CULTURAL MODERATION OF ONLINE HOTEL EVALUATIONS: A MULTILEVEL MODEL OF SERVICE ATTRIBUTE SALIENCE, WEIGHTING, AND SENTIMENT

KENAN MAHMUTOVIĆ*

UDK:
338.488.2:640.41]:339.138
338.487:659.1
338.487:339.13

Izvorni znanstveni članak
Original scientific paper
DOI 10.47960/2831-0322.2026.1.30.30

Abstract

The present study aims to investigate whether online hotel ratings can be seen as a culturally invariant measure of service performance or a culturally variant measure of an evaluative outcome. Using 346,463 multilingual reviews from Booking.com for 794 hotels in Croatia, we propose a fully moderated multilevel model that includes the effects of cultural region, 26 dimensions of service quality based on BERTopic modeling, and the interaction between sentiment and culture. Our results demonstrate the presence of systematic differences in the evaluation process. First, we find that rating levels vary significantly by cultural region after controlling for hotel-level heterogeneity. Second, all 26 dimensions of service quality exhibit at least one significant difference by culture. Third, we observe that the effects of service-quality dimensions vary by culture. Finally, we observe that the sentiment exhibits strong asymmetric shifts and significant heterogeneity in the translation of affect. Importantly, the model exhibits substantially improved performance with the addition of the cultural moderation effect (marginal R^2 = .477; conditional R^2 = .537). Our results support the notion of a three-layered cultural filter in the evaluation of digital reputation.

JEL classification: M31; L83; C38; D91

Keywords: hotel review; cultural moderation; hospitality marketing; booking; sentiment analysis; service quality; customer behavior

* Associate Professor Kenan Mahmutović, PhD, Faculty of Economics, University of Bihać, kenan.mahmutovic@gmail.com; kenan.mahmutovic@unbi.ba

Introduction

Online reviews have been found to be a critical source of information in the hospitality industry. In the online environment, for instance through Booking.com, guests rate their experiences numerically and provide a written review, thereby generating data that allow insights into how service quality is perceived and evaluated across markets. While past research has found the major drivers of hotel ratings to be cleanliness, staff quality, value for money, and location, there has been little investigation into the extent to which these drivers work universally across cultures. Indeed, the service attributes may be perceived and evaluated differently depending on the cultural background of the individual. If this is the case, ratings generated in online environments reflect not only the service performance but also the cultural evaluation standards. This study follows the thesis that cultural background serves as a perceptual filter for the evaluation of hotel services. Cultural influences may be found in the baseline rating severity, the importance of certain service dimensions, the strength of attribute-rating linkages, and the translation of emotional tone, particularly negative tone.

From an empirical perspective, the research examines 346,463 multilingual online reviews for hotels in Croatia. The cultural-background variable is operationalized according to the guests' countries of origin and is categorized into macro-cultural groups based on established approaches to cross-cultural studies. Using multilevel modeling to account for the nesting of reviews within hotels, the research investigates the extent to which cultural background influences online reviews. Four research questions are addressed in the research:

RQ1: Do the ratings vary between cultural groups?

RQ2: Do the cultural groups vary with respect to the importance of the dimensions of service quality mentioned in the online reviews?

RQ3: Does the impact of specific service-quality dimensions on ratings vary depending on the cultural background?

RQ4: Does the impact of negative sentiment vary depending on the cultural background?

This research contributes to the theoretical understanding of the role of cultural background in the evaluation of service quality and provides implications for hotels with respect to the evaluation of their quality from a multicultural perspective.

1. Literature Review and Hypothesis Development

1.1. Online Ratings as Behavioral Outcomes

Online hotel ratings are not simply a measure of service performance but, in fact, a measure of post-consumption evaluation embedded in digital platforms. According to Expectancy–Disconfirmation Theory (Oliver, 1980, 2015), satisfaction occurs when perceived performance is in line with expectations or exceeds them. The rating, in essence, can be viewed as a condensed form of the satisfaction outcome.

Studies have indicated that ratings are a function of performance, as well as various dynamics such as cross-hotel heterogeneity and reviewer behavior (Radojević et al., 2015, 2017; Mariani et al., 2019). The design of the platform also influences the expression of evaluation (Filieri et al., 2018). Cross-national studies have indicated that there are systematic response-style differences in scale usage, including variation in extreme-category selection (Harzing, 2006), which aligns with observed heterogeneity in online rating behavior (Radojević et al., 2017).

In the context of online ratings, there have been indications of a positivity bias and upper-scale skewness, which compress the variance of highly rated contexts (Hu et al., 2009). Therefore, small scale changes can be viewed as a function of group differences.

In the context of e-WOM systems, the rating can be viewed as a public market signal that influences demand (Chevalier & Mayzlin, 2006; Babić et al., 2016), with both expressive and informational motives (Hennig-Thurau et al., 2004). Therefore, the use of online ratings can be viewed as culturally and contextually embedded behavior, indicating cross-cultural differences in baseline rating levels.

1.2. Culture and Evaluation Standards

Culture influences perception, cognition, and evaluations. Hofstede (2001) refers to culture as the "collective programming of the mind" structured around value dimensions that shape expectations and patterns of behavior. Inglehart and Welzel (2005) argue that societies differ in the values they prioritize, which in turn shapes how people interpret and evaluate their experiences. They distinguish between survival-oriented values and self-expression–oriented values as two broad cultural orientations.

In services, culture has been found to influence how consumers perceive and evaluate various aspects of services (Mattila, 1999). Culture also influences how consumers perceive fairness and service recovery in collectivist vs. individualist cultures (Mattila & Patterson, 2004). A series of meta-analyses found that national

culture has persistent behavioral effects on consumers (Taras et al., 2010), which indicates cross-cultural variation in rating behaviors.

In tourism and hospitality, evaluations are also culturally influenced (Reisinger & Turner, 2003). Studies on hotel services found cross-national variation in consumers' perceived experiences and attribute evaluations (Zhang et al., 2015), while UGC analysis found cultural variation in consumers' ratings on hospitality services (Guan et al., 2022).

In conclusion, culture influences consumers' perceptions and evaluations of services, so it can be expected that baseline ratings also vary by cultural regions.

1.3. From Textual Topics to Service-Quality Dimensions

1.3.1. Topic Modeling in Hospitality Research

The increasing availability of online review data has also helped accelerate the use of computational text-based approaches for uncovering hidden patterns and topics in unstructured text. Topic modeling is a probabilistic approach for discovering topics or thematic structures in text-based data (Blei, 2012). Topic modeling has been extensively used in marketing and tourism for discovering topics related to customers' experiences and their associated outcomes for brand perceptions (Tirunillai & Tellis, 2014; Berezina et al., 2016; Li et al., 2025).

However, the topics often extracted through these models are typically too detailed and granular for econometric and managerial use. Therefore, these topics need to be aggregated and mapped to higher-order service-quality dimensions that conform to theoretical frameworks. To illustrate this, Wang et al. (2025) show the systematic approach for aggregating topics extracted through the BERTopic technique.

Building on this approach, recent studies have applied advanced topic modeling techniques such as BERTopic (Grootendorst, 2022) to identify service-related themes in multilingual review data. These themes can then be aggregated based on their semantic meanings and mapped to service-quality dimensions through an inductive-deductive approach (Mahmutović, 2025). In this way, detailed textual patterns are translated into interpretable service attributes that managers can interpret, in line with Expectancy–Disconfirmation Theory (Oliver, 1980, 2015) and Kano's asymmetric quality framework (Kano et al., 1984).

1.3.2. Salience as an Indicator of Evaluative Focus

From a salience perspective, what guests mention in the online reviews indicates what attributes are included in the judgments. Attribute accessibility theory indicates that cognitively accessible information has a better chance of influencing judgments (Feldman & Lynch, 1988). The mention of a service attribute in online reviews indicates that the attribute was salient enough to be included in the narrative. The use of the mention of attributes, as opposed to the rating of the attributes, becomes important in the cross-cultural context, as the direct assessment of the relative intensity of topics can be affected by the use of different styles and languages (Harzing, 2006; Liu, 2015). Differences in mention probabilities can therefore be understood as differences in evaluative focus - that is, what different guest groups find important enough to highlight in their reviews.

1.4. Service Quality Dimensions and Asymmetric Effects

Service quality is generally considered to be multi-dimensional, e.g., in SERVQUAL (Parasuraman et al., 1988). However, there is also evidence of asymmetric evaluation effects. The Kano model of attribute types distinguishes between must-be, performance, and excitement attributes, showing that certain attributes are mainly associated with high losses in their absence, while others are associated with disproportionate gains when present (Kano et al., 1984).

There is also psychological evidence of negativity bias, i.e., that negative information is generally more influential in forming judgments than equivalent positive information (Baumeister et al., 2001). Consistent with this, there is evidence of asymmetry in the effect of review valence in tourism studies (Park & Nicolau, 2015), as well as in studies of evaluation cues in the hospitality industry, in which satisfaction drivers have been found to vary depending on the context (Radojević et al., 2015, 2017; Berezina et al., 2016).

Moreover, attribute evaluation is culturally dependent, implying that there are cross-national differences in what is considered fundamental or exceptional in terms of service quality (Hofstede, 2001; Reisinger & Turner, 2003). Empirical evidence confirms systematic cross-national differences in the evaluation of hotel experiences (Zhang et al., 2015; Guan et al., 2022). Accordingly, both the salience and evaluative weight of service-quality dimensions may vary across cultural segments.

1.5. Emotional Tone and Sentiment in Service Evaluation

Emotions are the driving force behind evaluation in the post-consumption phase. The affect-as-information model proposes that people use affective states as cues in judgment formation (Schwarz & Clore, 1983), and marketing research has long

established emotions as key drivers of evaluative and behavioral outcomes (Bagozzi et al., 1999).

Online reviews are affective narratives. Hotel ratings on Online Travel Agency (OTA) websites co-occur with textual explanations, jointly constituting the evaluative outcome. The sentiment in reviews can be analyzed using sentiment analysis tools (Liu, 2015). The expression and evaluation of emotions are also shaped by their cultural context and the social norms established within that culture. People in different cultures may express the same emotion yet assign different ratings to the same experience (Hall, 1976; Markus & Kitayama, 1991), and this is supported by evidence on culture-dependent evaluations of service and perceptions of fairness in service settings (Mattila, 1999; Mattila & Patterson, 2004), as well as by large-scale UGC findings in the hospitality industry (Guan et al., 2022).

Accordingly, sentiment functions both as a direct driver of ratings and as a culturally moderated mechanism of affective translation.

1.6. Cultural Moderation of Evaluation Processes

Considering all previous approaches, a three-level evaluation mechanism can be inferred. First, the baseline cultural calibration level affects rating severity and rating scale usage (Harzing, 2006; Radojević et al., 2017). Second, dimension-contingent weighting reflects how cultural expectations influence which attributes of service are salient and how strongly they affect judgments (Hofstede, 2001; Guan et al., 2022; Zhang et al., 2015). Finally, culture-dependent affective translation reveals how emotions are converted into numerical ratings across cultures (Hall, 1976; Markus & Kitayama, 1991; Mattila, 1999). This approach aligns with multilevel theory, which situates individual evaluations within larger contexts (Snijders & Bosker, 2012), and provides a coherent basis for modeling cross-cultural moderation in hierarchically structured review data.

1.7. Conceptual Framework and Hypotheses Development

Synthesizing prior research, we propose a moderated multilevel evaluation framework in which culture operates as a systematic boundary condition of online rating formation. Cross-cultural differences in scale usage and evaluation standards (Harzing, 2006; Mattila, 1999; Radojević et al., 2017) imply that baseline rating levels are culturally calibrated:

H1: Overall hotel rating levels differ systematically across cultural regions.

Cultural research further suggests that guest groups differ in what they consider noteworthy in service encounters (Guan et al., 2022; Zhang et al., 2015), leading to cross-cultural variation in attribute salience:

H2: Cultural regions differ in the salience of specific service-quality dimensions mentioned in online reviews.

Building on asymmetric quality logic (Kano et al., 1984; Baumeister et al., 2001) and evidence of asymmetric review effects (Park & Nicolau, 2015), we expect cultural moderation in attribute weighting:

H3: The effect of service-quality dimensions on overall ratings is moderated by cultural background.

Finally, affect-as-information theory and culturally patterned emotional expression (Schwarz & Clore, 1983; Hall, 1976; Markus & Kitayama, 1991) suggest variation in affective translation:

H4: The effect of review sentiment on overall ratings is moderated by cultural background.

Overall, these hypotheses suggest a fully moderated multilevel model in which culture shapes baseline calibration, attribute salience and weighting, and the translation of affect into numeric evaluation.

2. Research Methods

2.1. Data, Unit of Analysis, and Measures

The dataset used in the current analysis included 346,463 hotel reviews of Croatian hotels. The unit of analysis was a single review, grouped by hotel based on the hotel's Booking.com identifier. This allowed for multilevel modeling of the data because the reviews were clustered at the hotel level.

Outcome variable. The dependent variable was the overall numerical review rating of a hotel on a scale from 1 to 10.

Cultural background. Information about the cultural backgrounds of the reviewers was represented by their countries of origin, grouped into macro-cultural clusters (cultural regions; see Appendix A, Table A1). This was included as a predictor in the analysis. The reference category was Central & Eastern Europe/Balkans (CEE_Balkans).

Sentiment. Review-level sentiment was derived using a transformer-based multilingual sentiment classifier (nlptown/bert-base-multilingual-uncased-sentiment), implemented using the Hugging Face framework. The model output five sentiment levels, which were recoded into three polarity categories: negative (1–2

stars), neutral (3 stars), and positive (4–5 stars). Sentiment classification was performed prior to econometric modeling and treated as a categorical predictor in subsequent multilevel analyses. Neutral was used as the reference category.

Service-quality dimensions. Service-quality dimensions were derived using BERTopic (Grootendorst, 2022) applied to the multilingual review corpus. The model initially extracted 98 latent topics, which were subsequently aggregated into 27 higher-order service-quality dimensions through an inductive–deductive mapping procedure aligned with established satisfaction theory. As in Mahmutović (2025), topic-to-dimension aggregation was validated through manual inspection of representative reviews per topic. In the present study, 26 dimensions were retained for modeling purposes, excluding the "Overall Satisfaction" category to avoid conceptual overlap with the dependent variable. For each review, binary salience indicators were constructed based on whether at least one topic belonging to a given dimension was present.

2.2. Modeling Approach: Multilevel Structure

Because multiple reviews were nested within the same hotel, all rating-model specifications were estimated as multilevel linear mixed-effects models, including a random intercept at the hotel level to capture unobserved, time-invariant heterogeneity between hotels, as well as dependence across reviews of the same hotel.

Formally, for review i in hotel j:

$Rating_{ij}=\beta_0+u_j+\epsilon_{ij}$

where β₀ is the fixed intercept representing the overall mean rating across hotels, u_j is the hotel random intercept and ϵ_{ij} is the idiosyncratic error term. Models were estimated via maximum likelihood (REML = FALSE) to enable likelihood-based comparisons across nested specifications.

2.3. Empirical Strategy and Model Specifications

The empirical strategy proceeds sequentially, aligning model blocks with hypotheses. Across all model specifications:

- $Rating_{ij}$ denotes the overall numerical rating given in review i for hotel j.
- $Dimension_{k,ij}$ represents a binary indicator equal to 1 if service-quality dimension k is mentioned in review i for hotel j, and 0 otherwise.

- $CulturalRegion_{ij}$ indicates the macro-cultural group of the reviewer.
- $Sentiment_{ij}$ represents the categorical sentiment classification of the review (positive, negative, neutral).
- β_0 is the fixed intercept (overall mean rating for the reference categories).
- β_k denotes fixed effects of service-quality dimensions.
- β_1 and $\beta_2$denote fixed effects of cultural region and sentiment, respectively.
- γ_k captures interaction effects between selected service dimensions and cultural region.
- δ captures the interaction between sentiment and cultural region.
- $u_j \sim N(0, \sigma_u^2)$ is the hotel-level random intercept capturing unobserved, time-invariant heterogeneity between hotels.
- $\epsilon_{ij} \sim N(0, \sigma_\epsilon^2)$ is the idiosyncratic error term at the review level.
- α_0 and α_1 denote intercept and cultural-region effects in the logistic salience models.
- $logit^{-1}(\eta)$ represents the inverse logit transformation mapping linear predictors to probabilities.
- K denotes the total number of service-quality dimensions included in the model.
- $Key \subset K$ denotes the theoretically selected subset of dimensions for which interaction effects are estimated.

Model 1: Cultural Differences in Rating Levels (H1)

To test the hypothesis of differences in overall rating levels between cultural regions, we propose the following model equation:

$$Rating_{ij}=\beta_0+\beta_1 CulturalRegion_{ij}+u_j+\epsilon_{ij}$$

This specification isolates cross-cultural mean differences while accounting for hotel clustering.

Model 2: Cultural Differences in Dimension Salience (H2)

To test the hypothesis of differences in the salience of individual service-quality dimensions between cultural groups, we proposed separate multilevel logistic regression equations for each of the binary dimensions:

$$Pr\left(Dimension_{ij}=1\right)=logit^{-1}\left(\alpha_0+\alpha_1 CulturalRegion_{ij}+u_j\right)$$

These models capture cross-cultural differences in attribute salience, net of hotel-level heterogeneity.

Model 3: Cultural Moderation of Dimension Effects (H3)

To test the hypothesis of differences in attribute–rating relationships between cultural groups, we propose the following multilevel rating model equation with:

1. main effects of all 26 service-dimension indicators,
2. main effects of cultural region, and
3. dimension × cultural-region interactions for a theoretically motivated subset of 12 key dimensions.

$$Rating_{ij}=\beta_0+\sum_{k=1}^{K}\beta_k\, Dimension_{k,ij}$$
$$+\beta_1 CulturalRegion_{ij}+\sum_{k\in Key}\gamma_k\left(Dimension_{k,ij}\cdot CulturalRegion_{ij}\right)$$
$$+u_j+\epsilon_{ij}$$

Key interaction dimensions include baseline drivers (e.g., cleanliness, staff/service, value, check-in/out) and culturally sensitive themes (e.g., food & beverage, destination perception, language/communication, noise, entertainment, pet-friendliness, EV charging, air conditioning/ventilation). Limiting interactions to these dimensions helps reduce the possibility of overparameterization.

Model 4: Cultural Moderation of Sentiment Effects (H4)

Model 4 extends the specification in model 3 by adding the main effects for sentiment and the entire set of sentiment x cultural region interactions:

$$Rating_{ij}=\beta_0+\sum_{k=1}^{K}\beta_k\, Dimension_{k,ij}$$
$$+\beta_1 CulturalRegion_{ij}+\beta_2 Sentiment_{ij}$$
$$+\sum_{k\in Key}\gamma_k\left(Dimension_{k,ij}\cdot CulturalRegion_{ij}\right)+\delta\left(Sentiment_{ij}\cdot CulturalRegion_{ij}\right)$$
$$+u_j+\epsilon_{ij}$$

This specification tests whether cultures differ in how positive and negative affect are translated into numeric ratings, conditional on identical service-attribute controls and hotel-level heterogeneity.

All statistical analyses were conducted in R, using the packages lme4 and lmerTest for multilevel modeling, and broom.mixed for model reporting. Logistic multilevel models in Model 2 were estimated using glmer with a logit link function. Model comparisons were performed using likelihood ratio tests via the anova() function.

2.4. Inference and Model Comparison

For Models 1, 3, and 4, fixed effects significance tests were performed using the lmerTest package with approximate p-values based on the Satterthwaite method. Tests of nested model improvements were performed using likelihood ratio tests in the anova() function:

- Model 1 vs. Model 3 tests the incremental contribution of service dimensions and dimension × culture interactions beyond baseline cultural differences.
- Model 3 vs. Model 4 assesses whether adding sentiment and sentiment × culture interactions improves explanatory power beyond dimension-level moderation.
- A reduced version of Model 4 (excluding sentiment × culture interactions) is compared with the full Model 4 to isolate the incremental contribution of cultural moderation of sentiment.

2.5. Multiple Testing Control in Model 2

The Model 2 analysis includes 26 multilevel logistic regression analyses, each of which includes several coefficients for cultural regions. To control for false positives in the results of multiple tests, we performed a Benjamini & Hochberg (BH) False Discovery Rate (FDR) adjustment to the p-values of the coefficients for cultural regions in each dimension.

2.6. Robustness Checks

The robustness checks involved: (1) different cultural reference category specifications, and (2) random slope models in which the sentiment effect is allowed to vary by hotel. These tests assess whether findings are sensitive to reference coding or hotel-level heterogeneity in affective translation.

3. Results

3.1. Sample characteristics and descriptives

The data set comprised 346,463 Booking.com reviews of 794 Croatian hotels, with a mean rating of M = 8.38 (SD = 1.88, range 1-10). Sentiment was predominantly positive at 81.97% (n = 284,003), negative 9.19% (n = 31,828), and neutral 8.84% (n = 30,632). Sentiment score, coded as positive = 1, neutral = 0, negative = -1, strongly correlated with ratings, r = .658 indicating that more positive sentiment was associated with higher ratings.

The sizes of the cultural-region groups varied, but all were substantial. CEE/Balkans had the largest group size, 33.58% (n = 116,332), followed by Germanic/Alpine 27.66% (n = 95,840), and Anglosphere 12.56% (n = 43,529). Remaining group sizes were: Western Europe 9.60% (n = 33,256), Mediterranean Europe 8.96% (n = 31,032), Nordic 3.10% (n = 10,745), Latin America/Caribbean 1.96% (n = 6,774), MENA/Turkey 1.41% (n = 4,884), Asia 0.74% (n = 2,573), and Other 0.43% (n = 1,498).

Mean ratings also differed descriptively by cultural region. The highest mean ratings were observed among reviewers from Latin America/the Caribbean (M = 8.67, SD = 1.68) and CEE/Balkans (M = 8.63, SD = 1.88), whereas lower means were observed for Mediterranean Europe (M = 8.08, SD = 1.91) and Germanic/Alpine (M = 8.14, SD = 1.90). Sentiment composition varied by region as well; for example, the MENA/Turkey segment contained the highest proportion of negative reviews (15.50%) and the lowest proportion of positive reviews (75.02%), while the Anglosphere had a comparatively low negative share (6.93%).

Table 1. Sample distribution and mean ratings by cultural region

Cultural Region (ref. = CEE/Balkans)	n	%	Mean Rating	SD
CEE/Balkans	116,332	33.6	8.63	1.88
Germanic/Alpine	95,840	27.7	8.14	1.90
Anglosphere	43,529	12.6	8.61	1.75
Western Europe	33,256	9.6	8.21	1.77
Mediterranean Europe	31,032	9.0	8.08	1.91
Nordic	10,745	3.1	8.16	1.94
Latin America/Caribbean	6,774	2.0	8.67	1.68
MENA/Turkey	4,884	1.4	8.31	2.01
Asia	2,573	0.7	8.37	1.88
Other	1,498	0.4	8.33	1.91

Source: author

Following BERTopic-based topic extraction (Grootendorst, 2022), individual topics were manually grouped into 27 higher-order service-quality dimensions based on conceptual similarity and established service-quality frameworks. The full list of dimensions is presented in Table 2.

Table 2. Frequency and Rating-Band Distribution of Identified Service-Quality Dimensions

Dimension	Frequency % (Total)	Rating >= 9 (Row %)	Rating 8-9 (Row %)	Rating < 8 (Row %)	Topics count
Food and Beverage Experience	53.2	55.8	22.7	21.5	11
TV and Digital Entertainment Availability	49.0	57.9	19.7	22.4	6
Value and Price Perception	45.9	69.4	18.7	11.9	6
Location and Accessibility	40.4	60.9	23.3	15.8	6
Infrastructure	21.4	40.2	22.6	37.2	8
Availability of Tennis Facilities	14.4	58.9	23.9	17.2	1
Noise	13.4	53.8	22.5	23.6	7
Pet Friendliness	13.1	66.4	20.8	12.9	2
Booking and Check-in/out Process	11.7	25.3	20.8	53.9	3
Language and Communication	11.1	63.2	26.0	10.8	5
Destination Perception	9.4	84.6	11.1	4.3	5
Staff Attitude & Professionalism	8.3	42.9	9.9	47.2	5
Insect Issues (Mosquitoes & Ants)	6.8	22.6	17.9	59.6	4
Evening Entertainment and Live Music	5.6	46.4	26.5	27.1	1
Access and Assistance for People with Disabilities	5.2	26.5	20.4	53.1	1
Overall Satisfaction	4.3	80.3	9.0	10.7	9
Cleanliness and Hygiene	3.3	66.4	18.0	15.6	3
Coffee Quality	2.9	60.8	22.1	17.1	1
Bed and Linen Comfort	2.8	31.8	22.7	45.5	3
Free Service Upgrade	2.8	75.7	15.7	8.6	2
Restaurant and Bar Operating Hours	1.6	38.6	28.8	32.6	2
Accommodation for Babies and Small Children	1.4	49.8	24.6	25.6	1
Air Conditioning and Ventilation	1.3	26.3	22.7	51.0	2
Availability of Electric Vehicle Chargers	1.2	59.4	24.3	16.3	1
Cocktail Bar Experience	1.1	86.0	10.5	3.5	1

Promotional Image vs Reality Mismatch	0.9	6.1	4.7	89.2	1
Motorcyclist and cyclist Friendliness	0.7	63.8	22.2	14.0	1

3.2. Multilevel structure (Model 0: null model)

To justify multilevel modeling, we estimated a null (intercept-only) model with random intercepts for hotels. The between-hotel variance was 0.673, and the residual variance was 2.978, yielding an intraclass correlation coefficient of ICC = 0.184. Thus, approximately 18.4% of the variance in ratings was attributable to differences between hotels, supporting the multilevel specification.

3.3. Cultural-region differences in ratings (Model 1)

Model 1 added cultural region (reference category: CEE/Balkans). Coefficients reflect differences within the analytic sample used for model estimation. Note that regression coefficients reflect within-hotel differences after accounting for hotel-level heterogeneity and therefore do not necessarily equal raw mean differences reported in descriptive statistics. Relative to CEE/Balkans, most cultural regions exhibited significantly lower ratings (all $p < .001$). Specifically, the largest negative differences were observed for Germanic/Alpine ($b = -0.611$, 95% CI [−0.626, −0.596]) and Western Europe ($b = -0.594$, 95% CI [−0.615, −0.572]). Differences were also negative for MENA/Turkey ($b = -0.548$, 95% CI [−0.598, −0.499]), Nordic ($b = -0.570$, 95% CI [−0.604, −0.535]), Mediterranean Europe ($b = -0.543$, 95% CI [−0.565, −0.521]), Asia ($b = -0.508$, 95% CI [−0.576, −0.441]), and the Anglosphere ($b = -0.312$, 95% CI [−0.333, −0.291]). The smallest difference was for Latin America/Caribbean ($b = -0.141$, 95% CI [−0.184, −0.098]).

Overall, cultural-region fixed effects improved fit compared with the null model (AIC decreased from 1,364,486 to 1,356,539), indicating meaningful between-region differences in average ratings.

3.4. Cultural variation in dimension salience (Model 2)

To examine cross-cultural differences in attribute salience, we estimated 26 multilevel logistic models (one per binary dimension) predicting mention likelihood from cultural region, with hotel-level random intercepts. All dimensions met the predefined prevalence criteria; none were excluded. After applying Benjamini–Hochberg correction within each dimension, all 26 showed at least one significant cultural-region effect (BH-adjusted $p < .05$), supporting H2.

The extent of cross-cultural heterogeneity varied across dimensions. The highest number of BH-significant contrasts (8 each) emerged for Access/assistance for disabilities, Coffee quality, and Infrastructure, followed by Cocktail bar experience, Destination perception, Insect issues, Language & communication, Location & accessibility, and Value & price perception (7 each). Restaurant and bar operating hours exhibited the fewest significant contrasts (2).

Substantively, several effects were substantial in size. For instance, Nordic reviewers were significantly less likely than CEE/Balkans reviewers to mention Cocktail bar experience (b = -0.946, OR = 0.39), whereas Germanic/Alpine reviewers were more likely to mention Motorcyclist/cyclist friendliness (b = 0.927, OR = 2.53). Furthermore, Anglosphere reviewers were less likely to mention Coffee quality (b = -0.706, OR = 0.49), whereas Western Europe reviewers were more likely to mention Tennis facilities (b = 0.484, OR = 1.62).

Overall, dimension salience differs systematically across cultural regions, indicating that guest groups vary in which service attributes enter the evaluative narrative.

3.5. Predicting ratings from thematic dimensions and moderated cultural effects (Model 3)

Model 3 examined ratings as a function of (a) all 26 dimensions (entered simultaneously), (b) cultural region main effects, and (c) culture-interaction effects, but only for a pre-specified set of 12 thematic dimensions: Cleanliness, Staff, Value/Price, Booking/Check-in/out, Food & Beverage, Destination perception, Language/communication, Noise, Evening entertainment, Pet friendliness, EV chargers, Air conditioning/ventilation. Model fit improved substantially compared to the culture-only specification (AIC = 1,247,553), while the random-intercept variance at the hotel level decreased from approximately 0.694 in Model 1 to 0.349 in Model 3, indicating that thematic dimensions account for a large share of between-hotel heterogeneity.

Main effects of dimensions (holding other dimensions constant)

Several dimensions exhibited strong negative ("penalty") effects when mentioned. The largest negative coefficient was associated with Promotional image–reality mismatch (b = −1.954, 95% CI [−2.011, −1.897]), followed by Booking/check-in/out (b = −1.602, 95% CI [−1.632, −1.572]) and Insect issues (b = −0.889, 95% CI [−0.912, −0.865]). Additional negative predictors included Staff & service quality (b = −0.804), Air conditioning/ventilation (b = −0.780),

Infrastructure (b = −0.646), Noise (b = −0.199), and Restaurant/bar operating hours (b = −0.189) (all p < .001).

Conversely, several dimensions were positively associated with ratings, including Free service upgrade (b = 0.562), Cleanliness & hygiene (b = 0.506), Value & price perception (b = 0.411), Tennis facilities (b = 0.389), TV/digital entertainment (b = 0.361), Pet friendliness (b = 0.349), Destination perception (b = 0.327), and Coffee quality (b = 0.317) (all p < .001). Accommodation for babies/children and Motorcyclist/cyclist friendliness were not significant; Language & communication showed a small positive effect (b = 0.033, p = .036).

Cultural-region main effects after accounting for dimensions

Even after controlling for all dimensions and interactions, several cultural regions rated lower than CEE/Balkans: Anglosphere (b = −0.385), Germanic/Alpine (b = −0.476), Mediterranean Europe (b = −0.451), MENA/Turkey (b = −0.462), Nordic (b = −0.638), Western Europe (b = −0.514), Asia (b = −0.528), and Other (b = −0.477) (all p < .001). Latin America/Caribbean did not differ significantly (b = 0.059, p = .191), suggesting that baseline cultural differences persist beyond attribute mention patterns.

Culture × dimension interactions (key dimensions)

Several key interactions indicated that the strength of certain dimension–rating relationships varies by cultural region compared with CEE/Balkans:

- Staff & service quality: Weaker (less negative) for the Anglosphere (b = 0.255) and Other (b = 0.591), but more negative for MENA/Turkey (b = −0.197).
- Value & price perception: Attenuated for Latin America/Caribbean (b = −0.212), Mediterranean Europe (b = −0.063), and Western Europe (b = −0.091), but stronger for Nordic (b = 0.106).
- Booking/check-in/out: Less negative for Germanic/Alpine (b = 0.167), Mediterranean Europe (b = 0.127), Nordic (b = 0.322), and Western Europe (b = 0.224).
- Food & beverage: More positive for Asia (b = 0.200) and Western Europe (b = 0.062), but more negative for Germanic/Alpine (b = −0.058).

- Destination perception: Stronger for Germanic/Alpine (b = 0.132) and Nordic (b = 0.162), weaker for Latin America/Caribbean (b = −0.155).
- Language & communication: Stronger for the Anglosphere (b = 0.160), weaker for Germanic/Alpine (b = −0.134), Mediterranean Europe (b = −0.127), and Latin America/Caribbean (b = −0.157).
- Noise: Slightly stronger for Anglosphere (b = 0.072) and Nordic (b = 0.130).
- Evening entertainment: Weaker for Anglosphere (b = −0.102) and Germanic/Alpine (b = −0.075).
- Pet friendliness: Weaker for Germanic/Alpine (b = −0.125), Latin America/Caribbean (b = −0.205), and Mediterranean Europe (b = −0.089), slightly stronger for Anglosphere (b = 0.078).
- EV chargers: Stronger for Asia (b = 0.472).
- Air conditioning/ventilation: More negative for Anglosphere (b = −0.279).

Overall, Model 3 demonstrates that (i) thematic dimensions strongly predict ratings and (ii) several of these effects vary meaningfully across cultural regions, consistent with culture-contingent attribute weighting.

3.6. Sentiment Effects and Culture-Dependent Affective Translation (Model 4)

Model 4 extends Model 3, in which sentiment (neutral reference), as well as its interaction with cultural region, is added to the model. A significant improvement in model fit is indicated by the likelihood ratio test, comparing model 4 to model 3, $\chi^2(20, N = 100{,}000) = 114{,}630.60$, $p < .001$, with AIC decreasing from 1,247,553 to 1,132,963, and BIC from 1,249,123 to 1,134,748. Moreover, explained variance increases, R^2 marginal = 0.268 to 0.477, R^2 conditional = 0.371 to 0.537, while between-hotel variance is reduced from ICC = .184 (null model) to ICC = .114, indicating that sentiment and its cultural moderation account for substantial heterogeneity in ratings.

Direct effects of sentiment, net of 26 thematic dimensions and interaction effects. Positive sentiment, compared to the neutral reference, increases ratings in CEE/Balkans by b = 1.125, 95% CI [1.096, 1.154], $p < .001$, while negative sentiment, compared to the neutral reference, decreases ratings by b = -1.516, 95% CI [-1.555, -1.478], $p < .001$. The larger magnitude of the negative coefficient ($|-1.516| > 1.125$) supports asymmetric, negativity-weighted affective translation.

Sentiment × culture interactions (culture-dependent "affective translation"). A nested LR test confirmed significant sentiment moderation by culture, $\chi^2(18, N = 100{,}000) = 2{,}389.68$, $p < .001$, indicating that positive sentiment effects vary across cultures, e.g., +1.065 in MENA/Turkey, +1.596 in Germanic/Alpine, versus +1.125 in CEE/Balkans. Compared to CEE/Balkans, positive sentiment is +0.471 points stronger in Germanic/Alpine, -0.060 points weaker in MENA/Turkey.

Negative sentiment effects varied across cultures, ranging from −1.159 (MENA/Turkey) to −1.997 (Other), compared with −1.516 in CEE/Balkans. Germanic/Alpine exhibited a 0.346-point stronger negativity translation (−1.862 vs. −1.516), whereas MENA/Turkey showed a 0.357-point attenuation (−1.159 vs. −1.516).

Overall, these results demonstrate pronounced cross-cultural heterogeneity in affective translation: identical emotional narratives do not produce uniform rating shifts across cultural groups.

3.7. Model Fit and Variance Explained

Likelihood ratio tests indicated significant improvement at each modeling step. Adding the cultural region (Model 1) significantly enhanced the model fit over the null model, $\chi^2(9) = 2{,}389.10$, $p < .001$. Adding service-quality dimensions and their cultural region interactions (Model 3) again significantly enhanced the model fit, $\chi^2(134)=7{,}965.67$, $p < .001$. Adding the sentiment and sentiment by culture (Model 4) components resulted in a third significant improvement in model fit, $\chi^2(20)=114{,}630.60$, $p < .001$. The sentiment-by-culture interaction block was again significant, $\chi^2(18)=109{,}253.30$, $p < .001$, validating the cross-cultural differences in affective translation.

Model fit, as measured by AIC, decreased from 1,364,486 for the null model to 1,132,963 for the full model, indicating that the full model had greater explanatory power. ICC decreased from 0.184 to 0.114, indicating that the cultural region, thematic dimensions, and sentiment components explain a significant proportion of the variance in hotel ratings across the sample of hotels.

As expected, the proportion of variance explained by the model increases across the steps. Marginal R^2 increases from 0.020 for Model 1 to 0.268 for Model 3, and then to 0.477 for Model 4, while conditional R^2 in the full model reached 0.537. Thus, more than half of total rating variance is explained when both fixed and random effects are considered, underscoring the explanatory value of the integrated cultural–thematic–affective framework.

3.8. Robustness Checks

To test stability, robustness checks were employed.

Alternative reference category

Model 4 was re-run with Germanic/Alpine as the reference category. Although the coefficient values changed as a result of the reparameterization, the overall results were the same. Cross-cultural baseline differences in rating levels persisted, as did significant cultural moderation of key dimensions (e.g., Cleanliness, Staff & Service Quality, Value and Price Perception, Booking Process) and of the sentiment–rating relationship. Thus, conclusions regarding cultural calibration, attribute weighting, and affective translation are not dependent on reference-category choice.

Random-slope specification

We next estimated a more flexible multilevel model allowing sentiment effects to vary across hotels by introducing a random slope for review sentiment at the hotel level:

$$(1+sentiment|hotel)$$

The random-slope model significantly improved fit ($\Delta\chi^2(5) = 5107.05$, $p < .001$), with AIC decreasing from 1,132,963 to 1,127,865 and log-likelihood improving from −566,315 to −563,762. This indicates meaningful hotel-level heterogeneity in sentiment translation. Importantly, fixed-effect estimates for cultural moderation of dimensions and sentiment remained substantively unchanged in magnitude, direction, and significance.

Overall, the findings are robust to alternative parameterizations and more flexible random-effects structures. Cultural differences in rating calibration, attribute weighting, and affective translation are not artifacts of model specification.

4. Discussion

4.1. Overview of Findings

This research examined whether online ratings are culturally contingent across four components: baseline calibration, attribute salience, attribute weighting, and affective translation. Using a fully moderated multilevel model with 346,463 reviews from 794 hotels, we controlled for cultural region effects and 26 dimensions of service quality. The results yielded four major findings:

1. As predicted in H1, overall ratings vary significantly between cultural regions, even after controlling for hotel-level heterogeneity, thus indicating that ratings are culturally embedded rather than neutral performance signals.

2. As predicted in H2, all 26 dimensions of service quality showed at least one culturally significant difference in salience.

3. As predicted in H3, several links between dimensions and ratings were culturally moderated, particularly for Staff & Service Quality, Value/Price, Booking/Check-in/Out, Destination Perception, Language & Communication, and Pet Friendliness. Thus, identical attributes do not carry identical evaluative weight across cultural segments.

4. As predicted in H4, sentiment had a very strong and asymmetric direct effect (positive: +1.125; negative: −1.516 in CEE/Balkans) and showed substantial cross-cultural heterogeneity (e.g., stronger negativity translation in Germanic/Alpine; attenuated negativity translation in MENA/Turkey). The integrated framework markedly improved explanatory power (R^2_marginal = .477; R^2_conditional = .537).

Overall, the results support the three-layered moderation model of the role of culture in online hospitality evaluations.

4.2. Theoretical Contributions

This study contributes to the field of cross-cultural hospitality research with the finding that online ratings are not culturally neutral performance indicators but are in fact culturally calibrated behavioral outputs. Even after controlling for hotel heterogeneity and thematic content, there remain systematic differences between cultures in the baseline level of ratings. By estimating these effects within a multilevel framework, the study moves beyond descriptive cross-national comparisons and provides adjusted evidence for cultural calibration in digital evaluation.

Furthermore, the fact that all 26 service dimensions exhibited at least one significant cross-cultural difference in mention probability adds to the body of research on cross-cultural topic salience in online content. The operationalization of salience as binary mention in attribute accessibility theory allows the research to capture what is salient to the guest rather than how salient it is to the guest. In terms of methodology, the combination of topic modeling with multilevel logistic analysis is shown to be a viable method for examining structured heterogeneity in cultures.

The results from Model 3 indicate that the evaluative weight of service attributes is not universal and thus extend asymmetric quality theory with the addition of

cultural contingency to attribute asymmetry: penalty and reward effects are not only attribute-inherent but also depend on the culture.

Finally, the results from the sentiment analysis indicate the presence of asymmetric affective translation effects (+1.125 vs. -1.516 in the reference group) as well as heterogeneity between cultures. The results indicate stronger negativity-translation effects among Germanic/Alpine cultures and lower negativity translation effects in MENA/Turkey cultures, which is consistent with the theory of culturally regulated emotional expression (Markus & Kitayama, 1991; Hall, 1976) and the theory of affect-as-information (Schwarz & Clore, 1983) with a moderation role played by culture: the same emotional tone is translated into numerically different values depending on the culture.

4.3. An Integrated Three-Layer Cultural Filter

Together, the research provides support for a three-layer model of the cultural filter in hotel rating formation:

- Baseline calibration: culture impacts the severity of average rating outcomes;
- Attribute salience and weighting: culture impacts what is mentioned and its impact on evaluation;
- Affective translation: the translation of emotional tone to rating outcomes varies across cultures.

This model integrates elements of expectancy-disconfirmation (Oliver, 1980, 2015), asymmetric quality logic (Kano et al., 1984), and affect-as-information theories (Schwarz & Clore, 1983) within a culturally mediated multilevel model.

In terms of the overall model, the ICC decreases from .184 to .114, suggesting that cultural and affective factors do indeed account for a significant amount of variance between hotels, and that rating differences partly reflect guest composition effects rather than purely objective performance differences.

4.4. Managerial Implications

The results have important implications for hotel management. Firstly, the ratings are not necessarily neutral with respect to cultures; rather, managers should account for cultural calibration effects, which can result in the same quality of service being associated with different ratings depending on the cultures of the guests.

Secondly, quality improvement priorities need to be segmented with respect to cultures because the penalty associated with each of the quality dimensions, such as Booking/check-in/out or Staff & Service Quality, differs across cultures.

Thirdly, there is a need to proactively manage negative ratings with respect to cultures characterized by higher levels of negativity translation effects, which can result in greater ratings losses with respect to poor service quality.

Fourthly, there is a need to account for the effects of cultures with respect to the sentiment analysis of the reviews; the same sentiment is not necessarily associated with the same ratings effects with respect to the two cultures.

For multinational hotel chains, these results suggest that centralized KPI systems may require cultural adjustment factors to avoid biased cross-destination performance comparisons.

4.5. Limitations and Future Research Directions

The current study has several limitations: (1) Culture is not measured as individual-level values, but rather as macro-regions. Future research can control for individual-level culture using the Hofstede or Schwartz scale. (2) Service quality is a binary measure of salience. Future research can control for intensity or sentiment. (3) Random slope analysis provides some evidence of robustness, but more research can be conducted to explore cross-level interactions of hotel trait variables and cultural makeup of hotel guests. (4) Future research can replicate the current study in other international settings.

Conclusion

The study investigated whether online hotel ratings can be considered culturally invariant performance indicators or culturally embedded evaluations. Drawing on a multilingual dataset of 346,463 hotel reviews from 794 hotels and employing a fully moderated multilevel model, the study found that culture plays a role in the formation of online hotel ratings at all stages of judgment.

The study found that baseline rating levels vary by culture even after controlling for hotel differences, content theme, and sentiment. All 26 dimensions of service quality exhibit at least one difference by culture. In addition, the weight of key dimensions of hotel quality, such as staff quality, value perception, booking processes, destination perception, and language/communication, also varies across regions. Moreover, the study found that the sentiment of the review exhibited strong asymmetric effects and culturally moderated translation to ratings.

In conclusion, the study validates a three-layered filter of culture in the evaluation of hotel quality. This includes baseline severity calibration across cultures, attribute salience/weighting by culture, and affective translation by culture.

By employing expectancy-disconfirmation theory, asymmetric quality theory, and affect-as-information theory in a hierarchical model of hotel ratings, the study provides a more comprehensive explanation of the role of culture in the post-consumption judgment of hotel quality.

By employing a multilevel model with a larger R-squared value, the study provides a more comprehensive explanation of the role of culture in the formation of hotel ratings. Methodologically, the research demonstrates the utility of combining transformer-based topic modeling, sentiment analysis, and multilevel modeling to disentangle content, affect, and contextual heterogeneity in large-scale review data. Textual and numeric components are shown not as redundant signals, but as jointly shaping evaluation in culturally differentiated ways.

The implications for hotel managers are clear: online hotel ratings are not culturally invariant performance indicators. In other words, the same hotel quality and even the same emotional expression can produce different rating outcomes. The study therefore finds that hotel ratings are socially constructed; they are a function of hotel performance and culturally entrenched standards of evaluation and expression.

References

- Babić Rosario, A., Sotgiu, F., De Valck, K., & Bijmolt, T. H. A. (2016). The effect of electronic word of mouth on sales: A meta-analytic review of platform, product, and metric factors. *Journal of Marketing Research, 53*(3), 297–318.
- Bagozzi, R. P., Gopinath, M., & Nyer, P. U. (1999). The role of emotions in marketing. *Journal of the Academy of Marketing Science, 27*(2), 184–206.
- Baumeister, R. F., Bratslavsky, E., Finkenauer, C., & Vohs, K. D. (2001). Bad is stronger than good. *Review of General Psychology, 5*(4), 323–370.
- Berezina, K., Bilgihan, A., Cobanoglu, C., & Okumus, F. (2016). Understanding satisfied and dissatisfied hotel customers: Text mining of online hotel reviews. *Journal of Hospitality Marketing & Management, 25*(1), 1–24.
- Blei, D. M. (2012). Probabilistic topic models. *Communications of the ACM, 55*(4), 77–84.

- Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research, 43*(3), 345–354.
- Feldman, J. M., & Lynch, J. G., Jr. (1988). Self-generated validity and other effects of measurement on belief, attitude, intention, and behavior. *Journal of Consumer Research, 15*(3), 421–435.
- Filieri, R., McLeay, F., Tsui, B., & Lin, Z. (2018). Consumer perceptions of information helpfulness and determinants of purchase intention in online consumer reviews of services. *Information & Management, 55*(8), 956–970. https://doi.org/10.1016/j.im.2018.04.010
- Grootendorst, M. (2022). BERTopic: Neural topic modeling with a class-based TF-IDF procedure (arXiv:2203.05794). *arXiv*. https://arxiv.org/abs/2203.05794
- Guan, C., Hung, Y.-C., & Liu, W. (2022). Cultural differences in hospitality service evaluations: Mining insights of user-generated content. *Electronic Markets, 32*(3), 1061–1081. https://doi.org/10.1007/s12525-022-00545-z
- Hall, E. T. (1976). *Beyond culture*. Anchor Press.
- Harzing, A.-W. (2006). Response styles in cross-national survey research: A 26-country study. *International Journal of Cross Cultural Management, 6*(2), 243–266.
- Hennig-Thurau, T., Gwinner, K. P., Walsh, G., & Gremler, D. D. (2004). Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the internet? *Journal of Interactive Marketing, 18*(1), 38–52.
- Hofstede, G. (2001). *Culture's consequences: Comparing values, behaviors, institutions and organizations across nations* (2nd ed.). Sage.
- Hu, N., Pavlou, P. A., & Zhang, J. (2009). Why do online product reviews have a J-shaped distribution? *Communications of the ACM, 52*(10), 144–147.
- Inglehart, R., & Welzel, C. (2005). *Modernization, cultural change, and democracy: The human development sequence*. Cambridge University Press.
- Kano, N., Seraku, N., Takahashi, F., & Tsuji, S. (1984). Attractive quality and must-be quality. *Journal of the Japanese Society for Quality Control, 14*(2), 39–48.

- Li, J., Lee, B., & Kim, J. (2025). Analyzing factors affecting overall customer satisfaction using hotel ratings and reviews with BERTopic and three-factor theory. *SAGE Open, 15*(3). https://doi.org/10.1177/21582440251335169
- Liu, B. (2015). *Sentiment analysis: Mining opinions, sentiments, and emotions*. Cambridge University Press.
- Mahmutović, K. (2025). Analyzing user-generated reviews to identify experience dimensions and their impact on satisfaction with mobile banking applications. *Ekonomski vjesnik: Review of Contemporary Entrepreneurship, Business, and Economic Issues*, 38(2), 283–297.
- Mariani, M. M., Borghi, M., & Gretzel, U. (2019). Online reviews: Differences by submission device. *Tourism Management, 70*, 295–298.
- Markus, H. R., & Kitayama, S. (1991). Culture and the self: Implications for cognition, emotion, and motivation. *Psychological Review, 98*(2), 224–253. https://doi.org/10.1037/0033-295X.98.2.224
- Mattila, A. S. (1999). The role of culture in the service evaluation process. *Journal of Service Research, 1*(3), 250–261. https://doi.org/10.1177/109467059913006
- Mattila, A. S., & Patterson, P. G. (2004). Service recovery and fairness perceptions in collectivist and individualist contexts. *Journal of Service Research, 6*(4), 336–346. https://doi.org/10.1177/1094670503262947
- Oliver, R. L. (1980). A cognitive model of satisfaction decisions. *Journal of Marketing Research, 17*(4), 460–469.
- Oliver, R. L. (2015). *Satisfaction: A behavioral perspective on the consumer* (2nd ed.). Routledge.
- Parasuraman, A., Zeithaml, V. A., & Berry, L. L. (1988). SERVQUAL: A multiple-item scale for measuring consumer perceptions of service quality. *Journal of Retailing, 64*(1), 12–40.
- Park, S., & Nicolau, J. L. (2015). Asymmetric effects of online consumer reviews. *Annals of Tourism Research, 50*, 67–83. https://doi.org/10.1016/j.annals.2014.10.007
- Radojević, T., Stanisic, N., & Stanic, N. (2015). Ensuring positive feedback: Factors that influence customer satisfaction in the contemporary hospitality industry. *Tourism Management, 51*, 13–21.

- Radojević, T., Stanisic, N., & Stanic, N. (2017). Inside the rating scores: A multilevel analysis of the determinants of customer satisfaction in the hotel industry. *Tourism Management, 59*, 142–157.
- Reisinger, Y., & Turner, L. W. (2003). *Cross-cultural behaviour in tourism: Concepts and analysis*. Butterworth-Heinemann.
- Schwarz, N., & Clore, G. L. (1983). Mood, misattribution, and judgments of well-being: Informative and directive functions of affective states. *Journal of Personality and Social Psychology, 45*(3), 513–523. https://doi.org/10.1037/0022-3514.45.3.513
- Snijders, T. A. B., & Bosker, R. J. (2012). *Multilevel analysis: An introduction to basic and advanced multilevel modeling* (2nd ed.). Sage.
- Taras, V., Kirkman, B. L., & Steel, P. (2010). Examining the impact of culture's consequences: A three-decade, multilevel, meta-analytic review of Hofstede's cultural value dimensions. *Journal of Applied Psychology, 95*(3), 405–439. https://doi.org/10.1037/a0018938
- Tirunillai, S., & Tellis, G. J. (2014). Mining marketing meaning from online chatter: Strategic brand analysis of big data using latent Dirichlet allocation. *Journal of Marketing Research, 51*(4), 463–479. https://doi.org/10.1509/jmr.12.0106
- Wang, B., Ge, J., & Lu, F. (2025). From review filtering to topic modeling: A deep learning-based approach for analyzing negative online reviews of five-star hotels. *Journal of Hospitality and Tourism Technology*. https://doi.org/10.1108/JHTT-04-2025-0328
- Zhang, Z., Li, H., & Law, R. (2015). Differences and similarities in perceptions of hotel experience: The role of national cultures. *Journal of Travel & Tourism Marketing, 32*(Suppl. 1), S2–S14. https://doi.org/10.1080/10548408.2014.959153

Appendix A

- *Table A1. Cultural Region–Country Mapping*

Cultural Region	Countries
CEE/Balkans	Albania; Belarus; Bosnia And Herzegovina; Bulgaria; Croatia; Czech Republic; Czechia; Estonia; Hungary; Kosovo; Latvia; Lithuania; Moldova; Montenegro; North Macedonia; Macedonia; Poland; Romania; Russia; Russian Federation; Serbia; Slovakia; Slovenia; Ukraine
Germanic/Alpine	Austria; Germany; Switzerland; Liechtenstein
Anglosphere	United Kingdom; United States; United States of America; Ireland; Canada; Australia; New Zealand
Western Europe	France; Netherlands; Belgium; Luxembourg; Monaco; Andorra; San Marino; Gibraltar; Jersey; Guernsey; Isle of Man
Mediterranean Europe	Italy; Spain; Portugal; Greece; Malta; Cyprus
Nordic	Sweden; Norway; Denmark; Finland; Iceland; Faroe Islands; Greenland; Åland Islands; Aland Islands
Latin America/Caribbean	Argentina; Bolivia; Brazil; Chile; Colombia; Costa Rica; Cuba; Dominican Republic; Ecuador; El Salvador; Guatemala; Honduras; Mexico; Nicaragua; Panama; Paraguay; Peru; Uruguay; Venezuela; Jamaica; Trinidad And Tobago; Bahamas; Barbados; Belize; Haiti; Guyana; Suriname; Grenada; Dominica; Saint Lucia; Saint Vincent And The Grenadines; Antigua And Barbuda; Saint Kitts And Nevis; Puerto Rico; Aruba; Bermuda; Cayman Islands; Curacao; Curaçao; Guadeloupe; Martinique; Anguilla; Turks And Caicos Islands; British Virgin Islands; Us Virgin Islands; United States Virgin Islands; Sint Maarten; Saint Martin; Saint Barthelemy; Saint Barthélemy; Bonaire, Sint Eustatius And Saba; Falkland Islands (Malvinas); French Guiana
MENA/Turkey	Turkey; Türkiye; Israel; United Arab Emirates; Uae; Saudi Arabia; Qatar; Kuwait; Bahrain; Oman; Jordan; Lebanon; Iraq; Iran; Iran (Islamic Republic of); State of Palestine; Palestinian Territory; Syria; Syrian Arab Republic; Yemen; Egypt; Morocco; Algeria; Tunisia; Libya; Sudan; Western Sahara; Armenia; Azerbaijan; Georgia
Asia	China; Hong Kong; Macao; Macau; Japan; South Korea; Republic of Korea; North Korea; Democratic People's Republic of Korea; Mongolia; Taiwan; China, Hong Kong Special Administrative Region; China, Macao Special Administrative Region; Singapore; Malaysia; Thailand; Vietnam; Viet Nam; Indonesia; Philippines; Cambodia; Brunei; Brunei Darussalam; Laos; Lao People's Democratic Republic; Myanmar; Timor-Leste; India; Pakistan; Bangladesh; Sri Lanka; Nepal; Maldives; Afghanistan; Bhutan
Other	All other countries not included in the lists above.

KULTURNA MODERACIJA *ONLINE* HOTELSKIH RECENZIJA: VIŠERAZINSKI MODEL ISTAKNUTOSTI, PONDERIRANJA I SENTIMENTA ATRIBUTA USLUGE

Sažetak

Cilj je ovoga istraživanja ispitati može li se *online* ocjena hotela promatrati kao kulturno invarijantna mjera uspješnosti usluge ili kao kulturno varijantna mjera evaluacijskoga ishoda. Koristeći 346 463 višejezične recenzije s platforme *Booking.com* za 794 hotela u Hrvatskoj, predlaže se potpuno moderirani višerazinski model koji uključuje učinke kulturne regije, 26 dimenzija kvalitete usluge identificiranih primjenom BERTopic modeliranja te interakciju sentimenta i kulture.

Rezultati ukazuju na postojanje sustavnih razlika u procesu evaluacije. Prvo, utvrđeno je da se razine ocjena značajno razlikuju prema kulturnoj regiji i nakon kontrole heterogenosti na razini hotela. Drugo, svih 26 dimenzija kvalitete usluge pokazuje barem jednu statistički značajnu razliku s obzirom na kulturu. Treće, uočeno je da se učinak dimenzija kvalitete usluge razlikuje ovisno o kulturnoj pripadnosti. Konačno, utvrđeno je da uloga sentimenta pokazuje snažne asimetrične pomake i značajnu heterogenost u prijenosu afekta u numeričku ocjenu.

Važno je istaknuti da model pokazuje znatno poboljšane performanse nakon uključivanja učinka kulturne moderacije (marginalni $R^2 = 0,477$; uvjetni $R^2 = 0,537$). Rezultati podupiru tezu o postojanju troslojnoga kulturnoga filtra u evaluaciji digitalne reputacije.

JEL klasifikacija: M31; L83; C38; D91

Ključne riječi: hotelska recenzija, kulturna moderacija, marketing u hotelijerstvu, rezervacija, analiza sentimenta, kvaliteta usluge, ponašanje potrošača