When Every Rating Is a Five: How Well Do Consumer Reviews Capture Service Quality in Credence Markets?

Sachin Sridhar
Indian Institute of Management Bangalore
Rex Du
University of Texas at Austin

Cite as:
Sridhar Sachin, Du Rex (2026), When Every Rating Is a Five: How Well Do Consumer Reviews Capture Service Quality in Credence Markets?. *Proceedings of the European Marketing Academy*, 55th, (133677)

Paper from the 55th Annual EMAC Conference, Bath, United Kingdom, June 2-5, 2026

When Every Rating Is a Five: How Well Do Consumer Reviews Capture Service Quality in Credence Markets?

Abstract

User ratings in credence services often inflate toward perfection, obscuring true provider quality. Using 73,236 verified real estate reviews linked to transaction-level outcomes, we test whether review text conveys diagnostic information beyond near-universal five-star ratings. Affective praise is abundant and predicts only faster sales. Experiential praise is common yet unrelated to price or speed and, in the case of empathy, signals worse outcomes. In contrast, rare references to credence outcomes and comparative performance reliably predict higher prices and shorter time on market. A MaxDiff study reveals a mismatch between objective and perceived diagnosticity: consumers overweight affective cues and undervalue comparative signals. Across analyses, specific textual cues, as opposed to numerical ratings, capture meaningful quality differences in credence markets.

Keywords: *Consumer Reviews, Service Quality, Credence Markets*
Track: *Service Marketing & Service Innovation*

1. Objective and Methodology

When selecting providers for high-stakes services, such as physicians, attorneys, financial advisors, or real estate agents, consumers often turn to user reviews as a primary source of information. These reviews are available not only on general-purpose platforms (e.g., Google, Yelp) but also on industry-specific sites (e.g., RateMDs for healthcare, Lawyers.com for legal services, WalletHub for financial planning, and Realtor.com and RateMyAgent for real estate).

User reviews can help inform consumer decisions only when customers are able to evaluate the quality of the services they receive. In *credence services*, where quality is difficult to judge even after consumption (e.g., a medical procedure, legal strategy, retirement plan, or home sale), such evaluations are especially challenging (Darby and Karni, 1973; Dulleck and Kerschbamer, 2006). Nevertheless, most platforms present user reviews as key indicators of provider quality. This tension raises a central question: Do user reviews accurately reflect service quality, or are they inherently unreliable measures in credence markets?

We examine this question in the context of residential real estate, focusing on home sellers' reviews of listing agents—the professionals responsible for marketing and selling properties. Real estate provides a compelling empirical setting for three reasons. First, the financial stakes are high: a home is often a seller's most valuable asset, and listing agents typically earn commissions of 2.5%–3% of the sale price in the United States, for example. Second, the service is credence-based. Although sellers can observe outcomes such as the final sale price and time to close, they cannot easily determine whether a different agent—or selling without representation—would have achieved better results. Third, because sellers face these constraints in evaluating service quality, it is unclear whether their reviews capture an agent's true contribution to the transaction, such as obtaining a higher price or facilitating a faster sale.

In this research, we adopt an empirics-first approach (Golder et al., 2022) for two reasons. First, existing theory provides limited guidance on whether user reviews accurately capture the underlying quality of credence service providers, rendering this a fundamentally empirical question. Second, our inquiry is motivated by patterns in a rich dataset comprising 73,236 seller reviews of 4,360 listing agents over a span of 14 years (2010 and 2023).

Each review is matched one-to-one with a verified transaction record, enabling us to link subjective evaluations with objective outcomes—specifically, sale price and days on market (DOM)—at the transaction level rather than the agent level. This distinction—agent-level versus transaction-level matching—is critical because the quality of service delivered by the same service provider can vary substantially across transactions. Without transaction-level matching, analyses are limited to cross-provider variation, overlooking within-provider differences across individual service encounters.

The reviews analyzed in this study originate from the digital platform of a Multiple Listing Service (MLS) in a major U.S. metropolitan area. The MLS, which is publicly accessible, serves as the most comprehensive and authoritative source of property and agent information in the regional housing market. Each seller review is verified by the platform and includes 1–5 star ratings on four dimensions: competency, market knowledge, communication, and overall experience. In 71% of cases (52,174 reviews), sellers also provide open-ended comments.

This granular linkage between subjective evaluations and objective, transaction-level outcomes enables us to extend prior research, which has largely relied on indirect proxies—such as helpfulness votes—to infer review informativeness. Although such proxies are useful, they do not fully capture the diagnostic value of review content (Ananthakrishnan et al., 2023; Lantzy et al., 2021).

Consistent with our empirics-first approach, the scope of this research is shaped by patterns that emerge from the data rather than by theory-driven priors. Among the 73,236 seller reviews in our dataset, 95% assign five stars on all four rated dimensions. The pattern is even stronger in the 52,174 reviews that include open-ended comments—97% of these also give perfect ratings. This extreme concentration of positive evaluations makes it difficult for prospective sellers to distinguish exceptional agents from average ones, a form of ratings inflation. Prior research attributes such inflation to selection bias: satisfied customers are more likely to post reviews, and service providers often solicit feedback only from satisfied clients (Aziz et al., 2022). Comparable patterns have been observed in other two-sided platforms, including Uber and Airbnb (Ke, 2017).

Noting the highly inflated numerical ratings, we next examine whether open-ended review texts contain diagnostic signals of service quality. To address this question, we link each review to its corresponding transaction and test whether common expressions of praise in review narratives are systematically associated with objectively superior outcomes in terms of higher sale prices and shorter DOM. In doing so, we provide a direct test of the diagnostic value of consumer review text in real estate services, a prototypical credence market.

Table 1 presents the ten textual attributes we examine—four experiential, three credence, and three meta-evaluative—along with their definitions and representative phrases. Our aim is not to propose a new or comprehensive taxonomy but to offer a structured set of attributes that (1) is grounded in established frameworks, (2) reflects the realities of credence services such as real estate, and (3) corresponds to the content most frequently observed in our data. This ten-attribute framework enables supervised text classification to extract interpretable signals from review texts and to test systematically their diagnostic value in assessing listing-agent service quality.

Table 1: Textual Attributes Extracted from Seller Reviews of Listing Agents

#	Attribute	Definition	Classification	Illustrative Examples
1	Responsiveness	Timeliness, availability, proactivity	Experiential	"Quick to respond," "Stayed on top of things," "Communication was excellent"
2	Assurance	Confidence-building, making clients feel secure	Experiential	"Made us feel confident," "Reassured," "Felt we were in great hands."
3	Empathy	Compassion, understanding, personalized care	Experiential	"Listens," "Personable," "Puts clients first"
4	Reliability	Dependability, consistency, delivering on promises	Experiential	"Everything went as planned," "Handled everything," "No surprises"
5	Technical Expertise	Knowledge, skill, expertise	Credence: Competency evaluation	"Professional," "Knowledge," "Expertise," "Skill," "Top quality," "Great negotiator"
6	Price Outcomes	Sale price	Credence: Outcome evaluation	"Good price," "Pleased with the price"
7	Speed Outcomes	DOM	Credence: Outcome evaluation	"Sold fast," "Quick closing"
8	Comparative Performance	Better than peers, exceeding expectations	Meta-Evaluative: Comparative evaluation	"Far better than expected," "Couldn't have asked for a better experience," "Went above and beyond"
9	Future Recommendation	Willing to recommend or reuse	Meta-Evaluative: Future intent	"Would totally use her services again," "Highly recommend her"
10	Positive Affect	Positive emotions	Meta-Evaluative: Affective praise	"Amazing," "Fantastic," "Excellent," "Outstanding," "Awesome"

In our empirical analyses, we include extensive controls and fixed effects to mitigate concerns about confounding in agent selection and transaction outcomes. For the property sold, we capture characteristics such as the number of bedrooms and bathrooms, lot size, and square footage. We also record transaction-specific details, including the initial listing price, subsequent price changes, and whether the property was relisted. Property addresses further enable linkage to neighborhood attributes such as ZIP code and subdivision.

Collectively, these variables serve as controls that help isolate the listing agent's influence on sale price and DOM, allowing for a more reliable assessment of service quality at the transaction level. Our regression models manage to account for up to, respectively, 96% of the variance in sale price and 67% of the variance in DOM, levels comparable to those reported in prior research on real estate agents and housing transactions (e.g., Agarwal et al., 2019; Goldsmith-Pinkham and Shue, 2023; Levitt and Syverson, 2008).

2. Major Results and Implications

Four key patterns emerge. First, consistent with ratings inflation, most review texts contain affective stand-alone adjectives (e.g., amazing, fantastic, excellent, outstanding, and awesome). Although this form of praise conveys little information about the sale price achieved by the agent, it is associated with faster sales. Thus, while affective language does not signal an agent's ability to secure a higher price, it may capture aspects of service quality that facilitate shorter DOM.

Second, sellers frequently highlight experiential aspects of the service encounter, praising agents for being responsive, reassuring, empathetic, or reliable. This emphasis is unsurprising, as such attributes are directly observable through interpersonal interaction (Lantzy et al., 2021). However, praise for these experiential qualities is unrelated to either sale price or DOM. More strikingly, all else equal, when an agent is praised for empathy, transactions close at lower prices and take longer to complete. This pattern suggests that experiential praise is not merely uninformative but may even mislead consumers regarding an agent's effectiveness in achieving superior transaction outcomes.

Third, sellers comment less frequently on credence attributes, such as an agent's technical expertise or effectiveness in achieving a high sale price or rapid closing. This relative scarcity likely reflects sellers' limited ability to assess counterfactuals—that is, what might have occurred had they selected a different agent or sold the property themselves. Nevertheless, when reviews explicitly praise an agent for securing a high price or completing a sale quickly, such statements are positively associated with superior transaction outcomes, holding other factors constant. This pattern challenges the common assumption that consumers cannot accurately assess the true contributions of credence service providers such as listing agents (Darby and Karni, 1973). Instead, it suggests that when consumers comment directly on measurable performance outcomes, their evaluations may be more diagnostic than previously assumed.

Fourth, sellers often express satisfaction by stating their willingness to recommend the agent or work with them again. To a lesser extent, they commend agents for exceeding expectations or outperforming others. Empirically, a seller's stated willingness to recommend or reuse an agent provides little diagnostic value. In contrast, holding other factors constant, reviews noting that an agent exceeded expectations or outperformed prior agents are consistently associated with both higher sale prices and shorter DOM.

Taken together, these results reveal systematic variation in the diagnostic value of review content. Affective praise, while abundant, primarily conveys enthusiasm but provides little insight into an agent's ability to achieve higher prices—though it does correspond to faster sales. Experiential praise, despite being common and easily observed, is unreliable and may even mislead prospective sellers. In contrast, performance- or credence-based praise, though relatively rare, is consistently associated with superior transaction outcomes. Similarly, reviews noting that an agent exceeded expectations or outperformed others—unlike generic statements of willingness to recommend or reuse—correlate with both higher sale prices and shorter DOM.

Collectively, these empirical patterns underscore a key implication for credence-based platforms: the diagnostic value of review text depends less on its frequency or valence and more on whether it explicitly references transaction outcomes or comparative performance.

Having established which common forms of praise in reviews are—and are not—associated with superior transaction outcomes, a natural next question is whether consumers can reliably distinguish signal from noise when interpreting review texts. Consistent with our empirics-first approach, we extend the analysis to test whether the *objective* diagnosticity of review content aligns with its *perceived* diagnosticity.

To this end, we field a MaxDiff (best–worst scaling) survey of U.S. homeowners (Louviere et al., 2013) to assess which forms of praise they perceive as most useful when screening listing agents. The results reveal both convergence with and divergence from the transaction-based analyses. Consistent with our findings on objective diagnosticity, performance-oriented praise referencing transaction outcomes is perceived as most diagnostic, whereas willingness to recommend or reuse an agent is rated least valuable—mirroring its limited actual diagnostic value. In contrast, affective praise, despite its weak association with objective outcomes, is perceived as more diagnostic than most experiential attributes (reliability excepted), while comparative praise, though more strongly associated with outcomes, carries relatively low perceived diagnosticity.

Our empirical findings—regarding both the relative prevalence of different forms of praise in reviews and the contrast between their objective and perceived diagnosticity—carry important implications for stakeholders across the credence services ecosystem.

For consumers (prospective home sellers in our context), the results clarify how to read between the lines, highlighting which elements of review text provide reliable indicators of provider quality and which may be misleading. For platforms, the findings identify opportunities to refine interface features that surface authentic quality signals embedded in open-ended comments—particularly in environments characterized by inflated numerical ratings. For service providers (listing agents in our context), the results offer guidance on interpreting textual feedback and using it to enhance service quality and client relationships.

3. Key References

Agarwal, Sumit, Jia He, Tien Foo Sing, and Changcheng Song (2019), "Do Real Estate Agents Have Information Advantages in Housing Markets?," *Journal of Financial Economics*, 134 (3), 715–735.

Ananthakrishnan, Uttara, Davide Proserpio, and Siddhartha Sharma (2023), "I Hear You: Does Quality Improve with Customer Voice?," *Marketing Science*, 42 (6), 1143–1161.

Aziz, Arslan, Hui Li, and Rahul Telang (2022), "The Consequences of Rating Inflation on Platforms: Evidence from a Quasi-Experiment," *Information Systems Research*, 34 (2), 590–608.

Darby, Michael R. and Edi Karni (1973), "Free Competition and the Optimal Amount of Fraud," *The Journal of Law and Economics*, 16, 67–88.

Dulleck, Uwe, and Rudolf Kerschbamer (2006), "On Doctors, Mechanics, and Computer Specialists: The Economics of Credence Goods," *Journal of Economic Literature*, 44 (1), 5–42**.**

Golder, Peter N., Marnik G. Dekimpe, Jake T. An, Harald J. van Heerde, Darren S. U. Kim, and Joseph W. Alba (2022), "Learning from Data: An Empirics-First Approach to Relevant Knowledge Generation," *Journal of Marketing*, 87 (3), 319–336.

Goldsmith-Pinkham, Paul and Kelly Shue (2023), "The Gender Gap in Housing Returns," *Journal of Finance*, 78 (2), 1097–1145.

Ke, Qing (2017), “Sharing Means Renting?: An Entire-Marketplace Analysis of Airbnb,” *Proceedings of the 2017 ACM on Web Science Conference*, 131–139.

Lantzy, Shannon, Rebecca W. Hamilton, Yu-Jen Chen, and Katherine Stewart (2021), “Online Reviews of Credence Service Providers: What Do Consumers Evaluate, Do Other Consumers Believe the Reviews, and Are Interventions Needed?,” *Journal of Public Policy and Marketing*, 40 (1), 27–44.

Levitt, Steven and Chad Syverson (2008), “Market Distortions When Agents Are Better Informed: The Value of Information in Real Estate Transactions,” *The Review of Economics and Statistics*, 90 (4), 599–611.

Louviere, Jordan, Ian N. Lings, Towhidul Islam, Siegfried P. Gudergan, and Terry N. Flynn (2013), “An Introduction to the Application of (Case 1) Best–Worst Scaling in Marketing Research,” *International Journal of Research in Marketing*, 30 (3), 292–303.