JOLLT Journal of Languages and Language Teaching
https://e-journal.undikma.ac.id/index.php/jollt
Email: jollt@undikma.ac.id
DOI: https://doi.org/10.33394/jollt.v14i3.19999
July 2026. Vol. 14, No. 3
p-ISSN: 2338-0810
e-ISSN: 2621-1378
pp. 994-1005

ALIGNMENT AND INCONSISTENCY BETWEEN TEXTUAL SENTIMENT AND STAR RATINGS IN ONLINE PRODUCT REVIEWS: A REFLEXIVE THEMATIC ANALYSIS

[1] *Sastika Seli, [1]Yulfi, [1]Episiasi
[1] English Language Education Program, Social Science and Humanities, PGRI Silampari University, Jln. Mayor Toha, Lubuklinggau City, South Sumatera, Indonesia
*Corresponding Author Email: sastikaseli@unpari.ac.id

Article Info
Article History
Received: March 2026
Revised: April 2026
Accepted: June 2026
Published: July 2026

Keywords
Online reviews;
Politeness;
Sentiment-rating alignment;
Reflexive thematic analysis;
Evaluative language:

Abstract
Online product reviews combine numerical star ratings with written reviews to communicate consumer evaluations. While recent studies highlight frequent mismatches between textual sentiment and numerical ratings, few have examined this phenomenon qualitatively. This research investigates how linguistic sentiment aligns or diverges from star ratings to understand the pragmatic functions behind these inconsistencies. The research analyzes 120 anonymized furniture reviews collected from the IKEA Australia website. A two-stage methodology was applied: a descriptive overview classifying sentiment-rating consistency, followed by Reflexive Thematic Analysis (RTA) to examine linguistic richness and rating alignment. The analysis reveals four dominant evaluative patterns: Polite Dissatisfaction, Understated Positivity, Expressive Dissatisfaction, and Low-Information Reviews. Reviewers frequently employ politeness strategies, contrastive structures, and evaluative restraint, resulting in partial or complete mismatches between expressed sentiment and numerical scores. These results indicate that star ratings and written texts serve distinct communicative purposes and should be interpreted together rather than in isolation. The research contributes to qualitative sentiment analysis and digital pragmatics by demonstrating that sentiment-rating inconsistency is a deliberate pragmatic strategy rather than an analytical measurement error.

How to cite: Seli, S., Yulfi, Y., & Episiasi, E. (2026). Alignment and Inconsistency between Textual Sentiment and Star Ratings in Online Product Reviews: A Reflexive Thematic Analysis. *JOLLT Journal of Languages and Language Teaching,* 14(3), 994-1005. Doi: https://doi.org/10.33394/jollt.v14i3.19999

Copyright© 2026, Seli et al.
This is an open-access article under the CC-BY-SA License.

INTRODUCTION

In this digital era, online shopping has become a large trading activity to buy products. Customers are also getting smarter to choose the items they wish to purchase through ratings and reviews. Product ratings indicate the quality of the products. High-rated products are more likely to be trusted and purchased by customers. Ratings, when combined with reviews, have a powerful effect on purchase decisions (Gao, 2023). Customers tend to trust reviews from verified purchasers and those that offer detailed, balanced opinions (Qin & Zeng, 2023). There are two well-known methods used by the customers to evaluate products online, namely through numerical star ratings and written reviews. Star ratings deliver a measurable overview of consumer satisfaction, whereas written reviews furnish deeper insights into user experiences, emotions, expectations, and social dynamics (Gupta et al., 2010; Lopez & Garza, 2022). Collectively, these two modalities cultivate consumer trust, affect purchasing behavior, and enhance brand reputation in manners that frequently surpass the efficacy of conventional advertising (Park et al., 2021; Sharma, 2023).

Online reviews of products indicate an interesting discourse of customers to evaluate publicly. Reviewers do not simply provide objective information; they actively make meaning within a social context by balancing honesty with politeness, criticism with fairness, and emotion with credibility (Lopez & Garza, 2022; Thomas & Weyerer, 2019). Consequently, the connection between the linguistic sentiment of a text and its numerical rating is not always linear. Reviews often contain positive statements combined with average ratings, or critical comments softened by high scores. In consumer behavior research, sentiment-rating inconsistency is increasingly recognized not as a user error, but as a deliberate communicative practice driven by impression management and digital cultural norms (Chang et al., 2020; Schneider et al., 2026).

To understand why these inconsistencies happen, we need to examine online customer discussions through digital pragmatics and politeness theory. From a pragmatic view, evaluating a product publicly is a complex social act (Dong et al., 2022). According to politeness theory, criticism has the potential to undermine both the reputation of the brand and the public image of the reviewer (Yoon, 2020). Even when reviewers are anonymous, they often reduce these threatening situations by using positive politeness strategies (Brown & Levinson, 2006), like being indirect, vague, or using praise-criticism sandwiches. This approach usually leads to a milder numerical rating that contrasts with a very critical written review.

Furthermore, from a digital pragmatics perspective, evaluative language is used to convey attitude, intensity, and engagement (Abierto & Fortanet-gómez, 2022). Reviewers do not evaluate products in positive or negative values. They rather strategically amplify or soften their emotion as a form of impression management. These theoretical viewpoints indicate that the relationship between text and rating is scalar and interactionally negotiated (Meuleman et al., 2019; Ziser et al., 2023). Therefore, sentiment-rating inconsistencies demonstrate how reviewers strategically utilize evaluative language to navigate the demands of product assessment and social politeness in an enduring digital environment.

Although computational sentiment analysis has made significant progress in identifying positive and negative language in reviews, there has been limited focus on how textual sentiment corresponds with or differs from the numerical ratings assigned by reviewers (Khedkar, 2018; Raghunathan & Kandasamy, 2023). Statistical analyses can efficiently quantify positive or negative lexicon, they provide limited insight into the pragmatic intent, such as cultural modesty, social accountability, or politeness strategies that drives these rating choices. In particular, few studies have explored this relationship from a qualitative discourse perspective, examining how reviewers construct evaluation through both language and rating choices (Kashiha, 2023; Ziser et al., 2023). Automated sentiment analysis systems still face challenges in dealing with subtle linguistic features like irony, understatement, and hedging (Liu & Mahmud, 2021). Consequently, linguistically nuanced evaluations are frequently misclassified or oversimplified when sentiment is processed into positive-negative relationship. There is a continuing need for comprehensive qualitative research that concentrates on the social significance, pragmatic purpose, and language used in online customer reviews.

This research examines English language IKEA furniture reviews to investigate how reviewers conduct their evaluation within a commercial context that intertwines product evaluation, emotional expression, and politeness. IKEA serves as an ideal context for this research because furniture products uniquely blend practical function (usability and durability) with luxurious value (comfort and home identity), eliciting a broad and complex scope of evaluative feedback. The platform attracts a large, diverse customers base and uses a

1-5 star rating system, making it a rich environment for examining the linguistic evaluation. The novelty of this study lies in its application of Reflexive Thematic Analysis (RTA) to interpret the pragmatic face-work and impression management strategies within these reviews, shifting the focus from computational error to human communicative strategy.

Although previous studies have explored sentiment analysis computationally, few have examined how linguistic choices shape alignment or mismatch between sentiment and rating from an interpretive, qualitative perspective. This research offers pragmatic contribution by analyzing sentiment-rating relationships through RTA. Therefore, this research addresses the following research questions: How do reviewers linguistically construct alignment between textual sentiment and numerical star ratings in online product reviews?; What linguistic and pragmatic patterns characterize cases of sentiment-rating inconsistency? And How do these patterns contribute to understanding evaluative meaning in online review discourse?

RESEARCH METHOD

Research Design

This research employs a qualitative-dominant method design. Reflexive Thematic Analysis (RTA) by (Braun & Clarke, 2023) serves as the primary qualitative analytical technique, supported by a supplementary descriptive quantitative overview. The descriptive statistics (frequency counts of consistency categories) function to map the general landscape of the data and contextualize the prevalence of alignment patterns. The qualitative RTA then enables a deep analysis of the linguistic and pragmatic choices driving these patterns. This research considers language as a social construct that is motivated by pragmatic factors. Sentiment viewed not treated as an objective property of the text, but rather as a result of social interaction driven by linguistic elements such as politeness markers, contrastive structures and irony. The researchers adopt a reflective role, examining how these linguistic features contribute to the alignment/consistency or misalignment/inconsistency between written sentiment and star ratings.

Data Source and Corpus Description

Because this research analyzes public digital discourse rather than human subjects, the dataset consists of a corpus of online consumer reviews. The raw dataset included 7726 publicly available reviews for three best-selling IKEA furniture products, specifically white MICKE desk, MARKUS Office chair, and BRIMNES bed frame, retrieved from the official IKEA Australia website. These specific product categories were selected because they are best-sellers generate high-volume, diverse consumer interactions, providing a rich linguistic corpus that balances utilitarian function with hedonic home identity. From this larger population, a final corpus of 120 reviews was selected using a stratified purposive sampling strategy (Patton, 2002). The inclusion criteria required reviews to have both a star rating (1-5) and a written comment in English. The exclusion criteria removed blank ratings, duplicate entries, and irrelevant spam. The selection of exactly 120 reviews was justified to ensure balanced representation across all three product types and all five rating levels. Specifically, 40 reviews were selected per product category. Within each product's subset, exactly 8 reviews were selected for each of the five star-rating levels (1-star to 5-star). This exact distribution ensured equal representation across the satisfaction spectrum, while keeping the dataset at a manageable size for deep, manual qualitative analysis until thematic saturation was reached.

Research Instruments

The researchers acted as the primary instrument of data analysis, supported by four specific tools to ensure consistency throughout the analytical process. First, a coding framework and guidelines were used as a set of criteria to identify and label evaluative language, contrastive markers, politeness strategies, and emotional expressions within the text. Second, a sentiment–rating classification rubric was applied to establish operational guidelines for categorizing reviews into one of four alignment types: consistent, partially consistent, inconsistent, or neutral. Third, a thematic analysis protocol based on Braun and Clarke's (2023) six-phase reflexive thematic analysis framework was employed to guide the systematic progression from raw data to finalized themes. Fourth, a reflexive journal was used as a continuous memoing tool to document interpretive shifts, recognize subjective biases, such as cultural interpretations of politeness, and record decision-making processes when coding ambiguous language, including irony and sincere praise.

Data Collection Techniques

Data retrieval was conducted manually from the official IKEA Australia platform in December 2025 to March 2026, covering reviews published from 2023 to 2026. Each review was extracted and copied into a secure, encrypted dataset file. Ethical guidelines for internet-mediated research were strictly followed: because the data was publicly available, informed consent was not required, but all data was thoroughly anonymized. Usernames and any potentially identifiable personal information were removed prior to analysis. The dataset was then systematically condensed to prepare for coding. The data condensation process involved filtering out reviews that contained typographical errors, doubled punctuations and simple emojis. The finalized corpus was then organized by product type and star rating.

Data Analysis Technique

Data analysis was conducted in two distinct but complementary stages: Descriptive Classification and Reflexive Thematic Analysis (RTA). First, the entire subset of 120 reviews underwent preliminary coding to identify patterns of sentiment-rating consistency. Using the classification rubric, each review was labeled based on the relationship between its text and its numerical score:

Table 1
Scoring Criteria of sentiment-rating Alignment

Consistent	Textual sentiment strongly matches the star rating.
Partially Consistent	Text and rating generally align, but the tone is softened or mitigated.
Inconsistent	Direct divergence between the linguistic sentiment and the numerical rating.
Neutral/Mixed	Minimal evaluative information or factual brevity.

This stage provided a descriptive statistical overview (proportions and counts) of the alignment patterns and allowed researchers to select representative examples for deeper qualitative investigation.

The secon stage is reflexive thematic analysis (RTA). It was employed to interpret the pragmatic intentions that created gaps between the written review texts and their numerical ratings. This stage moved beyond identifying whether the sentiment of a review aligned with its rating and focused instead on understanding how reviewers constructed meaning through language. Guided by Braun and Clarke's (2023) six-phase RTA framework, the analysis began with a process of familiarization. The researchers repeatedly read the corpus to gain a close understanding of the reviews, paying attention to tone, word choice, structural markers, and the ways evaluative expressions corresponded with or diverged from the assigned ratings. During this phase, initial impressions were recorded, particularly when a review appeared positive in wording but was paired with a low rating, or when a seemingly critical review was accompanied by a relatively high score.

After becoming familiar with the data, the researchers generated initial codes by systematically labeling relevant segments of the review texts. Both semantic and latent meanings were considered in this process. Semantic codes captured explicit evaluative expressions, such as direct praise, complaints, satisfaction, disappointment, or neutral descriptions. Latent codes focused on deeper pragmatic meanings, including politeness strategies, implicit dissatisfaction, understated approval, emotional intensity, and the use of contrastive markers such as “but,” “however,” and similar linguistic cues. Particular attention was given to how reviewers softened criticism, emphasized emotional responses, or shifted from praise to complaint within the same review. This coding process enabled the researchers to identify not only what reviewers said, but also how they framed their judgments in relation to the ratings they provided.

The next stage involved searching for broader themes by clustering related codes into meaningful pragmatic patterns. At this point, the researchers examined how reviewers constructed evaluations that were either consistent or inconsistent with their numerical ratings. Reflexive memoing through the reflexive journal played an important role during this phase. The researchers used the journal to reflect critically on how their own cultural assumptions, especially regarding politeness, indirect criticism, and acceptable forms of dissatisfaction, might influence the interpretation of the data. This reflexive process helped ensure that theme development remained grounded in the reviewers’ actual language rather than being shaped solely by the researchers’ prior assumptions.

The candidate themes were then reviewed against the coded extracts and the entire dataset to ensure that they accurately represented the linguistic realities of the reviews. This phase required careful comparison between individual review segments and the emerging thematic structure. The researchers examined whether each theme was supported by sufficient evidence and whether the boundaries between themes were clear. The reflexive journal was again used to document interpretive decisions, especially in cases where the meaning of a review was ambiguous. For example, some expressions required closer examination to determine whether they reflected subtle irony, genuine praise, polite dissatisfaction, or a neutral comment with limited emotional content.

After the review process, the researchers defined and named the final themes by identifying their distinctive linguistic and pragmatic features. Four themes were finalized: Polite Dissatisfaction, Understated Positivity, Expressive Dissatisfaction, and Low-Information Reviews. Each theme represented a different way in which reviewers negotiated the relationship between textual sentiment and numerical rating. Polite Dissatisfaction captured reviews that used softened or indirect language to express negative experiences. Understated Positivity referred to reviews that conveyed satisfaction in a restrained or modest manner. Expressive Dissatisfaction represented reviews with strong negative emotional expression, while Low-Information Reviews described brief or vague responses that provided limited evaluative detail.

The final stage involved producing the report by synthesizing the themes into a coherent interpretation of the findings. In this phase, the researchers integrated the analytical results with notes from the reflexive journal to explain how online consumer culture shaped the ways reviewers expressed satisfaction, dissatisfaction, politeness, and emotional judgment. The report did not simply present themes as isolated categories, but interpreted them as evidence of broader pragmatic practices in digital review writing. Through this process, the analysis revealed how consumers use language strategically to balance honesty, politeness, emotional expression, and rating behavior in online review contexts.

RESEARCH FINDINGS AND DISCUSSION

Research Findings

This research analyzed 120 IKEA online reviews to investigate the linguistic patterns of sentiment-rating alignment. The results indicate that while a majority of reviews exhibit a direct correlation between evaluative lexis and numeric scores, a significant subset demonstrates pragmatic inconsistency. Notably, 1-star reviews showed the highest stability (strong negative alignment), while 4-star reviews emerged as the primary site of sentiment-rating mismatch. Rather than reflecting general consumer behavior, these findings illustrate how reviewers utilize specific linguistic strategies, such as hedging, superlative lexis, and contrast markers, to negotiate the gap between their textual experience and the platform's numeric constraints. Overall, the descriptive overview shows that 4-star reviews contain the highest proportion of sentiment-rating inconsistencies, while 1-star reviews indicate the strongest alignment between textual sentiment and rating. From the 120 reviews analyzed, four major evaluative patterns emerged.

Table 2
Linguistic Patterns of Alignment in the IKEA Corpus

Consistency Type	Examples	Interpretation
Consistent	"perfect desk" 5 stars "Good quality and size, fits." 4 stars "The chair itself is alright, but the wheels don't turn well." 3 stars "Not worth it. Very heavy product, quality not very good." 2 stars "difficult to assemble…" 1 star	Strong alignment between positive lexical evaluation and rating
Partially	"Suitable for the job.Not a bad desk. Pretty easy to put together once all the pieces are sorted." 5 stars ""The drawer units are good although narrow." 4 stars "Perfect for my needs" 3 stars ""Junk furniture that serves its purpose." 2 stars	Negative tone matches low rating but softened tone prevents it from being harsher
Inconsistent	"Good product." 5 stars "Excellent!" 4 stars "Perfect for my needs." 3 stars "The draws are so good for the chidren run nice too, shelving great not too high." 2 stars	Strong positive language but not maximum rating
Neutral/Mixed	"good" 5 stars "had this desk before." 3 stars "Wiggly." 1 stars	Minimal evaluative information

Table 2 presents the linguistic patterns of alignment between textual review content and star ratings in the IKEA corpus. The table shows that consumer reviews do not always express evaluation in a direct or fully predictable way. In the consistent category, the language used in the review generally corresponds closely with the assigned rating. Reviews such as "perfect desk" with a five-star rating and "Good quality and size, fits" with a four-star rating demonstrate clear positive evaluations that match high ratings. Likewise, more critical comments such as "Not worth it. Very heavy product, quality not very good" and "difficult to assemble" are aligned with low ratings. This pattern suggests that in consistent reviews, the lexical choices, emotional tone, and rating value support one another, making the reviewer's overall judgment easy to interpret.

The partially consistent category shows a more complex relationship between language and rating. In these reviews, the textual evaluation and rating are generally related, but the tone is moderated by hedging, qualification, or mixed assessment. For example, the review "Suitable for the job. Not a bad desk. Pretty easy to put together once all the pieces are sorted" is paired with a five-star rating, although the phrase "not a bad desk" expresses approval in a restrained rather than strongly enthusiastic way. Similarly, "The drawer units are good although narrow" contains both positive and negative elements, indicating that the reviewer recognizes the product's usefulness while also noting a limitation. This pattern suggests that some consumers use moderate or softened language even when giving relatively high or low ratings, possibly because they want to provide a balanced evaluation rather than an absolute judgment.

The inconsistent category indicates a mismatch between the sentiment of the written text and the numerical rating. Several examples contain strongly positive language, such as “Good product,” “Excellent!,” and “Perfect for my needs,” but they are not always paired with the highest possible rating. The most striking example is the positive review “The draws are so good for the children run nice too, shelving great not too high,” which is paired with only two stars. This suggests that the written review alone may not fully explain the reviewer’s rating behavior. The inconsistency may result from missing contextual information, unmentioned product issues, misunderstanding of the rating scale, accidental rating selection, or cultural and pragmatic differences in how consumers express satisfaction. Therefore, inconsistent reviews require deeper interpretation because the surface-level sentiment does not fully correspond to the rating.

The neutral or mixed category consists of reviews with minimal evaluative information, making the relationship between text and rating more difficult to interpret. Short comments such as “good,” “had this desk before,” and “Wiggly” provide limited linguistic evidence about the reviewer’s full experience. Although “good” appears positive and “Wiggly” appears negative, these expressions are too brief to reveal the reasoning behind the rating. The phrase “had this desk before” is especially neutral because it describes prior ownership without clearly expressing satisfaction or dissatisfaction. This pattern shows that some reviewers use very brief comments that depend heavily on the star rating to communicate evaluation. As a result, neutral or low-information reviews are less useful for understanding the specific reasons behind consumer satisfaction or dissatisfaction.

The table demonstrates that alignment between review text and star ratings varies across the IKEA corpus. Some reviews show direct consistency, where positive or negative language clearly matches the rating. Others reveal partial consistency, where reviewers combine praise and criticism or use softened expressions. Inconsistent reviews show that positive language does not always lead to maximum ratings, suggesting that rating behavior may be influenced by factors not explicitly stated in the text. Meanwhile, neutral or mixed reviews highlight the limitations of brief review texts for sentiment interpretation. These findings indicate that analyzing online reviews requires attention not only to rating scores but also to linguistic features such as evaluative words, contrastive markers, emotional tone, politeness, and the amount of information provided.

Discussion

The underlying meaning of the findings indicates that the relationship between written evaluation and numerical star ratings is not a simple mirror image. In online review platforms, the star rating may appear to provide a clear and efficient summary of consumer satisfaction, but the written review often reveals a more layered and negotiable form of evaluation. Instead of functioning as a direct repetition of the rating, review texts frequently expand, qualify, soften, or even contradict the numerical score. This shows that online reviews operate as complex evaluative discourse in which numbers and language serve complementary, yet sometimes divergent, communicative purposes. A five-star rating may signal general satisfaction, but the accompanying text may still contain reservations. Similarly, a lower rating may be accompanied by polite or moderate wording that reduces the emotional harshness of the evaluation. Therefore, sentiment-rating inconsistency reveals that human consumer experience is multi-dimensional and cannot be fully compressed into a rigid 1-to-5 rating scale without a loss of nuance (Lopez & Garza, 2022; Sachdeva & Mcauley, 2026).

A critical takeaway from this study is that reviewers are not merely rating a product; they are also performing a social identity. When consumers write public reviews, they are aware, implicitly or explicitly, that their words are visible to other potential buyers and to the brand community. As a result, they often position themselves as reasonable, fair, and credible evaluators. The use of contrastive markers such as "but" and "however" indicates that reviewers do not always present a one-sided opinion. Instead, they often balance praise with criticism or criticism with acknowledgment of some positive qualities. Likewise, mitigated language such as "not very good," "could be better," or "not perfect" allows reviewers to express dissatisfaction without sounding overly emotional or unfair. This suggests that reviewers strive to appear as objective and rational actors rather than impulsive consumers. By acknowledging both strengths and weaknesses, reviewers build ethos and credibility within the IKEA community (Thomas & Weyerer, 2019). In this sense, ratings are socially negotiated judgments rather than purely spontaneous emotional reactions.

This phenomenon is closely related to politeness and evaluation. The Polite Dissatisfaction pattern found in the two-star reviews reveals a significant use of politeness strategies to save face. According to Brown and Levinson's politeness theory, a public complaint can be understood as a face-threatening act because it has the potential to damage the image of both the reviewer and the product provider. The reviewer risks appearing rude, unreasonable, or excessively demanding, while the company is publicly criticized. To manage this tension, reviewers often use softened expressions to communicate dissatisfaction in a socially acceptable way. Phrases such as "could be better" or "not perfect" reduce the directness of criticism and allow the reviewer to express a negative judgment without appearing aggressive. This strategy is especially important in public digital spaces where consumer comments are not private complaints but visible contributions to a wider community conversation. Through polite dissatisfaction, reviewers maintain their image as balanced consumers while still communicating disappointment (Baker & Hashimoto, 2024).

In contrast, the Understated Positivity observed in four-star reviews functions as a site of pragmatic modesty and impression management. These reviews often contain highly positive language, such as "excellent," "perfect for my needs," or "great," but the reviewer does not assign the maximum five-star rating. At first glance, this may appear inconsistent because the linguistic evaluation seems stronger than the numerical score. However, this pattern can be interpreted as a form of evaluative restraint. In online discussion, reviewers frequently adjust their language to soften or enhance their emotional expression depending on how they want to be perceived by others (Abierto & Fortanet-gómez, 2022). By using highly positive words while withholding the highest rating, reviewers may communicate that they are satisfied but still maintain strict evaluative standards. This scalar grading approach suggests that the reviewer reserves a perfect score for products that are not only useful or satisfying but flawless. In doing so, the reviewer presents themselves as careful, discerning, and credible rather than easily impressed (Meuleman et al., 2019).

These findings suggest that a numerical average may not always provide a reliable indicator of customer satisfaction. In many digital marketplaces, consumers and companies rely heavily on average ratings to evaluate product quality. However, this study shows that a four-star rating may not necessarily indicate moderate satisfaction or a minor deficiency. In some cases, it may represent an almost perfect experience filtered through the reviewer's desire to appear selective or balanced. Similarly, a three-star rating may not always represent neutrality, because the accompanying text may contain strong praise, strong criticism, or a mixture of both. Therefore, interpretation should prioritize the linguistic evidence contained in review texts rather than relying exclusively on numerical scores. The written component can reveal hesitation, politeness, irony, restraint, disappointment, or hidden satisfaction that the rating alone cannot capture. This is particularly important for businesses that use ratings to make product decisions, because the average score may obscure the actual reasons behind consumer evaluation.

Placing this study within previous literature, the findings reinforce Liu et al. (2022), who suggested that extreme ratings, particularly one-star and five-star reviews, tend to show the highest sentiment alignment because they are usually driven by strong affective intensity. Consumers who give one star often express clear dissatisfaction, frustration, or anger, while those who give five stars often communicate clear approval and satisfaction. The present findings support this pattern by showing that reviews at both ends of the scale are more likely to display stronger alignment between language and rating. However, this study also departs from Gupta et al. (2010) by showing that the three-star middle ground is often not neutral in a strict evaluative sense. Instead, it may represent functional brevity, mixed judgment, uncertainty, or conflicting experiences. In other words, the middle rating does not automatically indicate balanced neutrality. Aligning with Ruytenbeek and Decock (2024), this research argues that sentiment-rating mismatches are not merely data errors or noise, but intentional linguistic strategies that reflect how reviewers manage meaning in public review spaces.

To improve future qualitative sentiment analysis, researchers must account for pragmatic nuance. Sentiment analysis should not treat words such as “good,” “excellent,” “but,” or “however” as having fixed meanings across all rating contexts. The function of a contrast marker depends on where it appears, what it contrasts, and how it relates to the star rating. For instance, the contrast marker “but” in a four-star review may carry a lower negative value because it often introduces a minor reservation after an otherwise positive evaluation. In contrast, “but” in a two-star review may function as part of a face-saving strategy, where the reviewer softens dissatisfaction by acknowledging a small positive aspect before expressing a stronger criticism (Atteveldt & Velden, 2022). This means that algorithms must be trained to weigh evaluative constraints, rating contexts, and discourse patterns rather than simply detecting isolated sentiment words. A four-star rating may actually signal a five-star experience filtered through a reviewer’s desire to appear discerning, while a polite two-star review may conceal strong dissatisfaction behind moderated language.

Future research should expand the IKEA corpus to include broader and more varied contexts. Cross-platform analysis would be valuable because review behavior may differ across product types, brands, and service environments. For example, IKEA furniture reviews could be compared with reviews of luxury brands, hotels, restaurants, or digital services to examine whether politeness norms and rating practices change across consumer domains. Cross-cultural pragmatics should also be explored, especially to investigate whether Understated Positivity is more common in certain varieties of English or cultural contexts, such as British English compared with American English. Cultural expectations surrounding modesty, politeness, complaint behavior, and praise may influence how reviewers align or misalign language and rating. Product category analysis is also important because high-cost or high-effort items may trigger more expressive dissatisfaction than low-cost items. Furniture that requires assembly, for instance, may generate more emotionally loaded reviews because the consumer experience includes not only purchase but also delivery, effort, usability, and durability.

The core contribution of this manuscript is the establishment of a nuanced framework that redefines sentiment-rating alignment as a deliberate communicative act. Rather than treating review texts as simple explanations of star ratings, this study shows that written evaluations perform additional pragmatic work. One-star reviews may provide a platform for Affective Alignment, where strong negative emotion and low numerical rating reinforce each other. Four-star reviews, meanwhile, may serve as a platform for Evaluative Constraints, where reviewers communicate satisfaction while withholding perfection as a sign of high standards. By analyzing 120 IKEA reviews, this research provides a framework for understanding how consumers navigate the gap between subjective experience and the rigid numerical constraints of a platform. Ultimately, written text does not simply support ratings; it often corrects, qualifies, or adds detail to those ratings. This suggests that to truly understand the customer's voice, the word "but" in a review may be just as important as the numerical star attached to it.

CONCLUSION

This study concludes that sentiment-rating alignment in online consumer reviews is a complex pragmatic phenomenon rather than a simple correspondence between written evaluation and numerical score. The analysis of the IKEA review corpus shows that consumers use written language not only to explain their ratings, but also to negotiate meaning, manage social identity, and position themselves as credible evaluators within a public digital space. Consistent reviews demonstrate clear alignment between evaluative language and star ratings, particularly at the extreme ends of the scale, where satisfaction or dissatisfaction is expressed more directly. However, partially consistent, inconsistent, and neutral reviews reveal that numerical ratings often fail to capture the full range of consumer experience. Expressions of praise, complaint, hesitation, politeness, irony, and contrastive evaluation can either reinforce or complicate the meaning of a rating. Therefore, star ratings should not be interpreted as independent or complete indicators of customer satisfaction. The written review remains essential because it provides the linguistic and pragmatic context needed to understand why consumers evaluate a product in a particular way.

The findings also highlight the importance of pragmatic interpretation in both qualitative review analysis and automated sentiment analysis. Patterns such as Polite Dissatisfaction, Understated Positivity, Expressive Dissatisfaction, and Low-Information Reviews show that consumers strategically use language to balance honesty, politeness, emotional expression, and public self-presentation. A four-star review may reflect a highly positive experience filtered through evaluative restraint, while a two-star review may express dissatisfaction in softened language to avoid sounding overly harsh. Similarly, a three-star review may not represent true neutrality, but rather ambiguity, brevity, or conflicting evaluations. These findings suggest that future research and sentiment-analysis models should move beyond surface-level word polarity and consider discourse markers, rating context, politeness strategies, and pragmatic intention. In practical terms, companies should read consumer reviews as layered communicative acts, not merely as numerical feedback. By paying attention to both the rating and the language that accompanies it, businesses, researchers, and platform designers can gain a more accurate understanding of the customer's voice.

INFORMED CONSENT STATEMENT

This study analyzed publicly available online reviews from the IKEA Australia website. No personal identifying information was collected, and the study followed ethical guidelines for internet-mediated research.

DECLARATION OF USING AI TOOLS

The author utilized SciSpace for literature discovery and Grammarly for proofreading and language editing during the preparation of this manuscript. The author assumes full accountability for the final content and academic integrity of this paper.

REFERENCES

Abierto, A., & Fortanet-Gómez, I. (2022). Applying appraisal theory for the interpretation of experienced researchers' interviews on open access. *Revista Signos, 55*(109), 481–500. https://doi.org/10.4067/S0718-09342022000200481

Atteveldt, V., & Velden, V. D. (2022). The validity of sentiment analysis: Comparing manual annotation, crowd-coding, dictionary approaches, and machine learning algorithms. https://doi.org/10.1080/19312458.2020.1869198

Baker, M. J., & Hashimoto, B. (2024). Expression of customer (dis)satisfaction in online restaurant reviews: The relationship between adversative connective constructions and star ratings. https://doi.org/10.1177/23294884231200245

Braun, V., & Clarke, V. (2023). Using thematic analysis in psychology. *Qualitative Research in Psychology, 3*(2), 77–101. https://psychology.ukzn.ac.za/?mdocs-file=1176

Brown, P., & Levinson, S. C. (2006). Politeness theory. In A. Jaworski & N. Coupland (Eds.), *The discourse reader* (2nd ed., p. 547). Routledge.

Chang, J. P., Cheng, J., & Danescu-Niculescu-Mizil, C. (2020). Don't let me be misunderstood: Comparing intentions and perceptions in online discussions. *Proceedings of The Web Conference 2020*, 2066–2077. https://doi.org/10.48550/arXiv.2004.13609

Dong, L., Ji, T., & Zhang, J. (2022). Effects of conversation politeness on hiring decision in online labor markets: An inverted U-shaped relationship exploration.

Gao, H. (2023). The impact of topological structure, product category, and online reviews on co-purchase: A network perspective, 548–570.

Gupta, N., Di Fabbrizio, G., & Haffner, P. (2010). Capturing the stars: Predicting ratings for service and product reviews, 36–43.

Kashiha, H. (2023). Beyond words in evaluation: A study of research articles across disciplines. *Journal of English for Academic Purposes, 27*(2), 251–275.

Khedkar, B. (2018). Sentiment analysis of customer product reviews classification techniques: A review. *International Journal of Computer Sciences and Engineering, 5*(11), 148–150.

Liu, Z., Courant, R., & Kalogeiton, V. (2022). FunnyNet: Audiovisual learning of funny moments in videos. HAL. https://hal.science/hal-03839553

Liu, Z., & Mahmud, J. (2021). When and why does a model fail? A human-in-the-loop error detection framework for sentiment analysis, 170–177.

Lopez, A., & Garza, R. (2022). Do sensory reviews make more sense? The mediation of objective perception in online review helpfulness. https://doi.org/10.1108/JRIM-04-2021-0121

Meuleman, B., Moors, A., Fontaine, J. R. J., & Renaud, O. (2019). Interaction and threshold effects of appraisal on componential patterns of emotion: A study using cross-cultural semantic data. https://doi.org/10.1037/EMO0000449

Park, Y. J., Joo, J., Polpanumas, C., & Yoon, Y. (2021). "Worse than what I read?" The external effect of review ratings on the online review generation process: An empirical analysis of multiple product categories using Amazon.com review data.

Patton, M. Q. (2002). *Qualitative research and evaluation methods* (3rd ed.). SAGE Publications.

Qin, C., & Zeng, X. (2023). Do live streaming and online consumer reviews jointly affect purchase intention?

Raghunathan, N., & Kandasamy, S. (2023). Challenges and issues in sentiment analysis: A comprehensive survey. *IEEE Access, 11*, 1. https://doi.org/10.1109/ACCESS.2023.3293041

Ruytenbeek, N., & Decock, S. (2024). Expressing and responding to customer (dis)satisfaction online: New insights from discourse and linguistic approaches. https://doi.org/10.1177/23294884231199740

Sachdeva, N., & McAuley, J. (2026). How useful are reviews for recommendation? A critical review and potential improvements. https://doi.org/10.1145/3397271.3401281

Schneider, C., Weinmann, M., Mohr, P. N. C., & Brocke, J. (2026). When the stars shine too bright: The influence of multidimensional ratings on online consumer ratings.

Sharma, S. (2023). Boosting sales and building trust: Leveraging product reviews for marketing success. In *Managing product reviews: A comprehensive guide for brands and businesses* (pp. 115–134).

Thomas, M., & Weyerer, J. C. (2019). Determinants of online review credibility and its impact on consumers' purchase. *Journal of Electronic Commerce Research, 20*(1), 1–20.

Yoon, E. J. (2020). Polite speech emerges from competing social goals.

Ziser, Y., Webber, B., & Cohen, S. B. (2023). Rant or rave: Variation over time in the language of online reviews. *Language Resources and Evaluation, 57*(3). https://doi.org/10.1007/s10579-023-09652-5