Journal of
Theoretical and Applied Electronic Commerce Research

MDPI

Article

Cultural Cues and Sustainability-Relevant Consumer Experience in Handmade E-Commerce: Evidence from Consumer eWOM

Fan Wu [1], Congrong Xiao [2], Junbo Li [3] and Dongkwon Seong [4,*]

[1] School of Packaging Design & Art, Hunan University of Technology, Zhuzhou 412007, China; wufan@stu.hut.edu.cn
[2] College of Art and Design, Guangdong Baiyun University, Guangzhou 510900, China; xiaocongrong@baiyunu.edu.cn
[3] School of Art & Design, Guangdong University of Technology, Guangzhou 510006, China; jebli@baiyunu.edu.cn
[4] Department of X-Cultural Studies, Kookmin University Graduate School, Kookmin University, Seoul 02707, Republic of Korea
* Correspondence: dongkwonseong@kookmin.ac.kr

Abstract

We examined seller-generated cultural cues and consumer reports of product use in 6269 sampled reviews of 2000 products from the Handmade Products subset of Amazon Reviews'23. Most sustainability-related mentions concerned durability or longevity (S1). In the category-adjusted review-level model, generic cultural cues (C1) were associated with lower odds of a sustainability mention than no explicit cultural cue (C0) (OR = 0.494, 95% CI [0.323, 0.756]). The association varied across review caps and hypothetical coding-error scenarios. The equal-product-weighted C1–C0 estimate was imprecise. Because the proportional-odds assumption was rejected, we emphasized observed rating distributions and threshold-specific comparisons. Five-star ratings accounted for 94.10% of positive S1 reviews, 4.95% of negative S1 reviews, and 82.45% of no-mention reviews. Compared with positive S1, negative and mixed S1 had lower odds at each estimable threshold (≥3, ≥4, and ≥5). Neither contrast was estimable at ≥2 because of separation. The rating associations may reflect general dissatisfaction because overall sentiment was not independently measured. Positive helpful-vote counts were overdispersed and their associations remain exploratory. The study is observational and does not establish causality or measure environmental performance.

Keywords: e-commerce; cultural cues; sustainability-relevant experience; consumer eWOM; helpful-vote accumulation; handmade products

Check for updates

Academic Editors: Ting Chi, Ahmed Hassan Abdou and Shafique Ur Rehman

Received: 1 September 2026
Revised: 11 November 2026
Accepted: 12 November 2026
Published: 17 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.

1. Introduction

Online marketplaces reduce the visibility of product attributes that consumers can inspect before purchase. Durability, material performance, repairability, and long-term use are often difficult to evaluate from a listing alone. Consumers therefore rely on seller-generated product information and consumer-generated eWOM to form expectations and revise them after purchase [1–3].

The analytical tension is therefore not whether culture is "sustainable" in the abstract. It is whether an observed cultural cue in a consumer-facing product listing is associated with the sustainability-relevant product-use experience that consumers report in reviews, without treating either as an objective environmental-performance measure.

This question connects cultural sustainability with sustainable product experience. Cultural sustainability concerns the maintenance, adaptation, and communication of cultural meanings, whereas design for sustainability emphasizes attributes such as durability, repairability, circularity, and material efficiency. These domains may interact in marketplace communication, but they are not interchangeable constructs [4–8].

From an electronic-commerce perspective, the study treats listing cues, review content, ratings, and observed accumulated helpful votes as distinct measurement layers. The layers are linked by separately estimated associations; the available snapshot does not establish that a listing state was temporally prior to a consumer experience or outcome.

The literature has examined cultural goods, authenticity, sustainability signals, review helpfulness, and online trust in partly separate streams. What remains less clear is whether product-listing cultural cues and consumer-reported sustainability-related review content co-occur in a real marketplace corpus, and whether such content is associated with ratings or observed accumulated helpful votes [1–3,9–11].

These gaps motivate four research questions concerning cue distributions, sustainability-related mentions, rating associations, and observed helpful-vote accumulation.

The study makes three contributions: it distinguishes nominal cultural-cue categories from review-based product-use reports; estimates valence-specific rating associations rather than relying on a pooled contrast; and separates helpful-vote incidence from counts conditional on a positive vote.

2. Theoretical Background and Research Questions

2.1. Cultural and Heritage Cues in Digital Marketplaces

Experimental studies suggest that handmade cues can increase product attractiveness through symbolic associations with love, although responses vary by consumer segment [12,13].

Craft can be presented as a cultural good, design object, or ordinary market product; authenticity is socially constructed rather than guaranteed by a label [14–18]. We therefore treat C0/C1/C2 as categories of observable listing language, not verified provenance or authenticity.

A named place or technique may convey a more specific identity claim than a generic symbol, but neither verifies production conditions or material performance. The taxonomy measures listing language; it does not rank cultural value or establish the product attribute [19–23].

2.2. Digital Marketplace Signaling and Information Asymmetry

Listings signal selected information about products whose attributes consumers cannot fully inspect before purchase. Cultural narratives can shape this information environment, but do not verify product attributes [19–23].

Reviews can report durability, repair, or reuse after purchase, but remain situated accounts shaped by expectations, use conditions, product category, and vocabulary. Our coding therefore requires explicit product-use consequences and does not infer sustainability from praise, sentiment, or ratings. Positive reviews can describe a short useful life, and negative reviews can describe repair or reuse [1–3,24–26].

2.3. Sustainable Product Experience in Consumer eWOM

Design-for-sustainability research distinguishes durability, repairability, modularity, circularity, and material efficiency from generic product quality or esthetic approval. In this study, sustainability-related content means a consumer-reported product-use experience relevant to durability, longevity, repair, reuse, or circularity; it is not a direct measure of

objective environmental performance. Durability/longevity (S1) is the dominant observed dimension, while circularity (S2) is sparse [9–11,27].

2.4. Cultural Sustainability and the Heritage–Sustainability Boundary

Cultural sustainability provides the broader conceptual setting for examining how cultural meanings are maintained, adapted, and communicated in economic systems. The present study keeps that cultural-communication domain analytically distinct from consumer-reported product-use experience and from objective environmental performance.

2.5. Tradition-Washing as a Boundary Concept

Tradition-washing is used only as a boundary concept. Prior experimental work motivates a distinction between cultural/traditional framing and perceived greenness, but the present corpus does not measure perceived greenness, authenticity, or a causal perceptual mechanism [28].

2.6. Research Questions

RQ1. How are C0, C1, and C2 cultural cues distributed across handmade-product listings and reviews?

RQ2. Are observed product-listing cultural cues associated with the presence or proportion of sustainability-related consumer reviews?

RQ3. Are sustainability-related review experiences associated with ratings, and does the association vary by experience valence?

RQ4. Are cultural cues and sustainability-related review experiences associated with the observed accumulation of review helpful votes?

The relationships among listing cultural cues, consumer review records, and observed helpful-vote accumulation are illustrated in Figure 1.

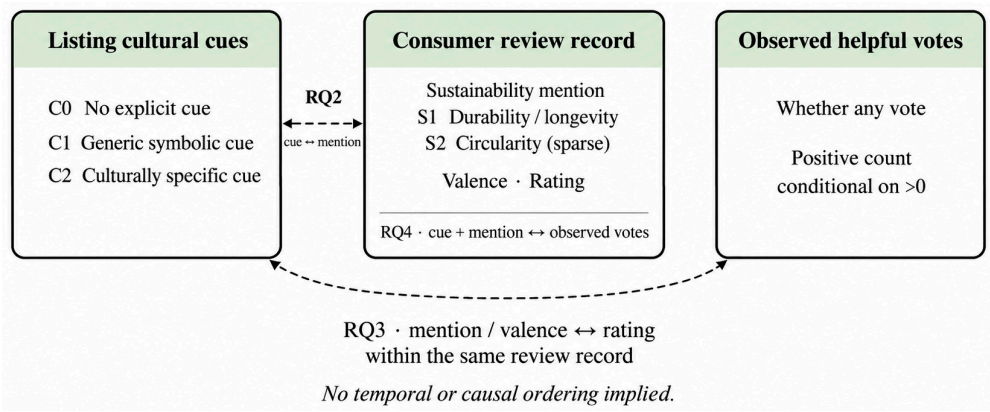


Figure 1. Observed e-commerce association framework. Listing cultural cues, the consumer review record, and observed helpful-vote accumulation are distinct measurement layers. RQ2, RQ3, and RQ4 identify separately estimated associations; the rating and review-text measures belong to the same review record. No temporal or causal ordering is implied.

3. Materials and Methods

3.1. Data Source and Research Context

We used the Handmade Products subset of the Amazon Reviews'23 corpus, a public corpus of Amazon product metadata and consumer reviews. The released metadata provides a product/review snapshot without historical listing versions or listing update dates. The analysis therefore links observed product metadata to historical review text; it cannot

verify which listing version a reviewer saw. The contemporaneous download record does not preserve a verifiable Hugging Face revision identifier or file hashes [29].

3.2. Sampling and Data Preparation

We sampled 2000 products from 164,817 unique parent records after excluding the 1200-product pilot (seed 20260825). Among 8916 matched review rows, empty-text records were removed, 98 exact duplicate (title, text) pairs within parents were dropped, and the remaining 8818 pairs were shuffled within parents using the same seed before retaining at most 20 per parent. The final sample contains 6269 reviews from 1999 represented products; one sampled product had no eligible review. The median was one review per represented parent (mean 3.136; maximum 20); review dates span 18 September 2014 to 4 September 2023 UTC, and no language filter was applied. Category is missing for 27 reviews from 15 products, which are excluded from category-fixed-effect models. Cap-5/10/20 sensitivities use the same frozen within-parent shuffle and are nested.

The taxonomy defines three nominally distinct cue categories rather than an ordered intensity scale. C0 provides the reference group because it contains no explicit cultural cue in the observed listing fields. C1 and C2 differ in cue content and are not ranked as higher or lower heritage.

The released source identifies parent_asin as the parent-product identifier. We derived cultural-cue exposure once from the parent-level metadata row and linked reviews through parent_asin, the linkage and clustering unit. The frozen source contains no child-level historical listing snapshots, so variant-specific cue agreement cannot be assessed. Fifty-two sampled parents had multiple child ASIN variants represented (604 reviews); all remain under one parent-level exposure. Under the frozen codebook, a place qualified as C2 only when it functioned as an explicit cultural referent; logistical or geographic mentions were excluded. The main and pilot product samples had no parent_asin overlap.

3.3. Cultural-Cue Operationalization

Cultural cues were coded at the product level from title, description, features, and details fields into three mutually exclusive, nominal categories. C0 denotes no explicit cultural cue. C1 denotes a generic cultural or symbolic cue without a clearly identifiable tradition, place, or named technique. C2 denotes a culturally specific cue that identifies a place, tradition, community, named technique, or comparable cultural referent. The categories are analytically distinct and are not treated as an ordinal scale.

3.4. Sustainability-Related Review Coding

Sustainability-related review content was coded from review titles and text. S1 captures durability and longevity expectations, observed product failure, continued use, and repair that restores use or extends useful life. S2 captures reuse, recycling, repurposing, or other explicitly described circular-use practices. Repair or repairability alone does not qualify for S2. A review receives both labels only when it independently meets the S1 and S2 definitions. The frozen counts remain S1 = 554, S2 = 24, overlap = 6, and union = 572. S1 is dominant and S2 is sparse; neither is a direct measure of environmental performance.

A separate inter-coder reliability exercise used 200 items from the existing validation holdout (100 listings and 100 reviews), not the main analysis sample or additional LLM gold cases. Coder 1 was the lead author/researcher familiar with the framework; Coder 2 was Congrong Xiao, the coauthor assigned to Validation. No verifiable specialist credentials or years of experience for Coder 2 were documented in the available project records. The interface showed task text and the codebook but withheld ratings, helpful votes, outcome

fields, model labels, and the other coder's labels. Coder 1 did not view Coder 2's item-level responses before coding, as reported by the author.

Agreement was evaluated post hoc by applying the archived DeepSeek coding configuration to the author-coded reference cases. Because these cases had previously informed prompt refinement, this exercise is not treated as an independent blind validation. The exercise improves traceability but may yield optimistic validation estimates because human annotations are an imperfect reference standard [30]; model outputs remain measurements with classification error.

3.5. LLM-Assisted Content Classification

The historical pipeline sent one item per request to the DeepSeek chat-completions route recorded in the project script (https://api.deepseek.com/chat/completions (accessed on 25 August 2026). The archived model string is "deepseek-v4-flash"; this is a project-recorded string, not a verified public model identifier. The script loaded the full v0.3 system prompt and appended one fixed task instruction plus the item text. Listing input used title, description, features, and details; review input concatenated review title and review text. The tasks returned JSON: H0/H1/H2 for listings and mentioned yes/no, S1/S2 dimensions, and positive/negative/mixed/neutral valence for reviews, following established opinion-mining distinctions [31]. The request configuration recorded temperature = 0, max_tokens = 2048, and stream = false. The optional thinking setting could be changed by an environment variable; the executed value and complete execution-date range were not preserved. The final JSONL files contain 2000 listing records and 6269 review records, each with a non-empty raw response and no recorded error. The parser removed optional Markdown fences and parsed the first JSON object. Network errors, selected server errors, and empty responses triggered up to six exponential-backoff attempts; malformed JSON or other exceptions were recorded as errors with missing labels, not silently assigned a substantive category. The human-coder response "Uncertain" is a workflow option, not a verified model output category, and was not mapped into a substantive category. Automated text-analysis results depend on the task and measurement choices [32]. LLM annotation performance is task-dependent, and classification error can affect downstream regression estimates [33,34].

3.6. Human Validation and Coding Reliability

Post hoc validation against the one-author reference cases yielded heritage macro-F1 = 0.789 (95% bootstrap CI [0.658, 0.892]; C1 F1 = 0.800; C2 F1 = 0.588), mention F1 = 0.886 [0.794, 0.950], and S1 F1 = 0.928 [0.853, 0.986]. S2 F1 was 0.545 [0.222, 0.875] (positive support = 8), so S2 remains descriptive/exploratory. Valence accuracy was 0.786 [0.643, 0.905] and macro-F1 was 0.565 [0.402, 0.723]; neutral support was zero. These post hoc cases informed prompt refinement and may overstate independent performance.

In the 200-item holdout exercise, pre-adjudication nominal Cohen's κ was 0.583 for cultural cues, 0.696 for mention, 0.699 for S1, 0.306 for S2, and 0.513 for valence across all reviews; conditional valence agreement among jointly identified mentions was 0.404. The intervals and confusion matrices are reported in the Supplementary Materials. These are human–human agreement estimates from the holdout, not main-sample accuracy estimates.

Fifty substantive differences were recorded as discussion consensus; the remaining three followed the author's selected values. Adjudicated labels were not used to calculate κ, and item-level rationales were not retained.

3.7. Statistical Analysis

RQ1 is descriptive. The RQ2 primary model is review-level logistic regression of sustainability mention on C1/C2 indicators (C0 reference) and product-category fixed effects; covariance is clustered by parent_asin. Product-level any-positive/proportion, listing-length, sampling-cap, and leave-one-category-out estimates are sensitivity analyses. For RQ3, the primary post-review analysis is an ordered logit contrasting positive S1, negative S1, mixed S1, and S2-only with no mention, adjusted for verified-purchase status and review age with parent-ASIN-clustered covariance. A joint threshold-by-predictor diagnostic tests parallel slopes; threshold-specific cumulative logits are sensitivity estimates, not a partial proportional-odds model. RQ4 retains the frozen hurdle model for any observed vote and positive count; the valence split is exploratory sensitivity. Product average rating and total rating count are excluded where their snapshot timing is unknown. Benjamini–Hochberg correction is applied within prespecified inferential families; all R2-only analyses are labeled exploratory.

4. Results

4.1. Descriptive Results (RQ1)

The sample contains 2000 products and 6269 reviews: 1723 C0 products (5436 reviews), 202 C1 products (597 reviews), and 75 C2 products (236 reviews). Review-level sustainability mentions occurred in 518/5436 C0 reviews (9.529%), 28/597 C1 reviews (4.690%), and 26/236 C2 reviews (11.017%), and 572/6269 overall (9.125%). These rates use sampled reviews as denominators and are distinct from the product-level any-mention rates. The product and review counts, together with the corresponding sustainability-mention rates, are summarized in Table 1.

Table 1. Cue-stratified product and review counts and review-level mention rates (RQ1).

Cue Category	Products	Sampled Reviews	Sustainability Mentions (Review-Level n, %)
C0	1723	5436	518 (9.529%)
C1	202	597	28 (4.690%)
C2	75	236	26 (11.017%)
Total	2000	6269	572 (9.125%)

Of the 572 sustainability mentions, 554 involved S1 and 24 involved S2, with six reviews in both groups. Valence among mentions was positive in 319, negative in 205, mixed in 45, and neutral in 3; durability/longevity was the dominant coded experience. The distributions of cultural cues, sustainability-related review content, and review valence are illustrated in Figure 2.

The cross-tabulation is non-monotonic: 306 of 1723 C0 products (17.76%), 24 of 202 C1 products (11.88%), and 16 of 75 C2 products (21.33%) had at least one sampled sustainability-related review. Because these are nominal categories and the product sample is compositionally heterogeneous, the descriptive rates are not interpreted as a graded cultural effect.

Anonymized paraphrases from the frozen reviews illustrate the rule. One reviewer wanted a more durable material, another described a chain failing within weeks, and a third described having a bracelet repaired after a stone detached. These illustrate S1. Repair or repairability alone does not qualify for S2, which requires reuse, recycling, repurposing, or another explicit circular-use practice. One review met both definitions: the reviewer reported repeatedly washing and reusing pads in place of disposable ones (S2) and said

that the pads held up well through repeated washing (S1). The examples illustrate the coding rule and provide no direct evidence of environmental performance.

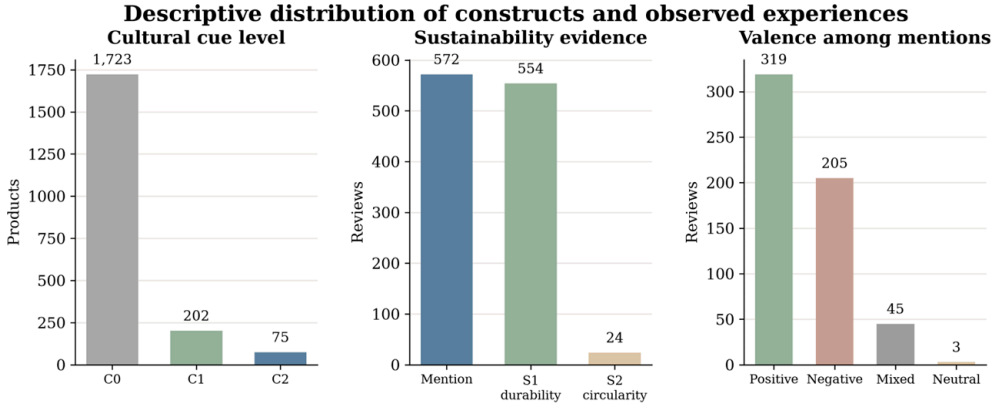


Figure 2. Descriptive distributions of C0/C1/C2 cultural cues, sustainability-related review content, and valence among mentioned reviews.

4.2. Cultural Cues and Sustainability Reviews (RQ2)

In the category-adjusted review-level model, C1 listings had lower mention odds than C0 (OR = 0.494, 95% CI [0.323, 0.756], $p = 0.00115$); C2 was imprecise (OR = 1.137, 95% CI [0.738, 1.752], $p = 0.560$). The category-adjusted regression results are presented in Table 2. The historical 5000-refit analysis propagated a reference-based error profile (Supplementary Materials Table S12). In the separate hypothetical diagnostic in Supplementary Materials Table S33, treating 29 C1 reviews as missed mentions raised the median adjusted OR above 1 to 1.057 (median 95% CI [0.772, 1.448]). The frozen sample originally contained 28 C1 mentions. Treating 125 C0 positive labels as false positives and removing them brought the median 95% CI across 1 [0.438, 1.029], while the median OR remained below 1 (0.673). All ten intervals included 1 in each scenario. The perturbations quantify sensitivity to assumed label changes. They provide neither observed misclassification rates nor a new blinded human validation.

Table 2. Category-adjusted review-level logistic model for sustainability mention (RQ2).

Predictor (Reference)	OR	95% CI	*p*
C1 vs. C0	0.494	[0.323, 0.756]	0.001154
C2 vs. C0	1.137	[0.738, 1.752]	0.56
Beauty and Grooming vs. Other	0.381	[0.195, 0.743]	0.004633
Clothing, Shoes and Accessories vs. Other	0.873	[0.517, 1.473]	0.6101
Home and Kitchen vs. Other	0.521	[0.336, 0.809]	0.003673
Jewelry vs. Other	0.631	[0.408, 0.974]	0.03754
Stationery and Party Supplies vs. Other	0.559	[0.308, 1.015]	0.05599

Note: Odds ratios with 95% confidence intervals. Reference categories are C0 for cultural cue and Other for product category. The model includes product-category fixed effects; covariance is clustered by parent_asin. $N = 6242$ reviews; 1984 product clusters; converged. The intercept is omitted.

Across the nested caps of 5, 10, and 20 reviews per product, the adjusted C1 ORs (95% CIs) were 0.688 [0.434, 1.091], 0.562 [0.359, 0.880], and 0.494 [0.323, 0.756], respectively. All three estimates were below 1, but magnitude and precision varied with the cap, and the cap-5 interval included 1.

At the product level, any-mention estimates were C1 OR = 0.641 [0.408, 1.009] and C2 OR = 1.168 [0.657, 2.077]. The grouped-binomial analysis weights products by their sampled-review denominators. The separate product sensitivity gave equal weight to

each of 1984 category-complete products with eligible reviews, using one row per product and category-adjusted OLS with HC3 covariance. The outcome was the within-product proportion of sampled reviews with a mention. The C1–C0 difference was −0.0128559 (95% CI [−0.0441618, 0.0184500], $p = 0.420888$) and the C2–C0 difference was 0.0134456 (95% CI [−0.0451346, 0.0720258], $p = 0.652805$). Both intervals included zero (Supplementary Materials Table S38). The main analysis uses the sampled review as its unit.

4.3. Sustainability Experience and Ratings (RQ3)

Archived Phase 6 ordered-logit fits reported pooled-mention OR = 0.842 (95% CI [0.609, 1.165]) and S1 OR = 0.593 [0.454, 0.776], but both were flagged as nonconverged. Converged R1 BFGS estimates were 0.287 [0.234, 0.352] and 0.285 [0.232, 0.350]; the pooled estimate was reproduced from the frozen R2 file. Rating order (1–5), sample (6269 reviews), and clusters (1999) were unchanged. Phase 6 included text length and snapshot rating count; R1 retained verified purchase and review age, moved text length to sensitivity analysis, and omitted rating count because its timing was unknown. The shift cannot be attributed to one cause. RQ2 and RQ4 use separate, independently reconstructed model families.

The observed one- through five-star distributions appear in Table 3, Panel B, with counts in Supplementary Materials Table S34. Negative S1 reviews were concentrated at one to three stars, while most positive S1 and no-mention reviews received five stars. The proportional-odds test rejected parallel slopes (W = 150.71, df = 18, $p < 0.001$). Panel A reports the main ordered-logit comparisons of each modeled S1 valence with no-mention reviews. The separate S1-only threshold analysis uses positive S1 as its reference and adjusts for verified purchase and review age ($N = 551,337$ product clusters). At rating ≥2, neither negative nor mixed S1 was estimable because of separation and non-convergence. At rating ≥3, the ORs (95% CIs) were 0.00178 [0.00024, 0.01302] for negative S1 and 0.04644 [0.00469, 0.45986] for mixed S1. At rating ≥4, they were 0.00119 [0.00035, 0.00399] and 0.02548 [0.00674, 0.09628]. At rating ≥5, they were 0.00324 [0.00142, 0.00742] and 0.05759 [0.02615, 0.12681]. Each estimate compares negative or mixed S1 with positive S1 at the stated threshold (Supplementary Materials Table S35).

The rating patterns cannot establish that sustainability-related content itself is penalized. Overall sentiment was not independently measured for all reviews, and a review star rating is not an independent overall-sentiment control. The observed correspondence between negative durability comments and lower ratings may reflect general dissatisfaction or sentiment-rating consistency. The adjusted associations across the primary regression models are summarized in Figure 3.

Average predicted probabilities for each rating category, with delta-method 95% confidence intervals using the product-cluster covariance, are reported in the R2 analysis supplement. The interval calculation and group-specific values are shown there; probability statements are not reduced to a sum-to-one sentence.

Table 3. Star ratings by review group (RQ3). Panel A. Ordered-logit contrasts for higher ratings, Panel B. Observed star-rating distribution by review group (%).

(**Panel A**)			
Predictor (Reference)	**OR**	**95% CI**	***p***
Positive S1 vs. no mention	3.651	[2.204, 6.047]	4.92×10^{-7}
Negative S1 vs. no mention	0.041	[0.033, 0.051]	1.49×10^{-180}
Mixed S1 vs. no mention	0.278	[0.185, 0.418]	7.88×10^{-10}

Table 3. *Cont.*

(Panel A)					
Predictor (Reference)	**OR**	**95% CI**	***p***		
S2-only vs. no mention (exploratory)	0.400	[0.164, 0.976]	0.0441		
Verified purchase	1.214	[0.876, 1.682]	0.2435		
Review age (years)	1.089	[1.043, 1.136]	8.54×10^{-5}		
(Panel B)					
Review Group	***N***	1 Star	2 Stars	3 Stars	4 Stars	5 Stars
No mention	5697	4.55	2.84	3.93	6.23	82.45
Positive S1	305	0.00	0.33	0.66	4.92	94.10
Negative S1	202	43.56	20.79	24.26	6.44	4.95
Mixed S1	44	2.27	4.55	20.45	25.00	47.73
Neutral S1	3	0.00	0.00	0.00	0.00	100.00
S2 only	18	11.11	0.00	11.11	11.11	66.67

Note: Panel A reports odds ratios for higher star ratings, with no sustainability mention as the reference group. Models adjust for verified-purchase status and review age, with standard errors clustered by parent_asin. $N = 6266$ reviews from 1999 product clusters; the proportional-odds assumption was rejected ($W = 150.71$, $df = 18$, $p < 0.001$). Panel B reports percentages within each review group. Neutral S1 ($N = 3$) is descriptive only, and S2-only ($N = 18$) is exploratory.

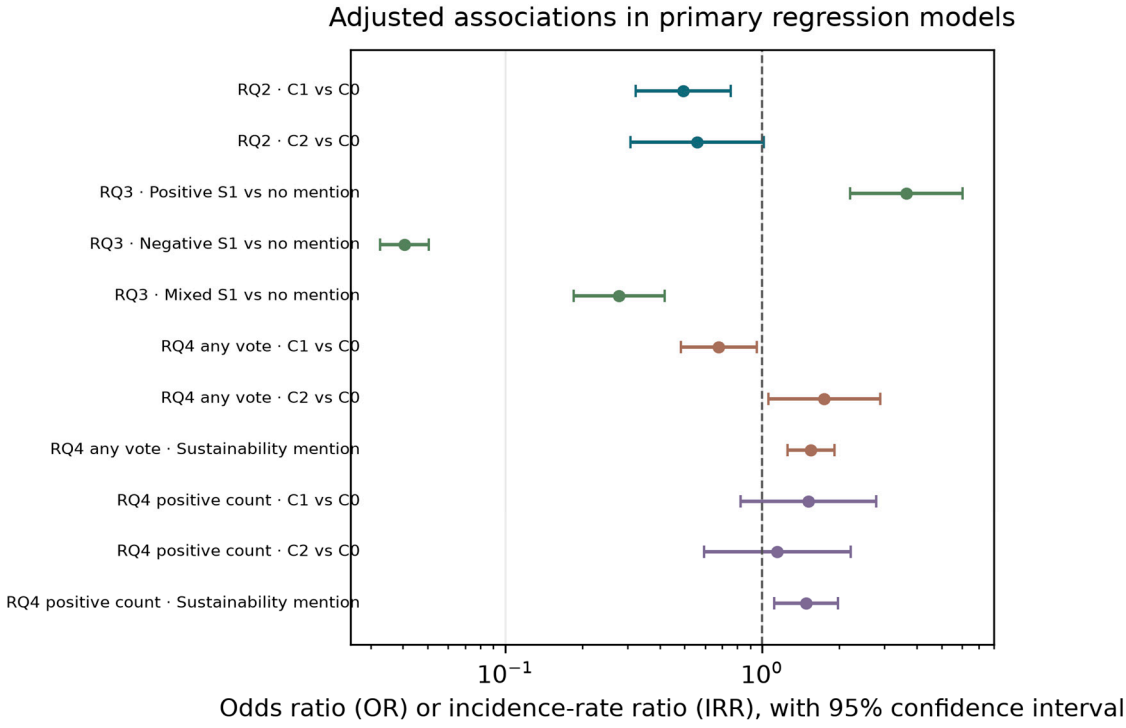


Figure 3. Adjusted associations in primary regression models.

4.4. Cultural Cues, Sustainability, and Observed Helpful-Vote Accumulation (RQ4)

In the category-complete sample (6242 reviews; 1984 products), 4780 reviews (76.58%) had no helpful votes. The any-vote logit gave C1 OR = 0.679 (95% CI [0.483, 0.955], $p = 0.026$), C2 OR = 1.747 [1.057, 2.887] ($p = 0.029$), and mention OR = 1.551 [1.256, 1.914] ($p < 0.001$). Among the 1467 reviews with a positive count, the mean was 2.214, the median was 1, the 90th, 95th, and 99th percentiles were 4, 6, and 12.34, and the maximum was 64. Ranked across the full samples, the top 1% of sampled reviews (63 reviews) accounted

for 25.31% of votes, and the top 1% of represented products (20 products) accounted for 27.12% (Supplementary Materials Table S36). Primary positive-count models retain ZTP residual overdispersion (Pearson chi-square/df = 5.97) and the ZTNB2 alternative has an extreme dispersion estimate.

In the separate influence specification (Supplementary Materials Table S37), the untrimmed ZTP C1 IRR was 1.453 (95% CI [0.794, 2.659]). After exclusion of the top 1% of category-complete reviews ranked by votes, the ZTP C1 IRR was 0.736 [0.455, 1.190] and ZTNB2 failed to converge. Excluding reviews of the top 1% of products by accumulated votes gave C1 IRRs of 0.779 [0.446, 1.358] for ZTP and 0.700 [0.348, 1.408] for ZTNB2. These C1 intervals include 1. The influence models omit sustainability mention. No comparable trimmed estimate from the joint Table 4 specification was available, so the sensitivity of its mention association to high-vote observations remains unresolved. Positive-count findings remain exploratory.

Table 4. Hurdle regression of observed helpful-vote accumulation (RQ4).

Hurdle Component	Predictor (Reference)	Effect	95% CI	p/BH q
Any-vote logit	C1 vs. C0	OR 0.679	[0.483, 0.955]	0.026016/0.029445
Any-vote logit	C2 vs. C0	OR 1.747	[1.057, 2.887]	0.029445/0.029445
Any-vote logit	Beauty and Grooming vs. Home and Kitchen	OR 1.447	[0.791, 2.648]	0.2303
Any-vote logit	Clothing, Shoes and Accessories vs. Home and Kitchen	OR 1.080	[0.743, 1.568]	0.6875
Any-vote logit	Jewelry vs. Home and Kitchen	OR 1.201	[0.956, 1.509]	0.116
Any-vote logit	Other vs. Home and Kitchen	OR 1.203	[0.891, 1.623]	0.2268
Any-vote logit	Stationery and Party Supplies vs. Home and Kitchen	OR 0.984	[0.657, 1.474]	0.9389
Any-vote logit	Sustainability mention	OR 1.551	[1.256, 1.914]	0.00004455/0.000134
Any-vote logit	Verified purchase	OR 1.195	[0.801, 1.783]	0.3833
Any-vote logit	log1p review age (days)	OR 1.384	[1.216, 1.575]	9.088×10^{-7}
Positive-count ZTP	C1 vs. C0	IRR 1.515	[0.826, 2.779]	0.179324/0.268987
Positive-count ZTP	C2 vs. C0	IRR 1.148	[0.594, 2.219]	0.681988/0.681988
Positive-count ZTP	Beauty and Grooming vs. Home and Kitchen	IRR 1.721	[1.018, 2.909]	0.04273
Positive-count ZTP	Clothing, Shoes and Accessories vs. Home and Kitchen	IRR 0.817	[0.528, 1.264]	0.3637
Positive-count ZTP	Jewelry vs. Home and Kitchen	IRR 0.923	[0.666, 1.279]	0.6294
Positive-count ZTP	Other vs. Home and Kitchen	IRR 0.866	[0.541, 1.385]	0.5469
Positive-count ZTP	Stationery and Party Supplies vs. Home and Kitchen	IRR 0.939	[0.583, 1.512]	0.7947
Positive-count ZTP	Sustainability mention	IRR 1.484	[1.114, 1.978]	0.007016/0.021047
Positive-count ZTP	Verified purchase	IRR 0.735	[0.299, 1.802]	0.5008
Positive-count ZTP	log1p review age (days)	IRR 1.175	[1.013, 1.362]	0.03271

Note. Part A is the product-clustered any-vote logit (N = 6242 reviews, 1984 products). Part B is the zero-truncated Poisson (ZTP) model (N = 1462 reviews, 568 products, Pearson dispersion = 5.97). Each Benjamini–Hochberg family contains three focal contrasts: C1 vs. C0, C2 vs. C0, and sustainability mention. The p/BH q column reports the unadjusted *p*-value followed by q for each focal contrast. Control variables and influence fits are outside the two families.

5. Discussion

5.1. Cultural Cues and Sustainability Experience Are Separably Measured

The category-adjusted C1 association does not support using cultural listing language as a positive proxy for reported product-use experience. C2 remains imprecise. Hypothetical label perturbations show how assumed error allocations could alter the estimate but are not independent validation and do not change frozen labels. Review-cap estimates changed

in magnitude; the cap-5 interval crossed 1. The equal-product-weighted contrasts were imprecise. The lower C1 mention odds in the sampled-review analysis do not establish a difference for the typical represented product. These results define an association boundary, not a mechanism.

5.2. Sustainability-Related eWOM and Durability

Durability was the dominant observed sustainability dimension, indicating that consumers more often described sustainability-relevant experience through whether an item lasted, failed, or remained usable than through circularity language. This pattern motivates a durability- and circularity-related product-use construct rather than a broad claim about environmental sustainability.

5.3. Experience Valence and Consumer Ratings

Because the proportional-odds assumption was rejected, the observed distributions and estimable thresholds provide the clearest evidence of the rating pattern. Negative S1 reviews were concentrated at low ratings. Within S1, negative and mixed valence were associated with lower odds than positive valence at thresholds from ≥3 through ≥5. The ≥2 models could not estimate either contrast. Without an independent measure of overall sentiment, the data cannot separate durability-specific evaluation from general dissatisfaction.

5.4. Cultural Cues, Sustainability, and Observed Helpful-Vote Accumulation

RQ4 separates the likelihood of receiving any observed vote from positive-count accumulation. The primary any-vote associations are supported by focal-family BH-adjusted values. Positive-count evidence is less stable: votes have a concentrated upper tail, ZTP retains overdispersion, ZTNB2 has an extreme dispersion estimate, and influence fits are imprecise or fail to converge. The influence models omit sustainability mention, leaving the sensitivity of the joint model's mention coefficient unresolved. The positive-count associations are exploratory, and neither component measures intrinsic review quality or establishes causal visibility.

The findings also provide a boundary condition for tradition-washing research. Experimental evidence about perceived greenness motivates the distinction, but this observational study does not test that perceptual mechanism, does not establish temporal precedence, and does not measure objective environmental performance.

5.5. Conclusions

In reviews of handmade products, durability and how long items lasted dominated sustainability-related comments. Reviewers seldom described circular practices such as reuse. Product pages show how sellers present cultural attributes, while reviews describe aspects of use that buyers chose to report. Cultural cues cannot substantiate claims about durability, reuse, or environmental performance. Ratings and helpful votes provide context about how buyers evaluated their experiences and how readers responded to reviews.

The study clarifies what these sources can tell us. The observational data cannot establish causal effects or confirm that the recorded listings preceded the reviews. We did not objectively measure environmental performance or verify product durability. Overall sentiment was not independently measured for all reviews, and some classifications and model estimates remain uncertain.

5.6. Implications

Academic implications. The findings support treating seller-side cultural cues and consumer-reported durability/circularity experience as distinct marketplace evidence rather

than assuming that cultural framing is a proxy for environmental performance. The category-adjusted C1 association is observational and does not identify a signaling mechanism; the imprecise C2 estimate does not demonstrate equivalence or absence of association.

Research implications. Future studies should measure listing states longitudinally, capture product-category and listing-variant differences, and directly assess perceived authenticity, greenness, consumer involvement, gifting context, and expectation-disconfirmation. A larger, separately sampled, rights-cleared human validation set with documented coder expertise is needed to better assess sparse circularity (S2) and the transportability of the coding results.

Managerial implications. Sellers should substantiate cultural claims with appropriate provenance information and keep any durability or circularity claim separate and supported by product-use evidence. Marketplaces may use review text and observed vote accumulation as descriptive monitoring signals, but neither is an environmental certification or a measure of exposure-adjusted review helpfulness.

5.7. Limitations

Interpretation is limited by the absence of retained historical listing snapshots, independent overall-sentiment labels for all reviews, sparse C2 and S2 support, and classification uncertainty. Adjusted label perturbation is hypothetical, not a blind human audit. Review-cap estimates vary in magnitude, and review-level, product-level any-mention, and equal-product-weighted proportion models target different estimands. Helpful-vote counts have a concentrated upper tail and unstable positive-count fits. The available influence models cannot establish whether high-vote observations drive the positive-count association with sustainability mention. This is a single-platform observational study.

These limitations define a constructive research agenda. A longitudinal marketplace design could link listing changes, review submission, rating snapshots, and helpful-vote accumulation. A mixed-method design could measure authenticity, perceived greenness, consumer involvement, and craftsmanship expectations directly, while independent human coding and label-perturbation sensitivity would strengthen the LLM-assisted classification evidence. Linking these records to independently verified product attributes could help distinguish what buyers report about durability and reuse from what can be measured about the products themselves.

Supplementary Materials: The following supporting information can be downloaded at: https://www.mdpi.com/article/10.3390/jtaer21110428/s1. The accompanying supplementary file contains additional methods, figures, tables, validation results, and sensitivity analyses. Further research materials may be obtained from the corresponding author upon reasonable request, subject to applicable data-use restrictions.

Author Contributions: Conceptualization, F.W.; methodology, F.W.; software, J.L.; validation, C.X.; formal analysis, F.W.; investigation, J.L.; writing—original draft preparation, F.W.; writing—review and editing, C.X. and D.S.; supervision, D.S.; project administration, D.S. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding.

Institutional Review Board Statement: Not applicable. This study analyzed an existing publicly accessible secondary dataset and involved no direct recruitment of, interaction with, or intervention involving human participants. No new data were collected directly from individuals.

Informed Consent Statement: Not applicable. This study did not involve direct recruitment of human participants or collection of new participant data.

Data Availability Statement: This study used the Handmade Products subset of Amazon Reviews'23, distributed by the McAuley Lab. The original review text is not redistributed by the authors. The

coding taxonomy, classification prompts, analysis scripts, model specifications, and non-identifying derived aggregate results are available from the corresponding author upon reasonable request, subject to the terms governing the source dataset.

Acknowledgments: During the preparation of this manuscript and the associated research work, the authors used AI-assisted tools, including OpenAI's GPT-5.6, to support research-related technical tasks, conceptual design refinement, graphical development, language editing, and manuscript preparation. The authors reviewed the relevant outputs, independently interpreted the findings, and took full responsibility for the accuracy, integrity, and final content of the manuscript.

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

References

1. Mudambi, S.M.; Schuff, D. What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.com[1]. *MIS Q.* **2010**, *34*, 185–200. [CrossRef]
2. Wang, L.; Che, G.; Hu, J.; Chen, L. Online Review Helpfulness and Information Overload: The Roles of Text, Image, and Video Elements. *J. Theor. Appl. Electron. Commer. Res.* **2024**, *19*, 1243–1266. [CrossRef]
3. Liu, X.; Lin, J.; Jiang, X.; Chang, T.; Lin, H. eWOM Information Richness and Online User Review Behavior: Evidence from TripAdvisor. *J. Theor. Appl. Electron. Commer. Res.* **2024**, *19*, 880–898. [CrossRef]
4. Nurse, K. Culture as the Fourth Pillar of Sustainable Development. *Small States Econ. Rev. Basic Stat.* **2006**, *11*, 32–48. Available online: https://www.bibalex.org/baifa/Resources/Document/449503 (accessed on 25 August 2026).
5. Soini, K.; Birkeland, I. Exploring the Scientific Discourse on Cultural Sustainability. *Geoforum* **2014**, *51*, 213–223. [CrossRef]
6. Soini, K.; Dessein, J. Culture-Sustainability Relation: Towards a Conceptual Framework. *Sustainability* **2016**, *8*, 167. [CrossRef]
7. Ceschin, F.; Gaziulusoy, I. Evolution of Design for Sustainability: From Product Design to Design for System Innovations and Transitions. *Des. Stud.* **2016**, *47*, 118–163. [CrossRef]
8. Mesa, J.A. Design for Circularity and Durability: An Integrated Approach from DFX Guidelines. *Res. Eng. Des.* **2023**, *34*, 443–460. [CrossRef]
9. Vilochani, S.; McAloone, T.C.; Pigosso, D.C.A. Management Practices for Sustainable Product Development: Insights from a Systematic Literature Review. *Proc. Des. Soc.* **2023**, *3*, 2505–2514. [CrossRef]
10. Cordella, M.; Alfieri, F.; Sanfelix, J.; Donatello, S.; Kaps, R.; Wolf, O. Improving Material Efficiency in the Life Cycle of Products: A Review of EU Ecolabel Criteria. *Int. J. Life Cycle Assess.* **2020**, *25*, 921–935. [CrossRef]
11. Mazzitelli, M.; Papile, F.; Del Curto, B. Materials Selection and Fashion Design: Strengthening Reflections on Fibre's Nature in Fibres and Textiles Selection. *Discov. Sustain.* **2024**, *5*, 180. [CrossRef]
12. Fuchs, C.; Schreier, M.; van Osselaer, S.M.J. The Handmade Effect: What's Love Got to Do with It? *J. Mark.* **2015**, *79*, 98–110. [CrossRef]
13. Degirmenci, T.; Mathmann, F.; Mortimer, G. Made with Love: Examining Consumer Engagement with Handmade Versus Machine-Made Production Cues. *J. Consum. Behav.* **2025**, *24*, 982–1002. [CrossRef]
14. Berta, O.G. Culture or Commerce? Craft as an Ambiguous Construction between Culture and Economy. *J. Cult. Econ.* **2025**, *18*, 92–108. [CrossRef]
15. Oyedele, A.; Saldivar, R.; Goenner-Munson, E. Consumers Purchase Decision for Cultural Products outside Their Countries of Origin: A Craft Retailing Context. *Tour. Hosp. Res.* **2025**, *26*, 782–798. [CrossRef]
16. Wherry, F.F. The Social Sources of Authenticity in Global Handicraft Markets: Evidence from Northern Thailand. *J. Consum. Cult.* **2006**, *6*, 5–32. [CrossRef]
17. Heller, M.; Pietikäinen, S.; da Silva, E. Body, Nature, Language: Artisans to Artists in the Commodification of Authenticity. *Anthropologica* **2017**, *59*, 114–129. [CrossRef]
18. Kendall, L. Intangible Traces and Material Things: The Performance of Heritage Handicraft. *Acta Koreana* **2014**, *17*, 537–555. [CrossRef]
19. Behre, B.; Cauberghe, V. Signaling Sustainability in Online Fashion Consumption: The Role of Credibility Induced by Certification Labels. *Clean. Responsible Consum.* **2025**, *17*, 100287. [CrossRef]
20. Tabelessy, W.; Nanere, M. How Eco-Badges Shape Online Purchase Decisions: The Roles of Perceived Diagnosticity and Label Credibility. *J. Trade Sci.* **2026**, *14*, 106–126. [CrossRef]
21. Zhang, S.; Wu, C.; Yan, X.; Chen, Y.; Shi, H. Unverifiable Green Signals and Consumer Response in E-Commerce: Evidence from Platform-Level Data. *Sustainability* **2025**, *17*, 5678. [CrossRef]
22. Choubey, V.; Joshi, P.; Chakraborty, D.; Khare, A.; Appolloni, A. Signals vs. Reality: Consumer Responses to Green Claims in Quick Commerce. *Bus. Strategy Environ.* **2025**, *34*, 3656–3671. [CrossRef]

23. Alshaketheep, K.; Shajrawi, A.; Al-husban, A.M.; Al-Sukkar, A.S.; Zraqat, O. Credible Sustainability Signals, Consumer Trust, and Purchase Intention in Selected Middle Eastern Online Retail Contexts. *SAGE Open* **2026**, *16*, 1–19. [CrossRef]
24. Hong, H.; Xu, D.; Wang, G.A.; Fan, W. Understanding the Determinants of Online Review Helpfulness: A Meta-Analytic Investigation. *Decis. Support Syst.* **2017**, *102*, 1–11. [CrossRef]
25. Park, J.; Park, H. Understanding the Impact of Inconsistency on the Helpfulness of Online Reviews. *J. Theor. Appl. Electron. Commer. Res.* **2025**, *20*, 80. [CrossRef]
26. Shih, H.-P.; Lai, K.-H.; Cheng, T.C.E. Complied by Belief Consistency: The Cognitive-Information Lens of User-Generated Persuasion. *J. Theor. Appl. Electron. Commer. Res.* **2023**, *18*, 372–393. [CrossRef]
27. De Souza, V.; Borsato, M. Sustainable Design and Its Interfaces: An Overview. *Int. J. Agil. Syst. Manag.* **2016**, *9*, 183–211. [CrossRef]
28. Larrañaga, A.; Valor, C.; Antonetti, P. The Risk of Tradition-Washing: Why Communicating Traditionality Increases Green Perceptions. *Environ. Commun.* **2025**, *19*, 833–853. [CrossRef]
29. Hou, Y.; Li, J.; Fu, X.; He, Z.; Yan, A.; Chen, X.; McAuley, J. Bridging Language and Items for Retrieval and Recommendation. *arXiv* **2024**, arXiv:2403.03952. [CrossRef]
30. Song, H.; Tolochko, P.; Eberl, J.-M.; Eisele, O.; Greussing, E.; Heidenreich, T.; Lind, F.; Galyga, S.; Boomgaarden, H.G. In Validations We Trust? The Impact of Imperfect Human Annotations as a Gold Standard on the Quality of Validation of Automated Content Analysis. *Politi. Commun.* **2020**, *37*, 550–572. [CrossRef]
31. Pang, B.; Lee, L. Opinion Mining and Sentiment Analysis. *Found. Trends Inf. Retr.* **2008**, *2*, 1–135. [CrossRef]
32. Grimmer, J.; Stewart, B.M. Text as Data: The Promise and Pitfalls of Automatic Content Analysis Methods for Political Texts. *Politi. Anal.* **2013**, *21*, 267–297. [CrossRef]
33. Gilardi, F.; Alizadeh, M.; Kubli, M. ChatGPT Outperforms Crowd Workers for Text-Annotation Tasks. *Proc. Natl. Acad. Sci. USA* **2023**, *120*, e2305016120. [CrossRef] [PubMed]
34. Qiao, M.; Huang, K.-W. Correcting Misclassification Bias in Regression Models with Variables Generated via Data Mining. *Inf. Syst. Res.* **2021**, *32*, 462–480. [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.