Customers' Review Content and Their Referral and (Re)Purchase Behaviors

Ruhan Liu, Assistant Professor of Marketing, Business School, Shanghai University of Finance and Economics, China, liuruhan@mail.shufe.edu.cn
Wanqing Zhang, Associate Professor of Marketing, Bayes Business School, City St George's, University of London, United Kingdom, Wanqing.Zhang@citystgeorges.ac.uk
Pradeep K. Chintagunta, Professor of Marketing, Booth School of Business, University of Chicago, United States, Pradeep.Chintagunta@chicagobooth.edu

Corresponding author: Wanqing Zhang
Author notes: All authors contributed significantly to this research. The authors have used generative AI to improve the manuscript's readability.
Acknowledgments: The authors sincerely thank the top executives and employees of our collaborator for their full support. The authors have benefitted from the advice of Jing Xu, Reid Hastie, Guofang Huang, Junhong Chu, Hongju Liu, Xing Li, Ying Lei, Qi Sun, Fang Wu, Xuezhen Tao, Jia Gai, Qiaowei Shen, Yang Pan, as well as seminar and class participants at Peking University, University of Chicago, University of London, Shanghai University of Finance and Economics, and INFORMS Marketing Science Virtual Conference 2021.

Statements and Declarations

Ethical Considerations

This study received ethical approval from the IRB of Guanghua School of Management, Peking University (approval #2024-12) on April 29th, 2024.

Consent to Participate

All participant information was de-identified, and participant consent was not required. Participant data will not be shared with third parties.

Declaration of Conflicting Interest

The authors declared no potential conflicts of interest regarding the research, authorship, and/or publication of this article.

Funding Statement

The authors disclosed receipt of the following financial support for the research, authorship, and/or publication of this article: This work was supported by *the Fundamental Research Funds for the Central Universities* and *the China Scholarship Council*.

Data Availability Statement

The data that support the findings of this study are available from the authors upon reasonable request and with the permission of our collaborator.

Customers' Review Content and Their Referral and (Re)Purchase Behaviors

Abstract

This research examines how customer review content influences review writers' subsequent decisions. Employing a mixed-methods approach, including a field experiment, a scenario experiment, and archival data analysis, the authors investigated the effects of affective content, cognitive content, and length of customers' reviews on their subsequent referral and (re)purchase behaviors across various contexts, such as household services, podcast trials, and airline services. The authors leveraged the random assignment of experimental interventions and other shifters to induce exogenous variation in review content features. Additionally, they employed instrumental and proxy variables to address endogeneity issues. Findings from the three studies consistently demonstrate that affective content in reviews enhances referral and repurchase behaviors, whereas cognitive content exerts adverse effects. Moderation analyses show that the effects of review length on these behavioral outcomes depend on individual and contextual factors that affect customers' elaboration likelihood during review writing. Overall, this research provides actionable insights for strategically shaping customer review content to drive critical business outcomes and enriches theoretical understanding of how content features of customer reviews affect review writers' decisions.

Keywords: customer review, referral, (re)purchase, affective processing, cognitive processing, elaboration likelihood, field experiments

Customer reviews are crucial to marketing strategies (Ananthakrishnan, Proserpio, and Sharma 2023; Kaul et al. 2025). The rise of digital platforms has broadened the reach and impact of reviews, affecting various customer behaviors, such as purchase decisions (Varga and Albuquerque 2024) and churn rates (Huang and Sudhir 2021), all of which affect company performance. In response, major platforms, such as *Amazon*, *Airbnb*, and *JD.COM*, are investing heavily in improving the quantity and quality of customer reviews (e.g., Fradkin and Holtz 2023; Gao et al. 2025; Web Appendix A summarizes various business practices related to soliciting customer reviews). Writing reviews has also become a common practice for many consumers: Over 70% contribute at least one review annually, indicating a substantial reviewer base (Ellis 2024; Paget 2025; Wu and Morwitz 2025). These reviewers provide strategic value to firms in terms of brand loyalty (Rocklage and Fazio 2020), brand advocacy (Fernández-Loría, Cohen, and Ghose 2023), informative feedback (Pocchiari, Proserpio, and Dover 2024), and social influence (Ma et al. 2022).

Despite extensive research on how review content affects *readers* (see Web Appendix B; e.g., Fradkin and Holtz 2023; Varga and Albuquerque 2024), its impact on review *writers* has received limited attention (Wu and Morwitz 2025). This gap limits our understanding of whether firms' investments in review generation are genuinely effective. To address this, we examine how the content features of reviews influence reviewers' subsequent referral and (re)purchase behaviors, which are key drivers of firm growth and profitability (Gershon and Jiang 2025). Drawing on the linguistic relativity (or Whorfian) hypothesis (Hunt and Agnoli 1991; Whorf and Carroll 1956), we propose that the linguistic content of reviews influences writers' decisions by shaping their perceptions and evaluations. Guided by the dual-process model (Kahneman and Frederick 2002), we focus on two dimensions of review content: (1) *affective content*, which includes expressions of emotions and affects, and (2) *cognitive content*, which involves descriptions of thought processes and objective features. We predict

that affective content fosters referral and (re)purchase behaviors by increasing reliance on emotions and heuristics in decision-making, whereas cognitive content inhibits these behaviors by prompting detailed and analytical processing (Chaiken 1980; Isen and Means 1983). This categorization of review content also aligns with the affective and cognitive orientations that businesses adopt when eliciting reviews. For instance, Netflix adopts an affective orientation by asking, "How did this movie make you feel?" while Amazon uses a cognitive approach with prompts like, "What are the pros and cons of this product? Be specific." (Additional examples are provided in Web Appendix A).

In addition, we examine the effects of review length, a straightforward indicator of review quality that marketers can easily influence through tools such as incentives (Burtch et al. 2018). Because review length is inherently correlated with affective and cognitive content, we incorporate it into our framework. Drawing on the elaboration likelihood model (ELM, Petty and Cacioppo 1986), we test how factors that affect consumers' motivation or ability to process information thoroughly (e.g., prior experience, product complexity, and price paid) moderate the relationship between customers' review length and their subsequent decisions. Our research framework is presented in Web Appendix C.

Methodologically, it is challenging to examine our research questions using observational data, because natural variation in review content is often endogenous. This endogeneity issue arises from multiple sources, such as omitted variables and measurement errors (Stock and Watson 2019; Web Appendix D details the primary sources of endogeneity and corresponding solutions). Typical examples of omitted variables are unobserved product quality and consumer expectations of quality. These variables can simultaneously affect consumers' review content features and their referral and (re)purchase decisions, thereby confounding the effects of consumers' review content on their subsequent decisions (Huang and Sudhir 2021). In terms of measurement errors, our focal measures of review content are

computed using textual mining methods and may, therefore, contain measurement errors, which can introduce biases in model estimation (Yang et al. 2018).

To address these challenges, we adopt a multi-method approach comprising three studies, as summarized in Table 1: (1) A field experiment (Study 1): In collaboration with an online household services platform, we employed interventions (e.g., monetary incentives) to exogenously alter consumers' review writing behaviors (Sajons 2020) and observed their real-world behavioral outcomes. Apart from experimental randomization, we leveraged the instrumental-variable (IV) method (Huang and Sudhir 2021; Sajons 2020) and the proxy-variable approach (Wooldridge 2006) to identify our effects of interest. (2) A scenario-based experiment (Study 2): To address the limitations of Study 1 (e.g., residual endogeneity concerns) and test the generalizability of our findings, we implemented a scenario-based experiment in a podcast trial context. This setup enabled us to directly manipulate and control product quality. Beyond the incentives offered in Study 1, Study 2 introduced additional interventions, i.e., review-writing guidance, generating exogenous variation in review content, to identify causality. (3) Archival data analyses (Study 3): To generalize our insights to unsolicited or unincentivized reviews, we analyzed data of verified consumers from an airline review platform. Together, these three studies leverage mixed methods and diverse data sources (e.g., firm databases, surveys, and secondary data) to triangulate our findings across distinct contexts. This approach distinguishes our work from prior research that primarily relies on a single method, such as surveys, interviews, or archival analyses (e.g., Yu, Khern-am-nuai, and Pinsonneault 2022; please see the Web Appendix B for a detailed comparison with existing studies).

Our research yields several key findings: (1) Affective content in customer reviews facilitates subsequent referral and (re)purchase behaviors by activating affective or heuristic decision-making. (2) Conversely, cognitive content in reviews suppresses these behaviors by

promoting systematic and analytical evaluation. These results align with the dual-process model (Kahneman and Frederick 2002; Shiv and Fedorikhin 1999) and provide empirical evidence that the content features of reviews can shape post-review decisions. (3) While the main effects of review length are not consistently significant, our moderation analyses reveal important boundary conditions. Specifically, writing longer reviews inhibits referral and (re)purchase behaviors when products are less complex, prices are higher, or consumers have more prior experience. These findings are consistent with the elaboration likelihood model (Petty and Cacioppo 1986) and suggest that firms should tailor their review solicitation strategies to individual and contextual factors.

Collectively, these results underscore an *informativeness-loyalty trade-off*: Writing more informative reviews, such as longer reviews with more cognitive content (Hou and Ma 2022), can hinder writers' referral and (re)purchase behaviors. Firms should, therefore, carefully balance the goal of eliciting informative feedback with the need to preserve reviewers' future engagement. For instance, if the objective is to collect informative reviews without compromising reviewers' loyalty, firms can solicit detailed feedback from customers who purchase more complex products, pay lower prices, or have limited experience, conditions under which increasing review informativeness is less likely to reduce writers' post-review engagement. Conversely, when targeting customers who purchase simpler or more expensive products or who are more experienced, firms should consider affect-oriented review elicitation strategies, such as emotional prompts or affective guidance for review writing, to increase the affective content in customer reviews and reinforce writers' loyalty behaviors.

Table 1: A Summary of the Design and Results of Studies.

Study	Purposes	Contexts	Data	DVs	Identification Strategies	Results
Study 1 (a field experiment)	Test the effects of customers' review content on their referral and repurchase decisions when reviews are solicited or incentivized.	Household services	3,374 reviews	Referral and repurchase decisions	(1) The proxy-variable method (2) The instrumental-variable method, with experimental interventions and employee baseline characteristics as instruments	Affective (cognitive) content in customer reviews positively (negatively) affects their referral and repurchase behaviors. The effects of review length on referral and repurchase are moderated by service complexity, prices paid, and users' prior service experience.
Study 2 (a scenario-based experiment)	(1) Test the generalizability of findings in Study 1 in a different context. (2) Address the limitations of Study 1 by directly manipulating and controlling for product quality and tracking all consumers' referral and purchase behaviors.	Podcast trials	1,899 reviews	Referral and purchase decisions	The instrumental-variable method, with experimental interventions as instruments	Affective (cognitive) content in customer reviews positively (negatively) affects their referral and repurchase behaviors. The effects of review length on referral and purchase are moderated by consumers' podcast listening frequency, gender, and age.
Study 3 (archival data analysis)	Test the generalizability of the findings in Studies 1 and 2 to a different context and to unsolicited and unincentivized reviews.	Airline services	8,790 reviews written by 2,219 consumers for 81 different airlines	Referral decisions	A largely correlational study that offers suggestive evidence, given the limitations of observational data	Affective (cognitive) content in customer reviews is positively (negatively) linked to their referral likelihood.

The rest of the paper proceeds as follows. We first lay out our theoretical underpinnings. We then introduce the context, design, and results of the three empirical studies. Following this, we discuss the theoretical and practical insights gained from our research, point out its limitations, and provide directions for future work.

Theoretical Underpinnings

Strategic Importance of Review Writers and Relevant Literature

Understanding the impact of review content on the *writers* themselves is crucial, given the large reviewer base and their significant influence on businesses (e.g., Pocchiari, Proserpio, and Dover 2024). Specifically, review writers often develop stronger loyalty toward businesses (Rocklage and Fazio 2020) and generate more valuable referrals (Fernández-Loría, Cohen, and Ghose 2023) than other consumers. They also accumulate product expertise, enabling them to provide meaningful feedback for quality improvement (Kaul et al. 2025). Additionally, they often command higher popularity (Goes, Lin, and Yeung 2014) or status (Ma et al. 2022) within their social networks, positioning them as opinion leaders who can influence others' decisions. In particular, they play a critical role in small businesses, where existing customers (e.g., review writers) are an important part of the customer base and substantially affect these firms' long-term profitability (Hawkins and Hoon 2020). Given the strategic importance of review writers, we examine how review content features affect writers' subsequent referral and (re)purchase behaviors, which are key determinants of a company's bottom line (Gershon and Jiang 2025).

Despite a large body of research on the effect of review content on *readers*' responses (Babić Rosario, de Valck, and Sotgiu 2020), there is limited evidence of its impact on *writers*, except for Wu and Morwitz (2025). Their study demonstrates that including both affective and cognitive content in reviews helps writers recover more effectively from negative consumption experiences than leaving purely emotional or rational reviews. Our research distinguishes itself from theirs in terms of *theoretical focus*, *practical implications*, and *methodologies*. First, we examine both positive and negative reviews, enhancing the generalizability of our findings across a broader range of contexts. In contrast, Wu and Morwitz (2025) focus exclusively on negative reviews. Second, beyond affective and

cognitive content, we incorporate review length into our framework and reveal *the informativeness-loyalty trade-off*, whereas Wu and Morwitz (2025) do not test the causal effect of review length on writers' decisions or discuss related managerial implications. Third, we investigate the moderating effects of individual and contextual factors, providing actionable insights for businesses to tailor review solicitation strategies. Wu and Morwitz (2025) do not explore such heterogeneous effects. Methodologically, our field experiment captures actual consumer behaviors, enables causal inference, and ensures external validity. In comparison, Wu and Morwitz (2025) rely on secondary data analyses and scenario experiments, primarily using intention measures.

In the following sections, we will elaborate on the theoretical connections between review content features, including affective and cognitive content and review length, and writers' referral and (re)purchase decisions.

Effects of Review Content on Writers' Referral and (Re)Purchase Decisions

Effects of affective and cognitive content on writers' decisions

There are two typical forms of reviews: numerical ratings and textual reviews. Different from numerical ratings, textual reviews capture unique variations in how review writers feel or think about their experiences (Wu and Morwitz 2025). Grounded in the linguistic relativity hypothesis, which posits that individuals' language use impacts how they perceive the social world (Hunt and Agnoli 1991; Whorf and Carroll 1956), we predict that the linguistic content of consumers' reviews can influence their subsequent decisions by shaping their perceptions and evaluations of their product-related experiences.

As previously noted, businesses' review elicitation strategies differ in affective and cognitive orientations. This categorization aligns with dual-process theories (Kahneman and Frederick 2002; Shiv and Fedorikhin 1999) and research on attitude formation (Petty, Fabrigar, and Wegener 2003), which propose that emotions and cognitions are two primary

bases consumers rely on to form evaluations and make decisions. Building on these real-world practices and theories, this research mainly focuses on two fundamental dimensions of review content: (1) affective content: expressions of emotions and feelings; and (2) cognitive content: descriptions of thought processes and objective features. When consumers evaluate products in the form of written reviews, they can either express their emotions/feelings or describe their thoughts/cognitions. These two types of content in reviews mirror and influence writers' engagement in the affective and cognitive processes. By nature, these processes differ in controllability, speed, and effort required. The affective process is relatively automatic, rapid, and effortless, while the cognitive process is more controlled, slow, and effortful (Kahneman and Frederick 2002).

When consumers engage in affective processing, they rely on feelings and heuristics, often paying little attention to product details. Consequently, they are less likely to detect product deficiencies during review writing (Isen and Means 1983). Moreover, as prior research reveals, emotions play a straightforward role in decision-making (Rocklage and Fazio 2020). When individuals have positive emotional experiences, they are often motivated to repeat those experiences to sustain or enhance their positive feelings. Therefore, if a consumer's experience evokes positive emotions, they are more likely to repurchase the services. Moreover, positive emotions can encourage customers to share their experiences with others and respond favorably to requests for referrals, as altruistic motives can drive customer referrals (Kornish and Li 2010; Jung et al. 2020).

Expressing negative emotions following adverse experiences can have therapeutic effects by mitigating such emotions and facilitating natural recovery (Greenberg and Stone 1992). Empirical research documents that expressing negative emotions in a structured manner and recasting emotional memories can generate health-promoting benefits (Littrell 1998). This highlights the role of emotional catharsis and linguistic representation in

channeling the favorable impacts of producing trauma narratives on recovery from negative experiences (Wu and Morwitz 2025). On the other hand, suppressing the natural expression of emotions after negative experiences requires significant energy and results in adverse health outcomes, as indicated by the behavioral inhibition system theory (Pennebaker and Beall 1986). Accordingly, we expect that affective content in customers' reviews fosters their referral and (re)purchase behaviors by promoting engagement in the affective process.

When consumers engage in cognitive processing, they invest effort in thoroughly evaluating products and base their decisions on message-based reasoning (Chaiken 1980). Research on negativity bias (Skowronski and Carlston 1989) and loss aversion (Tversky and Kahneman 1991) indicates that during detailed evaluations, negative aspects draw more attention and carry greater weight than positive ones. As a result, consumers are more likely to focus on the negative features of products when providing reviews. Prolonged rumination on these negative attributes makes the information more salient and easily retrievable from memory (Chan and Cui 2011). Consequently, consumers are more likely to rely on this information in their decision-making, leading to reluctance to repurchase the services.

Similarly, consumers may refrain from recommending services to others if they are involved in detailed evaluations that reveal product shortcomings. This is because referral decisions are often driven by a communal orientation, prioritizing others' welfare over self-interest (Kornish and Li 2010). Furthermore, referrals entail social costs and potential reputational risks, making consumers cautious and prudent when sharing recommendations (Jin and Huang 2014). Thus, we predict that cognitive content in reviews negatively impacts referrals and repurchase behaviors by increasing the likelihood of detailed processing.

Contingent effects of review length on writers' decisions

In addition to affective and cognitive content, we also incorporate review length in our theoretical framework, because review length and affective/cognitive content are inherently

interconnected. As an indicator of the informational richness of reviews, review length is straightforward to measure and can be easily influenced by marketers, for example, through incentives (Burtch et al. 2018). Companies such as *Amazon* and *JD.com* have invested heavily to obtain longer and more informative reviews, but they often overlook the downstream effects of providing in-depth feedback on writers (e.g., Gao et al. 2025).

While empirical evidence highlights the benefits of informative reviews for *readers* (Ananthakrishnan, Proserpio, and Sharma 2023), the existing literature offers an incomplete understanding of how review length influences *writers'* decisions. On the one hand, according to the commitment-consistency principle, a positive relationship exists between review length and consumer referral and repurchase decisions, especially after positive experiences. This principle states that individuals maintain consistency with their initial commitments (Kiesler 1971). In this vein, writing detailed reviews is seen as a commitment, and consumers tend to align their subsequent behaviors with their review activities. This is in line with the mere measurement effect, suggesting that the act of providing feedback itself can improve brand attitudes and increase brand loyalty (Bone et al. 2017).

On the other hand, according to the Elaboration Likelihood Model, the effect of review length on subsequent consumer decisions depends on the elaboration likelihood of consumers, which refers to individuals' motivation or ability to engage in detailed processing (Jayawardena et al. 2023; Petty and Cacioppo 1986). Consumers with a high elaboration likelihood, such as those making expensive purchases, are inclined to scrutinize product information more thoroughly when writing longer reviews (Bitner and Obermiller 1985). This could increase their sensitivity to and emphasis on product flaws, negatively affecting their referral and repurchase behaviors (Chan and Cui 2011). Conversely, if the elaboration likelihood is low, such as with complex products that consumers do not have the relevant knowledge to evaluate, they may turn to heuristics instead of exerting cognitive effort during

review provision, making review length a poor indicator of detailed processing and weakening its impact on future decisions (Swait and Adamowicz 2001).

In summary, consumers' elaboration likelihood moderates the effects of review length on subsequent decisions: The negative relationship intensifies as consumers' elaboration likelihood increases. Recognizing that consumers' elaboration likelihood varies by contextual and individual factors, this research aims to explore how these factors moderate the impact of review length on writers' decisions in subsequent empirical studies.

Overview of the Current Research

Quantifying the impact of customer review content features on their subsequent behaviors is inherently complex due to potential endogeneity. Endogeneity primarily arises from three sources: (1) omitted variables, (2) measurement errors, and (3) simultaneity, which present challenges in identifying causality and introduce potential biases in estimation (Angrist and Pischke 2009). For a detailed discussion of these endogeneity issues and corresponding solutions, please refer to Web Appendix D.

To address these identification challenges and enhance the generalizability of our findings, as summarized in Table 1, we adopted a mixed-methods approach, encompassing a field experiment (Study 1), a scenario-based experiment (Study 2), and archival data analysis (Study 3). The IRB approved the procedures of these studies. In Study 1, we conducted a field experiment on an online housekeeping services platform, employing interventions (such as incentives and nudges) to solicit reviews and collecting real-world consumer behavioral data. The randomization of these interventions served as a key strategy to generate exogenous variations in review content. Moreover, we addressed endogeneity issues using instrumental and proxy variables (Huang and Sudhir 2021; Sajons 2020; Wooldridge 2006). However, the field experiment had some limitations. For instance, while we used proxies for service quality, they may not fully resolve the omitted variable bias. Furthermore, measuring referral

behaviors within a random subsample introduced the potential for sample selection bias.

To address the limitations of Study 1, Study 2 utilized a scenario-based experiment in a podcast trial setting. In this study, we generated exogenous variation in review content through treatments (e.g., providing review writing guidance). We then employed the random assignment of treatments as instruments (Sajons 2020). Importantly, this setting enabled us to directly manipulate and control for actual product quality, as well as track referral behaviors of all participants, overcoming the constraints encountered in Study 1.

Lastly, because Studies 1 and 2 primarily focused on solicited or incentivized reviews, Study 3 explored the generalizability of our findings by analyzing unsolicited and unincentivized airline reviews. These reviews were submitted voluntarily by customers, without any solicitation or incentives. Unlike Study 1, this dataset provided access to most consumers' recommendation choices. Although observational data comes with inherent limitations, this analysis complements and extends our understanding of the relationships between review content and subsequent customer decisions across different contexts. Together, these studies address critical methodological challenges and expand the breadth and depth of our investigation in varied settings.

Study 1: A Field Experiment on a Household Services Platform

The primary goal of Study 1 is to identify the causal effects of customers' review content on their referral and repurchase behaviors in the real marketplace. We conducted a field experiment on an online platform for household services in China, incentivizing customers to write reviews and observing their behaviors afterward. The household services market is rapidly expanding globally and is projected to grow from USD 8.7 billion in 2024 to USD 12.8 billion in 2030 (Research and Markets 2025). Our partner utilizes mobile technology to streamline this service, enhancing accessibility for customers and operational efficiency for employees and managers.

Design and Procedure

To enhance the transparency of reporting, we followed the Consolidated Standards of Reporting Trials framework (Glasgow, Huebschmann, and Brownson 2018) to create a flow diagram in Web Appendix E. This diagram outlines the enrollment, allocation, and analysis phases of the study. Initially, we randomly selected 4,016 participants from approximately 18,000 customers on the platform, predominantly comprising new users. Upon service completion, each participant received a prompt to write a review. Those in the treatment groups were randomly exposed to either monetary incentives or emotional messages, while the control group received only neutral prompts. The type of prompts received was consistent across all orders of the same customer.

Following the interventions, all customers were asked to rate their satisfaction with the service on a 10-point scale (1 = "very dissatisfied," 10 = "very satisfied") and write down their textual comments about the services. The questions contained in the review request were the same across all conditions. We collected data on customers' satisfaction ratings (used as a metric of review valence and a proxy for service quality) and textual comments. After review provision, 20.5% of customers (across conditions) were randomly chosen to receive a referral request and decided whether or not to recommend the service.[1] Apart from tracking referral behaviors of this subsample, we also recorded repurchase behaviors of the entire sample during the experimental period; repurchase data were right censored at the endpoint.

Data Description

Throughout the experiment, we collected data on consumers' review, referral, and repurchase behaviors from the platform. Moreover, we complemented the dataset with information on customers (e.g., gender), employees (e.g., baseline work performance), and

[1] To avoid disrupting daily operations, our collaborator placed constraints on the proportion of customers who could receive referral requests. It allowed us to randomly select about 20% of customers as the targets for sending referral requests.

order specifics (e.g., price and category) from the company's database.

Outcome variables

Our focal outcome variables are consumers' referral and repurchase behaviors. We define referral behavior as a customer recommending the service to others. In our analyzed sample, 16% of customers recommended the service. Apart from the referral dummy, we used the number of referrals made by each customer as a supplementary measure of referral behavior. We also looked at the following aspects of customer repurchase behavior: (1) whether a customer repurchased the service within a period from when a customer entered the experiment to the end of the experiment; (2) repurchase hazard (or time to the next purchase); and (3) spending on repurchased orders. In the sample used in the analysis, 37% of consumers repurchased the services. For these repeat customers, the average time to the next purchase is 38 days (with a large SD of 64 days).

Explanatory variables

Over 95% of participants provided reviews (including ratings or comments), indicating that most customers leave reviews in our context. This could come from the specific context of household services and also from review requests going out right after service completion as instant messages to customers' mobile phones. We limited the sample to customers who left reviews and focused on the variation in review content. Since such a high proportion of customers wrote reviews, concerns about sample selection biases are mitigated.

Our focal review content characteristics include the affective score, cognitive score, and review length. The affective and cognitive scores capture customers' involvement in affective and cognitive processing during review provision. Based on the dual-process model (Kahneman and Frederick 2002; Shiv and Fedorikhin 1999), we extracted the affective and cognitive content in reviews using both dictionary-based and topic-modeling methods

(Zhang, Li, and Allenby 2024). We used the linguistic inquiry and word count (LIWC) program as the dictionary-based method to compute the affective and cognitive scores of customer reviews (Huang et al. 2012; Ludwig et al. 2013). The affective score is calculated based on the percentage of emotion-laden words (e.g., like, thank, and enthusiastic) that appear in a review and represents the degree to which people are engaged in affective processing. The cognitive score is computed based on the percentage of words related to cognitions (e.g., compare, problem, and detail) used in a review, indicating the extent to which people are involved in cognitive processing.

To ensure the validity of these textual measures, we also employed topic models, namely Seeded LDA (Latent Dirichlet Allocation, Watanabe and Baturo 2024) and Guided BERTopic (Grootendorst 2022), to construct alternative measures of affective and cognitive processing. These methods generate coherent and interpretable topics and make topic classification more theoretically grounded and content analysis more topic-specific (Watanabe and Baturo 2024; Zhang, Li, and Allenby 2024). Operationally, according to the dictionaries provided by LIWC and our specific contexts, we pre-defined seeded words that reflect affective and cognitive processing and then extracted relevant topics from review content with the two methods. Furthermore, we computed the affective and cognitive scores of reviews based on the predicted probabilities that reviews belonged to affect- and cognition-related topics, respectively. Finally, we averaged the scores calculated by these two methods as focal measures. The seeded words and topic words are listed in Web Appendix F. Regarding the associations between review features, review length is negatively correlated with the affective score ($r = -.357$, $p < .000$) and positively linked to the cognitive score ($r = .073$, $p < .000$), indicating that writing longer reviews is consistent with consumers relying more (or less) on cognitive (or affective) processing during service evaluation. Due to the covariation between review length and affective/cognitive content, we included all these attributes in our models.

Other variables

In addition to the above focal variables of interest, we gathered other information on employees, customers, and services/orders. In particular, we obtained employees' baseline measures of (1) job satisfaction, (2) industry experience, (3) service hours, and (4) service quality. We also collected the ratings of customers and supervisors for the current services. Customers' ratings capture their satisfaction with the focal services and review valence, solicited through the review requests. Supervisors' ratings are given based on their assessment of employees' service quality on the spot (an 8-point scale, 1 = "extremely low," 8 = "extremely high"). We also obtained data on customers' genders, number of previous orders, current order type (whether the current order belonged to a "package" order), service category (whether the current order involved complex services such as appliance repairing), price, promotion, location, and whether there were any lockdowns caused by COVID-19 as covariates. We present summary statistics of the main variables in Web Appendix G.

Identification

In the setting of service evaluations, a major challenge for causal inference is that unobservable confounders, such as service quality, simultaneously shape customers' review content features and their subsequent decisions. To address this challenge, we used a combination of two approaches, including the proxy variable approach (Wooldridge 2006) and the instrumental variable (IV) method (Huang and Sudhir 2021; Sajons 2020).

Proxy variables

The first is to approximate the unobserved service quality via observed proxy variables. The proxies of quality we include are customers' and supervisors' ratings of the current service as described in the "Other variables" section. However, the inclusion of customers' ratings is in and of itself a challenge, as they may be correlated with unobserved customer

characteristics. To address this endogeneity problem, we included the focal employees' baseline service quality (assessed by their supervisors before the experiment) as an instrument. Web Appendix H illustrates more details about the proxy variable approach.

Nevertheless, apart from actual quality, there may be other omitted variables that confound the effects of review content features on referral and repurchase decisions. To address such endogeneity issues, we utilized exogenous variation in review content induced by two types of randomizations across customers: (1) random assignment of experimental interventions to stimulate review behaviors; and (2) random assignment of employees to customers.[2] Conditional on these randomizations, we primarily used two types of instruments for endogenous review attributes: experimental interventions (Sajons 2020) and employees' baseline characteristics (Huang and Sudhir 2021). To ensure identification, each endogenous variable has at least one instrument closely tied to it (Hill et al. 2021).

Randomization of experimental interventions

For the first type of randomization, following the experimentally randomized instrumental variable (ERIV) approach (Sajons 2020), interventions were randomly assigned to customers to generate exogenous variation in customer review content. Specifically, we employed three types of interventions in the field experiment: (1) length-contingent monetary rewards, (2) emotional messages, and (3) a combination of these two types of interventions.[3] These interventions represent typical monetary and nonmonetary incentives commonly adopted by businesses for review creation (Millwood 2023). However, their effects on review generation have not been thoroughly examined in empirical research. Different from prior research (e.g., Burtch et al. 2018), monetary rewards used in our experiment are contingent

[2] We excluded orders where customers designated specific employees to ensure the random assignment of employees to customers.

[3] The experiment has a 2 (emotional messages: presence vs. absence) × 3 (monetary rewards: none vs. 5 yuan vs. 15 yuan) between-subjects design. There are six conditions in total.

on review length: Only when the word count of textual comments reached 20, could consumers get such rewards (in the form of coupons for future purchases).[4] Compared with other types of incentives (mentioned in Table W1), length-contingent rewards can increase review informativeness and offer a relatively honest representation of customer evaluations, for such rewards do not constrain the nature or sentiment of the reviews. Therefore, such rewards are used for review creation by multiple platforms, such as *JD.COM* (a major e-commerce platform in China) and *Trip.com* (one of the largest travel platforms worldwide).

Besides monetary incentives, we designed emotional messages to trigger customers' empathy with the service provider and facilitate review creation (Leung, Kim, and Tse 2020). These messages depict employees overcoming obstacles to deliver high-quality services to customers. Web Appendix I provides descriptions of the design.

Of note, the random assignment of monetary incentives did not directly impact customer referral decisions, because incentives were provided according to review length rather than referral behaviors, and they were too small to offset the reputational cost of inaccurate referrals (Jin and Huang 2014). However, since monetary incentives (i.e., coupons for future purchases) could directly influence repurchase decisions rather than through review writing, we turn to the next dimension of randomization: the random assignment of employees.

Randomization of employees

For the second type of randomization, we note that the assignment of employees to customers was automated in most cases, based on the employees' random availability and customers' geographical locations. Following Huang and Sudhir (2021), conditional on the randomization of service employees, we utilized the following employees' baseline characteristics as instruments for endogenous review content measures:

(i) *Employees' baseline job satisfaction as an instrument for the affective score of*

[4] Based on our collaborator's suggestions and baseline review data, we set 20 characters as the threshold for review length. Please note that Chinese sentences are generally shorter than English ones that convey the same meaning.

reviews: Drawing on the emotional contagion theory, interactions with satisfied employees evoke positive emotions in customers (Hennig-Thurau et al. 2006). These positive emotions further encourage customers to engage in affective processing during service evaluation, thereby increasing affective content (Isen and Means 1983).

(ii) *Employees' work experience as an instrument for the cognitive score of reviews*: Employees with greater expertise foster higher trust (Camacho, De Jong, and Stremersch 2014), reducing customers' tendency to evaluate service details thoroughly (Mayo 2016) and further decreasing cognitive content in reviews.

(iii) *Employees' baseline service hours as an instrument for review length*: Overworked employees are more prone to exhibit signs of fatigue and neglect service details (Singh 2000). Consequently, their customers are more likely to detect and elaborate on service flaws, naturally leading to longer reviews.

(iv) *Employees' baseline service quality as an instrument for customer satisfaction ratings:* Improving service quality can enhance customer satisfaction (Huang and Sudhir 2021). As long as employees' service quality is not very variable over time, we expect that employees' baseline service quality has a positive link with customer satisfaction with their current service.

To sum up, these instruments indicate employees' work-related abilities or motivation and can influence employees' performance in the current service. Therefore, they will further impact customers' service experiences and review provision process, translating into variation in customers' review content (Godes and Mayzlin 2004). Besides, since customers are informed of the employee assignment procedure used by the platform, they are aware that employees are randomly assigned based on availability in most cases. Further, because customers would expect a different random draw from the employee pool the next time, there is no direct impact of the focal employees' baseline characteristics on consumers' decisions to repurchase and recommend the service. Employee baseline characteristics are, therefore, orthogonal to customer repurchase (and referral) decisions. Web Appendix J presents randomization checks that ensure the exogenous assignment of interventions and employees.

Empirical Analysis

Our formal analyses consist of the following parts: (1) We first tested the relevance of instrumental variables (first-stage analyses). (2) We then identified the impacts of customer review content on their subsequent behaviors with these instruments (second-stage analyses). (3) Further, we conducted moderation analyses to examine the boundary conditions.

When analyzing repurchase behaviors, we used employees' baseline characteristics as instruments and controlled for incentives as covariates, due to the direct effects of incentives on repurchases. In contrast, when examining referral behaviors, apart from employees' baseline characteristics, we utilized the random assignment of experimental interventions as instruments. We mainly employed a control-function approach for model estimation, which accounts for multiple endogenous variables and the nonlinearity of models (Wooldridge 2015). Web Appendix K presents our model specifications. The units of analysis are customer-review observations. Because only the initial assignment of interventions was random, we mainly analyzed the first (re)order for each customer after interventions.

First-stage analyses: Effects of interventions and employee characteristics on review content

In the first-stage regressions with review content characteristics as DVs, we included the intervention dummies (i.e., the assignment of incentives and emotional messages), employees' baseline characteristics, and exogenous covariates as regressors. As shown in Table 2, with either the entire sample or the subsample who received referral requests, the instruments stimulate changes in review content. This is corroborated by the large F-statistics of these instruments (Angrist and Pischke 2009; Sanderson and Windmeijer 2016). For detailed explanations of first-stage regressions, please see Web Appendix L.

Table 2: Effects of Experimental Interventions and Employee Baseline Characteristics on Review Content Features (Study 1).

Columns	(1)	(2)	(3)	(4)	(5)	(6)	(7)	(8)
DVs	Affective Score (LIWC)	Cognitive Score (LIWC)	Review Length	Customer Rating	Affective Score (LIWC)	Cognitive Score (LIWC)	Review Length	Customer Rating
Units/Scales	Percentage	Percentage	Character	A 10-Point Scale	Percentage	Percentage	Character	A 10-Point Scale
Job Satisfaction	.043	-.008	-1.743	-.026	.045	-.016	-2.123	.006
	(.005, .000)	(.006, .153)	(.407, .000)	(.044, .558)	(.002, .000)	(.002, .000)	(.188, .000)	(.018, .751)
Industry Experience	-.003	-.020	.036	.005	.000	-.019	-.017	-.002
	(.002, .151)	(.002, .000)	(.116, .757)	(.011, .683)	(.001, .823)	(.001, .000)	(.051, .743)	(.005, .601)
I(Long Service Hours)	-.030	.016	12.414	-.018	-.034	.011	15.526	-.018
	(.013, .023)	(.016, .323)	(1.459, .000)	(.125, .885)	(.006, .000)	(.006, .069)	(.727, .000)	(.054, .733)
I(Monetary Reward Presence)	-.004	.040	5.699	.098	-.028	.029	7.475	.036
	(.014, .768)	(.015, .006)	(.942, .000)	(.106, .354)	(.006, .000)	(.006, .000)	(.465, .000)	(.047, .446)
I(Emotional Message Presence)	-.003	-.001	1.360	-.015	-.001	.003	-.349	.015
	(.012, .832)	(.013, .962)	(1.004, .176)	(.110, .892)	(.005, .864)	(.005, .597)	(.458, .446)	(.044, .742)
Service Quality	.005	-.013	-.551	2.016	.006	-.006	-1.826	2.021
	(.005, .307)	(.006, .038)	(.754, .465)	(.058, .000)	(.002, .002)	(.002, .015)	(.281, .000)	(.024, .000)
Control Variables	Y	Y	Y	Y	Y	Y	Y	Y
Observations	625	625	625	625	3374	3374	3374	3374
Excluded F-stat.	13.91	25.26	20.33	215.61	132.18	237.95	128.43	1770.68
SW F-stat.	24.68	39.92	28.59	360.39	363.29	576.39	450.30	2814.24
AP F-stat.	22.65	42.96	34.65	446.62	381.39	799.49	445.91	6789.02
R-squared	.259	.215	.452	.740	.187	.269	.367	.767

Notes: The first set of numbers in parentheses represents robust standard errors, while the second set indicates *p*-values. Columns (1)-(4) are the first-stage results of referral behavior analysis, while Columns (5)-(8) are the first-stage results of repurchase behavior analysis. Independent variables included employees' baseline characteristics and experimental interventions. Control variables included customer characteristics, service/order features, geographical location, COVID-19 lockdowns, etc. For multiple endogenous variables, we used Angrist-Pischke (AP) F-statistics and Sanderson-Windmeijer (SW) F-statistics in weak identification tests (Angrist and Pischke 2009; Sanderson and Windmeijer 2016). Most F-statistics exceed the Stock-Yogo critical values (referral sample: 11.12 for 10% and 19.28 for 5% maximal IV relative bias; repurchase sample: 10.27 for 10% and 16.85 for 5% maximal IV relative bias; Stock and Yogo 2002), indicating that the instruments meet the relevance criterion. Manipulation checks show that length-contingent incentives lead to an increase of 10 characters in review length ($M_{Incentive\text{-}Present}$ = 27, SD = 17; $M_{Incentive\text{-}Absent}$ = 17, SD = 14; t(3372) = 17.07, $p < .000$, Cohen's d = .59, r_{pb} = .28).

Second-stage analyses: Effects of the affective and cognitive content of reviews

From Table 3, we can see that the affective score of consumers' reviews positively influences their referral and repurchase behaviors, whereas the cognitive score has the opposite effect, supporting our predictions. Given the estimates, increasing the affective (cognitive) score by 1 SD can enhance (reduce) the number of referrals by .20 (.19) and repurchase expenses by 10.89 (30.25). These statistics underscore the importance of managing consumers' review provision process to drive pivotal business outcomes.

Table 3: Effects of Review Content Features on Referral and Repurchase Behaviors (Study 1).

Columns	(1)	(2)	(3)	(4)	(5)	(6)
DVs	Referral Likelihood	Number of Referrals	Number of Referrals	Repurchase Likelihood	Repurchase Hazard	Repurchase Expense
Units	Percentage	Individual	Individual	Percentage	Percentage	Yuan
Models	Linear Probability	Linear Regression	Poisson	Linear Probability	Cox	Linear Regression
Affective Score (LIWC)	1.128	1.276	6.608	.456	2.298	68.075
	(.280, .000)	(.334, .000)	(2.385, .006)	(.144, .002)	(.679, .001)	(29.772, .022)
Cognitive Score (LIWC)	-.872	-1.060	-5.585	-.821	-3.303	-168.069
	(.204, .000)	(.295, .000)	(1.147, .000)	(.089, .000)	(.368, .000)	(23.036, .000)
Review Length	.003	-.001	-.018	-.002	-.004	-.527
	(.002, .144)	(.003, .773)	(.020, .380)	(.001, .116)	(.005, .486)	(.218, .016)
Control Functions	Y	Y	Y	Y	Y	Y
Control Variables	Y	Y	Y	Y	Y	Y
Effect Size of Affective Score	.180	.204	.240	.073	.444	10.892
Effect Size of Cognitive Score	-.157	-.191	-.228	-.148	-.448	-30.252
Observations	625	625	625	3374	3374	3374
F-stat. (Chi-squared)	13.494	7.586	2622.926	194.820	1296.720	92.274
(Pseudo) R-squared	.364	.307	.460	.417	.074	.308

Notes: We used the control function approach to estimate the coefficients. Standard errors (the first set of numbers in parentheses) and *p*-values (the second set of numbers) were adjusted using a bootstrap approach (Ebbes, Papies, and van Heerde 2022). We used data on customers' review content when they were first exposed to our interventions to predict their subsequent referral and repurchase behaviors. Overidentification tests suggest instrumental variables are valid exclusions (Hansen J stat. = 4.203, $p = .122$). In addition, endogeneity tests indicate that the potential endogeneity issue of review content features should be accounted for (GMM distance test stat. = 13.99, $p = .007$). Control variables included service/order features, customer characteristics, geographical location, whether there were any lockdowns caused by COVID-19, etc. In robustness checks, we used alternative estimators, model specifications, textual measures, control variables, instrumental variables, samples, etc., and got similar results. In the main sample, the SDs of the affective and cognitive scores produced by LIWC are .16 and .18, respectively. Given these distribution statistics and the coefficient estimates, we calculated effect sizes, i.e., the change in the DV produced by a 1-SD increase in the independent variable. For the hazard model in Column (5), we calculated the percentage change in the hazard rate resulting from a 1-SD increase in the independent variable as the effect size metric.

Second-stage analyses: Contingent effects of review length

In contrast to the pronounced effects of the affective and cognitive content, the influences of review length on referral and repurchase are not always significant, implying the possible existence of boundary conditions. We then explored the moderators of the impacts of review length on referral and repurchase. According to ELM (Petty and Cacioppo 1986), we focus on product and individual characteristics that impact consumers' elaboration likelihood during review provision, primarily including (1) product complexity, (2) price, and (3) consumers' prior knowledge/experience (Krefeld-Schwalb, Sugerman, and Johnson 2024). Product complexity and consumers' prior knowledge mainly influence consumers' *ability* to process information in detail (Jayawardena et al. 2023), whereas price affects their *motivation* to elaborate on relevant information (Bitner and Obermiller 1985).

It has been well documented that only when relevant knowledge is accessible can individuals process information in a meaningful manner. In the absence of such knowledge, people usually resort to heuristics (Swait and Adamowicz 2001). Accordingly, as product complexity decreases, it becomes less difficult for consumers to comprehend and evaluate central aspects of the product, and consumers are more likely to elaborate on central cues (Bitner and Obermiller 1985). Therefore, we expect that when the product becomes less complex, writing longer reviews will increase deliberation, and the effect of review length on referral/(re)purchase decisions will become more negative.

Similar to product complexity, consumers' prior knowledge/experience can also moderate the impact of review length on subsequent decisions by influencing how consumers process information (Jayawardena et al. 2023). Based on the expert information processing model, experts outperform novices in intuitive judgment and problem recognition within their specific domain of expertise, because experts' domain-specific knowledge structures are larger and more easily accessed (Lord and Maher 1990). Due to such differences, more

experienced consumers are more sensitive to product deficiencies when writing longer reviews (Bitner and Obermiller 1985). Hence, we predict the relationship between review length and referral/(re)purchase will become more negative when consumers have more product-related knowledge and experience.

Different from the aforementioned factors related to knowledge accessibility, price, which indicates the importance of the purchase, influences consumers' sunk-cost consideration and their *motivation* to invest cognitive effort in information processing (Lee and Tsai 2014). Generally, for more expensive products, consumers will have more significant sunk-cost considerations and a stronger motivation to process information in detail (Bitner and Obermiller 1985). Therefore, when consumers pay a higher price for a product, they tend to be more involved in product evaluation. This leads to a more pronounced adverse effect of review length on referral and (re)purchase.

Table 4 presents the results of interaction effects. First, when service complexity increases, the detrimental effects of review length on referral and repurchase are attenuated. This implies that as task complexity increases, consumers move toward heuristics, consonant with the view of consumers as cognitive misers (Swait and Adamowicz 2001). Second, as the price paid increases, the relationships between review length and referral/repurchase become more negative, indicating that customers who pay higher prices are more involved in thoroughly evaluating services (Krefeld-Schwalb, Sugerman, and Johnson 2024; Lee and Tsai 2014). Third, customers' prior experience, operationalized as the number of previous orders before the experiment, negatively moderates the effects of review length. This is because customers with more relevant knowledge have higher abilities to detect service deficiencies when writing longer reviews (Jayawardena et al. 2023). To recap, these findings inform managers about the unintended consequences of encouraging customers to provide informative feedback, especially when the services are easier to evaluate, customers pay

higher prices, or customers have more prior experience.

Robustness checks

To ensure robustness, we have conducted multiple checks, including: (1) using alternative estimators (e.g., OLS, 2SLS, GMM, and LIML) (Goldfarb, Tucker, and Wang 2022); (2) separately examining each endogenous variable (York, Vedula, and Lenox 2018); (3) using alternative model specifications (Dong and Lewbel 2015); (4) using alternative measures calculated by different text mining methods (Guided BERTopic and Seeded LDA, Grootendorst 2022; Watanabe and Baturo 2024); (5) controlling for additional confounders (Duan and Mela 2009); (6) using different sets of instruments (Angrist and Pischke 2009); (7) accounting for measurement errors in review content characteristics (Yang et al. 2018); (8) limiting the sample to customers not assigned with monetary rewards to address the confounding effect of monetary incentives; (9) excluding observations during the lockdown period to eliminate the interference of the pandemic; (10) excluding reviews that include irrelevant content to rule out alternative explanations; (11) conducting seemingly unrelated regressions to account for correlations among DVs (Zellner 1963); and (12) adjusting *p*-values in multiple hypothesis testing (e.g., List, Shaikh, and Vayalinkal 2023). These robustness checks (summarized in Web Appendix M) support the validity of our results.

Discussion

Study 1 shows that affective (cognitive) content in reviews facilitates (deters) customer referral and repurchase behaviors. Furthermore, product and individual factors influence the impact of review length. These results reveal an *informativeness-loyalty trade-off*, i.e., soliciting informative feedback can impede review writers' referral and repurchase decisions.

Table 4: Contingent Effects of Review Length on Referral and Repurchase (Study 1).

Columns	(1)	(2)	(3)	(4)
DVs	Referral Likelihood	Repurchase Hazard	Referral Likelihood	Repurchase Hazard
Units	Percentage	Percentage	Percentage	Percentage
Models	Linear Probability	Cox	Linear Probability	Cox
Type of Examination	Separate Examination of Each Interaction Term		Joint Examination of All Interaction Terms	
I(Complex Service) × Length	.01882	.05446	.01903	.03854
	(.00614, .002)	(.00620, .000)	(.00640, .003)	(.00657, .000)
Price of the Current Order × Length	-.00002	-.00026	-.00002	-.00018
	(.00001, .030)	(.00004, .000)	(.00001, .019)	(.00004, .000)
Number of Previous Orders × Length	-.00016	-.00148	-.00016	-.00122
	(.00004, .000)	(.00034, .000)	(.00004, .000)	(.00035, .000)
Control Functions	Y	Y	Y	Y
Control Variables	Y	Y	Y	Y
Observations	625	3374	625	3374

Notes: Standard errors (the first set of numbers in parentheses) and *p*-values (the second set of numbers) were adjusted using a bootstrap approach (Ebbes, Papies, and van Heerde 2022). Due to the interaction terms, we employed the control function approach to address the potential endogeneity of review content features (Wooldridge 2015). Separate vs. joint examinations of the interaction effects produced consistent results. Using the number of referrals and repurchase expenses as DVs also yielded similar findings. Control variables included customer characteristics, service/order features, geographical location, COVID-19 lockdowns, etc.

To understand the financial implications of our findings, we performed back-of-the-envelope analyses based on how the affective and cognitive scores of reviews affect subsequent customer behaviors. These analyses showed that increasing the affective score by 1 SD leads to a profit gain of 331 yuan from referrals and a revenue increase of 11 yuan from repurchases per writer. In contrast, raising the cognitive score by 1 SD decreases referral profits by 316 yuan and reduces repurchase revenues by 30 yuan.[5] Additionally, we quantified the cost-effectiveness of the interventions. Such analyses indicated the consequences of incentivizing longer reviews: Financial incentives induce a profit decline of 96 yuan per writer, translating into an aggregate loss of 1,635 thousand yuan. Of note, these figures reflect only the influence on *writers* and do *not* account for the effect on *readers*. More details about these analyses can be found in Web Appendix N.[6]

Nevertheless, Study 1 has a few limitations that should be addressed. First, though we used proxy variables for quality, these proxies might not accurately reflect the actual quality, potentially leading to omitted variable bias in our findings. Second, only about 20% of customers received referral requests. Although we ensured that this subsample was randomly selected and comparable to the total sample (detailed in Web Appendix E), it reduced the statistical power of referral behavior analysis. Third, this field experiment was run on an online household services platform, and our findings may be context-specific. Whether the results of this study can be generalized to other contexts remains an open question.

Study 2: A Scenario-Based Experiment about Podcast Trials

Study 2 aims to address the limitations of Study 1 by introducing a controlled environment, where we can directly manipulate product quality and track *all* consumers'

[5] Although we did not directly manipulate the affective and cognitive content of reviews in this study, for illustrative purposes, we computed their financial impacts if we could directly change them.
[6] Based on historical transaction data, we assume a 5-year customer lifespan, a 600-yuan annual profit, an 80% retention rate, a 20% discount rate, a 20% profit margin, 17,113 reviewers, and 65% of reviewers receiving incentives.

referral behaviors. This study is an incentive-compatible scenario-based experiment in the context of podcast trials (pre-registered on AsPredicted: https://aspredicted.org/zdsz-f2wv.pdf), a medium chosen for its popularity and ability to evoke consumers' affective and cognitive responses. Industry data highlights the significance of podcasts, with over 584.1 million listeners worldwide and expected ad revenues reaching USD 4.46 billion by 2025 (Backlinko 2025).

Similar to Study 1, we employed treatments to induce exogenous changes in customers' review content and then used the random assignment of such treatments as instrumental variables (Sajons 2020). The underlying assumption is that these instruments can change consumers' review content while they are unlikely to directly influence consumers' referral and purchase decisions (other than through review content), since these interventions are in relation to review writing behaviors rather than referral/purchase decisions.

Design and Procedure

Study 2 employs a 2 (length-contingent incentives for reviews: present vs. absent) * 3 (guidance for review writing: affective vs. cognitive vs. unguided) * 2 (product quality: high vs. low) between-subjects design. Randomization checks (in Web Appendix O) show that there were no significant differences in covariates across groups. To ensure experimental realism, we framed the experiment as a podcast trial program launched by a content platform, involved participants in a real consumption experience, and later measured their actual behaviors. As shown in Figure W3 of Web Appendix P, participants first listened to a podcast episode (about enhancing sleeping quality), gave ratings based on their perceived quality and consumption experience, wrote a review, made their recommendation and purchase decisions, finished attention checks and manipulation checks, provided other basic information (e.g., review writing habits and demographics), and received explanations about the use of deception in the debriefing step.

When asking participants to write a textual review of the podcast, we randomly provided some participants with (1) incentives or (2) guidance for review writing. Regarding the requirement tied to incentives, only when reviews exceed 30 characters can writers get 5-yuan rewards.[7] Apart from incentives, we directly guided participants toward different information processing modes (Wu and Morwitz 2025). Specifically, participants in the affective (cognitive) guidance condition were asked to express their feelings and emotions (describe their thought processes) when writing reviews, and those in the unguided condition were not provided with any instructions for review writing. Such review-writing guidance also resembles businesses' review elicitation strategies that focus on either emotions or cognitions (listed in Web Appendix A). We also manipulated podcast quality (by changing the pronunciation and pacing of broadcasters while keeping the content of the podcast the same across conditions) and controlled for actual quality to eliminate any confounding effect of quality. We present interventions used in each condition in Web Appendix P.

Following review writing, *all* participants were requested to share/recommend the podcast. They indicated their sharing likelihood ("How likely are you to share the podcast with others") and referral likelihood ("How likely are you to recommend the podcast to others") on 7-point scales ("1 = "not at all," 7 = "extremely likely"), assuming that their friends and family need to improve sleeping quality. To complement these self-reported measures, we told participants that the platform would customize content based on users' characteristics and therefore requested participants to provide brief information about referees' surnames and gender (Sun et al. 2021). We then used (1) whether participants left any information about referees and (2) the number of referrals as proxies for participants' willingness to invest effort in product recommendation. Besides referral, we asked all

[7] The minimum length requirements for getting incentives vary across platforms, including 5 characters (e.g., VIPSHOP.COM), 10 characters (e.g., JD.COM), 20 characters (e.g., RED), over 50 characters (e.g., Dianping), etc. We took the average as the threshold used in this experiment.

participants to indicate their intentions to subscribe to the podcast on a 7-point scale ("How likely are you to subscribe to the podcast," 1 = "not at all," 7 = "extremely likely").

Afterwards, to capture real purchase behaviors beyond intentions and ensure incentive compatibility, we informed participants that at the end, we would run a lottery as a bonus for participation in this experiment. In this lottery, we would randomly draw several individuals and send them a coupon for subscribing to the podcast. The original subscription price is 10 yuan. The coupon offered an 80%-off discount and was only valid within 30 minutes. Participants were asked to decide whether to purchase the paid content with this coupon. Participants were told that if they were actually selected in the random draw and decided to purchase the content with this coupon, we would provide the podcast to them later and deduct the actual price paid by them (2 yuan) from their compensation for participation. Since consumers needed to make referral and purchase decisions upon review provision, there was *no* censoring for referral and purchase data.

Finally, all participants needed to finish attention checks (questions about the podcast content and our instructions for review writing) and manipulation checks ("To what extent do you pay attention to your feelings and emotions (thoughts and objective features) during the review provision process," a seven-point scale, 1 = "not at all," 7 = "very much," Wu and Morwitz 2025). To capture participants' review-writing habits in real life, we asked how many (incentivized) reviews they left in the past year. In addition, we measured their perceived familiarity with the podcast used in our experiment ("How familiar are you with the podcast," a seven-point scale, 1 = "not at all," 7 = "extremely familiar," Kim, Maslowska, and Tamaddoni 2019), perceived importance of improving sleeping quality ("How important is it for you to improve sleep quality," a seven-point scale, 1 = "not at all," 7 = "extremely important," Jensen and Yetgin 2017), podcast listening habits ("How often do you listen to podcasts in daily lives," a five-point scale, 1 = "never," 5 = "always"), as well as

demographic information (age, gender, education, and income). We list the main items used in Study 2 in Web Appendix P.

Data Description

We recruited 1,899 participants (M_{age} = 29 years, 67% female, 33% male) from the Credamo platform.[8] On average, participants had written 17 reviews and 6 incentivized reviews in the last year. Moreover, over 75% of participants have previously seen length-contingent incentives for reviews in real life, suggesting the prevalence of such incentives in business practices. Importantly, different from Study 1, in Study 2, all customers wrote reviews and made referral and purchase decisions, eliminating sample selection bias. Web Appendix Q provides variable descriptives.

Outcome variables

DVs of interest are both intentions and behaviors of referral and purchase, including (1) *Referral Intention*: the average of consumers' sharing and referral likelihood (Cronbach's alpha = .916); (2) *Referral Dummy*: whether consumers nominated others as referees; (3) *Number of Referrals*: the number of people nominated by consumers; (4) *Purchase Intention*: consumers' purchase likelihood; and (5) *Purchase Dummy*: whether consumers decided to purchase the podcast if they won a coupon in the lottery.

Explanatory variables and instrumental variables

Our focal measures of review content include the affective score, cognitive score, and length of reviews. Like Study 1, we used both dictionary-based and topic-modeling methods to capture the affective and cognitive content in reviews (Zhang, Li, and Allenby 2024). To deal with the potential endogeneity of these review content features, we employed the

[8] The platform automatically excluded 7 participants based on an attention check (please select "Good"). We pooled the samples of the pilot study (N = 120) and the formal study (N = 1,779). The designs of these two studies are the same.

random assignment of treatments as instruments. Specifically, we leveraged (1) the assignment of affective guidance (for review writing) as an instrument for the affective score of reviews, (2) the assignment of cognitive guidance as an instrument for the cognitive score of reviews, and (3) the assignment of length-contingent incentives as an instrument for review length. To recap, these three types of treatments are tightly linked to one of the three endogenous features of review content, ensuring identification (Hill et al. 2021).

Other variables

In addition, we controlled for consumers' perceived familiarity with the podcast, perceived importance of sleeping, podcast listening habits, and demographics as covariates, because these variables can impact consumers' information processing style during service evaluation and their subsequent decisions (e.g., Jensen and Yetgin 2017; Krefeld-Schwalb, Sugerman, and Johnson 2024).

Empirical Analysis

To test the associations between the aforementioned instruments and endogenous review features, we ran the first-stage regressions, which also serve as manipulation checks regarding the effectiveness of our treatments. Then, in second-stage analyses, we estimated the causal impacts of customer review content on their subsequent decisions. Further, we explored heterogeneity across consumers. We detail our model specifications in Web Appendix R. The unit of analysis is each review.

First-stage analyses: Effects of interventions on review content

As shown in Table 5, first-stage analyses demonstrate the effectiveness of interventions in exogenously shaping review content features, supporting the relevance of these ERIVs (Sajons 2020). We highlight several key findings below. (1) Affective (cognitive) guidance for review writing significantly enhances the affective (cognitive) score of reviews,

suggesting that affective (cognitive) guidance leads participants to rely on affective (cognitive) processing when writing reviews. Specifically, the presence of affective (cognitive) guidance for review writing increases the proportion of affective (cognitive) content by 3.3% (3.5%). Pairwise comparisons with how much participants focus on emotions vs. cognitions during review provision as DVs reveal a similar pattern (included in Web Appendix S). (2) Incentives bring about a boost of 30 characters in review length ($M_{Incentive\text{-}Present}$ = 74, SD = 37; $M_{Incentive\text{-}Absent}$ = 45, SD = 36; t(1897) = 17.64, $p < .000$, Cohen's d = .75, r_{pb} = .38). (3) The manipulation of quality is successful: Participants could distinguish between high-quality vs. low-quality podcasts, as evidenced by the significant differences in quality ratings between the high- and low-quality conditions ($M_{High\text{-}Quality}$ = 5.05, SD = 1.06; $M_{Low\text{-}Quality}$ = 3.92, SD = 1.35; t(1897) = 20.21, $p < .000$, Cohen's d = .84, r_{pb} = .42). Web Appendix S provides details about the first-stage analyses.

Table 5: Effects of Experimental Interventions on Review Content Features (Study 2).

Columns	(1)	(2)	(3)
DVs	Affective Score (LIWC)	Cognitive Score (LIWC)	Review Length
Units	Percentage	Percentage	Character
I(Affective Guidance)	.033	.017	-4.260
	(.004, .000)	(.004, .000)	(1.934, .028)
I(Cognitive Guidance)	-.015	.035	2.872
	(.003, .000)	(.004, .000)	(2.087, .169)
I(Review Incentives)	-.009	-.011	29.292
	(.003, .001)	(.003, .002)	(1.683, .000)
Control Variables	Y	Y	Y
Excluded F-stat.	66.969	30.606	112.705
SW F-stat.	122.462	67.392	135.863
AP F-stat.	150.578	67.590	205.160
R-squared	.133	.091	.161

Notes: N = 1899. The first set of numbers in parentheses represents robust standard errors, while the second set indicates *p*-values. Control variables included podcast quality, broadcaster gender, consumers' podcast listening frequency, perceived familiarity with the podcast, perceived importance of sleep, and demographics. For multiple endogenous variables, we used Angrist-Pischke (AP) F-statistics and Sanderson-Windmeijer (SW) F-statistics in weak identification tests (Angrist and Pischke 2009; Sanderson and Windmeijer 2016). All these F-statistics exceed the Stock-Yogo critical values (9.08 for 10% and 13.91 for 5% maximal IV relative bias, Stock and Yogo 2002), supporting the relevance of the instruments.

Second-stage analyses: Effects of review content on referral and purchase decisions

As summarized in Table 6, the *Affective (Cognitive) Score* has a positive (negative) effect on referral and purchase decisions. Aligning with Study 1, these results indicate that when consumers are more engaged in affective (cognitive) processing during review

provision, they are more (less) likely to recommend or purchase the product. In contrast, the coefficients of *Review Length* are sometimes insignificant. A possible reason is that the impacts of review length vary across circumstances, canceling each other out.

Accordingly, we explored heterogeneous effects of review length across consumers (presented in Table 7). Based on the elaboration likelihood model (Petty and Cacioppo 1986), we selected consumer characteristics that influence information processing during review writing, including (1) prior experience, as reflected in the frequency of listening to podcasts in daily life, (2) gender, and (3) age (Krefeld-Schwalb, Sugerman, and Johnson 2024).

First, when consumers listen to podcasts more frequently in their daily lives, review length has a more negative impact on their referral/purchase decisions. A possible account is that consumers who habitually listen to podcasts accumulate more expertise in this area and are more sensitive to the deficiencies in the podcast trial. Thus, when writing longer reviews, they are more likely to identify product flaws and less likely to recommend or purchase, echoing the expert information processing model (Lord and Maher 1990).

Second, regarding the moderation role of gender, relative to males, the impacts of review length on referral/purchase are more harmful to females, suggesting that women are more inclined to engage in detailed processing when writing longer reviews. Such explanations align with the "selectivity model," which states that females are comprehensive information processors who attend to subtle cues, while males are selective information processors who resort to heuristics and miss subtle cues (Darley and Smith 1995).

Additionally, compared with younger listeners, the relationships between review length and referral/purchase are more negative for older people. This is attributable to the elderly being more cautious and having more time to process information carefully (Botwinick 1984). Therefore, they are more likely to detect product shortcomings when writing longer reviews and are more reluctant to recommend or purchase it. As detailed in Web Appendix T,

we performed robustness checks similar to Study 1 and replicated key findings.

Discussion

Study 2 demonstrated the generalizability of what we found in Study 1 in a distinct context (podcast trials). Like Study 1, Study 2 leveraged experimental randomization to examine the causal effects of consumers' review content features on their subsequent referral and purchase decisions (Sajons 2020). It addressed the limitations of Study 1 by directly controlling product quality and involving all consumers in referral decisions, thus eliminating biases related to unobserved quality and sample selection. Moreover, richer demographic data of Study 2 allowed for a more nuanced analysis of how review length affects different groups, such as experienced consumers, females, and the elderly, highlighting the importance of customizing review solicitation approaches to accommodate diverse consumer profiles. To understand the financial implications, we performed back-of-the-envelope analyses, showing that a 1-SD increase in the affective (cognitive) score generates a profit gain (loss) of 219 (277) yuan from referrals and a 21.3% increase (13.3% decrease) in purchase rate per writer.[9]

In addition, cost-effectiveness analysis indicates the consequences of soliciting informative reviews with incentives or cognitive guidance. It also suggests the potential of affective guidance as a feasible tool for fostering review writers' loyalty: As we see in Web Appendix U, the net profit generated by affective guidance is 70.14 yuan per writer. As noted previously, these analyses reflect only the effect on *writers* and do *not* cover the impact on *readers*. While Studies 1 and 2 examined the focal relationships for incentivized or solicited reviews, it remains unclear if these findings apply to unsolicited and unincentivized reviews. Study 3 aims to fill this gap.

[9] According to industry reports on podcasts (e.g., Helmick 2021), the customer lifetime value of referred individuals is estimated at 318 yuan, assuming a 2-year customer lifespan, an annual profit of 200 yuan, a 50% churn rate, and a 20% discount rate.

Table 6: Effects of Review Content Features on Referral and Purchase Decisions (Study 2).

Columns	(1)	(2)	(3)	(4)	(5)	(6)
DVs	Referral Intention	Referral Likelihood	Number of Referrals	Number of Referrals	Purchase Intention	Purchase Likelihood
Units/Scales	A 7-Point Scale	Percentage	Individual	Individual	A 7-Point Scale	Percentage
Models	Linear Regression	Linear Probability	Linear Regression	Poisson	Linear Regression	Linear Probability
Affective Score (LIWC)	9.786	2.169	11.577	4.264	8.827	3.548
	(2.083, .000)	(.550, .000)	(2.634, .000)	(1.004, .000)	(2.093, .000)	(.581, .000)
Cognitive Score (LIWC)	-8.897	-1.199	-10.877	-4.382	-5.739	-1.659
	(2.788, .001)	(.709, .091)	(3.535, .002)	(1.424, .002)	(2.794, .040)	(.791, .036)
Review Length	-.005	-.001	-.007	-.002	.001	-.002
	(.003, .107)	(.001, .523)	(.004, .070)	(.002, .108)	(.003, .829)	(.001, .004)
Control Functions	Y	Y	Y	Y	Y	Y
Control Variables	Y	Y	Y	Y	Y	Y
Effect Size of Affective Score	.587	.130	.695	.671	.530	.213
Effect Size of Cognitive Score	-.712	-.096	-.870	-.919	-.459	-.133
F-stat. (Chi-squared)	80.659	65.740	58.804	550.732	80.838	133.841
(Pseudo) R-squared	.339	.276	.247	.101	.334	.361

Table 7: Contingent Effects of Review Length on Referral and Purchase Decisions (Study 2).

Columns	(1)	(2)	(3)	(4)	(5)	(6)
DVs	Referral Intention	Referral Likelihood	Number of Referrals	Number of Referrals	Purchase Intention	Purchase Likelihood
Models	Linear Regression	Linear Probability	Linear Regression	Poisson	Linear Regression	Linear Probability
Frequency of Listening to Podcasts × Length	-.0021	-.0006	-.0030	-.0012	-.0028	-.0003
	(.0008, .015)	(.0002, .006)	(.0010, .003)	(.0004, .006)	(.0009, .001)	(.0002, .166)
I(Female) × Length	-.0045	-.0010	-.0064	-.0022	-.0055	-.0010
	(.0021, .030)	(.0005, .056)	(.0026, .014)	(.0009, .015)	(.0020, .007)	(.0006, .090)
Age × Length	-.0002	-.0001	-.0002	-.0001	-.0002	-.00004
	(.0001, .093)	(.0000, .021)	(.0001, .082)	(.0001, .063)	(.0001, .076)	(.00003, .125)
Control Functions	Y	Y	Y	Y	Y	Y
Control Variables	Y	Y	Y	Y	Y	Y

Notes: N = 1899. We used the control function approach, with treatments as instruments (Wooldridge 2015). Standard errors (the first set of numbers in parentheses) and *p*-values (the second set of numbers) were adjusted using a bootstrap approach (Ebbes, Papies, and van Heerde 2022). Control variables included podcast quality, broadcaster gender, consumers' podcast listening frequency, perceived familiarity with the podcast, perceived importance of sleeping, and demographics. To alleviate collinearity, we examined a single interaction term separately. Joint examinations produced similar results. The SDs of the affective and cognitive scores produced by LIWC are .06 and .08, respectively. Given these statistics, we calculated effect sizes, i.e., the change in the DV produced by a 1-SD increase in the independent variable.

Study 3: Archival Data Analyses of Airline Reviews

In contrast to the previous studies, Study 3 tested the generalizability of our findings when *no* incentives or solicitations for reviews were present. It seems infeasible to use experiments for studying unsolicited reviews, because reviews collected in experiments are, by nature, solicited. So, we used secondary data from Skytrax, an online review platform for airline services, where consumers independently posted reviews without compensation. This dataset also includes data on whether consumers recommended the services after reviewing.[10] More information about this data can be found in Web Appendix V.

Data Description

The sample used in our analysis consists of 8,790 reviews written by 2,219 consumers for 81 different airlines from 2009 to 2019 (3,105 customer-airline dyads). The unit of analysis is customer-airline-review. Our sample is limited to customer-airline dyads with at least two available reviews within our observation window.

We used consumers' recommendation choices as our outcome measure of interest. In the analyzed sample, 79% of consumers recommended the services after writing reviews. We primarily focus on the following dimensions of review content as independent variables: (1) affective score, (2) cognitive score, and (3) review length. The operational definitions of these variables are similar to those of the prior two studies (listed in Web Appendix V). In addition to customer-airline fixed effects and time fixed effects (day-of-week, month-of-year, and year dummies), we controlled for a proxy of service quality (consumers' quality ratings) and the features of the focal trip, such as whether the consumer traveled in a premium seat (including first class, business class, and premium economy) and whether a consumer was a

[10] This dataset is accessible on Kaggle (https://www.kaggle.com/datasets/efehandanisman/skytrax-airline-reviews). Only recommendation data are available. Platform policies are accessible online (https://www.airlinequality.com/write-a-review). This dataset has also been used in relevant research on reviews (Wu and Morwitz 2025).

business traveler. This is because these variables can simultaneously impact consumers' review content and their referral decisions.

Empirical Analysis

To examine the associations between consumers' review content features and their referral likelihood, we mainly employed panel-data models, leveraging the variation across reviews of the same customer-airline pair. We acknowledge that these analyses are suggestive, due to the inherent limitations of archival data (e.g., no randomization). Web Appendix W provides details on our model specifications.

Associations between review content and referral choices

Results in Table 8 are consistent with prior findings: The *Affective* (*Cognitive*) *Score* positively (negatively) predicts referral likelihood, indicating the link between the way consumers process information during review provision and their referral choices. To ensure robustness, we conducted both joint and separate examinations with fixed-effects models, using measures computed by different textual mining methods (LIWC and Topic Modeling).

Table 8: Associations Between Review Content Features and Referral Likelihood.

Columns	(1)	(2)	(3)	(4)
Type of Examinations	Joint Examination	Separate Examination	Joint Examination	Separate Examination
Textual Measures	LIWC	LIWC	Topic Modeling	Topic Modeling
Affective Score	.467	2.580	4.455	8.689
	(.103, .000)	(.161, .000)	(.750, .000)	(1.092, .000)
Cognitive Score	-.034	-.938	-3.298	-6.572
	(.112, .759)	(.166, .000)	(.433, .000)	(.656, .000)
Review Length	-.000	-.000	.000	-.000
	(.000, .216)	(.000, .000)	(.000, .047)	(.000, .000)
Customer-Airline FE	Y	Y	Y	Y
Time FE	Y	Y	Y	Y
Control Variables	Y	Y	Y	Y
Effect Size of Affective Score	.014	.077	.089	.174
Effect Size of Cognitive Score	-.001	-.028	-.099	-.197
F-stat.	59.877	N/A	91.665	N/A
Within R-squared	.454	N/A	.593	N/A

Notes: N = 8790. The first set of numbers in parentheses represents standard errors clustered at the customer-airline pair level, while the second set indicates *p*-values. The analysis was restricted to customer-airline pairs with at least two reviews. We used fixed-effects models, incorporating control variables such as trip features (e.g., cabin type), customer-airline fixed effects, and time fixed effects. The SDs of the affective and cognitive scores produced by LIWC are .03; the SDs of the affective and cognitive scores computed through the topic modeling methods are .02 and .03, respectively. Effect sizes are the changes in the DV resulting from a 1-SD increase in the independent variable.

Robustness checks

We also replicated our results in the following robustness checks (summarized in Web Appendix X): (1) conducting instrumental-variable regressions (Stock and Watson 2019); (2) using alternative types of standard errors (two-way clustering standard errors, Cameron, Gelbach, and Miller 2011); (3) controlling for additional confounders (e.g., adding quality indicators for multiple facets of services as covariates, Stock and Watson 2019); and (4) accounting for measurement errors (Yang et al. 2018).

Discussion

Study 3 shows the applicability of our findings to unsolicited or unincentivized reviews through secondary data analyses of airline reviews on Skytrax. We acknowledge that this analysis is subject to the limitations of using observational data to assess causality. Nevertheless, in the presence of the other studies and given the objectives of this particular study, this analysis supports the generalizability of our findings across diverse contexts.

General Discussion

This research explores the effects of review content features on reviewers' subsequent decisions. To address the endogeneity issue, we employed a mixed-methods approach comprising a field experiment, a scenario experiment, and an archival data analysis. We introduced randomized interventions to create exogenous variations in review content. We analyzed the focal effects across contexts, including household services, podcast trials, and airline services. Our three studies consistently show that affective content facilitates, while cognitive content discourages, writers' referral and (re)purchase behaviors, suggesting the influence of review writing on customers’ subsequent decisions.

Theoretical and Substantive Takeaways

This research contributes to the customer review literature in three key ways: (1)

Shifting focus to review writers: While prior studies primarily emphasize how readers (e.g., consumers and firms) react to review (e.g., Fradkin and Holtz 2023; Varga and Albuquerque 2024), this study focuses on review writers. By comparing our findings with prior literature, we identify similarities and differences in how readers and writers process review content. Specifically, readers and writers exhibit *similar* positive responses to affective content (Guitart and Stremersch 2021). This can be explained by consumers' reliance on feelings and heuristics in decision-making (Lei, Yin, and Zhang 2024). However, readers and writers *differ* in how they process cognitive content. While cognitive content impedes writers' referral and (re)purchase decisions by enhancing deliberation on details (Chaiken 1980), it can improve readers' evaluations and facilitate their purchase decisions by increasing product knowledge (Akpinar and Berger 2017), reducing uncertainty (Guitart and Stremersch 2021), and building trust (Darke and Ritchie 2007). (2) *Expanding the scope of outcome variables*: Most research on review generation only examines review attributes like volume and valence as DVs (e.g., Gao et al. 2025). We extend beyond traditional metrics, investigating broader outcomes, including referral and repurchase behaviors. (3) *Exploring the effectiveness of common but understudied review solicitation strategies*: We empirically test commonly employed yet underexplored review elicitation methods, including length-contingent monetary rewards and affective/cognitive guidance. Moreover, this work contributes to the literature on referral and repurchase. Although existing studies have extensively examined predictors such as customer, product, or firm features (e.g., Jung et al. 2020), the impact of *review writing* on writers' referral and repurchase decisions remains largely unexplored.

Our findings also provide practical insights for marketers. First, we identify *an informativeness-loyalty trade-off*: Writing more informative reviews (i.e., longer reviews with more cognitive content, Hou and Ma 2022) can hamper writers' referral and (re)purchase behaviors. To manage this trade-off, companies should weigh the value of obtaining detailed

reviews against facilitating writers' loyalty based on their specific contexts, as the relative importance of the two objectives varies across contexts (Guitart and Stremersch 2021; Kaul et al. 2025). Generally, when the primary goal of review elicitation is to encourage writers' referral and (re)purchase behaviors, managers can leverage an affect-oriented approach, offering emotional prompts and guidance to review writers. In contrast, when the main purpose is to gather informative feedback, businesses can employ cognition-oriented tools, such as providing templates or examples of rational reviews. Moreover, our moderation analyses reveal that this trade-off intensifies when products are easier to evaluate, prices are higher, or customers have more prior experience. This suggests that businesses can balance the trade-off, i.e., obtaining informative reviews while minimizing harm to writers' loyalty, by soliciting detailed feedback from customers who purchase more complex products, pay lower prices, or are less experienced. Additionally, when targeting customers who purchase simpler or more expensive products or those who are more experienced, businesses can employ affect-oriented review elicitation strategies to guide the review writing process, in order to reinforce these customers' loyalty behaviors.

Limitations and Directions for Further Research

Our research has several limitations that present opportunities for future exploration. First, it would be valuable to investigate how customer referral decisions influence their review-writing behaviors by reversing the sequence of review and referral requests. Second, beyond short-term outcomes, tracking the long-term effects of review writing on customer loyalty would deepen our understanding of these dynamics. Third, we did not delve into how incentives shape review content due to inconsistent results across contexts. Future studies could explore how and when incentives impact consumers' review-writing behaviors and subsequent decisions. Fourth, we did not elaborate on the interactions between review content features, as there is limited guidance on analyzing interactions between endogenous variables

(Murtazashvili and Wooldridge 2016).

Regarding generalizability, our empirical studies were conducted within the services industry. Future research could test the applicability of our findings to other contexts. Additionally, we did not measure the effect of review content on readers, as this has been extensively examined in prior research (e.g., Ludwig et al. 2013; Rocklage and Fazio 2020). Besides, the reviews collected in our experiments are not public reviews (e.g., those shared on platforms like Amazon), resembling private feedback seeking where firms solicit reviews mainly for internal use to improve product/service quality (Ananthakrishnan, Proserpio, and Sharma 2023; Kaul et al. 2025). Nevertheless, we encourage further exploration into the diverse influences of review content features on multiple stakeholders, including firms and potential customers.

To conclude, through methodological triangulation, we demonstrate how a key contextual factor—customer review content features—shapes the subsequent referral and repurchase behaviors of review writers. As review writing becomes more prevalent among consumers, understanding the influence of review writing on reviewers offers increasing strategic value. Overall, by uncovering the nuanced effects of review content on writers' subsequent decisions, this research provides managerial insights into strategically shaping the review provision process to drive key business outcomes and balance the trade-off between improving review informativeness and developing customer loyalty.

References

Akpinar, Ezgi and Jonah Berger (2017), "Valuable Virality," *Journal of Marketing Research*, 54 (2), 318–30.

Ananthakrishnan, Uttara, Davide Proserpio, and Siddhartha Sharma (2023), "I Hear You: Does Quality Improve with Customer Voice?," *Marketing Science*, 42 (6), 1143–61.

Angrist, Joshua D. and Jorn-Steffen Pischke (2009), *Mostly Harmless Econometrics: An Empiricist's Companion*. Princeton University Press.

Babić Rosario, Ana, Kristine De Valck, and Francesca Sotgiu (2020), "Conceptualizing the Electronic Word-of-Mouth Process: What We Know and Need to Know about eWOM Creation, Exposure, and Evaluation," *Journal of the Academy of Marketing Science*, 48 (3), 422–48.

Backlinko (2025), "Podcast Statistics You Need to Know," (March 25), https://backlinko.com/podcast-stats.

Bitner, Mary J and Carl Obermiller (1985), "The Elaboration Likelihood Model: Limitations and Extensions in Marketing," in *Advances in Consumer Research*, Elizabeth C. Hirschman and Moris B. Holbrook, eds. Association for Consumer Research, 420–25.

Bone, Sterling A., Katherine N. Lemon, Clay M. Voorhees, Katie A. Liljenquist, Paul W. Fombelle, Kristen Bell Detienne, and R. Bruce Money (2017), "'Mere Measurement Plus': How Solicitation of Open-Ended Positive Feedback Influences Customer Purchase Behavior," *Journal of Marketing Research*, 54 (1), 156–70.

Botwinick, Jack (1984), *Aging and Behavior: A Comprehensive Integration of Research Findings*. Springer.

Burtch, Gordon, Yili Hong, Ravi Bapna, and Vladas Griskevicius (2018), "Stimulating Online Reviews by Combining Financial Incentives and Social Norms," *Management Science*, 64 (5), 2065–82.

Camacho, Nuno, Martijn De Jong, and Stefan Stremersch (2014), "The Effect of Customer Empowerment on Adherence to Expert Advice," *International Journal of Research in Marketing*, 31 (3), 293–308.

Cameron, A. Colin, Jonah B. Gelbach, and Douglas L. Miller (2011), "Robust Inference with Multiway Clustering," *Journal of Business & Economic Statistics*, 29 (2), 238–49.

Chaiken, Shelly (1980), "Heuristic Versus Systematic Information Processing and the Use of Source Versus Message Cues in Persuasion," *Journal of Personality and Social Psychology*, 39 (5), 752–66.

Chan, Haksin and Selina Cui (2011), "The Contrasting Effects of Negative Word of Mouth in the Post-Consumption Stage," *Journal of Consumer Psychology*, 21 (3), 324–37.

Darke, Peter R., and Robin J. B. Ritchie (2007), "The Defensive Consumer: Advertising Deception, Defensive Processing, and Distrust," *Journal of Marketing Research*, 44 (1), 114–27.

Darley, William K. and Robert E. Smith (1995), "Gender Differences in Information Processing Strategies: An Empirical Test of the Selectivity Model in Advertising Response," *Journal of Advertising*, 24 (1), 41–56.

Dong, Yingying and Arthur Lewbel (2015), "A Simple Estimator for Binary Choice Models with Endogenous Regressors," *Econometric Reviews*, 34 (1–2), 82–105.

Duan, Jason A. and Carl F. Mela (2009), "The Role of Spatial Demand on Outlet Location and Pricing," *Journal of Marketing Research*, 46 (2), 260–78.

Ebbes, P., D. Papies, and H. J. van Heerde (2022), "Dealing with Endogeneity: A Nontechnical Guide for Marketing Researchers," in *Handbook of Market Research*, Christian Homburg, Martin Klarmann, and Arnd Vomberg, eds. Springer Nature, 181–217.

Ellis, Miriam (2024), "The Habits of Review Writers," *MOZ* (accessed May 18, 2025), https://moz.com/local-business-review-survey-report/habits-of-review-writers.
Fernández-Loría, Carlos, Maxime C. Cohen, and Anindya Ghose (2023), "Evolution of Referrals over Customers' Life Cycle: Evidence from a Ride-Sharing Platform," *Information Systems Research*, 34 (2), 698–720.
Fradkin, Andrey and David Holtz (2023), "Do Incentives to Review Help the Market? Evidence from a Field Experiment on Airbnb," *Marketing Science*, 42 (5), 853–65.
Gao, Baojun, Jing Wang, Xiaojie Ding, and Yue Guo (2025), "The Pitfalls of Review Solicitation: Evidence from a Natural Experiment on TripAdvisor," *Management Science*, 71 (2), 1671–91.
Gershon, Rachel and Zhenling Jiang (2025), "Referral Contagion: Downstream Benefits of Customer Referrals," *Journal of Marketing Research*, 62 (1), 97–116.
Glasgow, Russell E., Amy G. Huebschmann, and Ross C. Brownson (2018), "Expanding the CONSORT Figure: Increasing Transparency in Reporting on External Validity," *American Journal of Preventive Medicine*, 55 (3), 422–30.
Godes, David and Dina Mayzlin (2004), "Using Online Conversations to Study Word-of-Mouth Communication," *Marketing Science*, 23 (4), 545–60.
Goes, Paulo B., Mingfeng Lin, and Ching-man Au Yeung (2014), "'Popularity Effect' in User-Generated Content: Evidence from Online Product Reviews," *Information Systems Research*, 25 (2), 222–38.
Goldfarb, Avi, Catherine Tucker, and Yanwen Wang (2022), "Conducting Research in Marketing with Quasi-Experiments," *Journal of Marketing*, 86 (3), 1–20.
Greenberg, Melanie A. and Arthur A. Stone (1992), "Emotional Disclosure about Traumas and Its Relation to Health: Effects of Previous Disclosure and Trauma Severity," *Journal of Personality and Social Psychology*, 63 (1), 75–84.
Grootendorst, Maarten (2022), "BERTopic: Neural Topic Modeling with a Class-Based TF-IDF Procedure," *arXiv* (March 11), https://doi.org/10.48550/arXiv.2203.05794.
Guitart, Ivan A. and Stefan Stremersch (2021), "The Impact of Informational and Emotional Television Ad Content on Online Search and Sales," *Journal of Marketing Research*, 58 (2), 299–320.
Hawkins, LeAllen and Stephanie Hoon (2020), "The Impact of Customer Retention Strategies and the Survival of Small Service-Based Businesses," *The IUP Journal of Marketing Management*,19 (2), 7–34.
Helmick, Casey (2021), "What's the Value of Your Podcast Listener?," *Medium* (April 26), https://terrafirmaaudio.medium.com/whats-the-value-of-your-podcast-listeners-46ef502dd22.
Hennig-Thurau, Thorsten, Markus Groth, Michael Paul, and Dwayne D. Gremler (2006), "Are All Smiles Created Equal? How Emotional Contagion and Emotional Labor Affect Service Relationships," *Journal of Marketing*, 70 (3), 58–73.
Hill, Aaron D., Scott G. Johnson, Lindsey M. Greco, Ernest H. O'Boyle, and Sheryl L. Walter (2021), "Endogeneity: A Review and Agenda for the Methodology-Practice Divide Affecting Micro and Macro Research," *Journal of Management*, 47 (1), 105–43.
Hou, Jinghui (Jove) and Xiao Ma (2022), "Space Norms for Constructing Quality Reviews on Online Consumer Review Sites," *Information Systems Research*, 33 (3), 1093–1112.
Huang, Chin-Lan, Cindy K. Chung, Natalie Hui, Yi-Zheng Lin, Yi-Tai Seih, Ben C. P. Lam, Wei-Quan Chen, Michael H. Bond, and James W. Pennebaker (2012), "The Development of the Chinese Linguistic Inquiry and Word Count Dictionary," *Chinese Journal of Psychology*, 54 (2), 185–201.
Huang, Guofang and K. Sudhir (2021), "The Causal Effect of Service Satisfaction on Customer Loyalty," *Management Science*, 67 (1), 317–41.

Hunt, Earl and Franca Agnoli (1991), "The Whorfian Hypothesis: A Cognitive Psychology Perspective," *Psychological Review*, 98 (3), 377–89.
Isen, Alice M. and Barbara Means (1983), "The Influence of Positive Affect on Decision-Making Strategy," *Social Cognition*, 2 (1), 18–31.
Jayawardena, Nirma Sadamali, Park Thaichon, Sara Quach, Ali Razzaq, and Abhishek Behl (2023), "The Persuasion Effects of Virtual Reality (VR) and Augmented Reality (AR) Video Advertisements: A Conceptual Review," *Journal of Business Research*, 160, 113739, https://doi.org/10.1016/j.jbusres.2023.113739.
Jensen, Matthew L. and Emre Yetgin (2017), "Prominence and Interpretation of Online Conflict of Interest Disclosures," *MIS Quarterly*, 41 (2), 629–44.
Jin, Liyin and Yunhui Huang (2014), "When Giving Money Does Not Work: The Differential Effects of Monetary Versus In-Kind Rewards in Referral Reward Programs," *International Journal of Research in Marketing*, 31 (1), 107–16.
Jung, Jaehwuen, Ravi Bapna, Joseph M. Golden, and Tianshu Sun (2020), "Words Matter! Toward a Prosocial Call-to-Action for Online Referral: Evidence from Two Field Experiments," *Information Systems Research*, 31 (1), 16–36.
Kahneman, Daniel and Shane Frederick (2002), "Representativeness Revisited: Attribute Substitution in Intuitive Judgment," in *Heuristics of Intuitive Judgment: Extensions and Applications*, Thomas Gilovich, Dale Griffin, and Daniel Kahneman, eds. Cambridge University Press, 49–81.
Kaul, Rupali, Stephen J. Anderson, Pradeep K. Chintagunta, and Naufel Vilcassim (2025), "Call Me Maybe: Does Customer Feedback Seeking Impact Nonsolicited Customers?," *Marketing Science*, 44 (1), 129–54.
Kiesler, Charles A. (1971), *The Psychology of Commitment: Experiments Linking Behavior to Belief*. Academic Press.
Kim, Su Jung, Ewa Maslowska, and Ali Tamaddoni (2019), "The Paradox of (Dis)trust in Sponsorship Disclosure: The Characteristics and Effects of Sponsored Online Consumer Reviews," *Decision Support Systems*, 116, 114–24.
Kornish, Laura J. and Qiuping Li (2010), "Optimal Referral Bonuses with Asymmetric Information: Firm-Offered and Interpersonal Incentives," *Marketing Science*, 29 (1), 108–21.
Krefeld-Schwalb, Antonia, Eli Rosen Sugerman, and Eric J. Johnson (2024), "Exposing Omitted Moderators: Explaining Why Effect Sizes Differ in the Social Sciences," *Proceedings of the National Academy of Sciences*, 121 (12), 1–9.
Lee, Leonard and Claire I. Tsai (2014), "How Price Promotions Influence Postpurchase Consumption Experience over Time," *Journal of Consumer Research*, 40 (5), 943–59.
Lei, Zhanfei, Dezhi Yin, and Han Zhang (2024), "Deliberative or Automatic: Disentangling the Dual Processes Behind the Persuasive Power of Online Word-of-Mouth," *MIS Quarterly*, 49(1), 331–346.
Leung, Fine F., Sara Kim, and Caleb H. Tse (2020), "Highlighting Effort Versus Talent in Service Employee Performance: Customer Attributions and Responses," *Journal of Marketing*, 84 (3), 106–21.
List, John A., Azeem M. Shaikh, and Atom Vayalinkal (2023), "Multiple Testing with Covariate Adjustment in Experimental Economics," *Journal of Applied Econometrics*, 38 (6), 920–39.
Littrell, Jill (1998), "Is the Reexperience of Painful Emotion Therapeutic?," *Clinical Psychology Review*, 18 (1), 71–102.
Lord, Robert G. and Karen J. Maher (1990), "Alternative Information-Processing Models and Their Implications for Theory, Research, and Practice," *Academy of Management Review*, 15 (1), 9–28.

Ludwig, Stephan, Ko De Ruyter, Mike Friedman, Elisabeth C Brüggen, Martin Wetzels, and Gerard Pfann (2013), "More Than Words: The Influence of Affective Content and Linguistic Style Matches in Online Reviews on Conversion Rates," *Journal of Marketing*, 77 (1), 87–103.

Ma, Dandan, Shuqing Li, Jia Tina Du, Zhan Bu, Jie Cao, and Jianjun Sun (2022), "Engaging Voluntary Contributions in Online Review Platforms: The Effects of a Hierarchical Badges System," *Computers in Human Behavior*, 127, 107042, https://doi.org/10.1016/j.chb.2021.107042.

Mayo, Ruth (2015), "Cognition is a Matter of Trust: Distrust Tunes Cognitive Processes," *European Review of Social Psychology*, 26 (1), 283–327.

Millwood, Aimee (2023), "How to Ask Customers for Reviews: Best Practices and Templates," *yotpo* (August 21), https://www.yotpo.com/resources/how-to-ask-customers-for-reviews/.

Murtazashvili, Irina and Jeffrey M. Wooldridge (2016), "A Control Function Approach to Estimating Switching Regression Models with Endogenous Explanatory Variables and Endogenous Switching," *Journal of Econometrics*, 190 (2), 252–66.

Pennebaker, James W. and Sandra K. Beall (1986), "Confronting a Traumatic Event: Toward an Understanding of Inhibition and Disease," *Journal of Abnormal Psychology*, 95 (3), 274–81.

Petty, Richard E. and John T. Cacioppo (1986), "The Elaboration Likelihood Model of Persuasion," in *Communication and Persuasion*. Springer, 1–24.

Petty, Richard E., Leandre R. Fabrigar, and Duane T. Wegener (2003), "Emotional Factors in Attitudes and Persuasion," in *Handbook of Affective Sciences*, R. J. Davidson, K. R. Scherer, and H. H. Goldsmith, eds. Oxford University Press, 752–72.

Paget, Sammy (2025), "Local Consumer Review Survey 2025," *Bright Local* (January 29), https://www.brightlocal.com/research/local-consumer-review-survey/#section-3-how-are-consumers-writing-business-reviews.

Pocchiari, Martina, Davide Proserpio, and Yaniv Dover (2024), "Online Reviews: A Literature Review and Roadmap for Future Research," *International Journal of Research in Marketing* (published online August 29), https://doi.org/10.1016/j.ijresmar.2024.08.009.

Research and Markets (2025), "Home Cleaning Services - Global Strategic Business Report," Research Report ID. 5519705, https://www.researchandmarkets.com/reports/5519705/home-cleaning-services-global-strategic#rela0-5674487.

Rocklage, Matthew D. and Russell H. Fazio (2020), "The Enhancing Versus Backfiring Effects of Positive Emotion in Consumer Reviews," *Journal of Marketing Research*, 57 (2), 332–52.

Sajons, Gwendolin B. (2020), "Estimating the Causal Effect of Measured Endogenous Variables: A Tutorial on Experimentally Randomized Instrumental Variables," *The Leadership Quarterly*, 31 (5), 101348, https://doi.org/10.1016/j.leaqua.2019.101348.

Sanderson, Eleanor and Frank Windmeijer (2016), "A Weak Instrument F-Test in Linear IV Models with Multiple Endogenous Variables," *Journal of Econometrics*, 190 (2), 212–21.

Shiv, Baba and Alexander Fedorikhin (1999), "Heart and Mind in Conflict: The Interplay of Affect and Cognition in Consumer Decision Making," *Journal of Consumer Research*, 26 (3), 278–92.

Singh, Jagdip (2000), "Performance Productivity and Quality of Frontline Employees in Service Organizations," *Journal of Marketing*, 64 (2), 15–34.

Skowronski, John J. and Donal E. Carlston (1989), "Negativity and Extremity Biases in

Impression Formation: A Review of Explanations," *Psychological Bulletin*, 105 (1), 131–42.
Stock, James H. and Mark W. Watson (2019), *Introduction to Econometrics*. Pearson.
Stock, James H. and Motohiro Yogo (2002), "Testing for Weak Instruments in Linear IV Regression," Technical Working Paper No. 0284, National Bureau of Economic Research (November), https://www.nber.org/papers/t0284.
Sun, Tianshu, Siva Viswanathan, Ni Huang, and Elena Zheleva (2021), "Designing Promotional Incentives to Embrace Social Sharing: Evidence from Field and Online Experiments," *MIS Quarterly*, 45 (2), 789–820.
Swait, Joffre and Wiktor Adamowicz (2001), "The Influence of Task Complexity on Consumer Choice: A Latent Class Model of Decision Strategy Switching," *Journal of Consumer Research*, 28 (1), 135–48.
Tversky, Amos and Daniel Kahneman (1991), "Loss Aversion in Riskless Choice: A Reference-Dependent Model," *The Quarterly Journal of Economics*, 106 (4), 1039–61.
Varga, Marton and Paulo Albuquerque (2024), "The Impact of Negative Reviews on Online Search and Purchase Decisions," *Journal of Marketing Research*, 61 (5), 803–20.
Watanabe, Kohei and Alexander Baturo (2024), "Seeded Sequential LDA: A Semi-Supervised Algorithm for Topic-Specific Analysis of Sentences," *Social Science Computer Review*, 42 (1), 224–48.
Whorf, Benjamin Lee and John Bissell Carroll (1956), *Language, Thought, and Reality: Selected Writings of Benjamin Lee Whorf.* The MIT Press.
Wooldridge, Jeffrey M. (2006), *Introductory Econometrics: A Modern Approach*. South Western.
Wooldridge, Jeffrey M. (2015), "Control Function Methods in Applied Econometrics," *Journal of Human Resources*, 50 (2), 420–45.
Wu, Alisa Yinghao and Vicki G Morwitz (2025), "Digital Therapy for Negative Consumption Experiences: The Impact of Emotional and Rational Reviews on Review Writers," *Journal of Consumer Research*, 51 (5), 937–58.
Yang, Mochen, Gediminas Adomavicius, Gordon Burtch, and Yuqing Ren (2018), "Mind the Gap: Accounting for Measurement Error and Misclassification in Variables Generated via Data Mining," *Information Systems Research*, 29 (1), 4–24.
York, Jeffrey G., Siddharth Vedula, and Michael J. Lenox (2018), "It's Not Easy Building Green: The Impact of Public Policy, Private Actors, and Regional Logics on Voluntary Standards Adoption," *Academy of Management Journal*, 61 (4), 1492–1523.
Yu, Yinan, Warut Khern-am-nuai, and Alain Pinsonneault (2022), "When Paying for Reviews Pays Off: The Case of Performance-Contingent Monetary Rewards," *MIS Quarterly*, 46 (1), 609–26.
Zellner, Arnold (1963), "Estimators for Seemingly Unrelated Regression Equations: Some Exact Finite Sample Results," *Journal of the American Statistical Association*, 58 (304), 977–92.
Zhang, Judy (Zijing), H. Alice Li, and Greg M. Allenby (2024), "Using Text Analysis in Parallel Mediation Analysis," *Marketing Science*, 43 (5), 953–70.