

Extracting Factors through Additional Impression Evaluation Experiment Assessing Both High-rated and Low-rated Reviews Posted at EC Sites

Yuya Yokoyama^{*}, Takaaki Hosoda^{*}, Tokuro Matsuo^{*}

Abstract

The global diffusion Internet has established electronic commerce (EC) sites where anyone can purchase online. In order to avoid a mismatch between user and products, users can write a review on the product they purchased, helping users refer to the review of the commodity and make decisions. Nevertheless, with more users and items flooded on EC sites, the issues of mismatches are becoming conspicuous. In order to solve these issues, the authors conducted impression evaluation experiment to extract the impression of low-rated reviews. However, the previous analysis yielded only three factors due to insufficient experimental materials. Therefore, this paper reports further experiment with appending high-rated reviews. As a result of the analysis, eight factors are obtained under the fifty impression words. It could be concluded that the approach of extracting impression from the statements can be applicable to the review statements of EC sites under the unbiased experimental materials.

Keywords: EC site, Factor analysis, Factor loading, High-rated Review, Impression evaluation experiment, Impression word

1 Introduction

Recently, the widespread diffusion of the Internet has brought about electronic commerce (EC) sites where people all over the world can purchase items online. According to FY2022 E-Commerce Market Survey carried out by Ministry of Economy, Trade and Industry, the Japanese market scale of BtoC-EC, standing for Business to Consumer-Electronic Commerce, reached 22.7 trillion yen and keeps increasing [1]. Among BtoC-EC, the scale of merchandising sector BtoC-EC such as Food, electrical appliances, books, etc. that can be available at EC sites e.g. Rakuten Ichiba [2] and Amazon [3], etc. amounted to 13.9 trillion yen. From these statistics, it can be easily presumed that EC transaction markets will continue to greatly flourish in future.

In order to prevent mismatches between users and products at EC sites, users themselves can write reviews on the product they made purchases of with a five-level rating. Other users can then refer to those reviews and rate them as “helpful” when they are content with their opinions. These EC site functions can help users use the reviews as references and contemplate the purchase of the

^{*} Advanced Institute of Industrial Technology, Tokyo, Japan

commodity. However, this helpful system may fail to function due to increasing number of users and products, thus the mismatches between users and products are becoming remarkable and important social issues. On the other hand, the methodology of introducing questioners to appropriate respondents at Question and Answer (Q&A) sites has been established [4-6]. This method was inaugurated by extracting impressions from the Q&A statements and has the possibility of being applied to other fields or datasets. [7, 8]. Thus, this paper aims to resolve the mismatches among EC sites by deriving impressions from the statements of reviews.

In contrast to high-rated reviews where only the satisfied comments are inclined to be included, low-rated ones tend to contain more concrete information e.g. what the issue is, why the store is not recommendable, how the store must improve, etc. Hence, it could be regarded that extracting the impression from the low-rated reviews could be the key to solving mismatches between users and products. As a first phase of the methodology, impression evaluation experiment using the low-rated review content posted to EC sites was conducted [9]. The analysis result has shown that three factors describing the style or content of statements were obtained. It could be implied that the factors could be affected according to the characteristics of statements.

Nevertheless, the experiment result mentioned above has left room for improvement. By using 50 impression words, nine factors were obtained for the Q&A statements [7, 8]. Meanwhile, based on merely 13 impression words, only three factors were to be extracted for the low-rated reviews posted at EC sites [9]. This result may result from selecting merely low-rated reviews as experimental materials. Hence, additional experiments appending high-rated reviews could remedy the previous analysis result. Therefore, in this paper, an additional experiment with high-rated reviews is conducted to obtain factors of statements of the reviews. The analysis result has shown that eight factors were extracted by using 50 impression words. It could be concluded that the methodology to extract impressions from the Q&A statements can be extended to the review statements of EC sites by using unbiased experimental materials, showing the possibility of generalization to other datasets or fields.

The remainder of this paper consists as follows. Related works are introduced in Section 2. Outline of Rakuten Ichiba data utilized for this work is summarized in Section 3. Previous impression evaluation experiments using low-rated reviews alone is stated in Section 4. An additional experiment appending high-rated reviews is described in Section 5. Considerations towards the analysis results are provided in Section 6. Finally, Section 7 concludes the paper.

2 Related Works

2.1 Researches on EC Sites

Li *et al.* proposed a system using an evaluation expression dictionary to create the assessment criteria for stores and perform automatic scoring [4]. Through its review classification and store evaluation comparison features, their system has shown the ability to clear store evaluations and would be useful for selecting desirable products and stores. With a view to comprehending the purchasing behavior the person who wrote the review took, Yoshida *et al.* associated the classified reviews with the customer ID of the purchasing data and analyzed them by identifying whether the person who posted each review would result in repurchase from the classified review data and purchase data [5].

Their analysis result has revealed the relationship between repurchasing and evaluation by topic. Horie *et al.* proposed a method to extract useful reviews mechanically by clarifying the useful reviews [6]. The interview surveys on 48 subjects who regularly used EC sites have clarified how greatly different product reviews were regarded useful in product selection. It was also shown that the information on reviews thought as useful could be different depending on consumers.

2.2 Methodology to Extract Impression from Q&A Statements

Yokoyama *et al.* have established a method using impression to introduce a questioner to appropriate respondents at Question and Answer (Q&A) sites [7, 8]. Through impression evaluation experiment, nine factors representing the style or content of Japanese Q&A statements were obtained [7]. As shown in Table 1, these nine factors were named as follows: *accuracy*, *displeasure*, *creativity*, *ease*, *persistence*, *ambiguity*, *moving*, *effort*, and *hotness* [7].

Table 1: Nine Factors Extracted from Japanese Q&A Statements [7]

Factors	Impression words						
1st (Accuracy)	Persuasive	Fluent	Important	Appropriate	Wonderful	Refreshing	Skillful
	Fulfilling	Beautiful	Favorable	Favorable	Courteous	Real	Accurate
2nd (Displeasure)	Uncomfortable	Resentful	Thoughtless	Disillusioning			
	Fearful	Amazing	Regrettable	Unjust			
3rd (Creativity)	Creative	Unexpected	Special	Original	Marvelous		
4th (Ease)	Easy	Clear	Difficult				
5th (Persistence)	Minute	Persistent	Long				
6th (Ambiguity)	Ambiguous	Insufficient					
7th (Moving)	Warm-hearted	Impressive					
8th (Effort)	Touching						
9th (Hotness)	Hot	Powerful					

With a view to generalization to other languages, this methodology was applied to English Q&A statements [8]. Similar to the case using Japanese, nine factors were obtained for English materials as well. The nine factors were named: *accuracy*, *evaluation*, *disappointment*, *discomfort*, *novelty*, *potency*, *difficulty*, *politeness*, and *nostalgia* [8].

From these results, both similarities and differences were observed between Japanese and English Q&A statements [8]. In terms of similarity, similar factors were obtained in both languages. Moreover, several major factors were obtained in common. Especially, a factor named “Accuracy” was extracted as the 1st factor in both languages. On the other hand, some differences were also observed [8]. Firstly, factors could be obtained as a different form in different languages. Specifically, a Japanese factor named “Displeasure” was extracted as the 2nd factor. This factor was subdivided into two English factors named “Disappointment” (3rd) and “Discomfort” (4th). Additionally, some factors extracted in one language may fail to be obtained in another language. For example, the 5th and 8th Japanese factors named “Persistence” and “Effort” did not appear in English, whereas the 8th and 9th English factors named “Politeness” and “Nostalgia” were not obtained in Japanese.

2.3 Principle of Methodology

As introduced in Section 2.1, there have been several earlier studies researching the reviews posted at EC sites. To our knowledge, however, there has not been any research using the impressions of the reviews at EC sites. Despite possible fakes or disguises, since ample amount of data will converge the certain quality, it is meaningful to analyze the contents of EC sites. Meanwhile, as explained in Section 2.2, a method using impressions of Q&A sites could also be generalized to the other datasets. Hence, our principle is to apply the methodology using impressions to the reviews of EC sites.

3 Rakuten Ichiba Data

The dataset of EC site used for this analysis is Rakuten Ichiba data, one of the datasets provided by the National Institute of Informatics [10]. The dataset includes product data, item review data, and shop review data during the period dating from 2015 to 2019. During the five-year session, 22,505,858 reviews were recorded. The contents of review data are as follows [11]: reviewer ID, shop name, shop ID, review point (ranging from 1 (worst) to 5 (best)), review content (statement of review), reference number, and review date. Reference number means that of “helpful,” which is given by users other than the reviewer who find the review helpful. This indicates how useful the review was for regarding the purchase of the product. Among all the 22,505,858 reviews, those with reference number no less than 20 were tentatively used for the subsequent analysis. Therefore, the amount of review data with reference number of at least 20 is 20,821.

4 Impression Evaluation Experiment Assessing Low-Rated Reviews of EC Sites

4.1 Aim

EC sites are flooded with various styles or impressions of reviews. All the reviews can possess any number of review points, with some having 5.0, while others being rated as just 1.0. For these reviews, high-rated reviews tend to end up showing their satisfied comments, e.g. “The quality of this item is great and I like it.” Meanwhile, low-rated reviews are apt to contain more specified opinions. For example, what the issue is, why this store is not recommendable, how the store must improve, etc. Hence, in order to avoid more mismatches between users and products, we tried to extract the impressions from low-rated reviews. Therefore, we inaugurated our research by conducting impression evaluation experiment using EC sites [9]. The basic procedures are substantially the same as the case of extracting the impressions of Q&A statements [7, 8].

4.2 Experimental Setups

In selecting experimental materials from 20,821 review contents whose reference numbers were no less than 20, we set the following two criteria. The first standard is to choose review contents with review points 1.0 (worst-rated) [9]. The second one is to select those free from grammatical errors, because applying syntactic analysis e.g. morphological analysis is planned for the subsequent pro-

cedure. As even trivial grammatical errors could prevent the proper result of syntactic analysis, it would be very vital to avoid such errors.

Through these criteria, the experimental materials selected were 20 review contents, 5 reviews each for 4 stores [9]. Taking an example of low-rated reviews posted to one store, their original Japanese statements and their English translations are shown in Tables 2 and 3, respectively. These statements are denoted as LR3-1, LR3-2, LR3-3, LR3-4, and LR3-5.

Table 2: Original Japanese Statements of Low-Rated Review Contents per Store [9]

LR3-1	1週間程度で発送とあったので購入を申し込んだら、「メーカー欠品」とのことで、1週間程度（8/21頃）納品が遅れるとのメール。その納品目途の日になって再度1週間延長、今日になってさらに1週間延長、「キャンセルするかどうか」とのメール…。さすがに3度目の延長メールでキャンセルさせてもらいました。「楽天24」=直営=信用できると思っていたが、商品を確保できないのなら、最初から注文を受けないで欲しい。極めて不誠実。
LR3-2	8月22日に購入した品物の中の1点が「入荷遅れになる」とのことで「全て揃ってから出荷になる。お届け予定日は9月8日～9月11日」とメールがきました。その商品のページを見ると納期が3日～5日と書いてあります。楽天の直営店なのに嘘の納期を書いています。今までは爽快ドラッグを使っていましたが、あまりの発送の遅さに嫌気がさし、こちらに変えました。楽天24に切り替えた最初の方は《あるものは先に発送、ないものは後日発送》してくれて助かってました。楽天の直営店として納期の表記をしっかりと、ないものは「在庫切れ」の表示の徹底をしてもらいたいです。楽天の他のショップの方がよほどきちんとしていると思います。
LR3-3	発送が驚く程遅いです。スタッフの対応も非常識でした。注文から2ヶ月以上待たされましたがその間の連絡等はなく、不安になり何度かご連絡差し上げましたが回答はなく、3ヶ月が経つ頃に在庫なし。との連絡が…。他のショップで頼めば、時間を無駄にすることも不快なることもありませんでした。どういう状況ご連絡いただけるか、せめて返信くらいはしてほしいです。無礼なショップ。
LR3-4	欠品は仕方ないが、とにかく対応が遅い！他の方のレビューをみると、明らかに自分より後に購入した人への連絡は1日ですのに対して、こちらへの連絡は購入の6日後。しかも到着予定の日になってもキャンセルとか失礼にも程がある。対応の悪さに呆れた。
LR3-5	毎回届くのが非常に遅い。だが買い回りの店舗稼働や近所がないものを注文するときには仕方なく利用する。最長で2か月待ったことがある。配送遅れているのだから、その間にセールのお知らせとかクーポンとか配信するな。そんな暇あったら在庫管理しろ。

Table 3: English Translation of Low-Rated Review Contents [9]

LR3-1	I ordered this because it says "shipment in about a week," and then I got an e-mail that the delivery would be delayed by a week (around Aug 21) due to "manufacture shortage." Then on the day delivery was supposed to be, another delay by a week, and today another week delay, and an e-mail saying, "If you want to cancel, go ahead." I had my cancellation on the third delay notification mail. I thought "Rakuten 24" is direct management so would be reliable, but they should not take an order if the item could not be secured. Extremely insincere of them.
LR3-2	I received an e-mail saying that because one of the items I purchased on Aug 22 had a shipping delay, your items would be shipped once the entire items were ready. The estimated delivery date was supposed to be between Sept 8 and 11. The order page of that item says delivery in 3 to 5 days. They show false delivery dates despite Rakuten claiming to be a direct management store. I used to use Soukai Drug, but I got fed up with their slow delivery, so I changed to this store. When I first moved to Rakuten 24, they were helpful enough as to ship immediately if they had the item, otherwise they would ship later. As a Rakuten direct management store, they should ensure the display of delivery date or say "out of stock" if there is none. I think the other Rakuten shops are much better at properly going about it.
LR3-3	Shipment is surprisingly slow. The stuff's responses were also thoughtless. I was kept waiting for over two months since I placed my order, but got no response etc. during that session, and I became so anxious that I contacted them several times but got no response, and then about three months later I got a message saying that that item went out of stock... I would not have wasted my time and become rather annoyed if I had made the order at other shops. At least they should have contacted me or at least replied to me about the present situation. What a rude shop!
LR3-4	It cannot be helped if an item is out of stock, but their action is so slow! According to the reviews from other users, they made contacts in one day to the users who obviously made purchases later than I, but in my case they did so six days after purchase. In addition, they were rude enough to cancel the delivery on the scheduled shipping day. I was disgusted at their terrible response.
LR3-5	Every time shipment is very slow. But I have no choice but to use this place since there were no discount stores in my shopping area and the items were unavailable in my neighborhood. At the longest I waited for two months. As their shipment is very slow, they should not distribute sale ads or coupons around here. If they have so much time, they ought to use it to manage their inventory.

4.3 Experimental Procedure

Impression evaluation experiment rating low-rated reviews was conducted with the cooperation of five subjects [9]. The subjects were asked to rate the 20 experimental materials using 50 impression words. The definition of impression word is the word that seems effective in expressing the style or content of the statements. The 50 Japanese impression words used are summarized in Table 4. There are several two English words listed like “Amazing/Shocking” due to conveying clearer translation of the original Japanese word. These impression words were used to rate the experimental materials, in a similar fashion as evaluating Q&A statements [9].

Table 4: 50 Japanese Impression Words Used for Assessment [7, 9]

Accurate	Amazing/Shocking	Ambiguous	Appropriate	Beautiful	Clear	Complicating
Courteous	Creative	Difficult	Disillusioning	Dull	Easy	Exaggerating
Faltering	Favorable	Fearful	Firm	Fluent	Fulfilling	
Fun	Hot/Intense	Important	Impressive	Inevitable	Insufficient	
Long	Marvelous	Minute/Detailed	Nostalgic	Original/Novel	Persistent	
Persuasive	Powerful	Real	Refreshing	Regretting	Resentful	
Sharp	Simple	Skillful	Special	Suspicious	Thoughtless	
Touching	Uncomfortable	Unexpected	Unjust	Warm-hearted	Wonderful	

4.4 Experimental Result

The application of factor analysis with Varimax rotation to the experimental result failed to yield any factors under 50 impression words. Thus, 50 impression words were reduced to 13 through trial and error. As a result of factor analysis performed under those 13 impression words, factor loadings with Varimax rotation are shown in Table 5. Impression words with absolute values of factor loadings over 0.5 are shaded to interpret the factors. From this result, three characteristic factors were obtained. These factors were named *Object*, *Creativity*, and *Undeserved* [9].

Table 5: Factor Loadings of Three Factors [9]

Impression Word	Factor 1	Factor 2	Factor 3
Disillusioning	0.948	-0.096	0.018
Appropriate	0.883	0.100	0.109
Regrettable	0.882	-0.239	-0.319
Resentful	0.839	-0.039	0.330
Uncomfortable	0.741	-0.117	0.362
Amazing	0.667	0.178	0.638
Unexpected	0.635	-0.644	0.085
Accurate	0.583	0.284	0.421
Real	0.574	0.159	0.213
Fulfilling	0.102	0.980	0.153
Creative	0.095	0.968	0.108
Important	-0.094	0.918	0.141
Unjust	0.097	0.164	0.864

5 Further Impression Evaluation Experiment with Additional Materials and Subjects

5.1 Requirement

As summarized in Section 4, no results were obtained under 50 impression words. The biggest cause was the characteristics of experimental materials. One of the standards of choosing Q&A statements was to eliminate the statements with abusive words, slander, or statements against public order and standards of decency [8, 9]. On the other hand, low-rated reviews were potentially written in unpleasant, angry, or critical styles. Therefore, impression words corresponding to the 1st factor shown in Table 5 were assessed with relatively high scores by subjects. Meanwhile, biased impression might have led the five subjects to rate the entirely same combinations of scores “1” or “2” for the impression words usually used for good evaluations e.g. “wonderful,” “persuasive,” “beautiful,” etc. This result produces many overlapping combinations for several impression words, diminishing the effectiveness of impression words for factor analysis.

It would also be notable to focus on the relationships between sample size and contents of experimental materials. Along with the Japanese Q&A statements (2,460 sample size; 60 statements and 41 subjects) [7], because of the sufficient sample size, it would be interesting to compare the English Q&A statements [8] (120 sample size; 30 statements and 4 subjects) with low-rated reviews at EC sites [9] (100 sample size; 20 statements and 5 subjects). In the absence of great differences, 9 factors were obtained for English Q&A statements, whereas no factor was extracted for low-rated reviews at EC sites under 50 impression words. This result could be directly attributed to the biased contents of their experimental materials, as pointed out earlier.

Therefore, it is necessary to conduct further experiment with additional subjects and contents. For the additional contents, high-rated review contents are appended as experimental materials. Using both high-rated and low-rated reviews would straighten out the biased contents besides deficient sample size.

5.2 Additional Experimental Procedure

The experiment was conducted in the same way as stated in Section 4.3. The different elements from the previous experiment were additional materials and subjects. Firstly, as for additional materials, similar to the low-rated reviews, the review contents with review points 5.0 (best-rated) were chosen as high-rated reviews. Another criterion of choosing with minimum grammatical errors, as explained in Section 4.2, was also taken into consideration. Similarly, 20 high-rated reviews (5 reviews each for 4 stores) are appended as the experimental materials. Taking an example of high-rated reviews on one store, their original Japanese statements and their English translations are shown in Tables 7 and 8, respectively. These statements are denoted as HR3-1, HR3-2, HR3-3, HR3-4, and HR3-5.

Table 6: Original Japanese Statements of High-Rated Review Contents per Store

HR3-1	沢山の商品が揃っているのので、購入するまでにかなり時間が掛かります。でも、いろいろな商品があるので一度に購入出来るのは嬉しいです。これから利用させていただきます。
HR3-2	4本足の杖と、洗濯洗剤を頂きました。スーパーセール中に注文、最終日には到着。洗剤は小箱に。杖はプチプチでくるまれ、茶の包装紙で、軽々と到着。ショップ様、助かりました。どうもありがとうございました。
HR3-3	5品注文。全てお届け目安を1〜3日以内で出荷予定の品に揃えて注文したので7日注文→11日着はまあ予定通りかなと。全く問題なく、丁寧に梱包され届いています。ありがとうございました。
HR3-4	当然といえば当然ですが、品揃え、対応の良さと どれをとっても文句のつけようがありません。クーポンの配布も頻繁ですし、ポイントの事も併せて考えると 結局安くつくのでまとめ買いしています。他店や公式サイトなどで売り切れの商品も探して見ると 在庫がある場合も多いのでチェックは欠かせません！ 頼りにしています。
HR3-5	いつもお世話になっています。10月購入分のレビューですが 今回の発送は早かったですね、注文より3日目に受け取っています。何時も急ぎの商品は有りませんが此くらいの日数で受け取れると嬉しいです。有難うございました。

Table 7: English Translation of High-Rated Review Contents

HR3-1	They have so many items that it takes me a lot of time to purchase. However, as there are various items, I'm glad I can buy a lot at a time. I will be using this store from now on.
HR3-2	I bought a walking stick with four legs and laundry detergent. I bought them during the super sale and got them on the final day. The laundry detergent was put in a small box. The walking stick covered with bubble wrap and brown paper was easily delivered. Dear shop, thanks a lot.
HR3-3	I ordered five items. Because I set all the estimated delivery within 1 to 3 days, my order on the 7th then its arrival on the 11th was almost on time. They arrived without problem, and wrapped neatly. Thank you.
HR3-4	It is of course as might be expected, but I have no complaints about the assortment of goods and nice services. Coupons are frequently offered, and I end up buying a lot at a time for cheaper, considering the points. I never fail to check an item beforehand as there is often its inventory online saying if it is out of stock at other stores or official sites! I rely on this store.
HR3-5	Thank you for your support always. This is my review for purchases in October, and this time shipment was immediate. I got the item on the third day from the order. Although I do not usually have urgent orders, I am glad I can get items as fast as this number of days. Thank you.

Meanwhile, as for subjects of the additional experiment, three out of the five subjects who joined the previous experiment explained in Section 4 rejoined the additional experiment. These three subjects were asked to assess 20 high-rated reviews in a similar procedure as low-rated ones. Moreover, an additional three subjects newly joined the experiment. For these new three subjects, they were asked to evaluate both 20 low-rated reviews and 20 high-rated ones. Overall, the additional experiment was carried out with the cooperation of six subjects.

5.3 Experimental Results

Similar to the previous experiment, factor analysis was conducted in the same way as stated in Section 4.3. In executing factor analysis, the criterion of determining the numbers of factors is either “their eigenvalue over 1.0” or “cumulative contribution ratio at least 80%”. Of these two criteria, by adopting the former standard the number of factors was set to 8. Eigenvalues, contribution ratio, and cumulative contribution ratio are summarized in Table 8. Factor loadings with Varimax rotation are shown in Table 9. As explained in Section 4.4, the factor loadings with absolute values over 0.5 are shaded to underscore them since they indicate significant interpretations. Their minute interpretations will be provided in the subsequent section.

Table 8: Eigenvalue, Contribution Ratio, and Cumulative Contribution Ratio for 8 Factors

Factor	Eigenvalues	Contribution Ratio [%]	Cumulative Contribution Ratio [%]
1	13.5	17.4	17.4
2	9.47	16.7	34.2
3	5.31	16.3	50.5
4	3.11	6.2	56.7
5	1.99	5.2	61.9
6	1.45	2.6	64.6
7	1.25	2.3	66.9
8	1.11	2.2	69.0

6 Considerations

6.1 Interpretations of Factors

As summarized in Section 5.3, eight factors expressing the style or content of reviews posted at EC sites were obtained. According to the impression words with absolute values of factor loadings are

Table 9: Factor Loadings of Eight Factors (Additional Experiment)

Impression Word	Factor 1	Factor 2	Factor 3	Factor 4	Factor 5	Factor 6	Factor 7	Factor 8
Amazing/Shocking	0.916	0.026	-0.059	-0.165	0.009	0.057	0.147	0.064
Uncomfortable	0.901	-0.014	-0.228	-0.063	0.059	0.092	-0.024	0.078
Thoughtless	0.865	-0.016	-0.023	-0.151	-0.030	0.037	0.158	0.119
Resentful	0.838	-0.003	-0.129	0.164	0.102	0.087	0.085	0.188
Regretting	0.815	-0.157	-0.237	0.104	0.321	0.007	-0.176	-0.088
Insufficient	0.759	0.254	0.186	0.040	-0.214	-0.122	-0.023	-0.187
Suspicious	0.727	0.261	0.162	0.203	-0.254	0.014	-0.065	-0.160
Disillusioning	0.721	0.053	-0.083	0.509	0.158	0.096	-0.028	0.134
Sharp	0.648	0.069	-0.237	0.217	0.155	0.065	0.100	0.081
Inevitable	0.629	0.000	-0.102	0.581	0.275	0.092	-0.042	0.060
Exaggerating	0.033	0.834	0.087	-0.016	-0.032	-0.051	-0.033	-0.131
Original/Novel	0.018	0.801	0.249	-0.093	0.097	-0.033	0.000	-0.071
Special	0.008	0.760	0.285	-0.119	0.110	0.063	0.083	0.177
Firm	-0.104	0.739	0.360	-0.043	0.150	0.034	0.085	0.002
Ambiguous	0.112	0.690	0.153	0.026	0.072	0.023	-0.021	-0.305
Fearful	0.218	0.687	-0.184	-0.137	-0.055	0.193	-0.016	0.237
Creative	0.060	0.654	0.167	-0.080	0.159	0.161	0.067	0.222
Faltering	0.098	0.608	-0.029	-0.080	-0.098	0.396	-0.188	0.133
Dull	-0.113	0.593	0.329	0.018	-0.026	-0.014	0.101	-0.012
Hot/Intense	-0.054	0.574	0.148	0.087	0.339	0.201	0.064	0.124
Minute/Detailed	0.144	0.571	0.159	0.242	0.232	-0.024	-0.009	0.045
Fulfilling	-0.159	0.562	0.524	-0.106	0.247	-0.022	0.205	0.087
Special	0.457	0.559	0.235	-0.030	-0.252	-0.008	-0.021	-0.092
Persistent	0.165	0.544	-0.289	0.259	0.371	0.321	-0.156	0.066
Beautiful	-0.031	0.077	0.922	0.086	-0.002	-0.108	-0.152	-0.002
Warm-hearted	-0.151	0.103	0.870	0.080	-0.113	-0.055	-0.192	0.030
Easy	0.241	0.095	0.821	0.217	-0.121	-0.041	0.107	0.084
Favorable	-0.400	0.194	0.757	-0.022	0.136	0.066	0.001	-0.129
Wonderful	-0.283	0.247	0.748	-0.054	0.173	0.189	0.101	-0.122
Impressive	-0.222	0.447	0.709	-0.085	0.233	0.033	-0.060	-0.051
Fun	-0.254	0.483	0.674	-0.078	0.208	0.046	-0.005	-0.083
Refreshing	0.101	0.204	0.598	0.190	-0.103	0.102	0.206	0.122
Important	0.193	0.285	0.597	0.061	0.092	-0.030	0.291	0.172
Courteous	-0.331	0.322	0.576	0.104	0.216	-0.012	0.033	-0.069
Accurate	0.275	-0.062	0.570	0.439	0.046	0.044	0.458	-0.011
Clear	0.342	0.039	0.509	0.495	0.134	0.050	0.370	-0.019
Simple	0.320	0.026	0.280	0.582	-0.025	-0.054	-0.036	0.003
Appropriate	0.462	0.026	0.439	0.514	0.139	-0.031	0.271	-0.117
Fluent	-0.024	0.286	-0.009	0.369	0.756	-0.040	-0.097	0.123
Unexpected	0.073	0.372	0.115	-0.150	0.569	-0.029	0.004	-0.055
Persuasive	0.201	-0.005	0.256	-0.029	0.519	-0.092	0.233	0.049
Touching	0.210	0.376	0.047	0.019	-0.059	0.777	0.060	-0.049
Powerful	0.445	0.131	0.166	0.364	0.195	-0.047	0.000	0.604

over 0.5, interpretations against those eight factors were as follows:

- 1st factor is composed of negative impressions of the sentences, such as “amazing/shocking,” “uncomfortable,” “thoughtless,” etc. as well as assessing unavoidable situations such as “sharp” and “inevitable.” Regarding both these axes, this factor is named “critical attitude.”
- 2nd factor consists of expressing unique viewpoints such as “exaggerating,” “original/novel,” “special,” etc. and perplexing words such as “ambiguous,” “fearful,” “faltering,” etc. Considering these elements, this factor is named “confusion on unknown.”
- 3rd factor is dominated by positive impression of the sentences, e.g. “beautiful,” “warm-hearted,” “easy,” etc. Hence, this factor is named “impressive experience.”
- 4th factor is made up of honesty evaluation of contents such as “simple,” “inevitable,” “disillusioning,” and “appropriate.” Therefore, this factor is named “realistic assessment.”
- 5th factor consists of words expressing how convenient items are, such as “fluent,” “unexpected,” and “persuasive.” Thus, this factor is named as “convenience.”
- 6th factor is composed of a word representing difficulty or effort such as “touching.” This factor is named as “sympathy.”
- 7th factor happens to have no impression words whose absolute value of factor loadings are over 0.5. Therefore, no interpretation is given to this factor.
- 8th factor consists of the pressure from sentences such as “powerful.” Hence, this factor is named as such as “impact.”

Compared with the previous result under 13 impression words, the further experiments with additional materials and subjects were effective. Specifically, appending high-rated statements has played a vital role in balancing the biased result where only three factors were obtained.

6.2 Comparison of Weight of Factors

It would be noteworthy to focus on the different tendencies of contribution ratio (CR) of factors among EC sites and Japanese Q&A statements. Factor names, eigenvalue, CR, and cumulative contribution ratio (CCR) are summarized in Table 10. Eigenvalue, CR, and CCR are shown in the columns entitled “Eigen,” “CR [%],” and “CCR [%],” respectively. For the cases of Japanese Q&A sites, two factors mainly account for representing the impression of Q&A statements. Meanwhile, for the case of EC sites, the major three factors dominate the expression of the impression of EC sites. In terms of CCR, the main two factors for Q&A site accounted for 25.9% [7], while CCR of the major three factors for EC site reached as high as 50.4%. These comparisons have shed light on the different tendencies of main factors among EC sites and Q&A sites.

From another perspective, the characteristics of their main three factors are substantially similar despite their different orders. Specifically, the 1st factor “critical attitude” for EC and the 2nd factor “displeasure” are the same. Likewise, the 2nd factor “confusion on unknown” for EC resembles the 3rd factor “creativity,” whereas the 3rd factor “impressive experience” for EC corresponds to the 1st factor “Accuracy.” The possible reason for the different orders of three major factors is that although high-rated reviews were appended to experimental materials, the effect of impressions from low-rated reviews in negative evaluation must have lingered on. Nevertheless, it would not be too much to say that this was a definite improvement over the previous experimental result rating only low-rated reviews and obtaining just three factors.

Table 10: Comparison of Eigenvalue (Eigen), Contribution Ratio (CR), and Cumulative Contribution Ratio (CCR) between EC Sites and Japanese Q&A Sites [7]

Factor	EC Site				Japanese Q&A Site [7]			
	Name	Eigen	CR [%]	CCR [%]	Name	Eigen	CR [%]	CCR [%]
1	Critical attitude	13.5	17.4	17.4	Accuracy	11.1	14.5	14.5
2	Confusion on unknown	9.47	16.7	34.1	Displeasure	6.78	11.4	25.9
3	Impressive experience	5.31	16.3	50.4	Creativity	2.69	6.1	32.0
4	Realistic assessment	3.11	6.2	56.6	Ease	2.24	3.7	35.7
5	Convenience	1.99	5.2	61.8	Persistence	1.56	3.6	39.3
6	Sympathy	1.45	2.6	64.4	Ambiguity	1.50	3.5	42.8
7	-	1.25	2.3	66.7	Moving	1.43	3.1	45.9
8	Impact	1.11	2.2	68.9	Effort	1.18	2.1	48.0
9					Hotness	1.10	2.0	50.0

7 Conclusion

In order to reduce the mismatches between users and items at EC sites, in this paper we conducted an additional impression evaluation experiment assessing both high-rated and low-rated reviews. Factor analysis was then applied to the experimental result to extract eight factors. It has been shown that the methodology to extract impression from the Q&A statements was able to be extended to the review statements of EC sites. It could also be implied that the generalization to other datasets or fields would be possible with selection of biased experimental results.

For future work, in order to establish the methodology to clear mismatches between users and items at EC sites, factor scores of review statements will be estimated using feature values of statements, in a similar fashion as the case of the Q&A sites. It would be vital to inspect if using reviews can reveal the characteristics of products under evaluation. Whether this methodology can be extensive to other field or dataset such as hotel reviews must be conducted. Besides this methodology, a different approach such as large language model (LLM) needs to be considered.

Acknowledgement

In this paper, we used “Rakuten Dataset” [11] provided by Rakuten Group, Inc. via IDR Dataset Service of National Institute of Informatics.

References

- [1] Results of FY2022 E-Commerce Market Survey Compiled (URL, in Japanese), - Ministry of

- Economy, Trade and Industry, <https://www.meti.go.jp/press/2023/08/20230831002/20230831002.html>, 2025-01-10.
- [2] Rakuten Ichiba (URL, in Japanese), <https://www.rakuten.co.jp>, 2025-01-10.
- [3] Amazon (URL), <https://www.amazon.com>, 2025-01-10.
- [4] D. Li, H. Nishikawa, M. Ueda, and S. Nakajima, “A Store Evaluation System using Automatic Scoring of Retail Stores Based on Product Review Analysis”, Proc. of the 2023 IEEE International Conference on Big Data (BigData), pp.2124-2130, 2023.
- [5] M Yoshida, T Tabata and T Hosoda, “A Study on Relationship Between Consumer Review Behavior and Purchasing in EC site,” Proc. of the 9th International Congress on Advanced Applied Informatics (AAI 2020), pp.791-796, 2020.
- [6] K. Horie, M. Onoda and T. Kaneko, “A proposal of review extraction method considering the knowledge amount of consumers and the purchase stage,” Total Quality Science, Vol.4, No.2, pp.83-91. 2018.
- [7] Y. Yokoyama, T. Hochin, H. Nomiya and T. Satoh, “Obtaining Factors Describing Impression of Questions and Answers and Estimation of their Scores from Feature Values of Statements,” Software and Network Engineering, Springer, Volume 413, pp.1-13, 2013.
- [8] Y. Yokoyama, T. Hochin and H. Nomiya, “Estimation of Factor Scores from Feature Values of English Question and Answer Statements,” Proc. of IEEE/ACIS 15th International Conference on Computer and Information Science (ICIS 2016), pp.741-746, 2016.
- [9] Y. Yokoyama, T. Hosoda and T. Matsuo, “Extracting Impression from the Low-Rated Statements Posted at EC Sites,” Proc. of the 16th International Congress on Advanced Applied Informatics (AAI 2024), pp.9-14, 2024.
- [10] Rakuten Group, Inc. (2020): Rakuten Ichiba data. Informatics Research Data Repository, National Institute of Informatics. (dataset). (URL), <https://doi.org/10.32130/idr.2.1>, 2025-01-10.
- [11] Rakuten Data Release (URL), https://rit.rakuten.com/data_release/, 2025-01-10.