Proceedings of ICEMGD 2026 Symposium: Data-Driven Business Management, Corporate Governance, and Organizational Performance
DOI: 10.54254/2754-1169/2026.LD35313

Comparing Likes and Favorites as Signals of Purchase-Oriented Behavior in Social E-Commerce Beauty Content

Xinran Liu

School of International Exchange, Guangzhou Maritime University, Guangzhou, China
sunnyliu_0755@163.com

Abstract. With the development of Social Commerce, consumers increasingly rely on User-Generated Content, platform interaction data and engagement metrics for product evaluation and purchase-related decision-making. In the beauty-related Social E-Commerce context, likes, comments, shares and favorites/collects not only represent content popularity, but also may reflect different levels of consumer interest. This article mainly studies whether favorites/collects are closer to purchase-oriented behavioral signals than likes. This article uses Alibaba Cloud Tianchi Social E-Commerce User Purchase Behavior Dataset as the main public dataset and focuses on beauty and personal care category. Research methods include descriptive statistics, correlation analysis, verification of derived behavioral indicators, stepwise Ordinary Least Squares regression, standardized coefficient comparison, Wald test and robustness checks. The results show that favorites/collects present a more stable positive association with derived purchase-oriented behavioral index, while likes weaken or turn into a negative coefficient after controlling favorites/collects. However, the binary purchase label check did not show significant results. Therefore, the conclusion of this article should be interpreted as exploratory association evidence, not causal evidence of actual purchase conversion.

Keywords: Social Commerce, Engagement Quality, Favorites, Behavioral Signal

1. Introduction

Social Commerce has changed the way consumers obtain product information and form purchase-related decisions. In Social E-Commerce platforms, consumers not only rely on the information provided by sellers or brands, but also refer to User-Generated Content, peer reviews and engagement metrics. Especially in beauty-related content, consumers usually compare product reviews, price, discount, ingredients, user experience and platform interaction signals to form purchase-related consideration [1, 2].

However, different Social Media Engagement behaviors do not necessarily represent the same level of consumer interest. Likes is usually a kind of low-cost approval, which may reflect more general popularity or quick reaction. In contrast, favorites/collects have saving-based and future-oriented characteristics, which may indicate that users want to check product information or compare again later. Therefore, if likes, favorites/collects, comments and shares are all regarded as unified engagement metrics, the difference in engagement quality may be ignored [3-5].

DOI: 10.54254/2754-1169/2026.LD35313

This article focuses on Social E-Commerce beauty content and studies whether favorites/collects are closer to purchase-oriented behavioral indicators than likes. This article uses Alibaba Cloud Tianchi public dataset and regards purchase_intent as derived purchase-oriented behavioral index, not directly observed psychological purchase intention. The goal of this article is not to prove the causal effect, but to test the association pattern between different engagement behaviors and purchase-oriented behavioral index. This article can help brands and platform operators better understand Engagement Quality, instead of relying only on visible popularity metrics such as likes. This article can also provide a reference for researchers who follow up on Social Commerce consumer behavior, that is, different engagement behaviors should be understood differently, rather than simply combined into total engagement.

2. Literature review and hypotheses

User-Generated Content (UGC) plays an important role in Social Commerce. Compared with traditional advertising, UGC usually comes from the user's product experience, review, recommendation or discussion, so it is more likely to be used as a reference for product evaluation by other consumers. In the Social Commerce environment, content, interaction and product information are closely connected, and engagement behaviors also become important observable signals to understand consumer interest [1-3].

Social Media Engagement generally includes likes, comments, shares and favorites/collects. Existing studies often use overall engagement or total interaction to measure content performance, but this practice may ignore the behavioral cost and psychological meaning of different behaviors. Likes has a low operating cost and may mainly represent content attractiveness or general popularity; favorites/collects means that users save the content for later viewing or reference. Therefore, favorites/collects may reflect a higher level of consumer involvement than likes [4, 5].

In public datasets, directly observed psychological purchase intention is often difficult to obtain. Therefore, many studies will use behavioral indicators or proxy variables to analyze consumer interest, such as click behavior, add-to-cart behavior, coupon receiving, engagement metrics or purchase-related index [2, 3, 6, 7]. The purchase_intent in this article is not interpreted as a purchase intention at the real psychological level, but is regarded as a derived purchase-oriented behavioral index. Based on this point, this article focuses on analyzing the association of likes and favorites/collects with the derived index, rather than the causal prediction of actual purchase behavior.

Based on the above literature and research gap, this article puts forward the following research questions and hypotheses:

RQ1: The relationship between likes and favorites/collects and derived purchase-oriented behavioral index respectively.

RQ2: The relative strength of the association between favorites/collects and the derived purchase-oriented behavioral index compared with likes.

H1: Likes are positively associated with the derived purchase-oriented behavioral index.

H2: Favorites/collects are positively associated with the derived purchase-oriented behavioral index.

H3: Favorites/collects show stronger alignment with the derived purchase-oriented behavioral index than likes.

DOI: 10.54254/2754-1169/2026.LD35313

3. Methodology

3.1. Dataset and variable measurement

This article uses Alibaba Cloud Tianchi Social E-Commerce User Purchase Behavior Dataset as the main public dataset [8]. The dataset contains user-content interaction records, engagement variables, product-related variables and purchase-related behavioral indicators. In order to keep the research context consistent with beauty-related Social Commerce, this article selects the beauty and personal care category as an analysis sample.

The dependent variable of this article is purchase_intent. Since its exact construction formula is not fully disclosed, this article regards it as derived purchase-oriented behavioral index, not directly observed psychological purchase intention. Key independent variables include like_num and collect_num. Among them, like_num means likes, which is interpreted as a low-cost approval or general popularity signal; collect_num means favorites/collects, which is interpreted as a saving-based engagement signal. Control variables include comment_num, share_num, price, discount rate, title length, title emotional score, image count, video format, page views and freshness score. The variable measurement and expected roles are shown in Table 1.

Table 1. Variable measurement and expected roles

Variable	Measurement	Expected role
purchase_intent	Derived purchase-oriented behavioral index	Dependent variable used as a behavioral proxy
like_num	Number of likes	Low-cost approval and popularity signal
collect_num	Number of favorites or collects	Saving-based and future-oriented engagement signal
comment_num	Number of comments	Higher-effort interaction control
share_num	Number of shares	Diffusion-oriented interaction control
price	Product price	Product-level control
discount_rate	Discount rateLength of content title	Promotion control title_lengthContent presentation control
title_emo_score	Emotional score of title	Content sentiment control
img_count	Number of imagesVideo indicatorPage views	Media richness control has_videoMedia format control pv_countExposure control
freshness_score	Freshness score	Time-related content control

3.2. Model design and additional checks

Before the formal analysis, this article conducts a missing value check and an extreme observation check. Since Social Commerce engagement data usually has a right-skewed distribution, this article uses $\log(x + 1)$ transformation for the main count variables. This article adopts descriptive statistics, correlation analysis, stepwise Ordinary Least Squares (OLS) regression, standardized coefficient comparison, Wald test and robustness checks.

Stepwise OLS regression is used to observe how the coefficients of likes and favorites/collects change with model specification. Model 1 only adds likes, Model 2 only adds favorites/collects, Model 3 adds likes and favorites/collects at the same time, and Model 4 adds other engagement variables and control variables. The basic form of the model is as follows:

DOI: 10.54254/2754-1169/2026.LD35313

$$\log\_purchase\_intent = \beta 0 + \beta 1 \log\_like\_num + \beta 2 \log\_collect\_num + \beta 3 Controls + \varepsilon \quad (1)$$

Among them, β1 represents the association of likes and derived purchase-oriented behavioral index, and β2 represents the association of favorites/collects and the index. This article focuses on comparing β1 and β2, rather than explaining the causal effect.

In addition, this article also conducts additional checks. First of all, this paper verifies whether some behavioral indicators can be approximately reconstructed by observed variables. The results show that interaction_rate can be almost completely reconstructed by engagement variables, and R2 is close to 1; however, the explanatory power of the test formula for purchase_intent is low, and R2 is about 0.4591, indicating that the formula cannot completely reconstruct purchase_intent. Secondly, Variance Inflation Factor (VIF) is used to detect multicollinearity. Thirdly, the binary purchase label is used as an alternative outcome. Finally, this paper uses sample sensitivity check and exposure-adjusted engagement rates to test the stability of the results.

4. Results and discussion

4.1. Descriptive and diagnostic results

Descriptive statistics show that there is obvious skewness in Social Commerce engagement variables. The average value of Likes, favorites/collects, comments and shares is usually higher than the median, indicating that a few high-hot content may get a lot of interactions. This distribution pattern supports the use of log(x + 1) transformation.

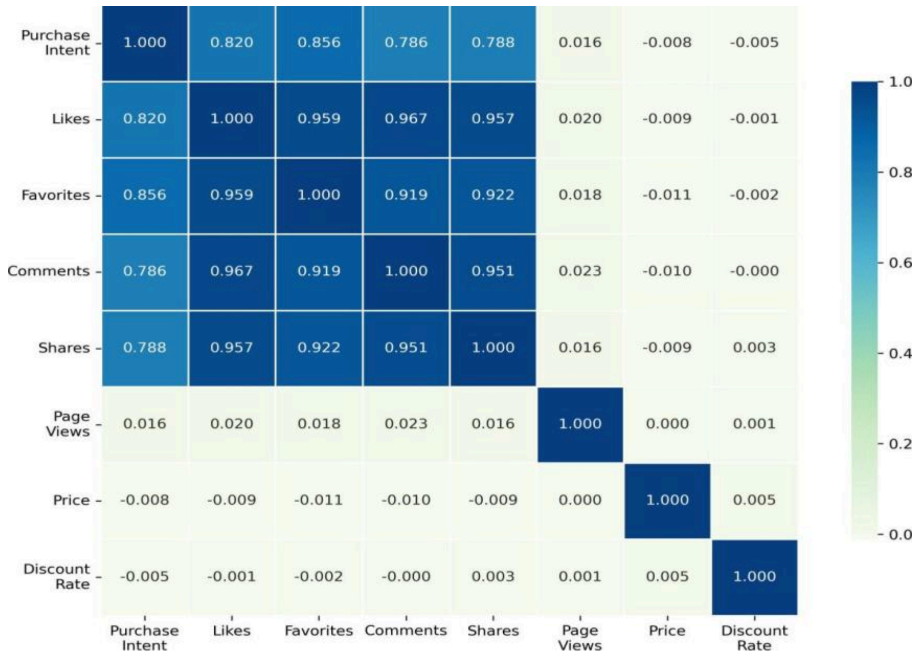

Figure 1. Correlation heatmap of main variables

Correlation analysis shows that favorites/collects and derived purchase-oriented behavioral index show a strong positive correlation, as shown in Figure 1. Likes also has a positive correlation with the index, but the correlation results can only be used as preliminary evidence. VIF results further show that the main engagement variables have multicollinearity. For example, the VIF of log_like_num is about 13.513, the VIF of log_collect_num is about 18.470, and the VIF of

DOI: 10.54254/2754-1169/2026.LD35313

log_comment_num is about 14.382. Therefore, this article is cautious about the interpretation of coefficients. The VIF results and verification results are shown in Table 2.

Table 2. Diagnostic and verification results

Diagnostic item	Indicator	Result
Formula verification	interaction_rate reconstruction R2	1.000
Formula verification	interaction_rate reconstruction MAE	0.000
Formula verification	purchase_intent reconstruction R2	0.459
Formula verification	purchase_intent reconstruction MAE	0.765
Multicollinearity	VIF of log_like_num	13.513
Multicollinearity	VIF of log_collect_num	18.470
Multicollinearity	VIF of log_comment_num	14.382

4.2. Regression results

Stepwise regression results show that in Model 1, the coefficient of log_like_num is 0.255 and statistically significant. This means that when likes enter the model alone, it presents a positive association with the derived purchase-oriented behavioral index. In Model 2, the coefficient of log_collect_num is 0.320 and statistically significant, indicating that favorites/collects are also positively related to the index when entering the model alone. The stepwise OLS regression results are shown in Table 3.

However, after adding likes and favorites/collects at the same time in Model 3, the coefficient of log_like_num becomes -0.125, while the coefficient of log_collect_num rises to 0.458. After Model 4 adds complete control variables, log_like_num is still a negative coefficient, about -0.153, while log_collect_num still maintains a positive coefficient, about 0.299. This pattern shows that favorites/collects may absorb more purchase-oriented behavioral signals, and likes are more like residual popularity signals after controlling favorites/collects.

Table 3. Stepwise OLS regression results

Variable	Model 1	Model 2	Model 3	Model 4
Constant	0.309***	0.508***	0.631***	0.737***
Log likes	0.255***		-0.125***	-0.153***
Log favorites		0.320***	0.458***	0.299***
Log comments				0.046***
Log shares				0.209***
Log price				-0.002
Discount rate				-0.034
Title length				-0.001
Title emotional score				-0.049
Image count				-0.002
Video indicator				-0.003
Log page views				0.001
Freshness score				0.057**

DOI: 10.54254/2754-1169/2026.LD35313

Table 3. (continued)

Observations	21,143	21,143	21,143	21,143
R^2	0.181	0.214	0.218	0.228

Notes: Robust standard errors are used. ***, **, and * indicate significance at the 1%, 5%, and 10% levels.

4.3. Coefficient comparison and robustness checks

In order to further compare the relative association strength of likes and favorites/collects, this paper uses standardized coefficient comparison and Wald test. The results show that the standardized coefficient of log_collect_num is about 0.433, while the standardized coefficient of log_like_num is about -0.256. Wald test shows that there is a significant difference between the coefficients of likes and favorites/collects; $\chi 2$ is about 231.151, $p < 0.001$. Therefore, H3 has received strong support.

Binary purchase label check shows that log_like_num and log_collect_num are not significant for binary purchase occurrence. This shows that the association observed on the derived purchase_intent index is not necessarily directly converted into binary purchase occurrence. Robustness checks provide further support. Sample sensitivity check shows that after eliminating extreme viral observations, log_collect_num still maintains a positive coefficient of about 0.270, while log_like_num is about -0.097. The results of the Alternative engagement measure show that after using exposure-adjusted engagement rates, log_favorite_rate is still significantly positive, and the coefficient is about 0.526, $p < 0.001$; log_like_rate is a negative coefficient, about -0.266, $p < 0.001$. The coefficient comparison and robustness results are shown in Table 4.

Table 4. Coefficient comparison and robustness checks

Check	Variable or statistic	Estimate	p-value
Standardized full model	log_like_num	-0.256	< 0.001
Standardized full model	log_collect_num	0.433	< 0.001
Wald test	Difference between likes and favorites/collects	231.151	< 0.001
Binary purchase label	log_like_num	-0.056	0.128
Binary purchase label	log_collect_num	0.024	0.637
Trimmed sample	log_like_num	-0.097	< 0.001
Trimmed sample	log_collect_num	0.270	< 0.001
Exposure-adjusted rates	log_like_rate	-0.266	< 0.001
Exposure-adjusted rates	log_favorite_rate	0.526	< 0.001

4.4. Discussion of findings

Overall results show that H1 can only get partial support, H2 can get stronger support, and H3 can get clear support under the derived index framework. For H1, likes present a positive association with derived purchase-oriented behavioral index when entering the model alone, but after controlling favorites/collects, the coefficient turns negative. Therefore, H1 only exists in the simple model, not in the full model. For H2, favorites/collects maintain a positive association in different model specifications. For H3, both standardized coefficients and Wald test show that the alignment of favorites/collects and derived index is stronger than likes.

DOI: 10.54254/2754-1169/2026.LD35313

It should be noted that the negative coefficient of likes in the full model should not be overexplained. This result does not mean that likes will reduce purchase-oriented behavior. A more reasonable explanation is that likes and favorites/collects represent different levels of engagement quality. Likes may mainly reflect general popularity or low-cost approval, while favorites/collects are more likely to reflect saving-based and future-oriented engagement.

5. Contributions and limitations

The theoretical contribution of this article is to split Social Media Engagement into different behavioral signals, rather than treating engagement as a single overall metric. By comparing likes and favorites/collects, this article emphasizes the difference between low-cost approval and saving-based engagement. This helps to enrich Engagement Quality research in Social Commerce, and also shows that there may be different degrees of alignment between different engagement behaviors and purchase-oriented behavioral indicators.

The practical contribution of this article is to provide more subdivided engagement evaluation ideas for brands, marketers and platform operators. Compared with relying only on likes as a popularity metric, marketers can pay more attention to favorites/collects, because they may be closer to saving-based interest and future reference behavior. This discovery can be used for content evaluation, campaign analysis and platform recommendation strategy.

This article still has limitations. First, purchase_intent is a derived behavioral index, and the exact construction formula is not fully disclosed. Therefore, this article cannot claim to have measured the directly observed psychological purchase intention. Second, the binary purchase label check does not show the significant effect of likes or favorites/collects. This means that the association observed in the derived index may not be directly converted into actual purchase occurrence. Third, this article uses a cross-sectional public dataset, so causal inference is limited. Future research can use user-level longitudinal data, actual conversion records or experimental design to further test whether favorites/collects can predict real purchase behavior.

6. Conclusion

This article mainly studies whether favorites/collects are closer to purchase-oriented behavioral indicators in Social E-Commerce beauty content than likes. Based on Alibaba Cloud Tianchi public dataset, this article uses descriptive statistics, correlation analysis, stepwise regression, coefficient comparison, additional outcome checks and robustness checks to compare likes and favorites/collects.

The results show that favorites/collects and derived purchase-oriented behavioral index present a more stable and stronger positive association. Likes are positively related to the index when entering the model alone, but weaken or turn into a negative coefficient after controlling favorites/collects. This shows that likes may mainly reflect general popularity, while favorites/collects may represent saving-based and future-oriented engagement.

However, the conclusion of this article needs to be explained carefully. Since purchase_ intent is a derived index and the binary purchase label results are not significant, this article cannot claim that favorites/collects can directly predict actual purchase conversion. Future research can use more transparent platform-level data or user-level purchase records to further test this relationship. In general, this article provides a more segmented perspective for Engagement Quality research in Social Commerce by distinguishing between favorites/collects and likes.

DOI: 10.54254/2754-1169/2026.LD35313

References

[1] Hajli, N. (2015). Social commerce constructs and consumer's intention to buy. *International Journal of Information Management, 35*(2), 183-191.

[2] Erkan, I., & Evans, C. (2016). The influence of eWOM in social media on consumers' purchase intentions: An extended approach to information adoption. *Computers in Human Behavior, 61*, 47-55.

[3] Zhao, Y., Wang, L., Tang, H., & Zhang, Y. (2020). Electronic word-of-mouth and consumer purchase intentions in social e-commerce. *Electronic Commerce Research and Applications, 41*, 100980.

[4] Trunfio, M., & Rossi, S. (2021). Conceptualising and measuring social media engagement: A systematic literature review. *Italian Journal of Marketing, 2021*, 267-292.

[5] Dessart, L., Veloutsou, C., & Morgan-Thomas, A. (2015). Consumer engagement in online brand communities: A social media perspective. *Journal of Product & Brand Management, 24*(1), 28–42.

[6] Chevalier, J. A., & Mayzlin, D. (2006). The effect of word of mouth on sales: Online book reviews. *Journal of Marketing Research, 43*(3), 345–354.

[7] Godes, D., & Mayzlin, D. (2004). Using online conversations to study word-of-mouth communication. *Marketing Science, 23*(4), 545–560.

[8] Alibaba Cloud Tianchi. (2025). Social e-commerce user purchase behavior dataset. *Alibaba Cloud Tianchi Dataset*. https://tianchi.aliyun.com/dataset/215680/comment