequency distribution of the scores as percentages, which allows us to observe how consistently both evaluators rated the level of familiarity among respondents. The diagonal values in the table represent cases where the evaluators gave identical scores, totaling 67.7% of all responses. This indicates that in about two-thirds of the evaluations, the two evaluators were in full agreement regarding the respondents' level of familiarity. In addition to this, the computed correlation coefficient between the evaluators' scores is 0.628, which suggests a moderate to strong positive relationship between their assessments. This correlation is statistically significant at the 5% level (p-value<0.001), reinforcing the reliability of the evaluation process. This distribution demonstrates that, despite some variability in scores, there is strong evidence of agreement between the evaluators, both through the percentage of identical scores (67.7%) and the positive correlation coefficient. These results indicate that the evaluators were largely consistent in assessing the respondents' familiarity levels, lending credibility to the evaluation process.

Table 6. Frequency distribution of awareness scores by two evaluators (the values are percentage)

		Evaluator 1				
		0	1	2	3	4
Evaluator 2	0	11.20	3.40	0.20	0.00	0.00
	1	1.40	41.00	13.10	0.90	0.00
	2	0.30	6.60	12.50	2.30	0.00
	3	0.20	1.30	2.60	2.00	0.00
	4	0.00	0.20	0.40	0.40	0.00

On a scale from zero to one hundred percent, the overall familiarity index (FI_C) of students using the

classic approach was 26 percent, with a standard deviation of 18.2 percent. The large standard deviation suggests substantial heterogeneity in the respondents' familiarity, with some respondents possibly having very low familiarity and others much higher.

The Bayesian approach implementation requires additional computational resources. As outlined in the theoretical section, we utilize the VGAM package in R for parameter estimation. Using this package, the MLE of the hyperparameters are α_0=10.86 and β_0=27.96. These values yield the following Bayes estimator for two evaluators:

$$\hat{\theta}_{i,j} = \frac{B_{i,j,1} + B_{i,j,2} + 10.86}{46.82},$$

with an average of 26 percent and a standard deviation of 1.6 percent, indicated slightly higher familiarity among respondents.

Figure 1 presents a boxplot comparison of the familiarity index (awareness) for all responses, estimated using both the Classic and Bayesian approaches. The boxplot for the Classic method shows a wider spread in familiarity index values, with a median 0.25. The interquartile range (IQR) is relatively large, indicating greater variability among respondents. There are numerous outliers, especially at both the lower and upper ends, suggesting that some responses reflect either very low or very high familiarity. In contrast, the Bayesian approach yields a noticeably narrower distribution. The median familiarity index is 0.27 that is slightly larger than that of the Classic approach, and the IQR is much smaller, reflecting reduced variability. The Bayesian estimates also display fewer and less extreme outliers, indicating that this method produces more conservative and stable estimates of familiarity across respondents. Overall, the figure illustrates that the Bayesian approach tends to shrink extreme values toward the center.

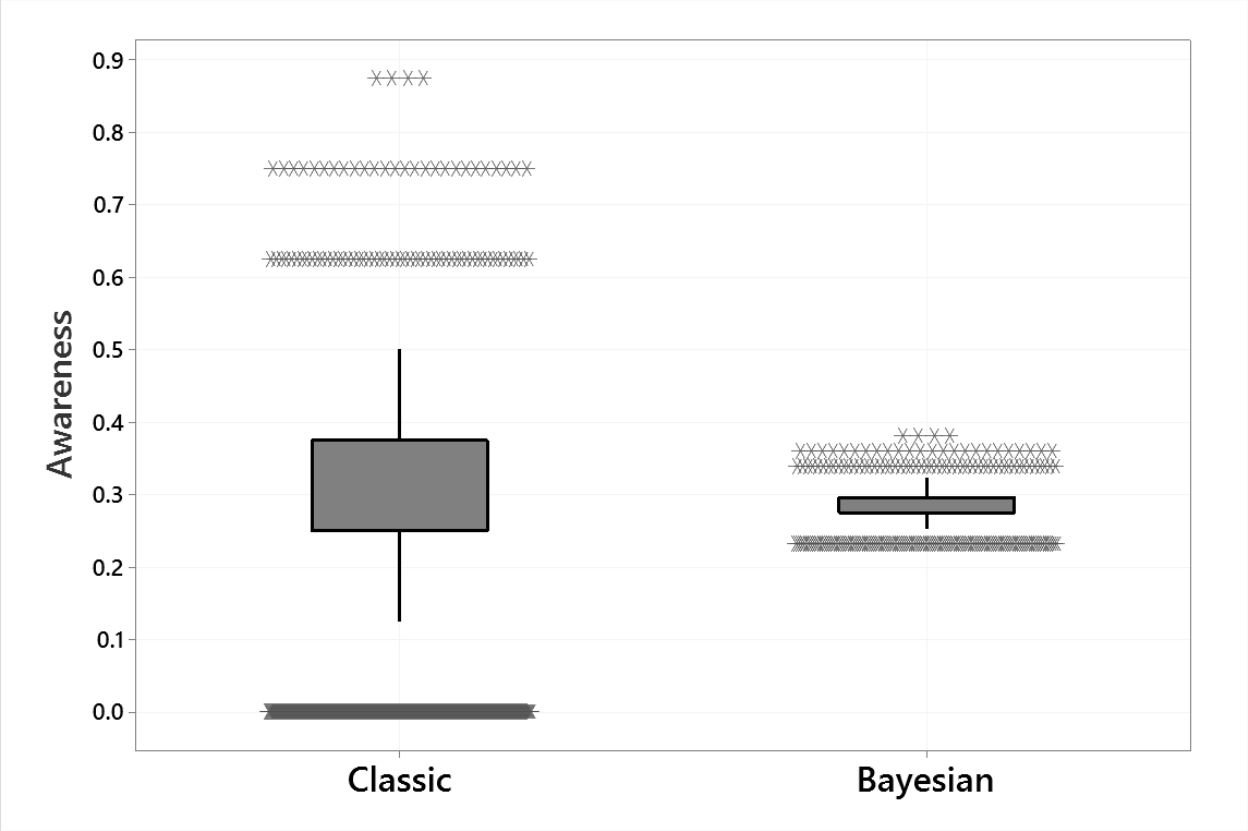


Figure 1. Comparison of familiarity index distributions estimated by Classic and Bayesian approaches.

Figure 2 present the distribution of average familiarity scores among respondents, comparing estimates obtained from both the Bayesian and classic approaches.

Both distributions exhibit a right-skewed pattern, indicating that the majority of respondents demonstrate low to moderate levels of familiarity. However, the Bayesian approach introduces a notable shift, reflecting a slightly higher concentration of respondents with moderate familiarity (between 0.25 and 0.35). This suggests that the Bayesian model, by incorporating prior information, captures a more optimistic view of respondent familiarity compared to the classic (frequentist) approach.

Additionally, both approaches confirm that only a small fraction of respondents demonstrated high levels of familiarity (above 0.45). The slight variations between the two methods emphasize how different statistical approaches can provide nuanced insights. These findings highlight the importance of implementing more effective information dissemination strategies to enhance respondents' understanding of the university's services and processes.

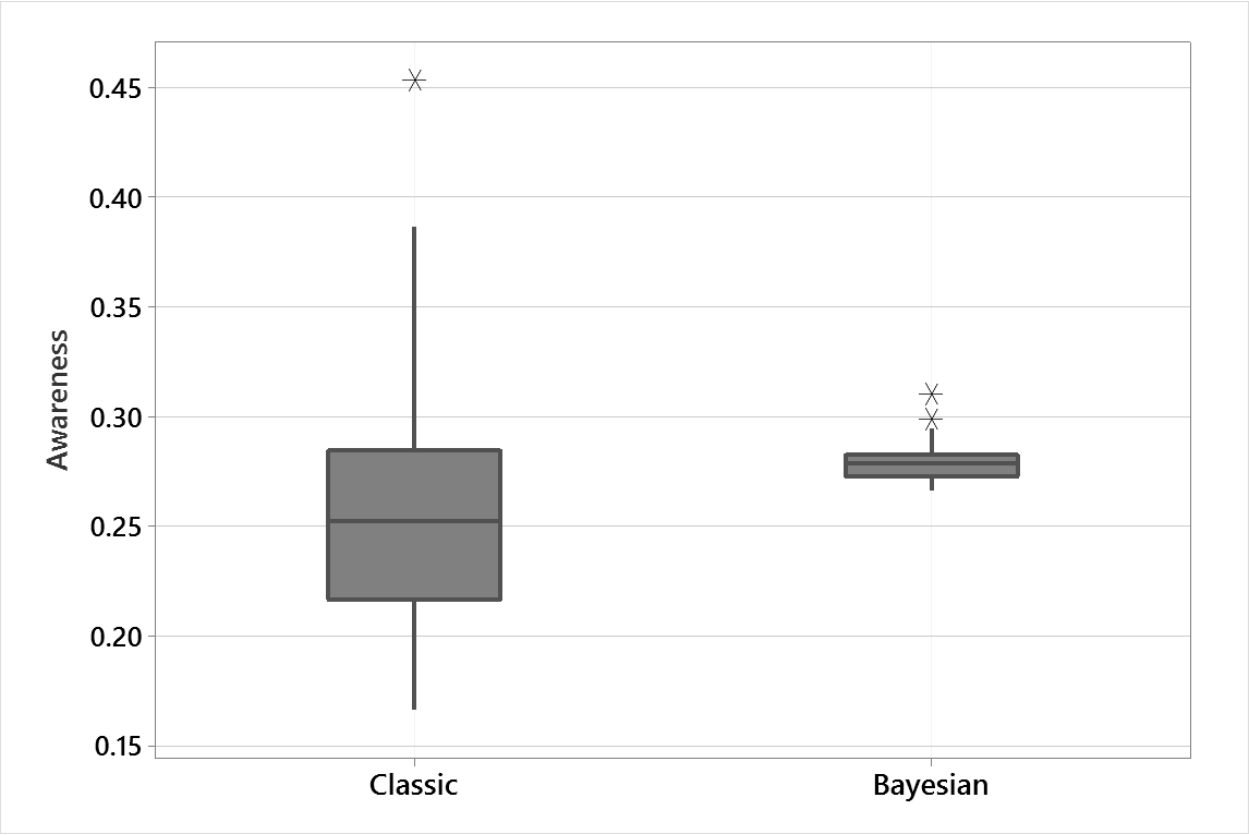


Figure 2. Distribution of familiarity scores among respondents estimated by Classic and Bayesian approaches

To address potential group membership bias, group differences based on gender, educational level, and domain of university services were examined. The normality of the FI_C score was tested for gender and educational level, with both distributions showing significant deviation from normality (p-value < 0.001), indicating the need for non-parametric statistical methods. The Wilcoxon test was subsequently applied to compare FIC scores between gender and educational level groups. Results suggested that familiarity with university services varies minimally by gender or education, with mean familiarity levels of approximately

26% in both categories (see Table 7). In contrast, the Kruskal-Wallis test revealed significant differences in familiarity levels according to the domain of university services, highlighting a meaningful association between service type and user awareness (see Table 7).

Table 7. Statistical analysis of cognition component in terms of different factors

Variable	Summary values	Testing the Hypothesis of Equality
Sex	M (mean=0.25; sd=0.18)	Test statistic= 272374
	F (mean=0.27; sd=0.19)	p-value= 0.089
Level	MSc (mean=0.25; sd=0.18)	Test statistic=246676
	Ph.D. (mean=0.27; sd=0.19)	p-value= 0.099
Domain	Rector's office (mean=0.19; sd=0.13)	Kruskal-Wallis Test =75.08
	Student affairs (mean=0.38; sd=0.17)	p-value=0.000
	Research affairs (mean=0.30; sd=0.17)	
	Education affairs (mean=0.26; sd=0.16)	
	Administrative affairs (mean=0.25; sd=0.19)	
	Cultural affairs (mean=0.21; sd=0.13)	

The above results were computed using the classic approach. Since we reached similar conclusions with the Bayesian approach, we have opted not to report those separately.

5 Effect of familiarity on the results

The primary focus of this paper is not to provide a full presentation of the interview analysis results. However, a summary of selected findings is presented here to demonstrate the relevance and importance of the familiarity index introduced in this study. Considering the qualitative nature of the data, thematic analysis (Kiger and Varpio, 2020) was employed to systematically identify key themes and explore how varying levels of familiarity with the office influence participants' opinions and the reliability of their responses.

To provide clear results, we focus on the services provided by the Rector's Office. Since there are five units (see Table 3) from which students receive services, we expected to collect 300 responses for this section. However, we received 287 responses from the interviewees.

We used a familiarity index to categorize the answers into three groups:

- The first group includes responses with a low familiarity index (less than 0.33).
- The second group includes responses with a medium familiarity index (between 0.33 and 0.66).
- The third group consists of responses with a high familiarity index (greater than 0.66).

This categorization resulted in 211, 70, and 6 responses in the low, medium, and high familiarity groups, respectively. We then applied thematic analysis within each category.

Results for the Low Familiarity Group

The low familiarity group consists of 211 responses. The thematic analysis of this group revealed three main themes with the following frequencies:

- **No direct interaction or limited knowledge (187 responses):**
  Most respondents indicated that they had no direct contact with the services or limited understanding of what the units do. Typical statements included: *"No interaction," "They are good," "No problems,", "I do not know exactly what they do,"* and *"I had no involvement."* This suggests a general lack of engagement or awareness among the majority of respondents in this group.
- **Negative perceptions about service quality and communication (13 responses):**
  A smaller subset expressed dissatisfaction with laboratory and research services, poor communication—especially regarding defense announcements—and perceived unresponsiveness or disorganization. Examples of comments include:
  "Laboratory and research services are not good," "I had no direct interaction but based on past experience, they do not behave well," "Communication about defenses is poor; announcements are not seen," "Subordinate staff do not complete tasks," and *"If we have problems with professors, we do not know where to turn."*
- **Positive remarks on supervision and responsiveness (11 responses):**
  A few respondents noted good supervision and responsiveness from faculty members and administrative leaders. Sample comments include:
  "Supervision is good," "The behavior of professors, the dean, and the university president is good," "Supervision over professors is adequate," and *"Responsiveness is good."*

Overall, the majority of respondents in the low familiarity group either had no direct experience or limited knowledge of the services, while a minority expressed clear positive or negative opinions about service quality and responsiveness.

Results for the Medium Familiarity Group

The medium familiarity group includes 70 responses. The thematic analysis identified two main themes with the following frequencies:

- **Positive feedback on communication and services (41 responses):**
  Many respondents appreciated the quality of communication and certain services. Comments highlighted effective information dissemination through the website, SMS, and Telegram, with suggestions to improve platforms for new students. Positive remarks included:
  "Information dissemination has been good via the website, SMS, and Telegram; it could be improved for new students by using more accessible platforms,"
  "The PhD interview exam process was good,"
  "The selection process for teaching permission was handled well,"
  "We are satisfied with the performance of professors and the university; they support talented students and address problems,"
  and *"The security guards are polite."*
- **Concerns about management, communication, and security (29 responses):**
  Some respondents expressed concerns regarding organizational issues, supervision, and communication inefficiencies. Key points included the need for better planning to manage interview crowding and documentation, insufficient supervision tailored to different fields, and limited flexibility for PhD students' attendance. Security-related complaints involved guards being uncooperative or overly strict, frequent changes in security personnel causing confusion, and inappropriate focus on dress codes. Communication issues were also noted, such as difficulty reaching university offices by phone, ineffective phone systems, delayed or poor information dissemination, and reliance on hard-to-navigate websites instead of direct messaging. Sample comments were:

"More organized planning is needed to control interview problems like crowding and delays,"
"Supervision is not thorough and should be adapted to different disciplines,"
"PhD students' problems are not given special attention, and attendance flexibility is limited,"
"Security guards often cause trouble, are sometimes unhelpful, and frequently change, which creates confusion,"
"It is difficult to contact the university by phone; calls are often transferred without proper guidance,"
and *"Information dissemination is weak; messages do not reach us timely, and announcements are often last-minute."*

Overall, the medium familiarity group shows a mix of appreciation for communication efforts and services alongside notable concerns about management, security, and information flow.

Results for the High Familiarity Group

The high familiarity group consists of 6 responses. The thematic analysis revealed the following key points:

- **Strong criticism of communication and supervision (5 responses):**
  Respondents in this group expressed significant dissatisfaction with the quality and timeliness of information dissemination. They noted that announcements about events such as seminars, cancellations, or other important matters were often delayed or communicated only after the fact. There was a strong perception of inadequate supervision over faculty members, with dissatisfaction regarding professors' performance and a belief that no effective action is taken to address issues. The evaluation process for professors was described as merely formal, lacking genuine attention to performance concerns. Respondents also emphasized the need for supervision not only of faculty but also of administrative managers, highlighting that negative consequences of some managerial decisions on students are overlooked. Additionally, concerns were raised about the disciplinary committee's regulations, particularly that accusations can be confirmed based on testimonies from as few as two witnesses, even if they are friends or classmates, which may lead to unfair outcomes.
- **Positive note on responsiveness (1 response):**
  One respondent mentioned satisfactory responsiveness and follow-up until resolution.

Overall, this group reflects a high level of critical awareness regarding institutional communication, supervision, and governance, coupled with a call for more effective management and fair disciplinary procedures.

To further clarify the relationship between familiarity levels and the nature of participants' opinions, Table 8 presents a joint frequency distribution of familiarity groups against the thematic sentiment categories—negative, neutral, and positive—derived from the interview responses related to the services provided by the Rector's Office.

Table 8. Distribution of Thematic Sentiment by Familiarity Level in Responses Concerning Rector's Office Services

Familiarity Level	Negative Themes	Neutral Themes	Positive Themes	Total Responses
Low	13	187	11	211
Medium	29	0	41	70
High	5	0	1	6

As shown in Table 8, respondents with low familiarity predominantly provided neutral or vague feedback, with a large majority of neutral themes and few positive or negative comments. This aligns with their limited direct experience or knowledge of the services. The medium familiarity group exhibited a more balanced distribution, with a substantial number of both positive and negative themes and no neutral responses, reflecting their more informed and critical engagement. The high familiarity group, although small in sample size, contributed mostly negative themes, indicating a deeper critical insight into institutional challenges, alongside a few positive remarks.

The analysis clearly demonstrates that the level of familiarity significantly influences both the depth and usefulness of participants' opinions. Respondents with low familiarity primarily provided limited or superficial feedback, often indicating no direct experience or general impressions without detailed insights. Their responses tended to be less informative and more neutral or vague.

In contrast, those with medium familiarity offered more balanced and nuanced perspectives, combining positive feedback with constructive criticism. This group demonstrated a better understanding of the services and processes, allowing them to identify specific strengths and weaknesses, which enhance the practical value of their input.

Finally, the high familiarity group, although small in number, contributed the most critical and in-depth evaluations. Their responses reflected a heightened awareness of institutional challenges, including governance and procedural shortcomings, and called for systemic improvements. This group's feedback, while sometimes negative, is particularly valuable for informing targeted policy and operational changes.

Overall, the familiarity index proves to be a useful tool for interpreting the reliability and richness of qualitative data in this study. It highlights the importance of considering respondents' knowledge levels when analyzing and applying feedback, ensuring that conclusions and recommendations are grounded in informed and meaningful perspectives.

6 Discussion and conclusion

It is well understood that inferences drawn from surveys may sometimes diverge from reality due to various sources of uncertainty. However, researchers aim to identify these uncertainties and enhance the reliability of their conclusions through rigorous survey methodologies. This paper has addressed one significant factor often overlooked in service quality surveys: respondent familiarity, which we refer to as the cognitive component. We proposed a theoretical framework to quantify respondents' familiarity with the topic, introduced a familiarity index, and derived two estimators for it.

We examined this index in a case study conducted at a public university in Iran. The results of our study revealed that the average level of students' familiarity of the university's processes was only 26 percent. Familiarity levels varied depending on factors such as the service-providing unit, while variables such as gender and educational level did not show significant effects. The findings from this case study are applicable to many qualitative surveys and suggest that researchers should consider incorporating an option in service quality assessment questionnaires allowing respondents to indicate if they feel unqualified to provide informed feedback due to limited awareness.

To improve the validity of service quality assessments in Iranian public universities, it is essential to consider respondent familiarity. This may involve refining survey instruments to include questions that gauge respondents' understanding of the services and institutional processes being evaluated. By assessing

familiarity alongside perceptions of service quality, researchers can obtain a more nuanced understanding of the factors influencing survey results.

Incorporating expert evaluations of respondent familiarity could also provide valuable insights. Recent studies suggest that expert assessments can serve as benchmarks for evaluating the alignment between student perceptions and the actual quality of services. By combining expert evaluations with student feedback, universities can create more comprehensive and accurate evaluations of service quality.

Evaluating service quality in public universities in Iran presents unique challenges, particularly because survey methodologies often neglect the role of respondent familiarity. As educational institutions seek to improve their services, addressing the gap in student awareness is essential. Future research should focus on developing integrated approaches that assess both service quality and respondent understanding, thereby providing more reliable and actionable insights for institutional improvement.

The low levels of familiarity found in this study also suggest that universities need to focus on improving the awareness of students, faculty, and staff regarding institutional missions and processes. Researchers and practitioners have proposed several strategies to address this issue:

- Orientation Programs: Universities should develop comprehensive orientation programs that clearly communicate their missions and the processes supporting them. This could include workshops or informational materials that explain not just academic programs, but also administrative and operational services.
- Transparent Communication: Enhancing internal communication channels—through newsletters, online platforms, and open forums—can help make students and faculty more aware of the institutional goals and service processes. This transparency helps to build trust and encourages more accurate feedback during service quality assessments.
- Engagement with Stakeholders: Universities can foster awareness by involving students and faculty in decision-making processes, such as participatory governance models. Involving stakeholders in service delivery processes can increase their understanding of these services and lead to more informed assessments.

Acknowledgement. The authors thank Zahra Seifi from AGNA Company for her assistance in preparing the file of respondent explanations. The first author was partially funded by the European Union - Next GenerationEU, in the framework of the GRINS -Growing Resilient, INclusive and Sustainable project (GRINS PE00000018 – CUP F53C22000760007). The views and opinions expressed are solely those of the authors and do not necessarily reflect those of the European Union, nor can the European Union be held responsible for them.

Reference

Bozbay, Z., Baghirov, F., Zhang, Y., Rasli, A. and Karakasoglu, M. (2020). International students' service quality evaluations towards Turkish universities. *Quality Assurance in Education*. 28(3): 151-164.

Braun, V., Clarke, V., Boulton, E., Davey, L. and McEvoy, C. (2021). The online survey as a qualitative research tool. *International journal of social research methodology*, 24(6): 641-654.

Creswell, J. W. and Poth, C. N. (2016). *Qualitative inquiry and research design: Choosing among five approaches*. Sage publications.

Daryazadeh, S., Yavari, M., Sharif, M. R., Azadchahr, M. J., Hoseini, S. V. Z. and Akbari, H. (2022). Assessing the quality of educational services from the viewpoint of clinical teachers and medical students using servqual

model. *Educational Research in Medical Sciences*. 11(2): e119417.

Garside, R. (2014). Should we appraise the quality of qualitative research reports for systematic reviews, and if so, how?. *Innovation: The European Journal of Social Science Research*. 27(1): 67-79.

Gilavand, A. and Torabipour, A. (2022). The quality of services of Iran University hospitals based on SERVQUAL's evaluation model: a systematic review and meta-analysis. *Frontiers in Public Health*. 10: 838-359.

Kang, L., Kang, X., Deng, X. and Jin, R. (2018). A Bayesian hierarchical model for quantitative and qualitative responses. *Journal of Quality Technology*. 50(3): 290-308.

Karunarathna, I., Gunasena, P., Hapuarachchi, T. and Gunathilake, S. (2024). Comprehensive data collection: Methods, challenges, and the importance of accuracy. *Uva Clinical Research*. 1-24. 10.13140/RG.2.2.13134.47689

Kiger, M. E. and Varpio, L. (2020). Thematic analysis of qualitative data: AMEE Guide No. 131. *Medical teacher*, *42*(8), 846-854.

Mejia, J., Mankad, S. and Gopal, A. (2021). Service quality using text mining: Measurement and consequences. *Manufacturing & Service Operations Management*. 23(6): 1354-1372.

Mellinger, C. D. and Hanson, T. A. (2020). Methodological considerations for survey research: Validity, reliability, and quantitative analysis. *Linguistica Antverpiensia, New Series–Themes in Translation Studies*. 19: 172-190

Nooripoor, M. and Sharifi, Z. (2020). Assessment of the Quality of Academic Services at Yasouj University: Application of SERVQUAL Model. *International Journal of Agricultural Management and Development (IJAMAD)*. 10(4): 347-359.

Osman, A.R. and Saputra, R.S. (2019). A pragmatic model of student satisfaction: a viewpoint of private higher education. *Quality Assurance in Education*. 27(2): 142-165.

Rafati, F., Arbabisarjou, A. and Dastyar, N. (2021). Analyzing the gap between perceptions and expectations of students about the quality of educational services in Southern of Iran: SERVQUAL Model. *Pakistan Journal of Medical & Health Sciences*. 15(4): 1334-1340.

Revilla, M. (2016). Impact of raising awareness of respondents on the measurement quality in a web survey. *Quality & Quantity*. 50(4): 1469-1486.

Sohail, M. S. and Hasan, M. (2021). Students' perceptions of service quality in Saudi universities: the SERVPERF model. *Learning and Teaching in Higher Education: Gulf Perspectives*. 17(1): 54-66.

Tourangeau, R. and Yan, T. (2007). Sensitive Questions in Surveys. *Psychological Bulletin.* 133(5): 859-883.

Tourangeau, R., Yan, T. and Sun, H. (2021). Who can you count on? Understanding the determinants of reliability. *Journal of Survey Statistics and Methodology*. 8(5): 903-931.