Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

Expectation-Related Online Consumer Reviews of Activewear Leggings: Understanding Consumer Satisfaction and Post-Purchase Evaluations

Eonyou Shin [1,*], Chreston Miller [2] and Behnam Malmir [3]

[1] Fashion Merchandising and Design, School of Design, Virginia Tech, Blacksburg, VA 24061, USA
[2] University Libraries, Virginia Tech, Blacksburg, VA 24061, USA; chmille3@vt.edu
[3] Gies College of Business, University of Illinois at Urbana-Champaign, Champaign, IL 61820, USA; bmalmir@illinois.edu
* Correspondence: eonyous7@vt.edu

Abstract

Online consumer reviews (OCRs) enable consumers to publicly communicate post-purchase evaluations, satisfaction, and electronic word-of-mouth (eWOM). This study investigates the associational relationships between multi-dimensional product and service evaluations, expectations, and overall consumer ratings. The unit of analysis is the individual OCR. A sample of 2300 OCRs of 20 best-selling leggings was drawn from a brand-owned website. Data were analyzed using topic modeling with human validation of topic labels, sentiment analysis, and regression models to account for product-level dependence and statistical uncertainty. Topic modeling extracted themes related to expectations, multidimensional product and service performance, confirmation/disconfirmation, satisfaction, and eWOM expressions. Sentiment analysis indicated that expectations, aesthetic product performance, and recommendations leaned positive, whereas service and extrinsic product performance exhibited neutral to negative sentiment. The regression results demonstrated positive associations between overall satisfaction and four dimensions: product performance, service performance, confirmation/disconfirmation, and post-purchase outcomes.

Keywords: online consumer reviews; electronic word-of-mouth; user-generated content; online ratings; consumer satisfaction; topic modeling; sentiment analysis; expectation confirmation theory

1. Introduction

Online consumer reviews (OCRs) have become one of the most influential information sources in electronic commerce, shaping consumer decision-making and firm strategies across digital marketplaces. As consumers increasingly rely on peer-generated information when shopping online, OCRs provide rich insights into customer experiences, expectations, and satisfaction with products and services [1]. In addition to guiding purchasing decisions, review platforms generate large volumes of user-generated content that firms can analyze to understand consumer preferences, identify product issues, and improve service performance [2]. Consequently, analyzing OCRs has become an important research area in electronic commerce, offering opportunities to better understand how consumers evaluate products in online environments.

A particularly relevant context for examining these evaluations is the rapidly growing market for activewear. The COVID-19 pandemic accelerated the popularity of the 'athleisure' trend by blurring the lines between activewear and fashion [3]. In the United States,

Check for updates

Academic Editors: Patrícia Dias and Alexandre Duarte

Received: 6 October 2026
Revised: 5 November 2026
Accepted: 9 November 2026
Published: 20 November 2026

Copyright: © 2026 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license.

activewear has become one of the fast-growing apparel categories, with U.S. women's activewear alone generating $35.57 billion in revenue in 2021—surpassing men's and children's activewear combined [3]. This rapid growth reflects consumers' increasing demand for versatile apparel that supports both athletic and lifestyle needs. These dual demands make consumer satisfaction in activewear more complex—and more strategically important—than in many other apparel categories.

In today's retail environment, OCRs have become a critical resource for evaluating activewear. Because OCRs are generated by consumers following product experiences, they provide naturally occurring expressions of the evaluations and expectations of consumers who choose to post reviews [1]. For performance-based attributes such as fit, comfort, and fabric quality that are difficult to judge online, OCRs serve as an especially valuable source of information for consumers' purchase decisions [4]. At the same time, OCRs often include expressive and aesthetic aspects such as versatility, appearance, and style coordination in addition to functional attributes [4].

Taken together, OCRs reveal a multidimensional view of how consumers evaluate activewear, reflecting both functional performance and fashion-driven considerations. For brands competing in digital markets, analyzing OCRs can reveal which product- and service-related evaluations are associated with favorable and unfavorable overall evaluations and which experiences are publicly discussed alongside recommendations, complaints, or transactional responses such as returns.

Researchers have long examined factors associated with consumer satisfaction [5] because satisfaction represents an important dimension of consumers' post-purchase evaluations. Satisfaction, often expressed through star ratings, is key to success, as it is positively related to brand loyalty [6] and, in turn, revenue [7]. The textual content of OCRs provides details about the products and services that shape satisfaction, offering companies valuable insights for improving their product and services [2]. For activewear brands, understanding how consumers articulate satisfaction and dissatisfaction in OCRs is particularly critical, given the category's dual emphasis on performance and style.

Given the large volume of review data generated on digital platforms, text mining techniques have emerged as powerful tools for extracting insights from unstructured textual information [8]. Methods such as topic modeling and sentiment analysis enable researchers to identify dominant themes within reviews and quantify the emotional tone associated with each theme [8]. These analytical approaches allow firms to prioritize issues based on consumer sentiment, identify product strengths and weaknesses, and make data-driven decisions regarding product development, feature enhancements, and marketing strategies.

Although prior studies have explored evaluative criteria in OCRs on customized clothing using qualitative approaches (e.g., [9]) and adaptive clothing using topic modeling and sentiment analysis (e.g., [10]), less is known about how multidimensional product and service evaluations are publicly expressed in brand-hosted reviews and how these evaluations are associated with overall ratings and subsequent forms of brand-related eWOM. This gap is particularly relevant to electronic commerce because OCRs serve simultaneously as records of individual post-purchase evaluation and as publicly accessible user-generated content that informs subsequent consumers. Activewear provides a useful context for examining this process because consumers evaluate products across functional, aesthetic, expressive, and extrinsic dimensions while also communicating those evaluations through OCRs.

There remains a dearth of academic research examining the content of OCRs on activewear to elucidate customer satisfaction using machine learning techniques (e.g., topic modeling analysis and sentiment analysis), in conjunction with the functional, expressive, and aesthetic consumer needs (FEA, [11]) model and expectation-confirmation theory

(ECT, [12]). Because ECT requires observable information about consumers' prior expectations, this study focuses specifically on reviews in which expectation-related expressions are explicitly stated. These reviews provide the textual context needed to examine how prior expectations are discussed alongside product and service evaluations, confirmation/disconfirmation, satisfaction, and subsequent responses. Accordingly, the study examines expectation-related review narratives rather than attempting to represent the full population of product reviews.

To address this gap, the purpose of this study is two-fold: (1) to identify the dominant features discussed in expectation-related OCRs of activewear and (2) to examine how evaluations of these features are associated with overall consumer evaluations. Because OCRs are a form of eWOM, consumers' accounts of expectations, product performance, and satisfaction are not only evaluations, but messages communicated to prospective buyers. This study therefore examines the content of activewear eWOM through the lens of ECT.

2. Literature Review

2.1. Expectation-Confirmation Theory

ECT explains how consumer satisfaction is formed at the post-purchase stage [12,13]. Consumer satisfaction is a function of expectations, perceived performance, and confirmation of belief. Expectations refer to the attributes or characteristics of a product or service that an individual predicts at the pre-purchase stage. Perceived performance reflects one's perception of the actual performance of the product or service and mediates the relationship between expectations and disconfirmation of belief or satisfaction. Confirmation of belief is defined as the evaluation based on the comparison between one's original expectations and the actual performance of the product or service. When perceived performance is consistent with prior expectations, expectations are confirmed; when performance exceeds expectations, consumers experience positive confirmation, whereas performance below expectations produces negative confirmation. Positive confirmation is positively related to satisfaction, which refers to the consumer's positive feelings toward the product or service [12,13]. Previous studies have confirmed the antecedents and consequences of satisfaction. The antecedents include initial expectations at the pre-purchase stage, and perceived performance and confirmation at the post-purchase stage, while repurchase intentions are a consequence of satisfaction [13,14].

ECT has been further applied in studies on consumer satisfaction using OCRs in different ways. Oh et al. [15] used deep learning techniques to elucidate customer satisfaction in hospitality services, based on expectations and confirmation in ECT. The researchers used OCRs and seller-provided information on hotels (both textual and images) to predict customer satisfaction with hotel services. Park [16] employed sentiment analysis using the Linguistic Inquiry and Word Count (LIWC) software to analyze OCRs, examining the effects of UX elements (i.e., hedonic values, user burdens, expectation confirmation, pragmatic values, and social values) on satisfaction. However, Jang and Yi [17] noted a limitation of the LIWC software—it only allows the use of a single keyword to categorize it into a dimension of UX elements.

In a recent study, Kumar et al. [18] used Amazon customer review data regarding appliances to examine the effects of product- and service-related quality attributes mentioned in OCRs on repurchase intentions through customer satisfaction with products and services. The study provides useful evidence of the value of textual review data in understanding consumer behavior. However, the study did not examine sentiment values associated with individual attributes or include confirmation as part of the model, which could have offered a more holistic view of perceived quality. In addition, the omission of confirmation, a central element in expectation-confirmation theory, leaves the model less

comprehensive in explaining satisfaction and repurchase behavior. Building on these limitations, the present study incorporates multi-dimensional quality and integrates confirmation with sentiment values to provide a more robust understanding of customer satisfaction and repurchase intentions.

2.2. Functional, Expressive, and Aesthetic Consumer Needs (FEA) Model

Lamb and Kallal [11] proposed a model of functional, expressive, and aesthetic consumer needs (FEA model) to help designers understand target consumers' needs. In the FEA model, consumers are placed at the center, encircled by culture, as "culture acts as a mediator or filter between the intended users of apparel and their requirements or desires in their apparel items" (p. 43). The design criteria consist of functional, expressive, and aesthetic components. Functional criteria relate to the utility of apparel products, including protection, thermal comfort, fit, and ease of movement (i.e., physical comfort). Expressive criteria concern the communicative and symbolic aspects of dress, such as values, roles, status, and self-esteem. Aesthetic criteria involve the use of design and art elements, as well as appearance [11].

The FEA model has also been applied mainly to plus-sized exercise apparel to explore female consumers' needs. Greenleaf et al. [19] found that plus-size women were not happy with their experiences shopping for exercise apparel due to unfashionable designs, inaccurate sizing, limited variety, and poor size availability. In the case of swimwear for plus-size women, functional needs (i.e., poor fit, low quality, difficulty putting on and taking off, and insufficient bust support) were often unmet, followed by aesthetic needs (i.e., style, color, and print) and expressive needs (i.e., too revealing cut) [20].

Recently, researchers applied the FEA model to develop a taxonomy of consumer needs for pants based on OCR data from the largest online fashion platform [21]. Functional attributes included ease, weight, durability, maintenance, and performance features such as waterproof and windproof. Values among expressive needs were most frequently mentioned, particularly in relation to specific occasions such as workouts and travel. Aesthetic needs were described in terms of appearance on the leg, waist, and hip, as well as coverage (e.g., concerns about transparency) and coordination with shoes [19]. While this taxonomy provides a useful categorization, the study largely emphasized surface-level mentions rather than deeper sentiment. Without weighting or sentiment analysis, it is unclear how strongly these needs drive purchase decisions relative to functional or aesthetic attributes.

2.3. Evaluative Criteria of Apparel Products

Researchers have focused on identifying features that consumers use to evaluate apparel products. Eckman et al. [22] found that a total of fifteen features of apparel products were used, either positively and/or negatively, when conducting specific garment evaluations: Color/pattern, styling, fit, fabric, price, comfort, versatility, matching, appearance, uniqueness, care, workmanship, appropriateness, brand, and utility. These features were grouped into four criteria: aesthetic, usefulness, performance and quality, and extrinsic criteria [20].

Abraham-Murali and Littrell [23] identified conceptual dimensions of apparel attributes at the expectation and post-purchase evaluation stages. At the expectation stage, consumers mentioned having expectations regarding fabric and garment construction, care, value, style, appearance on the body, individuality, and expression. At the post-purchase stage, Abraham-Murali and Littrell [21] evaluated the product in terms of self-confidence, fit, appearance, fabric, care, and individuality.

Ghalachyan and Karpova [24] developed the Apparel Product Evaluation (APE) framework to clarify and update evaluative criteria for apparel products, categorizing them into four dimensions: intrinsic, marketing, functional, and socio-communicative. Intrinsic attributes refer to product characteristics determined by designers and producers, such as color/pattern, materials, construction, design features, shape/silhouette, style, smell, and even sound. Marketing attributes include price, brand, country of origin, and service—factors external to the garment itself but influential in consumer evaluations. Functional attributes align with earlier [22,23], encompassing performance-related aspects such as comfort, durability, and care. Finally, socio-communicative attributes correspond to what previous studies labeled expressive attributes [22,23], capturing how apparel communicates identity, social meaning, and self-expression. The APE framework illustrates the ongoing evolution of apparel evaluation models.

2.4. Online Consumer Reviews as Electronic Word of Mouth and Analytics

Electronic word of mouth (eWOM) refers to positive or negative statements that consumers communicate about products or companies through the Internet [25]. OCRs posted on retail websites are a prominent form of eWOM because they make consumers' post-purchase evaluations publicly available to other consumers. Hennig-Thurau et al. [25] identified several motivations for eWOM, including social benefits, economic incentives, concern for other consumers, and self-enhancement, indicating that online reviews serve communicative functions beyond simply recording a product rating.

OCRs can also influence consumer decision making and marketplace outcomes. Chevalier and Mayzlin [26] found that online reviews were associated with relative sales, while Mudambi and Schuff [27] showed that review depth and rating extremity can influence perceived helpfulness. In the context of ECT, OCRs may contain evidence of consumers' post-purchase evaluations and confirmation or disconfirmation of prior expectations [12,13], while also providing information to prospective consumers.

However, OCRs should not be treated as a representative cross-section of all consumer experiences. Hu et al. [28] showed that online ratings often follow a J-shaped distribution and may reflect purchasing and reporting biases. This issue is particularly relevant to the present study because the analysis focuses specifically on reviews containing explicit expectation-related expressions. Accordingly, the retained reviews are treated as an expectation-focused subset rather than as representative of all purchasers or activewear consumers.

The growing volume of OCRs has also increased the use of automated text analysis in marketing research. Berger et al. [29] reviewed approaches including dictionary-based methods, topic modeling, and machine learning, while emphasizing issues of measurement and validity. Hartmann et al. [30] further showed that sentiment-analysis methods differ substantially in accuracy and performance. In the present study, VADER was used to derive continuous sentiment-valence scores from short consumer-generated text, while the topic-modeling approach was selected to accommodate the short-text structure of OCRs, as described in Section 3.3.2.

Recent research has expanded the use of advanced text analytics and artificial intelligence for understanding OCRs. For example, Kim et al. [31] developed a topic-based approach to identify textual drivers of star ratings from unstructured online reviews, while Praveen et al. [32] demonstrated how large language models can be used to extract topics, emotions, and sentiment from large-scale consumer-review data. More recent research has further examined the relationship between textual sentiment and numerical ratings and the implications of such consistency for review usefulness [33]. Emerging evidence also suggests that generative AI is changing the production of eWOM itself by influencing review

volume, length, subjectivity, and sentiment expression [34]. Together, these developments demonstrate the growing sophistication of computational approaches to OCR analysis and highlight the value of examining how text-derived evaluations relate to consumers' overall post-purchase judgments.

2.5. Content of Online Consumer Reviews of Apparel Products

In recent studies, researchers have identified evaluative features shown in the context of OCRs. McKinney and Shin [35] found that consumers evaluated rental dresses in OCRs in terms of fit, garment styling, social feedback, appropriateness, color and pattern, fabric, and physical comfort. For custom-fit T-shirts, consumers used similar evaluative criteria in OCRs, including care, comfort, fabric feel, fit, quality, workmanship, matching, social feedback, uniqueness, usage situation, appearance, color, weight, country of origin, and price [10].

Lang et al. [8] investigated features found in positive and negative subsets of OCRs regarding mass-customized apparel products. In the positive OCRs, the most frequently mentioned features were functional value (e.g., fit, quality), ease of use services, and emotional value (e.g., enjoyment and pleasure). In the negative OCRs, consumers mentioned features related to poor service, product performance (e.g., fit, quality), and financial risk (e.g., high price).

In OCRs, because consumers' needs in the activewear are likely to be more functional than expressive or aesthetic [20], they may use different attributes for activewear than for formal wear when evaluating activewear. For plus-size women, their functional needs in exercise apparel included a longer inseam on pants and shorts for their larger thighs [4]. Female consumers also expressed their needs in terms of more variety of size (functional) and style (aesthetic) options for activewear, in addition to the high cost of the activewear as an extrinsic factor [4,19].

Recent research on OCRs confirms that the evaluative criteria for apparel products, originally identified in traditional shopping contexts, are also evident in online reviews. For example, Sharma [36] found that consumers frequently discussed aesthetic, functional, expressive, performance-related, and extrinsic aspects, along with return policies, when reviewing women's apparel online—mirroring many of the classic criteria outlined by Eckman et al. [22]. Similarly, Shin et al. [10] demonstrated that OCRs for custom-fit T-shirts revealed functional attributes (e.g., care, comfort, fit, fabric feel, quality, workmanship), expressive attributes (e.g., matching, social feedback, uniqueness, usage situation), aesthetic attributes (e.g., appearance, color, weight), and extrinsic attributes (e.g., price, country of origin), echoing the dimensions proposed by Abraham-Murali and Littrell [22] and Eckman et al. [21]. In addition, Shin et al. [9] identified themes related to expectations, confirmation, satisfaction, and outcomes of satisfaction in OCRs, which are consistent with the processes described in expectation-confirmation theory (ECT; [12,13]). Together, these studies suggest that OCRs not only capture consumers' overall satisfaction judgments but also provide rich evidence of product performance, service attributes, and confirmation processes that drive evaluations in contemporary apparel markets. However, little is known about whether these evaluative criteria and ECT-related processes appear similarly in OCRs for activewear, highlighting the need for further investigation.

Previous studies have examined activewear needs and applied text-mining approaches to apparel reviews. However, limited research has examined how multidimensional product and service evaluations and ECT-related processes are jointly reflected in expectation-related OCRs of activewear products. Accordingly, the current study used text mining to identify key attributes discussed in activewear OCRs and examined how sentiment associated with these attributes and ECT-related evaluations was related to consumers' overall evaluations,

as reflected in star ratings. These relationships are further developed in the conceptual framework and hypotheses presented in the following section.

2.6. Conceptual Framework and Hypotheses Development

Based on the theories and concepts from previous studies, we propose a conceptual framework regarding the content of online consumer reviews when evaluating activewear (see Figure 1). This conceptual framework categorizes features into major themes in OCRs, which encompass expectations, perceived performance, dis/confirmation, dis/satisfaction, and outcomes of dis/satisfaction.

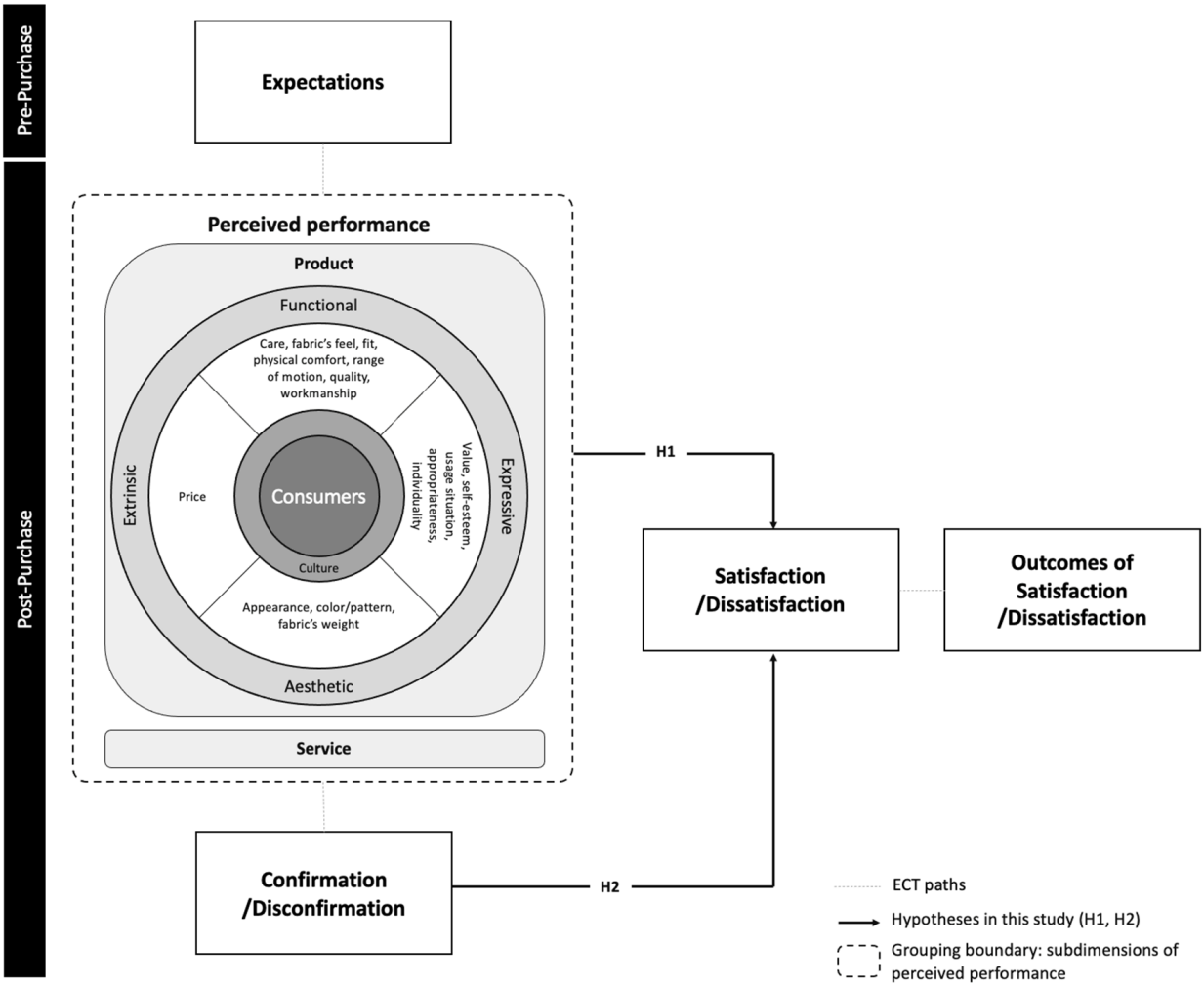


Figure 1. A Conceptual Framework for the Content of OCRs regarding Activewear.

Building on Shin et al. [9], expectation is conceptualized as a holistic, preexisting belief that consumers hold prior to purchase. Consumers frequently articulate skeptical, positive, or neutral expectations about apparel products before evaluating the performance, which aligns with the ECT perspective that expectation precedes experience and guides subsequent evaluations [12,13].

In contrast, perceived performance occurs post-purchase [12,13] when consumers evaluate products in relation to functional qualities (e.g., fit, comfort, durability), expressive qualities (e.g., versatility), aesthetic qualities (e.g., appearance, style), and extrinsic attributes (e.g., brand reputation, or value for money) [11,22,23]. These dimensions are well documented as evaluative criteria for apparel products and are directly observable in OCRs [9]. In addition, service performance (e.g., purchase process, return and exchange

policies) plays a critical role in shaping consumers' overall evaluations in online contexts [8]. Thus, our framework includes expectation as a general construct, while performance is evaluated through its multidimensional subdimensions. This articulation maintains theoretical alignment with ECT [12,13] and previous studies while also accounting for the exploratory nature of OCR data (e.g., [9]).

Building upon the FEA model, ECT, and previous studies, we propose a conceptual framework (Figure 1) to explore evaluative criteria for activewear within OCRs, as outlined in prior research (e.g., [9]). In a study of OCRs [9], consumers highlighted both the antecedents (expectation, perceived performance, confirmation) and consequences (compliments, suggestions, recommendation, repurchase intentions/behavior, regret, return intentions/behavior, complaints, product disposal) of satisfaction when evaluating custom-fit T-shirts. This evidence supports the ECT framework [12,13], which views confirmation or disconfirmation as a critical mechanism linking expectation and performance to satisfaction. Similarly, in OCRs for activewear leggings products, consumers may express the antecedents and consequences of their satisfaction. In the conceptual model, expectations, perceived performance, and confirmation act as antecedents of satisfaction, while the outcomes (positive or negative) are its consequences. Importantly, consumer satisfaction level in OCRs for activewear leggings may vary in strength depending on which subdimension of performance is emphasized, reflecting the multidimensional nature of consumer evaluation.

Drawing on FEA, prior apparel-evaluation research, and ECT, the framework further proposes that consumers evaluate apparel across multiple dimensions, including functional, expressive, aesthetic, and extrinsic product performance. Although these dimensions were used as a priori categories to guide interpretation of the review content, whether evaluations within these dimensions are associated with consumers' overall evaluations remains an empirical question.

Hypothesis 1. *More positive evaluations of functional, expressive, aesthetic, and extrinsic product performance and service performance will be associated with higher satisfaction, as reflected in star ratings.*

Based on ECT, consumers' overall evaluations are also expected to vary according to whether their product experiences confirm or disconfirm prior expectations. Accordingly, more positive confirmation expressed in OCRs is expected to be associated with more favorable overall evaluations.

Hypothesis 2. *More positive confirmation expressed in OCRs will be associated with higher satisfaction, as reflected in star ratings.*

3. Methods

In this section, we explain the process of collecting relevant OCRs, filtering, cleaning, and preprocessing the data, and organizing it for regression analyses. Figure 2 provides an overview of the entire procedure, including data collection, cleaning, preprocessing, and analysis.

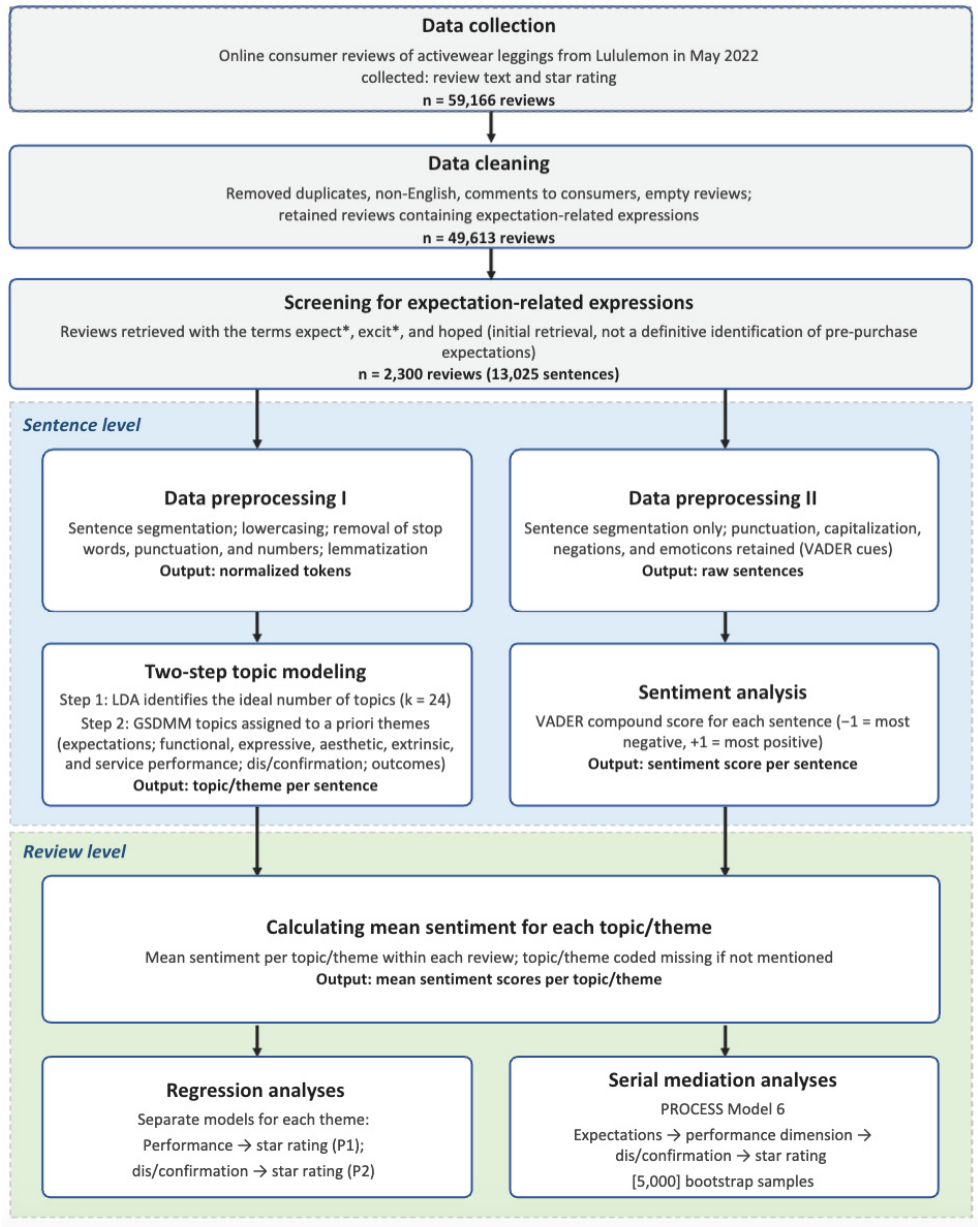


Figure 2. Overall Procedure in Data Collection, Cleaning, Preprocessing, and Data Analysis. * indicates that multiple word forms sharing the same stem were included in the screening.

3.1. Data Collection

Using the Selenium library in Python 3.9, we crawled OCRs for twenty leggings listed under the best-seller category from the Lululemon website. Leggings were chosen because they have become a staple in activewear wardrobes. Lululemon was selected because it is renowned for high-quality leggings (Tighe, [37]) and offers a wide variety of styles catering to different activities and body types. A complete list of the 20 analyzed products is provided in Supplementary Table S1.

OCRs were collected in May 2022 from the Lululemon website, covering reviews posted between March 2017 and May 2022. Country information was not available. During extraction, we collected the following fields for each review (e.g., review text, star rating, review title). The initial crawl yielded a total of 59,166 raw reviews. After deleting duplicates, non-English, and comments from the customer service, a total of 49,613 reviews had remained. Because the study focused on consumers stating pre-purchase expectations and subsequent evaluations, reviews were screened for expectation-related expressions. The

screening terms included expect*, excit*, and hoped. This screening served as an initial retrieval procedure rather than a definitive identification of pre-purchase expectations. The screening process resulted in 2300 reviews containing 13,025 sentences, which were retained for further analyses.

To examine potential differences associated with the expectation-based screening procedure, we compared the filtered analytic sample with the broader eligible review corpus (see Supplementary Table S1). The expectation-filtered reviews had a lower mean star rating than the broader corpus (2.88 vs. 3.51) and were longer on average (73.3 vs. 50.1 words). Five-star reviews represented 18.0% of the filtered sample compared with 43.6% of the broader corpus, whereas 1- and 2-star reviews accounted for 42.9% and 32.9%, respectively. These differences indicate that reviews containing explicit expectation-related expressions tended to be longer and less positive than reviews in the broader corpus. However, the relative representation of the major products was generally similar across the two datasets.

To assess the validity of the keyword-based screening procedure, 200 retained reviews and a stratified random sample of 100 excluded reviews were examined. Among the retained reviews, 98.5% contained broadly expectation-related expressions. In the excluded sample, four reviews (4.0%) were identified as containing personal prior expectation-related statements, indicating that some relevant content may not have been captured by the predefined keywords.

The keyword screening and topic classification represented two separate analytic steps. The screening procedure was used to identify reviews containing language potentially related to pre-purchase expectations, whereas topic modeling was subsequently used to identify the substantive themes discussed within those reviews. Accordingly, a review could pass the initial screening criterion without being classified primarily under the expectations theme. The resulting analytic sample therefore represents expectation-related review narratives rather than the full population of product reviews.

3.2. Data Cleaning and Preprocessing

We performed several steps of data cleaning and preprocessing. During the data cleaning process, the OCRs were divided into a sentence-level dataset (n = 13,025) to facilitate preprocessing for topic modeling and sentiment analysis. Since topic modeling algorithms use a bag-of-words model, stop words (e.g., "and", "the", and "how") and custom stop words (e.g., "Lululemon" and specific product names like "Align") were removed, as they do not add significant value or context to the text being modeled. Given our focus on consumer opinions about products, we retained certain "keep" words commonly used to express opinions, such as "like" and "significant", as well as descriptive product-related phrases like "see through" and "thin". Additionally, we removed punctuation, numbers, and any web links.

Although we conducted a rigorous data preprocessing procedure for topic modeling analysis, a separate preprocessing approach was used for sentiment analysis. This involved deleting repetitive reviews while retaining punctuation, numbers, and the original case of words in OCRs, as these factors can influence sentiment classification results [34].

3.3. Data Analysis

3.3.1. Topic Modeling Analyses

Two steps of topic modeling analysis, using unigrams and bigrams, were performed following the guidance of Ridhwan and Hargreaves [38]: (1) A series of Latent Dirichlet Allocation (LDA) models [39] were run, starting with 1 topic and incrementing by 1 up to 30 to identify the ideal number of topics; and (2) the Gibbs Sampling Dirichlet Multinomial

Mixture (GSDMM) model [40] was used because it is specifically designed for short texts, based on the assumption that such texts contain only a single topic. Using the pyLDAvis tool [41], we reviewed 30 keywords for each topic with a lambda value of 0.6, as well as the top five keywords with a lambda value of 0.2, to qualitatively label each topic.

The results of the LDA analyses showed coherence scores for each topic, as shown in Figure 3. We reviewed the data and groups of keywords extracted from the GSDMM model [40] for topics 24, 28, and 30, all of which had coherence scores of 0.48 or higher. After closely examining the actual OCR data and keywords within these topics, 24 topics were identified as the ideal number, with a coherence score of 0.48. Then, using the pyLDAvis tool [41,42], we analyzed the keywords in each topic to label them (see Figure 4).

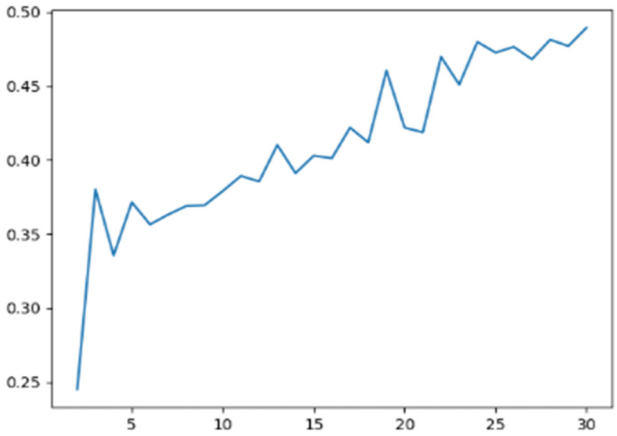


Figure 3. Coherence Scores for Each Number of Topics for LDA.

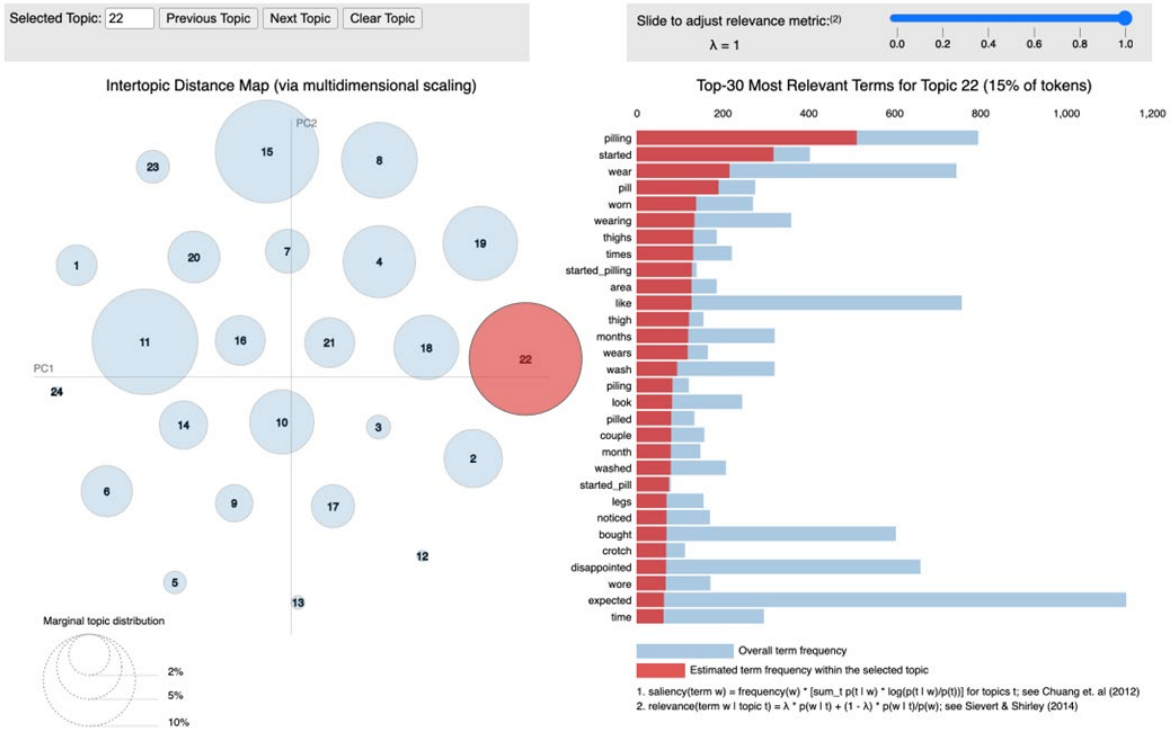


Figure 4. A Snapshot of the pyLDAvis Interface to Visualize the 24-topic GSDMM Model. Circle size indicates topic prevalence; the selected topic is shown in red and all other topics in blue in the (**left panel**). In the (**right panel**), red bars show term frequency within the selected topic, and blue bars show overall term frequency in the corpus [41,42].

More recent approaches, such as BERTopic, use transformer-based embeddings and clustering to capture contextual semantic similarity in text [43]. However, GSDMM was retained because it is designed for sparse short texts and assigns each short document to a single dominant topic, which aligns with our sentence-level topic assignment and subsequent review-level aggregation. Recent comparative studies also suggest that GSDMM can provide more interpretable topics in some short-text contexts despite BERTopic's higher coherence scores [44].

To validate the topic interpretation, the researchers conducted a human-coded review of the topic-modeling results. Two coders independently reviewed the randomly selected twenty sentences assigned to the identified topics with the associated keywords and labeled each topic. Then, they evaluated whether the proposed topic labels accurately represented the underlying review content. Disagreements were discussed until consensus was reached. The final topic labels were retained or revised based on this validation process.

Because GSDMM assigns one topic to each sentence, and reviews typically contain several sentences addressing different topics, we did not assign a single dominant topic to each review. Instead, a review was counted under a theme/topic if at least one of its sentences was assigned to a topic within that theme/topic.

The higher-level categories were informed by the conceptual framework developed a priori from FEA and prior apparel-evaluation literature. These categories were used to guide interpretation of the identified topics rather than to test whether FEA categories would independently emerge from the data. Topics that did not clearly correspond to the proposed categories were retained and interpreted according to their observed content rather than being forced into a predetermined category.

3.3.2. Sentiment Analysis

Valence Aware Dictionary for Sentiment Reasoning (VADER) [45], a powerful tool for gauging public opinion and sentiment on a large scale, was employed. Lexicon-based sentiment analysis approaches do not require labeled data to calculate sentiment scores. In our case, we do not have any labeled data. Thus, we use the VADER algorithm, which relies on predefined dictionaries of words and their associated sentiments. VADER encompasses four main phases: (1) Tokenization: in this phase, the text is broken into words, known as tokens; (2) Sentiment scoring: based on a predefined lexicon, sentiment scores are calculated for the extracted tokens; (3) Aggregation: the overall sentiment score for the entire text is produced by aggregating the individual sentiment scores. This score consists of positive, negative, and neutral scores, along with a compound score representing the overall sentiment; (4) Normalization: the aggregated scores are normalized to ensure they fall within a specific range, typically from −1 (extremely negative) to +1 (extremely positive). VADER is supported by the vaderSentiment library in Python, making it accessible for implementation in various projects. This library is utilized in this study. VADER sentiment scores were calculated for each sentence by summing the valence scores of each word from the lexicon, adjusted according to rules, and normalized between −1 (extremely negative) and +1 (extremely positive) to obtain a compound score.

After sentiment scores and topics were assigned at the sentence level, sentence-level sentiment scores within each theme were averaged to obtain review-level mean sentiment scores. This allowed us to capture comprehensive evaluations of the product and service based on the textual reviews provided by each consumer. For the review-level analyses, topic- and theme-specific sentiment scores were calculated only when the corresponding topic and theme were present in a review. If a topic/theme was not mentioned, the value for that theme was coded as missing rather than zero. Therefore, a sentiment score of zero represents neutral sentiment toward a topic/theme only when that theme was actually

present, whereas topic/theme nonmention represents the absence of observable information for that dimension.

3.3.3. Regression Analysis

The OCR data was organized at the review level, with sentiment scores ranging from −1 to 1 for each theme and star ratings from 1 to 5 in SPSS 29. To examine the effects of evaluative features (i.e., expectations, perceived product/service performance, and confirmation/disconfirmation) on satisfaction, Pearson's correlation and a series of regression analyses were performed. Because the prevalence of themes varied substantially across reviews, the number of observations available for each theme-specific analysis also varied. Separate models were therefore estimated for each theme to avoid restricting the analysis to the much smaller subset of reviews containing all themes. The effective sample size (N) is reported for each statistical model. Additionally, PROCESS analyses with Model 6 [46] were conducted to examine sequential relationships based on the ECT. For each dimension, mean sentiment scores for each theme at the review level were calculated. Based on the ECT, the model consisted of expectations as an independent variable (IV), product/service performance as the first mediator (M1), confirmation/disconfirmation as the second mediator (M2), and satisfaction/dissatisfaction as a dependent variable (DV) (see Figure 5). Since Model 6 included two sequential mediators (M1 and M2) with IV and DV, testing was conducted separately for each sub-dimension of perceived product performance (M1a□d) (i.e., 1a: functional, 1b: expressive, 1c: aesthetic, and 1d: extrinsic) and perceived service performance (M1e).

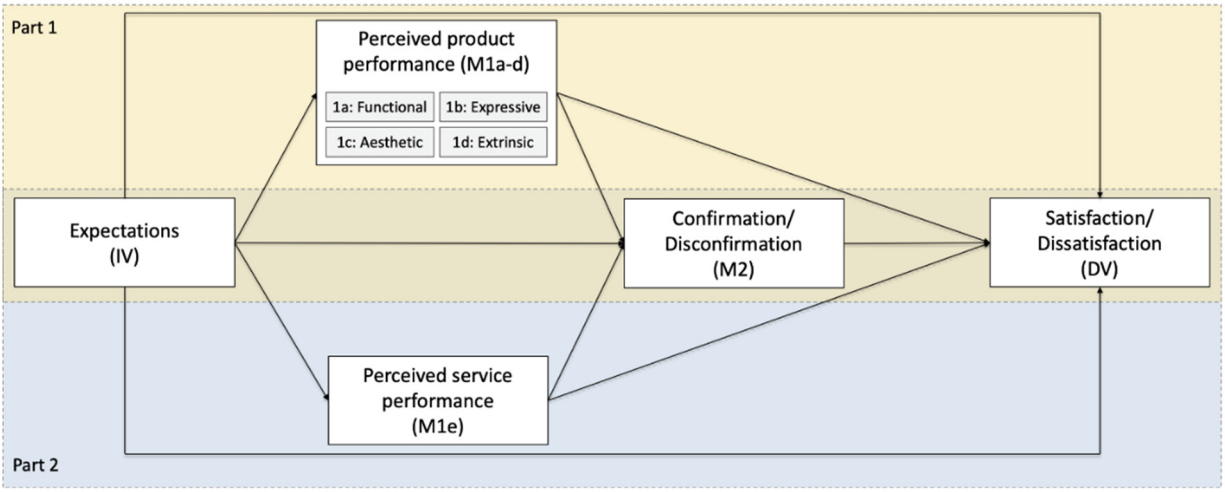


Figure 5. Testing the ECT using Model 6 (Part 1, Part 2) through the PROCESS Analysis with Model 6.

4. Results

4.1. Sample Characteristics Dominating Features Expressed in OCRs

A total of 2300 OCRs containing 13,025 sentences were used for further analysis. Among 13,025 sentences, 302 sentences (2.3%) were not used for further analysis because they contained no informative words after preprocessing, and the remaining 12,723 were assigned to 24 topics. The mean number of sentences per OCR was 5.66, ranging from 1 to 40 sentences. The average star rating for OCRs was 2.88, distributed across different star ratings as follows: 1-star reviews (f = 449), 2-star reviews (f = 538), 3-star reviews (f = 569), 4-star reviews (f = 331), and 5-star reviews (f = 413).

4.2. Dominant Features Expressed in OCRs

According to the results of topic modeling analyses, a total of 24 topics were labeled based on keywords with lambda values of 0.6 and 0.2, along with examples of sentences from the OCRs. Researchers determined the topic label assignments to ensure the proposed labels reflected the underlying review content. The assigned sentences were reviewed before the final topic label was determined. After finalizing the topic labels, upper-level themes emerged based on the conceptual framework shown in an earlier section (Figure 1).

Themes were derived from sentence-level topic assignments (see Table 1); n counts reviews at the sentence-level, adding to the theme-level. Ten upper-level themes emerged from the 24 topics, supporting the FEA and ECT frameworks: (1) expectations ($n = 988$), (2) functional ($n = 6655$), (3) expressive ($n = 465$), (4) aesthetic ($n = 1218$), and (5) extrinsic aspects of perceived product performance ($n = 1829$), (6) perceived service performance ($n = 65$), (7) confirmation/disconfirmation ($n = 2890$), (8) satisfaction/dissatisfaction ($n = 6745$), (9) outcomes of satisfaction/dissatisfaction ($n = 1632$), and (10) descriptive information for others ($n = 441$).

Most expectation and perceived product performance themes were classified as mutually exclusive based on their dominant meaning. However, selected overlap was retained when the same topic clearly represented more than one theoretically relevant dimension. For example, Topic 15 was classified under both extrinsic product performance and confirmation/disconfirmation because the review content combined price–quality evaluation with expectation-related language. In addition, post-purchase expressions, including satisfaction/dissatisfaction and outcomes such as recommendation, complaint, return, and refund, were coded separately and could overlap with other themes when they appeared in the same review content. To reduce overlap between the text-derived satisfaction/dissatisfaction expressions and the explanatory variables used in the regression analyses, star rating was used as an independently observed proxy for overall evaluative satisfaction rather than the text-derived satisfaction/dissatisfaction theme.

Under the expectation theme, topic 19 captured the initial excitement and enthusiasm consumers felt when purchasing leggings, often influenced by friends' recommendations and online reviews. This topic reflected a high level of anticipation and decision-making driven by external opinions.

For the perceived product performance (PPP) theme, there were four sub-dimensions supporting the FEA model [11] and the evaluative criteria of apparel products [22,23], as suggested in the conceptual framework (Figure 1): Functional, Expressive, Aesthetic, and Extrinsic. The functional dimension included quality concerns, such as pilling (topic 6) in specific areas (i.e., crotch and thighs) (topic 22), durability concerns (topic 8), fit in terms of length (topic 7), tightness (topic 10), sizing during body size changes (topic 16), comfort and fabric feel (topic 11) in certain conditions (topic 13), comfort during workouts (topic 10), the functionality of pockets (topic 12), overall fabric quality (topic 24), and care instructions (topic 18). In the expressive dimension, consumers frequently highlighted the versatility of leggings, suitable for both athletic activities and casual wear (topic 2). For the aesthetic dimension, consumers commented on the fabric's appearance, including sweat wicking and sweat marks during exercise (topic 1), as well as sweat, squat, and stain resistance (topic 23). They also mentioned issues with lint attraction and visibility of underwear (topic 3), and their preferences for a variety of colors and patterns (topic 4). Lastly, in the extrinsic dimension, customers evaluated the leggings' quality in relation to their price, often expressing disappointment when the product did not meet their expectations (topic 15).

Table 1. Themes, Topics, and Selected Keywords Emerged in OCRs regarding Leggings.

Themes	#	*n*	Topic Labels	Selected Keywords ($\lambda = 0.6$ & $\lambda = 0.2$)
Expectations	19	988	Pre-purchase excitement and purchase decisions influenced by reviews and others	**excited**, bought, **purchase,** super_excited, **heard,** decided, **purchase**, **friends,** disappointed [b], love [a]
Perceived product performance (PPP)—Functional	6	431	Quality concerns identified through other customers' reviews	reviews, **customer,** issue, pilling, read, returning [d], **hassle, listened, regretted [d], complain [d]**
	7	287	Fit—length	**length,** perfect**, ankle, shorter, hit,** inseam, fit, long, **petite,** crop
	8	695	Quality and durability concerns	**seam, hole, stitching**, noticed, seams, **coming_undone, thread**, loose, leg, small
	9	185	Comfort issues during workouts	**constantly_pull, worry, rolled, tugging,** adjusting, roll, floor, waistband, stay, **breathe**
	10	569	Comfort and fit issues—tightness and compression	**waist_band,** tighter, expected, thicker, uncomfortable, drawstring, high, **roll, slip, stomach**
	11	2025	Comfort and fabric's feel (softness)	**love [a], comfortable, feel,** fit, like, fabric, material, **amazing,** super, **buttery_soft**
	12	15	Fit, body shape, and pocket functionality	**Max,** iphone, **dinner, sightseeing,** hourglass, fair, **shifted, complain [d],** curvey, belt
	13	18	Fabric's feel and environmental performance	**West, heartedly, sense, east, fence,** comforting, weighted, wind, sandpaper, smoothcover
	16	381	Fit and sizing during body size changes (e.g., pregnancy)	Size, wear, fit, big, pregnancy, **lost, grow, relaxed, mama, burpees**
	18	415	Care	**Dry,** wash, **cold, washing_instructions, air, hang_dry,** care, **delicate_cycle, hand,** follow
	22	1612	Pilling concerns on thighs and crotch areas	**started_pilling, thighs, wears,** times, **crotch,** washed, legs, look, like, **crazy**
	24	12	Fabric quality	**Bingo,** pun, disappointed [b], claw, anchorage, catching, knox, rosa, roseville, shout,
PPP—Expressive	2	465	Usage situations (e.g., active and everyday wear)	Yoga, running, wear, **lounging,** errands, workout, **intensity,** hot, gym, **barre, fitness, weights**
PPP—Aesthetic	1	193	Appearance of fabric (e.g., sweat wicking, sweat marks) when exercising	Sweat, marks, catch, patterns, wicking, **integrity, minimum, absorbing, knocked, fluffing**
	3	85	Appearance and material concerns	Lint, see, underwear, **lighting,** fuzz, attract, bend, **nude, fur, roller, printed**
	4	820	Color and pattern	Color, black, love [a], different, **blue, green, beautiful, pink, double,** camo
	23	120	Sweat management and stain resistance	Sweat, visible, proof, stain, squat, wicking, **desk,** rub, **car, intensive, delivered**

Table 1. *Cont.*

Themes	#	*n*	Topic Labels	Selected Keywords (**λ = 0.6 & λ = 0.2**)
PPP—Extrinsic	15	1829	Perceived value of the product (price-quality)	**Expected, quality, price, better,** disappointed [b], **better_quality,** expect, longer, paid, price_point
Perceived service performance	5	66	Purchase experience and customer service	**Final,** sale, return [d], bra, **shops, ship, exchange** [d]**,** tags, service, **matched**
Confirmation/ disconfirmation	14	390	Pre-purchase excitement and post-purchase emotions (love, bummed) & actions (return, refund) at post-purchase	**Super_excited,** gift, **mail, received,** love [a], pockets, favorite, bummed [b], **collective,** return [d], refund [d]
	15	1829	Pre-purchase quality expectations in relation to the price they paid and post-purchase disappointment	**Expected, quality, price, better,** disappoint [b], **better_quality,** expect, longer, paid, price_point
	17	288	Pre-purchase hesitation and post-purchase disappointment based on the value-price at post-purchase	**Money, worth,** buying, **waste,** buy, afraid, regret [d], disappoint [b], **penny,** recommend [c], **earned**
	20	383	Pre-purchase expectation and exceeding expectations — positive experiences at post-purchase	**months_ago,** bought, **fell_love, weeks_ago,** love [a], expectations, **gift, exceeded, surpassed,** excited
Satisfaction/ dissatisfaction	Satisfaction: Topics containing a keyword of "love"—Topics 4 [a], 11 [a], 14 [a], 19 [a], 20 [a] Dissatisfaction: Topics with a keyword of "disappointment"—Topics 15 [b], 17 [b], 19 [b], 24 [b]			
Outcomes of satisfaction/dissatisfaction	Positive outcome: Public positive eWOM expression: Topics with a keyword of "recommend"—Topics 17 [c], 21 [c] Negative outcomes: Public negative eWOM expression: Topics with a keyword of "complaint"—Topics 6 [d], 12 [d]; Transactional responses: Topics with a keyword of "exchange"—Topics 5 [d], 21 [d]; Topics with a keyword of "return" or "refund"—Topics 5 [d], 6 [d], 14 [d] Affective responses: Topics with a keyword of "regret"—Topics 6 [d], 17 [d]			
Descriptive information	21	441	Fit and sizing recommendations for others	**Size, lbs,** ordered, **sizing, reference,** usual, **pounds,** smaller, recommend [c], big, exchange [b]
Total		12,723		
Not modelled			302	
All sentences			13,025	

Note. Bold types are keywords at λ = 0.2 in the pyLDAvis tool. [a,b] Topics under satisfaction/dissatisfaction where keywords included "love" and/or "disappointment". [c] Topics regarding positive outcomes of satisfaction where keywords included "recommend". [d] Topics regarding negative outcomes of dissatisfaction where keywords included "complaint", "exchange" "regret", "return" and/or "refund".

In the theme of perceived service performance, topic 5 covered the purchasing process and customer service experiences, including returns and exchanges. Many consumers mentioned that they had purchased leggings as final sales, which made them unable to return or receive a refund, leading to disappointment.

Under the dis/confirmation theme, consumers often mentioned their expectations and emotional reactions within the same sentence, describing the gaps between their expectations, excitement, and hesitations at the pre-purchase stage and their perceived product performance at the post-purchase stage, often expressed through emotions (e.g., love,

bummed, disappointed), which were indicators of satisfaction or dissatisfaction. When expectations were exceeded, positive disconfirmation was expressed through terms such as "love" or "thrilled," signaling satisfaction and delight (topic 20). Conversely, when expectations were not met, negative disconfirmation appeared in statements combining unmet expectations with negative emotions, such as being "bummed" or "disappointed," particularly when high prices or value–quality gaps were involved, such as the high price point (topic 15), a value-price discrepancy (topic 17), or stemming from their initial excitement (topic 14).

Sentences for the two themes—dis/satisfaction and outcomes of dis/satisfaction—were double-counted, as shown in Table 1. As mentioned earlier, for satisfaction and dissatisfaction in OCRs, consumers often used the word "love" to express a high level of satisfaction with a product. However, when they reported feeling "disappointed" at the post-purchase stage, it typically reflected unmet high expectations, adding an emotional layer to their evaluations. These expressions were frequently tied to antecedents (e.g., expectations, perceived product and service performance, dis/confirmation) and consequences (e.g., outcomes of dis/satisfaction). Because satisfaction-related expressions and pre- and post-purchase responses can occur within the same sentence, these categories were allowed to overlap. For example, a reviewer could express dissatisfaction while also describing a return, complaint, or regret. Retaining both classifications captures the evaluative response and the behavioral or communicative response expressed in the same review content.

The outcomes of dis/satisfaction were either positive or negative. As a positive outcome, when consumers were highly satisfied with a product, they were more likely to recommend it to others in their OCRs. Recommendations are valuable indicators of a product's success, as they suggest that the product not only met but exceeded consumer expectations, prompting them to endorse it to friends, family, and potential new customers.

Negative outcomes were expressed with terms like "complaint", "exchange", "regret", "return" or "refund" in OCRs. Dissatisfied consumers frequently voiced their grievances through complaints (topics 6 and 12), often detailing specific issues such as quality defects, discomfort, or poor customer service. Complaints can harm the brand's reputation and deter potential customers from making a purchase. They highlight areas where the product or service failed to meet consumer expectations and can serve as critical feedback for improving product quality and customer service. Requests for exchanges indicated that the product did not meet consumers' expectations or requirements in terms of size, fit, or quality (topics 5 and 21). While exchanges can still retain the customer within the brand, they reflect initial dissatisfaction that requires rectification. Expressions of regret signaled a profound level of dissatisfaction, where consumers wished they had not made the purchase (topics 6 and 17). Regret often leads to decreased brand loyalty and reluctance to purchase from the brand in the future. Returns and refunds were clear indicators of dissatisfaction (topics 5, 6, and 14) in the OCRs, showing that consumers found the product unacceptable and chose to reverse their purchase.

Through sentiment analysis, a sentiment score was assigned to each sentence based on its associated topic. Topic-/theme-specific sentiment values were coded as missing when the topic/theme was not mentioned in a review. At the review level, sentiment scores for themes were calculated by averaging the scores when multiple topics within the same theme were present in a review. As shown in Table 2, the average sentiment score for all 2300 OCRs was 0.19, ranging from 1-star ratings (M = 0.08) to 5-star ratings (M = 0.34). Ten themes with their mean sentiment scores are shown in Table 2: (1) expectations (M = 0.34), (2) functional (M = 0.18), (3) expressive (M = 0.25), (4) aesthetic (M = 0.23), and (5) extrinsic aspects (M = 0.06) of PPP, (6) perceived service performance (M = −0.02), (7) confirmation/disconfirmation (M = 0.14), (8) satisfaction/dissatisfaction

(M = 0.28), (9) outcomes of satisfaction/dissatisfaction (M = 0.28), and (10) descriptive information for others (M = 0.12).

Table 2. Descriptive Statistics for Sentiment Scores across Themes and Sub-themes.

Themes		*f*	%	*M*	*SD*	*Var*	Min	Max
Expectations		774	33.7	0.34	0.37	0.14	−0.78	0.95
Perceived product performance (PPP)	Functional	2176	94.6	0.18	0.29	0.08	−0.78	0.93
	Expressive	395	17.2	0.25	0.37	0.13	−0.80	0.92
	Aesthetic	790	34.3	0.23	0.34	0.12	−0.78	0.96
	Extrinsic	1498	65.1	0.06	0.37	0.13	−0.80	0.94
Perceived service performance		61	2.7	−0.02	0.40	0.16	−0.80	0.90
Confirmation/ Disconfirmation		1730	75.2	0.14	0.36	0.13	−0.82	0.94
Satisfaction/ Dissatisfaction		2242	97.5	0.28	0.27	0.08	−0.71	0.95
Satisfaction: love		2011	87.4	0.41	0.30	0.09	−0.66	0.98
Dissatisfaction: disappointment		1759	76.5	0.14	0.34	0.12	−0.86	0.94
Outcomes of satisfaction/ dissatisfaction		1228	53.4	0.22	0.36	0.13	−0.80	0.98
Positive outcome: recommend		557	24.2	0.22	0.39	0.16	−0.86	0.94
Negative outcomes: complaint, exchange, regret, return, refund		1046	45.5	0.18	0.37	0.14	−0.81	0.98
Descriptive information		351	15.3	0.12	0.27	0.07	−0.67	0.87
Star-ratings		2300	100.0	2.88	1.36	1.86	1	5
Mean sentiment scores	1-star	449	19.5	0.08	0.18	0.03	−0.41	0.65
	2-star	538	23.4	0.13	0.18	0.03	−0.59	0.85
	3-star	569	24.7	0.18	0.15	0.02	−0.27	0.93
	4-star	331	14.4	0.25	0.18	0.03	−0.48	0.92
	5-star	413	18.0	0.34	0.20	0.04	−0.40	0.94
	Total	2300	100.0	0.19	0.20	0.04	−0.59	0.94

f = the number of reviews containing at least one sentence assigned to the theme (a review is counted once per topic/theme regardless of how many of its sentences fall in it, % = the percentage of reviews (f/2300), *M* = mean, *SD* = standard deviation, *Var* = Variance, Min = Minimum, Max = Maximum.

4.3. The Effects of Evaluative Features on Satisfaction

To examine the effects of evaluative features on satisfaction, this study used star ratings as a measure of satisfaction for three reasons: (1) it is widely employed as a measure of customer satisfaction in product and service review contexts [47]; (2) it is often aligned with detailed written feedback [47]; and (3) the dis/satisfaction theme contained topics that were also assigned to other themes (e.g., PPP-functional), which disqualified it from being used as a dependent variable.

Because topic-/theme-specific sentiment scores were available only when the corresponding topic/theme was present, the effective sample size varied across analyses. For example, functional product performance was represented in 2176 reviews, expressive performance in 395 reviews, and service performance in 61 reviews. The sample

size for each correlation, regression, and mediation model is therefore reported with the corresponding results.

Regression analyses showed that perceived product performance in functional ($\beta = 0.39$, $p < 0.001$, $n = 2176$), expressive ($\beta = 0.16$, $p = 0.001$, $n = 395$), and aesthetic ($\beta = 0.17$, $p < 0.001$, $n = 790$) and extrinsic ($\beta = 0.17$, $p < 0.001$, $n = 1498$) aspects, and perceived service performance ($\beta = 0.42$, $p < 0.001$, f = 61) were positively associated with satisfaction, supporting H1. Consistent with H2, confirmation ($\beta = 0.19$, $p < 0.001$, $n = 1727$) was also positively associated with satisfaction. Furthermore, satisfaction was also positively associated with outcomes of satisfaction ($\beta = 0.08$, $p < 0.01$, $n = 1228$). There was a marginally significant effect of expectation ($\beta = 0.07$, $p = 0.07$, $n = 774$) on satisfaction.

These associations were generally consistent with the proposed conceptual framework, although the cross-sectional review data do not establish temporal or causal relationships. Separate regression analyses were conducted for each evaluative theme. Theme-specific sentiment was coded as missing when a theme was not mentioned; therefore, the number of observations available varied across themes. Separate models allowed each relationship to be examined using the reviews containing information on the relevant theme without restricting the analysis to the much smaller subset of reviews containing all themes. Accordingly, coefficients from the separate models were interpreted as theme-specific associations and were not compared as measures of relative importance.

Ad hoc mediation analyses were conducted using PROCESS Model 6 to explore whether product- and service-performance evaluations and confirmation mediated the relationship between expectations and star ratings. Complete results for all models are provided in Supplementary Table S2.

In the functional-performance model (N = 566), the direct effect of expectations on star ratings was positive and statistically significant (B = 0.26, SE = 0.13, $p < 0.05$, 95% CI [0.001, 0.52]). The total indirect effect was not significant; however, the specific indirect effect through confirmation was significant (B = −0.06, BootSE = 0.03, 95% bootstrap CI [−0.12, −0.001]). The indirect effect through functional performance alone and the sequential pathway through functional performance and confirmation were not significant.

No significant indirect effects were observed in the expressive-, aesthetic-, extrinsic-, or service-performance models. The service model was based on a very small analytic sample (N = 16) and should therefore be interpreted with caution. Overall, the mediation findings provide limited evidence for indirect effects, with confirmation emerging as the only significant mediator in the functional-performance model.

5. Discussions and Implications

This study provides a comprehensive understanding of consumer satisfaction in the activewear leggings context by integrating topic modeling, sentiment analysis, and regression analysis of quantified online consumer review (OCR) data. The findings demonstrate that consumers' post-purchase evaluations reflect multiple dimensions of product and service performance rather than a single overall assessment. Functional product performance was the most frequently discussed dimension and was positively associated with star ratings, an indicator of consumer satisfaction. Consumers particularly valued positive functional attributes such as material softness and comfort, whereas quality and durability problems, including pilling, unraveling stitching, holes, and discomfort during exercise, were associated with less favorable evaluations. These findings illustrate how consumers use OCRs to articulate detailed evaluations of activewear leggings based on their actual product experiences.

5.1. Theoretical Implications

The findings extend understanding of multidimensional apparel evaluation in digital commerce. The product-evaluation topics were interpreted in relation to the functional, expressive, aesthetic, and extrinsic dimensions specified in the a priori conceptual framework. Partially, the findings demonstrate how the theoretically derived FEA dimensions are represented in naturally occurring OCRs in the context of online activewear leggings consumption. Consumers evaluated leggings products not only according to functional characteristics but also through expressive, aesthetic, and extrinsic considerations, demonstrating the multifaceted nature of post-purchase product evaluation. Although functional performance was most frequently discussed, expressive, aesthetic, and extrinsic dimensions also contributed positively to satisfaction. This suggests that satisfaction with activewear emerges from an integrated evaluation of utilitarian performance, self-expression, appearance, and perceived value. The findings therefore extend the application of the FEA framework beyond product design and evaluation by demonstrating its relevance to naturally occurring consumer-generated content in digital commerce.

Importantly, the frequency of discussion did not necessarily correspond to the strength of an attribute's relationship with satisfaction. Perceived service performance was the least frequently mentioned theme in the OCRs, but service-related sentiment was positively associated with star ratings among reviews in which service was discussed. Because the theme-specific models were estimated using different analytic samples, this coefficient should not be interpreted as evidence that service was more important than other product-performance dimensions.

Nevertheless, the finding suggests that relatively infrequent aspects of the consumption experience can still be meaningfully associated with consumers' overall evaluations when they are mentioned. In digital commerce, customer service, returns, exchanges, and other service encounters may therefore represent important components of the review experience even when they occur less frequently in the corpus. More broadly, the results illustrate the value of considering both topic prevalence and evaluative sentiment, because frequency alone does not indicate how strongly a particular theme is associated with overall ratings.

The results are also consistent with expectation-confirmation theory (ECT) [12,13]. Consumers expressed prior expectations, evaluated subsequent product and service experiences, and described whether those experiences confirmed or disconfirmed what they had anticipated. More positive confirmation was also associated with higher star ratings, supporting the proposed hypothesis 1 that consumers evaluate post-purchase experiences partly in relation to prior expectations rather than solely on the basis of objective product attributes. OCRs thus provide a particularly useful context for examining expectation-confirmation processes because they contain retrospective accounts of both anticipated and experienced product performance.

The ad hoc mediation analyses further illustrate the role of confirmation in this process, although the evidence for mediation was limited. In the functional-performance model, the specific indirect relationship between expectations and star ratings through confirmation was significant, whereas the indirect effects through functional performance alone and through the sequential functional-performance and confirmation pathway were not significant. No significant indirect effects were observed in the expressive, aesthetic, extrinsic, or service models. These findings suggest that confirmation may contribute to the relationship between expectations and overall evaluation in some review contexts, while also indicating that the broader sequential mechanisms proposed by ECT were not consistently observed in the present data.

This study also contributes to the eWOM literature by showing that OCRs convey not only an overall evaluation but also the expectation-confirmation process that underlies it.

Reviewers frequently described what they had expected, how the leggings performed, and whether they met those expectations. Because these evaluations are publicly available, one consumer's account of confirmation or disconfirmation can provide expectation-relevant information to prospective consumers. OCRs therefore connect post-purchase evaluation with the broader information environment of digital commerce, in which consumers' experiences become part of the information available to prospective buyers.

Finally, the findings demonstrate the value of combining topic modeling, sentiment analysis, and regression analysis when examining OCRs. Topic modeling identifies which aspects of the consumer experience are discussed most frequently, while sentiment analysis captures the evaluative direction of those discussions. Attribute-level analysis of review text therefore provides diagnostic information that star ratings alone cannot reveal. Regression analysis further allows the valence of evaluations for each identified feature to be examined in relation to consumers' overall evaluations, as reflected in star ratings. This suggests that detailed, experience-based reviews may be particularly valuable to readers in this category [29].

5.2. Managerial Implications

From a managerial perspective, these results suggest that OCRs can be used to manage expectations as well as to monitor satisfaction for retailers and brands. Because dissatisfaction often reflects a gap between expected and actual performance, retailers can reduce negative OCRs by setting more accurate expectations before purchase. Possible measures include detailed size guidance, information on fabric opacity and stretch, and photos or videos of the garments in use. Retailers could also prompt reviewers to rate specific attributes such as fit, support, and comfort, which would make OCRs more informative for prospective buyers and easier for firms to analyze. Finally, rather than relying on average star ratings, managers should monitor attribute-level sentiment in reviews to find and fix the product features that most often lead to negative OCRs.

The findings indicate that activewear retailers should prioritize both product performance and service performance when managing consumer satisfaction and online review experiences. Quality, durability, fit, and comfort remain fundamental, particularly for products such as leggings that consumers expect to perform across multiple activities. The positive evaluation of versatility further suggests that retailers can emphasize multipurpose use in product communication while ensuring that such claims accurately reflect actual product performance. Expressive and aesthetic attributes should likewise remain important considerations in product development because consumers evaluate activewear in terms of appearance and self-expression as well as functionality.

The strong relationship between service performance and satisfaction further indicates that investments in customer service, transparent return and exchange procedures, and effective resolution of product problems may have substantial effects on consumers' overall experiences. Because dissatisfied consumers can communicate these experiences publicly through OCRs, service recovery has implications not only for the focal customer but also for the broader digital audience exposed to the review. Finally, retailers should manage pre-purchase expectations carefully. Accurate product descriptions, realistic performance claims, and consistent product quality can reduce negative disconfirmation and increase the likelihood that consumers' actual experiences correspond with the expectations established during the online shopping process.

5.3. Future Studies and Limitations

The expectation-based screening identified a focused subset of reviews in which expectations were explicitly articulated. Because such language may be more common in reviews

involving stronger evaluative reactions, the retained sample is not intended to represent the full distribution of product reviews. The findings should be interpreted within the context of the expectation-related reviews analyzed in this study and should not be generalized to all activewear consumers. In addition, demographic characteristics such as reviewer gender could not be reliably determined from the available review data; therefore, no gender-based assumptions are made regarding the reviewers.

Because topic-/theme-specific sentiment was analyzed only when a topic/theme was present, the study focuses on the valence of expressed evaluations rather than on the likelihood that a particular topic/theme is mentioned. Topic/theme presence and topic/theme valence may capture different aspects of consumer evaluation, and future research could examine these processes separately.

Using sentiment analysis on dominant topics identified through topic modeling provides a methodological framework for quantifying and interpreting OCRs, offering insights into consumer emotions and attitudes at scale. The positive correlations between star ratings and sentiment scores further validate this approach. This supports the application of sentiment analysis in consumer research to gain real-time insights into consumer perceptions and behavior. Future studies can refine these techniques to improve accuracy and applicability across different contexts.

Because OCRs are retrospective narratives, theoretically distinct constructs may be expressed within the same sentence or topic, creating overlap among some theme-level measures. Although this overlap reflects how consumers naturally describe their experiences, it limits the extent to which the text-derived categories can be interpreted as fully independent psychological constructs.

This study supports the idea that managing and aligning customer expectations with the actual product experience, leading to positive confirmation, is crucial. Future research can explore effective marketing strategies to manage consumer expectations and reduce negative confirmation.

6. Conclusions

This study demonstrates that consumer satisfaction in online activewear retailing emerges from the interplay of multidimensional product performance, service performance, expectations, confirmation/disconfirmation, and affective responses. By integrating topic modeling, sentiment analysis, and regression analysis of OCRs, the findings show that the aspects consumers discuss most frequently are not necessarily those most strongly associated with overall evaluations. Functional product performance dominated consumer discussions, whereas service performance was mentioned far less frequently but was still positively associated with star ratings when discussed. These findings highlight the importance of considering both topic prevalence and evaluative sentiment when interpreting consumer review content.

The findings extend the FEA model and ECT to consumer-generated review data and illustrate how affective responses connect post-purchase evaluations with subsequent behaviors such as recommendation, complaint, and return intentions. More broadly, the study demonstrates the value of OCRs as naturally occurring expressions of consumers' product, service, and brand experiences in digital commerce. Understanding these multidimensional evaluations can help online retailers better align product and service performance with consumer expectations and foster more favorable post-purchase experiences and consumer-generated brand discourse. While this study provides valuable insights into how expectations and multidimensional performance shape user-generated content, these findings are specifically derived from brand-hosted reviews of selected best-selling leggings. Consequently, our conclusions most directly reflect this specific activewear context. We

propose that the broader implications for generalized digital commerce and other retail categories serve as propositions for future research, requiring testing across diverse brands, platforms, and product types to confirm their wider applicability.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jtaer21110434/s1, Supplementary Table S1. Star rating, review length, and product category from the full dataset to the analytic sample; Supplementary Table S2. Summary of PROCESS Model 6 Path and Mediation Results.

Author Contributions: Conceptualization, E.S. and C.M.; methodology, C.M.; software, C.M. and B.M.; validation, E.S., C.M. and B.M.; formal analysis, E.S., C.M. and B.M.; investigation, E.S., C.M.; resources, C.M.; data curation, C.M. and B.M.; writing—original draft preparation, E.S. and C.M.; writing—review and editing, E.S., C.M. and B.M.; visualization, E.S. and C.M.; supervision, E.S.; project administration, E.S.; funding acquisition, E.S. and C.M. All authors have read and agreed to the published version of the manuscript.

Funding: This research was funded by University Libraries, Virginia Tech, and the APC was funded by Virginia Tech.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: The data that support the findings of this study are available from the corresponding author upon reasonable request.

Conflicts of Interest: The authors declare no conflicts of interest.

Abbreviations

The following abbreviations are used in this manuscript:

DV	Dependent variable
ECT	Expectation-confirmation theory
FEA model	Functional, Expressive, and Aesthetic Consumer Needs model
GSDMM	Gibbs Sampling Dirichlet Multinomial Mixture
IV	Independent variable
OCR	Online consumer reviews
VADER	Valence Aware Dictionary for Sentiment Reasoning

References

1. Zhao, L.; Min, C. The rise of fashion informatics: A case of data-mining-based social network analysis in fashion. *Cloth. Text. Res. J.* **2019**, *37*, 87–102. [CrossRef]
2. Yan, Z.; Zhang, D. EXPRS: An extended PageRank method for product feature extraction from online consumer reviews. *Inf. Manag.* **2015**, *52*, 850–858. [CrossRef]
3. Smith, P. Size of the Global Activewear Market 2021 to 2028. Statista 2024. Available online: https://www.statista.com/statistics/613169/size-of-the-global-sportswear-market/ (accessed on 21 August 2025).
4. Zhou, X.; Hanlon, C.; Robertson, J.; Spaaij, R.; Westerbeek, H.; Hossack, A.; Funk, D.C. Dress for fit: An exploration of female activewear consumption. *Sport Manag. Rev.* **2018**, *21*, 403–415. [CrossRef]
5. Liu, Y.; Wan, Y.; Shen, X.; Ye, Z.; Wen, J. Product customer satisfaction measurement based on multiple online consumer review features. *Information* **2021**, *12*, 234. [CrossRef]
6. Quaye, E.S.; Taoana, C.; Abratt, R.; Anabila, P. Customer advocacy and brand loyalty: The mediating roles of brand relationship quality and trust. *J. Brand Manag.* **2022**, *29*, 363–382. [CrossRef]
7. Sayfuddin, A.T.M.; Chen, Y. The signaling and reputational effects of customer ratings on hotel revenues: Evidence from TripAdvisor. *Int. J. Hosp. Manag.* **2021**, *99*, 103065. [CrossRef]
8. Lang, C.; Xia, S.; Liu, C. Style and fit customization: A web content mining approach to evaluate online mass customization experiences. *J. Fash. Mark. Manag.* **2021**, *25*, 224–241. [CrossRef]
9. Li, M.; Zhao, L.; Srinivas, S. It's about inclusion! Mining online reviews to understand the needs of adaptive clothing customers. *Int. J. Consum. Stud.* **2023**, *47*, 1157–1172. [CrossRef]

10. Shin, E.; Kincade, D.; Han, J. Exploring online consumer reviews of customized apparel products. *J. Fash. Mark. Manag.* **2024**, *28*, 139–160. [CrossRef]
11. Lamb, J.M.; Kallal, M.J. A conceptual framework for apparel design. *Cloth. Text. Res. J.* **1992**, *10*, 42–47. [CrossRef]
12. Oliver, R.L. Effect of expectation and disconfirmation on postexposure product evaluations: An alternative interpretation. *J. Appl. Psychol.* **1977**, *62*, 480–486. [CrossRef]
13. Oliver, R.L. A cognitive model of the antecedents and consequences of satisfaction decisions. *J. Mark. Res.* **1980**, *17*, 460–469. [CrossRef]
14. Oliver, R.L. *Satisfaction: A Behavioral Perspective on the Consumer*, 2nd ed.; Routledge: New York, NY, USA, 2014.
15. Oh, S.; Ji, H.; Kim, J.; Park, E.; Pobil, A. Deep learning model based on expectation-confirmation theory to predict customer satisfaction in hospitality service. *Inf. Technol. Tour.* **2022**, *24*, 109–126. [CrossRef]
16. Park, E. Motivations for customer revisit behavior in online review comments: Analyzing the role of user experience using big data approaches. *J. Retail. Consum. Serv.* **2019**, *51*, 14–18. [CrossRef]
17. Jang, J.; Yi, M.Y. Modeling user satisfaction from the extraction of user experience elements in online product reviews. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems, Denver, CO, USA, 6–11 May 2017*; Association for Computing Machinery: New York, NY, USA, 2017; pp. 1718–1725.
18. Kumar, R.; Mukherjee, S.; Choudhary, D. Uncovering the roots of customer dissatisfaction via Amazon reviews: A hybrid ensemble-deep learning approach for E-commerce quality management. *Ann. Oper. Res.* **2025**, 1–30. [CrossRef]
19. Greenleaf, C.; Hauff, C.; Klos, L.; Serafin, G. "Fat people exercise too!": Perceptions and realities of shopping for women's plus-size exercise apparel. *Cloth. Text. Res. J.* **2020**, *38*, 75–89. [CrossRef]
20. Christel, D.A.; O'Donnell, N.H. Assessment of women's plus-size swimwear for industry applications. *Fash. Pract.* **2016**, *8*, 257–278. [CrossRef]
21. Park, H.; Kim, S.; Choi, Y.; An, H. Application of the FEA consumer needs model in e-commerce: A conversational recommendation system for fashion products. *Fash. Text.* **2025**, *12*, 24. [CrossRef]
22. Eckman, M.; Damhorst, M.L.; Kadolph, S.J. Toward a model of the in-store purchase decision process: Consumer use of criteria for evaluating women's apparel. *Cloth. Text. Res. J.* **1990**, *8*, 13–22. [CrossRef]
23. Abraham-Murali, L.; Littrell, M.A. Consumers' conceptualization of apparel attributes. *Cloth. Text. Res. J.* **1995**, *13*, 65–74. [CrossRef]
24. Ghalachyan, A.; Karpova, E. Development of Apparel Product Evaluation (APE) framework: A systematic classification of evaluative criteria. *Int. J. Fash. Des. Technol. Educ.* **2021**, *14*, 243–252. [CrossRef]
25. Hennig-Thurau, T.; Gwinner, K.P.; Walsh, G.; Gremler, D.D. Electronic word-of-mouth via consumer-opinion platforms: What motivates consumers to articulate themselves on the Internet? *J. Interact. Mark.* **2004**, *18*, 38–52. [CrossRef]
26. Chevalier, J.A.; Mayzlin, D. The effect of word of mouth on sales: Online book reviews. *J. Mark. Res.* **2006**, *43*, 345–354. [CrossRef]
27. Mudambi, S.M.; Schuff, D. What makes a helpful online review? A study of customer reviews on Amazon.com. *MIS Q.* **2010**, *34*, 185–200. [CrossRef]
28. Hu, N.; Zhang, J.; Pavlou, P.A. Overcoming the J-shaped distribution of product reviews. *Commun. ACM* **2009**, *52*, 144–147. [CrossRef]
29. Berger, J.; Humphreys, A.; Ludwig, S.; Moe, W.W.; Netzer, O.; Schweidel, D.A. Uniting the tribes: Using text for marketing insight. *J. Mark.* **2020**, *84*, 1–25. [CrossRef]
30. Hartmann, J.; Huppertz, J.; Schamp, C.; Heitmann, M. Comparing automated text classification methods. *Int. J. Res. Mark.* **2019**, *36*, 20–38. [CrossRef]
31. Kim, S.; Lee, S.; McCulloch, R. A topic-based segmentation model for identifying segment-level drivers of star ratings from unstructured text reviews. *J. Mark. Res.* **2024**, *61*, 1132–1151. [CrossRef]
32. Praveen, S.V.; Gajjar, P.; Ray, R.K.; Dutt, A. Crafting clarity: Leveraging large language models to decode consumer reviews. *J. Retail. Consum. Serv.* **2024**, *81*, 103975. [CrossRef]
33. Winter, C.G.H.; Zacharias, N.A.; Hartwig, M. Using LLMs as sentiment analyzers to predict review helpfulness: First insights to open the black box. *Mark. Lett.* **2026**, *37*, 30. [CrossRef]
34. Gu, W.K.M.; Spann, M. The impact of generative artificial intelligence on consumer reviews: Insights from a quasi-experiment. *Int. J. Res. Mark.* **2026**, *in press*. [CrossRef]
35. McKinney, E.; Shin, E. Exploring criteria consumers use in evaluating their online formal wear rental experience: A content analysis of online reviews. *Cloth. Text. Res. J.* **2016**, *34*, 272–286. [CrossRef]
36. Sharma, H. Using topic modeling for extracting customers' expectations: A case of women apparel. *Bus. Perspect. Res.* **2025**, *13*, 454–466. [CrossRef]
37. Tighe, D. Net Revenue of Lululemon Worldwide from the Fiscal Year of 2008 to 2022. Statista 2024. Available online: https://www.statista.com/statistics/291238/net-revenue-of-lululemon-worldwide/ (accessed on 21 August 2025).

38. Ridhwan, K.; Hargreaves, C.A. Leveraging Twitter data to understand public sentiment for the COVID-19 outbreak in Singapore. *Int. J. Inf. Manag. Data Insights* **2021**, *1*, 100021. [CrossRef]
39. Blei, D.M.; Ng, A.Y.; Jordan, M.I. Latent Dirichlet allocation. *J. Mach. Learn. Res.* **2003**, *3*, 993–1022.
40. Yin, J.; Wang, J. A Dirichlet multinomial mixture model-based approach for short text clustering. In Proceedings of the 20th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, New York, NY, USA, 24–27 August 2014; pp. 233–242.
41. Sievert, C.; Shirley, K. LDAvis: A method for visualizing and interpreting topics. In Proceedings of the Workshop on Interactive Language Learning, Visualization, and Interfaces, Baltimore, MD, USA, 27 June 2014; pp. 63–70.
42. Chuang, J.; Manning, C.D.; Heer, J. Termite: Visualization techniques for assessing textual topic models. In Proceedings of the International Working Conference on Advanced Visual Interfaces, Capri Island, Italy, 21–25 May 2012; pp. 74–77.
43. Grootendorst, M. BERTopic: Neural topic modeling with a class-based TF-IDF procedure. *arXiv* **2022**, arXiv:2203.05794.
44. Feng, Y.; Hou, B.; Klein, A.; O'Connor, K.; Chen, J.; Mondragón, A.; Yang, S.; Gonzalez-Hernandez, G.; Shen, L. Analyzing dementia caregivers' experiences on Twitter: A term-weighted topic modeling approach. *AMIA Annu. Symp. Proc.* **2024**, *2024*, 407–416. [PubMed]
45. Hutto, C.J.; Gilbert, E.E. VADER: A parsimonious rule-based model for sentiment analysis of social media text. In Proceedings of the 8th International AAAI Conference on Web and Social Media, Ann Arbor, MI, USA, 1–4 June 2014; pp. 216–225.
46. Hayes, A.F. *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-Based Approach*, 2nd ed.; Guilford Press: New York, NY, USA, 2017.
47. Yoon, Y.; Kim, A.; Kim, J.; Choi, J. The effects of eWOM characteristics on consumer ratings: Evidence from TripAdvisor.com. *Int. J. Advert.* **2019**, *38*, 684–703. [CrossRef]

Disclaimer/Publisher's Note: The statements, opinions and data contained in all publications are solely those of the individual author(s) and contributor(s) and not of MDPI and/or the editor(s). MDPI and/or the editor(s) disclaim responsibility for any injury to people or property resulting from any ideas, methods, instructions or products referred to in the content.