

Space Norms for Constructing Quality Reviews on Online Consumer Review Sites

Online Appendices

Supplementary analyses, methodological details, and experimental materials etc. are presented.

Online Appendix A. Literature Review on Review Writing Behavior

Context	Antecedents	Outcomes	Reference
IMDB and Douban online movie rating sites	Motivation factor: Extreme opinions (e.g., like vs. dislike); Cultural influences (China, U.S., and Singapore)	Self-reported frequency of rating movies	Koh et al. 2010
Online product reviews from Amazon	Number of reviews and distribution of ratings; Self-reported degree of passion about one's product review; Extreme opinions - very satisfied vs. very disgruntled	Self-reported willingness to voluntarily post reviews	Hu et al. 2006
Laboratory experiment on factious and real brands	Need for uniqueness	Self-reported willingness to recommend products	Cheema and Kaikati 2010
An online survey among users of a German web-based opinion-platform	Self-reported motivation factors: extroversion; positive self-enhancement; economic incentives; concern for other consumers; and social benefits	Number of comments written by the survey respondents	Hennig-Thurau et al. 2004
Yahoo! Movies and BoxOfficeMojo	Movie characteristics (e.g., marketing budget)	User population's propensity to rate a movie online	Dellarocas and Narayan 2006
Yahoo! Movies and BoxOfficeMojo	Movie's market popularity; Volume of previous reviews of a movie	User population's propensity to review movie (i.e., number of people who reviewed movie /number of people who watched movie)	Dellarocas et al. 2010
Airbnb reviews	Financial incentives	Probability that a review is submitted	Fradkin et al. 2018
A large online retailer in China that sells children's apparel	Financial incentives	Percentage of users who submit reviews;	Burtch et al. 2017

		Review quality (i.e., review length)	
The e-commerce website of a large American retailer	Financial incentives	Total number of reviews; Number of new reviewers; Review quality (i.e., review length)	Khern-am-nuai et al. 2018
Epionions.com	Reviewer popularity (i.e., number of incoming ties)	Number of ratings submitted; Whether any reviews were written	Goes et al. 2014
Book reviews from Amazon and BarnesandNoble.com	Books type (e.g., popular but less-reviewed); Reviewers' reputation	Number of reviews posted	Shen et al. 2015
MovieLens: an online recommender system	Normative message showing the median user's total number of movie ratings	Number of monthly movie ratings submitted by users	Chen, Harper, Konstan, & Li 2010
Google Answers	Price offered to award answers; Answerer's reputation	Length of the answers; The overall quality of the answers rated by human coders	Chen, Ho, and Kim 2010
Dianping.com	Virtual rewards; Number of "likes" votes Number of followees; Number of followers	Number of newly written reviews in a certain period	Dong et al. 2020
Shoppers of mobile fashion/lifestyle retailers in Mumbai	Customer engagement; Trust in the retailer; Customer satisfaction with the retailer	Self-reported intention to write online reviews	Thakur 2018
Online Reviews for restaurants in Delhi and Gurugram	Ego involvement; Perceived behavioral control; Taking vengeance	Self-reported intention to write online reviews	Dixit et al. 2019
TipAdvisor users	Online expectation confirmation; Offline expectation conformation; Satisfaction	Self-reported whether a reviewer had written any positive, negative, or no review	Nam et al. 2020

Note. We cannot exhaustively list all findings in the above studies. We highlight the findings that directly inform our analyses, and to discuss the gap in the existing literature that we seek to fill.

References


Burtch, G., Hong, Y., Bapna, R., and Griskevicius, V. 2017. "Stimulating Online Reviews by

- Combining Financial Incentives and Social Norms,” *Management Science* (64:5), 2065-2082.
- Cheema, A. and Kaikati, A. M. 2010. “The Effect of Need for Uniqueness on Word of Mouth,” *Journal of Marketing Research* (47:3), 553-563.
- Chen, Y., Harper, F. M., Konstan, J., and Li, S. X. 2010. “Social Comparisons and Contributions to Online Communities: A Field Experiment on MovieLens,” *American Economic Review* (100:4), 1358-1398.
- Chen, Y., Ho, T.-H., and Kim, Y.-M. 2010. “Knowledge Market Design: A Field Experiment at Google Answers,” *Journal of Public Economic Theory* (12:4), 641-664.
- Dellarocas, C. and Narayan, R. 2006. “A Statistical Measure of a Populations Propensity to Engage in Post-Purchase Online Word-of-Mouth,” *Statistical Science* (21:2), 277-285.
- Dellarocas, C., Gao, G., and Narayan, R. 2010. “Are Consumers More Likely to Contribute Online Reviews for Hit or Niche Products?” *Journal of Management Information Systems* (27:2), 127-158.
- Dixit, S., Badgaiyan, A. J., and Khare, A. 2017. “An Integrated Model for Predicting Consumer's Intention to Write Online Reviews,” *Journal of Retailing and Consumer Services* (46:January), 112-120.
- Dong, L., Huang, L., Hou, J. and Liu, Y. 2020. “Continuous Content Contribution in Virtual Community: The Role of Status-Standing on Motivational Mechanisms,” *Decision Support Systems* (132:May), 1-12.
- Fradkin, A., Grewal, E., Holtz, D., and Pearson, M. 2018. “Bias and Reciprocity in Online Reviews: Evidence from Field Experiments on AirBNB,” *working paper*, accessed at https://andreyfradkin.com/assets/reviews_paper.pdf
- Goes, P. B., Lin, M., and Yeung, C. A. 2014. “Popularity Effect” in User Generated Content: Evidence from Online Product Reviews,’ *Information Systems Research* (25:2), 222-238.
- Hennig-Thurau, T., Gwinner, K. P., Walsh, G. and Gremler, D. D. 2004. “Electronic Word-of-Mouth via Consumer-Opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet?” *Journal of Interactive Marketing* (18:1), 38-52.
- Hu, N., Pavlou, P. A., and Zhang J. 2006. “Can Online Reviews Reveal a Product’s True Quality?: Empirical Findings and Analytical Modeling of Online Word-of-Mouth Communication,” In *Proceedings of the 7th ACM Conference on Electronic Commerce* (ACM, New York), 324-330.
- Khern-am-nuai, W., Kannan, K., Ghasemkhani, H. 2018. “Extrinsic Versus Intrinsic Rewards for Contributing Reviews in an Online Platform,” *Information Systems Research* (29:4), 871-892.
- Koh, N. S., Hu, N., and Clemons, E. K. 2010. “Do Online Reviews Reflect a Product’s True Perceived Quality? An Investigation of Online Movie Reviews Across Cultures,” *Electronic Commerce Research and Applications* (9:5), 374-385
- Nam, K., Baker, J., Ahmad, N., and Goo, J. 2020. “Determinants of Writing Positive and Negative Electronic Word-of-mouth: Empirical Evidence for Two Types of Expectation Confirmation,” *Decision Support Systems* (129:October), 113168.
- Shen, W., Hu, Y. J., and Ulmer, J. R. 2015. “Competing for Attention: An Empirical Study of Online Reviewers Strategic Behaviors,” *MIS Quarterly* (39:3), 683-696.
- Thakur, R. 2018. “Customer Engagement and Online Reviews. *Journal of Retailing and Consumer Services*, (41), 48-59.

Online Appendix B. Screenshots of the Experimental Site

Calculator, Low-Norm Condition


Today's Deals
Free Shipping on orders over \$49
Cart



Texas Instruments TI-36X Pro Engineering/Scientific Calculator
Price: \$19.78 & Free Shipping on orders over \$49.
In Stock. Gift-wrap available.
Size: Handheld

Product Description
 The Texas Instruments TI-36X Pro 4-line Scientific Calculator is one of TI's most advanced scientific calculators. It features a large, four-line display and higher-level math and science functionality that is great for advanced algebra, trigonometry, computer science, calculus, engineering and statistics courses. Its Multiview™ display shows multiple calculations at the same time on screen. You can scroll through previous entries, see the equation you input, and modify data to derive new answers. This solar-powered calculator can perform 1- and 2-variable statistics, including linear regression and trendline analysis. It can perform vectors and matrices using a vector and matrix entry window. It can also perform fractional equations and calculate angle units (radians, grads, degrees) when performing trigonometric functions.

The TI-36X Pro includes 8 physical constants and 10 English/metric conversions functions. This calculator combines features for statistical analysis, Boolean logic operations, and unit conversions in one sleek package. Ideal for students at every level, from junior high school algebra to college calculus, the Texas Instruments TI-36X packs a lot of punch in a small, inexpensive unit. It is ideal for computer science and engineering courses in which graphing technology may not be permitted. The calculator comes with a Slide-On cover. Free instruction manuals and instructional videos are available online.



Some users reported that the TI-36X's LCD screen is a little difficult to read when the unit is lying on a desk. The keyboard layout for common functions like adding and subtracting is a little unusual, and the silvered keys are difficult to read. Some users found there to be a delay when running complex calculations. Users in need of graphing functions will have to look elsewhere. Overall, however, for a calculator in this price range, the TI-36X provides enough power for most high-level math and science needs.

Reviews

By **Krystal** on September 30, 2016 from United States
 The calculator is one of the models on the approved calculator list for my FE exam.

By **Jill L** on September 23, 2016 from United States
 Got me through my engineering classes just fine.

By **Bruno** on September 22, 2016 from United States
 Good calculator and does everything I need it to do. It does not have graphing functions, tho.

By **Farhan Cleric** on September 4, 2016 from United States
 Use this as a college student studying engineering. It is an efficient scientific calculator with a lot of features.

Add a Review

May 18, 2018 United States

Posted publicly as *

Write your review *

Some users reported that the TI-36X's LCD screen is a little difficult to read when the unit is lying on a desk. The keyboard layout for common functions like adding and subtracting is a little unusual, and the silvered keys are difficult to read. Some users found there to be a delay when running complex calculations. Users in need of graphing functions will have to look elsewhere. Overall, however, for a calculator in this price range, the TI-36X provides enough power for most high-level math and science needs.

Reviews

By **Krystal** on September 30, 2016 from United States
 The calculator is one of the models on the approved calculator list for my FE exam.

By **Jill L** on September 23, 2016 from United States
 Got me through my engineering classes just fine.

By **Bruno** on September 22, 2016 from United States
 Good calculator and does everything I need it to do. It does not have graphing functions, tho.

By **Farhan Cleric** on September 4, 2016 from United States
 Use this as a college student studying engineering. It is an efficient scientific calculator with a lot of features.

Today's Deals

Submit a Review

Calculator, High-Norm Condition

Today's Deals Free Shipping on orders over \$49 Cart



Texas Instruments TI-36X Pro Engineering/Scientific Calculator

Price: **\$19.78** & Free Shipping on orders over \$49.

In Stock. Gift-wrap available.

Size: **Handheld**

Product Description

The Texas Instruments TI-36X Pro 4-line Scientific Calculator is one of TI's most advanced scientific calculators. It features a large, four-line display and higher-level math and science functionality that is great for advanced algebra, trigonometry, computer science, calculus, engineering and statistics courses. Its Multiline™ display shows multiple calculations at the same time on screen. You can scroll through previous entries, see the equation you input, and modify data to derive new answers. This solar-powered calculator can perform 1- and 2-variable statistics, including linear regression and trendline analysis. It can perform vectors and matrices using a vector and matrix entry window. It can also perform fractional equations and calculate angle units (radians, grads, degrees) when performing trigonometric functions.

The TI-36X Pro includes 8 physical constants and 10 English/metric conversions functions. This calculator combines features for statistical analysis, boolean logic operations, and unit conversions in one sleek package. Ideal for students at every level, from junior high school algebra to college calculus, the Texas Instruments TI-36X packs a lot of punch in a small, inexpensive unit. It is legal for use in engineering courses in which graphing technology may not be permitted. The calculator comes with a Slide-On cover. Free PDF manuals are available online.



Some users reported that the TI-36X's LCD screen is a little difficult to read when the unit is lying on a desk. The keyboard layout for common functions like keys are difficult to read. Some users found there to be a delay when running complex calculations. Users in need of graphing functions will have to look elsewhere. In other words, the TI-36X provides enough power for most high-level math and science needs.

Some users reported that the TI-36X's LCD screen is a little difficult to read when the unit is lying on a desk. The keyboard layout for common functions like keys are difficult to read. Some users found there to be a delay when running complex calculations. Users in need of graphing functions will have to look elsewhere. In other words, the TI-36X provides enough power for most high-level math and science needs.

Reviews

By **Krystal** on September 30, 2016 from United States
 This non-graphic calculator is one of the models on the approved calculator list for my FE exam. I think this calculator is a good choice over previous ones even beyond the last 4. Also it has many functions and is easy to use. Another reason for choosing this model over the others is that it is solar powered. On the other hand, its keys are a little difficult for me to read. Lots of glare on the silver operation keys.

By **Jill L** on September 23, 2016 from United States
 I got this scientific calculator as a gift and I quite like it! The TI-36X Pro is a great calculator for everyday and school use. One feature that my friends get frustrated with their 1 line displays and I realize this is actually quite useful. Another good thing is the TI-36X Pro supports a lot of functions like generate tables, do matrix calculations, convert units, and etc. Overall, it makes a great scientific calculator to have around! The only downside is because it's not a computer :)

By **Bruno** on September 23, 2016 from United States
 Got this scientific calculator to use with my college math and chemistry courses. We are not permitted to use graphing calculators so this is the best choice. It displays previous entries just like a typical graphing calculator. It is solar powered so you do not have to worry about changing batteries, which was a concern with a similar Casio model but went with the TI in the end due to it's ease of use and constants within the menu.
 Now let's talk about the negative side. To me, there are two major annoyances. First, this TI has unusual keyboard layout for common functions like addition and etc. are silver-colored with no highlighting. I continually make mistakes by pressing the wrong operation key due to the peculiarity of the layout. But still, I have to say this is a good calculator which has a lot of functions for the price.

By **Ferthan Cleric** on September 4, 2016 from United States
 This TI is user friendly in terms of solving complex numbers and matrices. Simply having a multi-line display and being able to check your work is a huge benefit versus the one line display I was used to. It has easily accessible physical constants and unit converter. I am impressed with the many functions it includes, such as performing derivatives, equation solving, data tables, and some memory devoted to 8 different variables. For \$20 it has much of the capability of the much more expensive graphing calculators. The calculator also has a nice thin cover case that clips on to the back when using the calculator.
 I really have no complaints about it. There is a minor display problem (as mentioned in the description) which may concern some, but I regard it as trivial.

Reviews

By **Krystal** on September 30, 2016 from United States
 This non-graphic calculator is one of the models on the approved calculator list for my FE exam. I think this calculator is a good choice over previous ones even beyond the last 4. Also it has many functions and is easy to use. Another reason for choosing this model over the others is that it is solar powered. On the other hand, its keys are a little difficult for me to read. Lots of glare on the silver operation keys.

By **Jill L** on September 23, 2016 from United States
 I got this scientific calculator as a gift and I quite like it! The TI-36X Pro is a great calculator for everyday and school use. One feature that my friends get frustrated with their 1 line displays and I realize this is actually quite useful. Another good thing is the TI-36X Pro supports a lot of functions like generate tables, do matrix calculations, convert units, and etc. Overall, it makes a great scientific calculator to have around! The only downside is because it's not a computer :)

Add a Review

May 18, 2018 United States

Posted publicly as *

Write your review *

Today's Deals

Submit a Review

Headphone, Low-Norm Condition

Today's Deals Free Shipping on orders over \$49 



BOHM Bluetooth Wireless Noise Cancelling Headphones with In-line Microphone - Black/Silver

Price: ~~\$169.99~~
\$84.99 & Free Shipping
 In Stock. Gift-wrap available.
 Color: Black, Silver

Product Description

BOHM Wireless Bluetooth headphones use the active noise cancellation technology and are comparable to Bose and Sennheiser in this area. With a simple flick of a switch, the outside world simply disappears, replaced by a vacuum of your favorite music or perhaps just precious silence, if you prefer. The aluminum housing and luxurious leather finish of the ear cups are comfortable while looking like they belong on headphones that cost 2-3 times as much. The simple-to-use controls are located on the right earpiece, including the USB charge point, with the noise cancelling indicator on the left.

This entirely wireless setup and advanced Bluetooth 4.0 enables you to link up to any music source that is in range. Whether you use a PC/Mac at home, your tablet outside, or your phone when you are out and about, BOHM wireless headphones can keep up with the pace and switch between your personal devices. The total range is 33 feet (10 meters), so you will not feel tied down to a small location while you enjoy your superior connection. They also have a cable if you don't wish to use them wirelessly, or if your battery dies. Another handy feature is that you can switch between music and phone calls, without needing to unplug, or even turn the volume down. Once your call ends, your music will resume automatically.

Charging the headphones for 3 hours provides you with approximately of sixteen hours of battery life for playing music or a talk time battery life of about eighteen hours. Also, it has 40mm premium quality stereo drivers to provide the users with remarkable audio experience. All thanks to the Premium Stereo Drivers in the headphones. The bass, middle and high tones all come off great on this quality headset's 20 Hz - 20 KHz frequency range.

The lightweight headphones come in a sturdy zippered case, which has ample room for the Micro-USB charging cable and stereo patch cable. However, there is no play/pause button on the headphones. The metal frame of the headphones also looks firm. And some people have reported that the active noise cancelling feature and the battery quality could be better.



Reviews

By **Robert P.** on October 2, 2016 from United States
 Overall a great purchase. Would recommend to anyone.

By **Ronda Ory** on September 28, 2016 from United States
 Not sure if my head is big, but the headphones fit really tightly around my ears.

By **Molly M** on September 26, 2016 from United States
 Sound and overall build quality is what I expected at this price point.

By **Santeria** on September 11, 2016 from United States
 Fine so far and the price is way better than Bose.

Add a Review

May 18, 2018 United States

Posted publicly as *

Write your review *

lightweight headphones come in a sturdy zippered case, which has ample room for the Micro-USB charging cable and stereo patch cable. However, there is no play/pause button on the headphones. The metal frame of the headphones also looks firm. And some people have reported that the active noise cancelling feature and the battery quality could be better.

Reviews

By **Robert P.** on October 2, 2016 from United States
 Overall a great purchase. Would recommend to anyone.

By **Ronda Ory** on September 28, 2016 from United States
 Not sure if my head is big, but the headphones fit really tightly around my ears.

By **Molly M** on September 26, 2016 from United States
 Sound and overall build quality is what I expected at this price point.

By **Santeria** on September 11, 2016 from United States
 Fine so far and the price is way better than Bose.

Today's Deals

Submit a Review

Headphone, High-Norm Condition

Today's Deals Free Shipping on orders over \$49 



BOHM Bluetooth Wireless Noise Cancelling Headphones with Inline Microphone - Black/Silver

Price: ~~\$469.99~~
\$84.99 & Free Shipping
 In Stock. Gift-wrap available.
 Color: Black, Silver

Product Description

BOHM Wireless Bluetooth headphones use the active noise cancellation technology and are comparable to Bose and Sennheiser in this area. With a simple flick of a switch, the outside world simply disappears, replaced by a vacuum of your favorite music or perhaps just precious silence, if you prefer. The aluminum housing and luxurious leather finish of the ear cups are comfortable while looking like they belong on headphones that cost 2-3 times as much. The simple-to-use controls are located on the right earpiece, including the USB charge point, with the noise cancelling indicator on the left.

This entirely wireless setup and advanced Bluetooth 4.0 enables you to link up to any music source that is in range. Whether you use a PC/Mac at home, your tablet outside, or your phone when you are out and about, BOHM wireless headphones can keep up with the pace and switch between your personal devices. The total range is 33 feet (10 meters), so you will not feel tied down to a small location while you enjoy your superior connection. They also have a cable if you don't wish to use them wirelessly, or if your battery dies. Another handy feature is that you can switch between music and phone calls, without needing to unplug, or even turn the volume down. Once your call ends, your music will resume automatically.

Charging the headphones for 3 hours provides you with approximately of sixteen hours of battery life for playing music or a talk time battery life of about eighteen hours. Also, it has 40mm premium quality stereo drivers to provide the users with remarkable audio experience, all thanks to the Premium Stereo Drivers in the headphones. The bass, middle, and high tones all come off great on this quality headset's 20 Hz - 20 kHz frequency range.

The lightweight headphones come in a sturdy zippered case, which has ample room for the Micro-USB charging cable and stereo patch cable. However, there is no play button on the headphones. The metal frame of the headphones also looks firm. And some people have reported that the active noise cancelling feature and the battery could be better.

Reviews

By **Robert P.** on October 2, 2016 from United States
 Sound: The headphones sound great, and plugged-in is better than via the Bluetooth connection. The noise cancellation works. Does it eliminate all exterior headphones.

Comfort: These headphones are not around the ear but on the ear. The cushions are soft on the ears. With the long battery life, I don't have to worry about Bluetooth. The pairing was simple, and the range is good.
 Design: They look expensive and cool. Material looks to be of high quality and durable.
 Overall a great purchase. Would recommend to anyone.

By **Ronda Ory** on September 28, 2016 from United States
 I'm not sure if my head is big, but the headphones fit really tightly around my ears. The audio, I will say, is very good in my honest opinion. Noise canceling is engine noise or white noise, but not the "human voice" noise. Functions buttons easy to navigate and are easy to differentiate by touch. Carrying case is nice and scratched or dented.

By **Molly M.** on September 26, 2016 from United States
 I have divided my review into appearance, sound, and functionality.
 1. Appearance and Comfort: This pair of headphones has a nice minimalist look for those of us that don't like the bulkiness and extravagance of most of headphones in the wish they made them over the ear, not on the ear.
 2. Sound: Sound is good. Good mids and no distortions in highs. Crap sound without using active noise canceling. I noticed no difference in sound quality using the headphones on my phone/computer using the cable provided.
 3. Functionality: They have quick and flawless Bluetooth pairing/connection. Active noise canceling is good considering it is an on ear headset. When noise canceling is on, sound quality is in short, the so far and the price is way better than Bose.

By **Santana** on September 11, 2016 from United States
 I want to list some pros and cons of the headphones:
 Pros: These headphones have great sound and good looking design. The Bluetooth works great, there is no interference, and it is easy to pair and easy to connect. And the headset is lightweight.
 Cons: In terms of comfort, I was a bit disappointed because it presses my ear. Noise canceling does not work well especially for mid- or high-frequency sounds. The battery capacity could be better.
 The bottom line, sound and overall build quality is what I expected at this price point.

Add a Review

May 18, 2018 United States

Posted publicly as *

Write your review *

Lightweight headphones come in a sturdy zippered case, which has ample room for the Micro-USB charging cable and stereo patch cable. However, there is no play button on the headphones. The metal frame of the headphones also looks firm. And some people have reported that the active noise cancelling feature and the battery could be better.

Reviews

By **Robert P.** on October 2, 2016 from United States
 Sound: The headphones sound great, and plugged-in is better than via the Bluetooth connection. The noise cancellation works. Does it eliminate all exterior headphones.
 Comfort: These headphones are not around the ear but on the ear. The cushions are soft on the ears. With the long battery life, I don't have to worry about Bluetooth. The pairing was simple, and the range is good.
 Design: They look expensive and cool. Material looks to be of high quality and durable.
 Overall a great purchase. Would recommend to anyone.

By **Ronda Ory** on September 28, 2016 from United States
 I'm not sure if my head is big, but the headphones fit really tightly around my ears. The audio, I will say, is very good in my honest opinion. Noise canceling is engine noise or white noise, but not the "human voice" noise. Functions buttons easy to navigate and are easy to differentiate by touch. Carrying case is nice and scratched or dented.

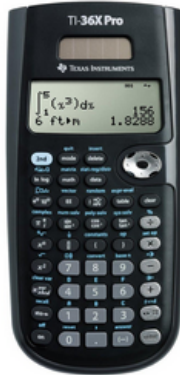
Today's Deals

Submit a Review

Calculator, Control Condition

Today's Deals

Free Shipping on orders over \$49.



Texas Instruments TI-36X Pro Engineering/Scientific Calculator

Price: \$19.78 & Free Shipping on orders over \$49.

In Stock. Gift-wrap available.

Size: Handheld

Product Description

The Texas Instruments TI-36X Pro 4-line Scientific Calculator is one of TI's most advanced scientific calculators. It features a large, four-line display and higher-level math and science functionality that is great for advanced algebra, trigonometry, computer science, calculus, engineering and statistics courses. Its MultView™ display shows multiple calculations at the same time on screen. You can scroll through previous entries, see the equation you input, and modify data to derive new answers. This solar-powered calculator can perform 1- and 2-variable statistics, including linear regression and trendline analysis. It can perform vectors and matrices using a vector and matrix entry window. It can also perform fractional equations and calculate angle units (radians, grads, degrees) when performing trigonometric functions.



The TI-36X Pro includes 6 physical constants and 10 English/metric conversions functions. This calculator combines features for statistical analysis, Boolean logic operations, and unit conversions in one sleek package. Ideal for students at every level, from junior high school algebra to college calculus, the Texas Instruments TI-36X packs a lot of punch in a small, inexpensive unit. It is ideal for computer science and engineering courses in which graphing technology may not be permitted. The calculator comes with a Slide-On cover. Free instruction manuals and instructional videos are available online.

Some users reported that the TI-36X's LCD screen is a little difficult to read when the unit is lying on a desk. The keyboard layout for common functions like adding and subtracting is a little unusual, and the silvered keys are difficult to read. Some users found there to be a delay when running complex calculations. Users in need of graphing functions will have to look elsewhere. Overall, however, for a calculator in this price range, the TI-36X provides enough power for most high-level math and science needs.

Add a Review

May 18, 2018 United States

Posted publicly as *

Write your review *

Today's Deals

Submit a Review

Headphone, Control Condition

Today's Deals

Free Shipping on orders over \$49



BOHM Bluetooth Wireless Noise Cancelling Headphones with Inline Microphone - Black/Silver

Price: ~~\$459.99~~

\$84.99 & Free Shipping

In Stock. Gift-wrap available.

Color: **Black, Silver**

Product Description

BOHM Wireless Bluetooth headphones use the active noise cancellation technology and are comparable to Bose and Sennheiser in this area. With a simple flick of a switch, the outside world simply disappears, replaced by a vacuum of your favorite music or perhaps just precious silence, if you prefer. The aluminum housing and luxurious leather finish of the ear cups are comfortable while looking like they belong on headphones that cost 2-3 times as much. The simple-to-use controls are located on the right earpiece, including the USB charge point, with the noise cancelling indicator on the left.

This entirely wireless setup and advanced Bluetooth 4.0 enables you to link up to any music source that is in range. Whether you use a PC/Mac at home, your tablet outside, or your phone when you are out and about, BOHM wireless headphones can keep up with the pace and switch between your personal devices. The total range is 33 feet (10 meters), so you will not feel tied down to a small location while you enjoy your superior connection. They also have a cable if you don't wish to use them wirelessly, or if your battery dies. Another handy feature is that you can switch between music and phone calls, without needing to unplug, or even turn the volume down. Once your call ends, your music will resume automatically.

Charging the headphones for 3 hours provides you with approximately of sixteen hours of battery life for playing music or a talk time battery life of about eighteen hours. Also, it has 40mm premium quality stereo drivers to provide the users with remarkable audio experience, all thanks to the Premium Stereo Drivers in the headphones. The bass, middle and high tones all come off great on this quality headset's 20 Hz - 20 KHz frequency range.

The lightweight headphones come in a sturdy zippered case, which has ample room for the Micro-USB charging cable and stereo patch cable. However, there is no play/pause button on the headphones. The metal frame of the headphones also looks flimsy. And some people have reported that the active noise cancelling feature and the battery quality could be better.



Add a Review

May 18, 2018 United States

Posted publicly as *

Write your review *

Today's Deals

Submit a Review

Online Appendix C. Manipulation Pretest

Table. Pretest and Manipulation Check Questionnaire of the Treatment Reviews

Pretest judges were shown the treatment reviews...

To what extent do you think the reviews...

...are helpful reviews for potential buyers? not helpful at all (1)/extremely helpful (7)

...offer information relevant to the product? not relevant at all (1)/extremely relevant (7)

...are informative/diagnostic? not informative at all (1)/extremely informative (7)

To what extent do you think the reviewers had been thoughtful in writing the reviews?

...not thoughtful at all (1)/extremely thoughtful (7)

How would you rate the overall valence (positive, neutral, negative) of the reviews? In other words, do reviewers in general think positively or negatively about the products?

...extremely negative(1)...neutral(4)...extremely positive (7)

Online Appendix D. Treatment Reviews and Manipulation Pretest

Table. Treatment Reviews and Manipulation Pretest Results

#	Content of the Treatment Reviews	Quality Reviewing <i>M (SD)</i>	Valence <i>M (SD)</i>
Calculator: High-Norm Treatment Reviews		6.25 (.70)	5.80 (.56)
1	<p>This non-graphic calculator is one of the models on the approved calculator list for my FE exam. I think this calculator is a good choice because it allows you to see 4 lines of calculations and you can scroll to previous ones even beyond the last 4. Also it has many functions and is easy to use. Another reason for choosing this model over the other contenders is the availability of free instructional video for the TI 36X Pro. On the other hand, its keys are a little difficult for me to read. Lots of glare on the silver operation keys.</p>		
2	<p>I got this scientific calculator as a gift and I quite like it! The TI-36X Pro is a great calculator for everyday and school use. One feature that stands out is the large 4-line display and the calculation history. I see my friends get frustrated with their 1 line displays and I realize this is actually quite useful. Another good thing is the TI-36X Pro supports a multitude of functions and features. For example, you can use constants, generate tables, do matrix calculations, convert units, and etc. Overall, it makes a great scientific calculator to have around! The only downside is the lack of speed when solving equations, but I think it's expected because it's not a computer ;)</p>		
3	<p>Got this scientific calculator to use with my college math and chemistry courses. We are not permitted to use graphing calculators so this is the next best thing. The most useful function for me is that it saves and displays previous entries just like a typical graphing calculator. It is solar powered you do not have to worry about changing batteries, which was a huge plus for me. I had been going back and forth between this TI and a similar Casio model but went with the TI in the end due to its ease of use and constants within the menu.</p> <p>Now let's talk about the negative side. To me, there are two major annoyances. First, this TI has unusual keyboard layout for common functions like addition and subtraction. Second, the keys for addition, subtraction, and etc. are silver-colored with no highlighting: I continually make mistakes by pressing the wrong operation key due to the poor visibility.</p> <p>But still, I have to say this is a good calculator which has a lot of functions for the price.</p>		
4	<p>This TI is user friendly in terms of solving complex numbers and matrices. Simply having a multi-line display and being able to check your work is a huge benefit versus the one line display I was used to. It has easily accessible physical constants and unit converter. I am impressed with the many functions it includes, such as performing derivatives, equation solving, data tables, and some memory devoted to 8 different variables. For \$20 it has much of the capability of the much more expensive graphing calculators. The calculator also has a nice thin cover case that clips on to the back when using the calculator.</p>		

	I really have no complaints about it. There is a minor display problem (as mentioned in the description) which may concern some, but I regard it as trivial.	
Calculator: Low-Norm Treatment Reviews		2.40 (.73)
		5.67 (.81)
1	The calculator is one of the models on the approved calculator list for my FE exam.	
2	Got me through my engineering classes just fine.	
3	Good calculator and does everything I need it to do. It does not have graphing functions, tho.	
4	Use this as a college student studying engineering. It is an efficient scientific calculator with a lot of features.	
Headphone: High-Norm Treatment Reviews		6.32 (.55)
		5.40 (.51)
1	<p>Sound: The headphones sound great, and plugged-in is better than via the Bluetooth connection. The noise cancellation works. Does it eliminate all exterior sound - no - but it does as good a job as top of the line headphones.</p> <p>Comfort: These headphones are not around the ear but on the ear. The cushions are soft on the ears. With the long battery life, I didn't have to worry about recharging them anytime soon.</p> <p>Bluetooth: The pairing was simple, and the range is good.</p> <p>Design: They look expensive and cool. Material looks to be of high quality and durable. Overall a great purchase. Would recommend to anyone.</p>	
2	<p>I'm not sure if my head is big, but the headphones fit really tightly around my ears. The audio, I will say, is very good in my honest opinion. Noise canceling is just satisfactory, because it seems to eliminate only the engine noise or white noise, but not the "human voice" noise. Functions buttons easy to navigate and are easy to differentiate by touch. Carrying case is nice quality and keeps the aluminum housing from getting scratched or dented.</p>	
3	<p>I have divided my review into appearance, sound, and functionality.</p> <p>1. Appearance and Comfort: This pair of headphones has a nice minimalist look for those of us that don't like the bulkiness and extravagance of most of headphones in the market today. They are very light as well. I wish they made them over the ear, not on the ear.</p> <p>2. Sound: Sound is good. Good middles and no distortions in highs. Crisp sound without using active noise cancelling. I noticed no difference in sound quality using the headphones in Bluetooth or wired to phone/computer using the cable provided.</p> <p>3. Functionality: They have quick and flawless Bluetooth pairing/connection. Active noise cancelling is good considering it is an on ear headset. When noise cancelling is on, sound quality is a little bit reduced.</p> <p>In short, fine so far and the price is way better than Bose.</p>	
4	<p>I want to list some pros and cons of the headphones:</p> <p>Pros: These headphones have great sound and good looking design. The Bluetooth works great: there is no interference, and it is easy to pair and easy to connect. And the headset is lightweight.</p>	

	<p>Cons: In terms of comfort, I was a bit disappointed because it presses my ear. Noise cancelling does not work well especially for mid- or high-frequency sounds. The battery capacity could be better.</p> <p>The bottom line, sound and overall build quality is what I expected at this price point.</p>		
	<p>Headphone: Low-Norm Treatment Reviews</p>	<p>2.70 (.84)</p>	<p>5.47 (.83)</p>
1	Overall a great purchase. Would recommend to anyone.		
2	Not sure if my head is big, but the headphones fit really tightly around my ears.		
3	Sound and overall build quality is what I expected at this price point.		
4	Fine so far and the price is way better than Bose.		

Online Appendix E. Postexperiment Questionnaire Used in Study 1

Social presence (Kumar and Benbasat 2006)

Please indicate to what extent you agree with the following statements.

This is a sense of human contact in the website.

There is a sense of sociability in the website.

There is a sense of human warmth in the website.

There is a sense of human sensitivity in the website.

... strongly disagree (1)/strongly agree (7)

Manipulation checks

To what extent do you think others' reviews...

...are helpful reviews for potential buyers? not helpful at all (1)/extremely helpful (7)

...offer information relevant to the product? not relevant at all (1)/extremely relevant (7)

...are informative? not informative at all (1)/extremely informative (7)

To what extent do you think the reviewers had been thoughtful in writing the reviews?

...not thoughtful at all (1)/extremely thoughtful (7)

How would you rate the overall valence (positive, neutral, negative) of the reviews? In other words, do reviewers in general think positively or negatively about the products?

...extremely negative (1)/extremely positive (7)

Experience

To what degree have you had experience with writing online product reviews?

...I have no experience with writing online product reviews (1)/I have a lot of experience with writing online product reviews (7).

Background questions

- Gender, Age

References

- Kumar, N. and Benbasat, I. 2006. "Research Note: The Influence of Recommendations and Consumer Reviews on Evaluations of Websites," *Information Systems Research* (17:4), 425-439.

Online Appendix F. Perceived Social Presence

In this appendix, we provide explanations on why social presence was not manipulated in our research. Social presence (and presence) is a natural cognitive effect of mediated communication (see Lee 2004, for a detailed discussion on the evolutionary origin of the feeling of social presence, and Reeves and Nass 1996, the media-equation phenomena). Social presence cannot be *directly manipulated* because it is a cognitive effect; certain media characteristics have been shown to influence the feeling of social presence (e.g., Social Presence Theory, Short, Williams & Christie 1976; Media Richness Theory, Daft & Lengel 1986; Media Synchronicity Theory, Dennis & Valacich 1999; the Computers Are Social Actors Paradigm, Reeves & Nass 1996).

In the table below, we list some key references that examine social presence in relevant contexts. Altogether, in the laboratory experimental studies (Zhu et al. 2010; Kumar and Benbasat 2006; Lee and Nass 2005; Qiu and Benbasat 2005), social presence is a dependent variable that is influenced by website (i.e., the medium) features, such as voice chat and shared navigation support, provision of consumer reviews, and computer voice. In the survey studies, social presence is an independent variable (Gefen & Straub 2003; Shen 2012) or an intermediary variable (Hajli et al. 2017; Zhang et al. 2014; Animesh et al. 2011; Ou et al. 2014) in structural models.

Altogether, social presence is a measured variable instead of a manipulated variable.

Table. Research on Social Presence in E-commerce Related Contexts

Method	Role of SP	Effects	Context	References
Lab experiment: Support for voice chat and shared navigation (via Microsoft MSN 8)	Dependent Variable	Support for chat and shared navigation affects SP	Collaborative online shopping: Amzaon.com	Zhu, Benbasat, & Jiang 2010
Lab experiment: A website's support for	Dependent Variable	The website's support affects SP	e-commerce: Amazon.com	Kumar & Benbasat 2006

the provision of consumer reviews				
Lab experiment: Extrovert vs. introvert computer voice (for book descriptions)	Dependent Variable	Voice affects SP	e-commerce: a book-buying website	Lee and Nass 2005
Lab experiment: Text-to-speech voice and 3D avatars	Dependent Variable	Neither website features significantly affects SP	e-commerce: a shopping website for digital cameras and accessories	Qiu and Benbasat 2005
Survey	Independent Variable	SP affects perceived enjoyment	e-commerce: Kaboodle.com	Shen 2012
Survey	Independent Variable	SP affects trust	Commercial e-service: city travel site	Gefen & Straub 2003
Survey	Intermediary variable	SP affects purchase intention	Social commerce on Facebook	Hajli et al. 2017
Survey	Intermediary variable	SP affects intention to share commercial information	Commercial activities via social media: Sina Weibo and RenRen	Zhang et al. 2014
Survey	Intermediary variable	SP affects flow experience	Virtual world: Second Life	Animesh et al. 2011
Survey	Intermediary variable	SP affects trust	e-commerce: TaoBao	Ou et al. 2014

While social presence cannot be directly manipulated, it could be “manipulated” indirectly. For example, we may manipulate the voice feature in our research website (e.g., based on the finding from Lee & Nass 2005), which might result in participants in one condition reporting a higher level of mean perceived social presence than those in another condition. Nevertheless, using an indirect “manipulation” would almost inevitably introduce confounding mechanism(s) in the present study. For example, the voice feature can bring in factors such as likability of the voice narrator, perceived source of the communication medium, perceived similarity with the voice narrator, and so on (Lee & Nass 2005).

Conclusion. Moderators can be either manipulated variables or measured variables (Baron and Kenny 1986). Considering that: 1) social presence cannot be directly manipulated, and 2) the

confounding mechanism(s) induced by indirect manipulation can severely invalidate our conclusions, we thus do not manipulate social presence in our research.

References

- Animesh, A., Pinsonneault, A., Yang, S.-B., & Oh, W. 2011. "An Odyssey into Virtual Worlds: Exploring the Impacts of Technological and Spatial Environments on Intention to Purchase Virtual Products," *MIS Quarterly* (35:3), 789-810.
- Daft, R. L. & Lengel, R. H. 1986. "Organizational Information Requirements, Media Richness and Structural Design," *Management Science* (32:5), 554-571.
- Dennis, A. R. & Valacich, J. S. 1999. "Rethinking Media Richness: Towards a Theory of Media Synchronicity," in *Proceedings of the 32nd Annual Hawaii International Conference on Systems Sciences. HICSS-32. Abstracts and CD-ROM of Full Papers*, Maui, HI, USA.
- Gefen, D. & Straub, D. 2003. "Managing user trust in B2C e-services," *E-Service Journal* (2:2), 7-24.
- Hajli, N., Shanmugam, M., Papagiannidis, S., Zahay, D., and Richard, M. O. 2017. "Branding Co-creation with Members of Online Brand Communities," *Journal of Business Research* (70), 136-144.
- Kumar, N. and Benbasat, I. 2006. "Research Note: The Influence of Recommendations and Consumer Reviews on Evaluations of Websites," *Information Systems Research* (17:4), 425-439.
- Lee, K. M. & Nass, N. 2005. "Social-psychological Origins of Feelings of Presence: Creating Social Presence with Machine-generated Voices," *Media Psychology* (7:1), 31-45.
- Lee, K. M. 2004. "Presence, Explicated," *Communication Theory* (14:1), 27-50.
- Ou, X., Pavlou, P. A., & Davison, R. M. 2014. "Swift guanxi in online marketplaces: The role of computer-mediated communication technologies," *MIS Quarterly* (38:1), 209-230.
- Qiu, L. and Benbasat, I. 2005. "Online consumer trust and live help interfaces: The Effects of Text-to-Speech Voice and Three-dimensional Avatars," *International Journal of Human-computer Interaction* (19:1), 75-94.
- Reeves, B., & Nass, C. I. 1996. *The Media Equation: How People Treat Computers, Television, and New Media Like Real People and Places*. Center for the Study of Language and Information; Cambridge University Press.
- Shen, J. 2012. "Social Comparison, Social Presence, and Enjoyment in the Acceptance of Social Shopping Websites," *Journal of Electronic Commerce Research* (13:3), 189-212.
- Short, J., Williams, E., and Christie, B. 1976. *The Social Psychology of Telecommunications*, London: Wiley.
- Zhang, H., Lu, Y., Gupta, S., & Zhao, L. 2014. "What motivates customers to participate in social commerce? The impact of technological environments and virtual customer experiences" *Information & Management* (51:8), 1017-1030.
- Zhu, L., Benbasat, I., & Jiang, Z. 2010. "Let's Shop Online Together: An Empirical Investigation of Collaborative Online Shopping Support," *Information Systems Research* (21:4), 872-891.

Online Appendix G. Inclusion Criteria for the Dimensions in Main Measurement

Guided by the definition of the behavior of writing quality reviews in the main manuscript, the following key dimensions are included in our two-factor measurement:

- Quality-1: Textual Analysis Measures
 - Readability
 - Length
 - Objectivity
 - Number of product- or service-relevant points
- Quality-2: Human-rated Measures
 - Relevance
 - Thoughtfulness
 - Informative-ness
 - Helpfulness

It should be noted that we do not intend to measure review quality per se; instead, we aim to measure the behavior of writing quality reviews. This difference is important, and it also illustrates the conceptual basis of our measurement. We do not aim to develop an exhaustive measurement. The items we include are guided by two criteria:

- (1) They have been well-documented in the literature;
- (2) They reflect the influential mechanism of the space norms—they can exemplify an expectation that subsequent review writers can conform to (do in the same manner).

In this Appendix, we elaborate on certain review characteristics that one may be tempted to include, but some of them may not qualify based on these criteria.

Many *metadata* of a review—e.g., timeliness, reviewer reputation, the ratings a review received, albeit good indicators of review quality—are not about ways of writing quality content; hence, they are not included.

In contrast, *length* is included as a key dimension; writing a longer review represents a way to write higher quality reviews, because longer reviews allow more information to be conveyed than shorter ones (Kwon and Sung 2015; Ghose and Ipeirotis 2011). As such, seeing others write long reviews can exemplify an expectation that subsequent review writers can conform to.

Relevance. Relevance (Wang and Chaudhry 2018) is included in our measures. Writing content that is relevant to the focal product or service (e.g., features and functions) is a good way to construct a high-quality review (Shoham et al. 2017). Seeing others write content that focuses on the key features and functions of the focal product or service can set up the expectation about what subsequent writers *should* also focus on in writing their reviews. Thus, relevance, measured as the number of product- or service-relevant points (Wang and Chaudhry 2018), is included.

Subjective vs. objective description of features. Review objectivity is included as one of the indicators in our measure because a review that contains more objective content is regarded as more helpful and convincing (Ghose and Ipeirotis 2007, 2011; Otterbacher 2009).

Emotional expressiveness. Emotional expressiveness may not be included in our measures because it does not meet the second criterion. Emotional expressiveness, such as expressing anger (Yin et al. 2014), can positively influence perceived helpfulness of a review. Thus, it may be a good measure to identify good quality reviews. However, it is less sensible to say that expressing anger is an apt way to write quality review content, at least not a common or advisable way (e.g., *How to write a good quality review? Do it by expressing anger*). Essentially, seeing others express anger is not likely to exemplify an expectation that all subsequent review writers can conform to. It cannot be socially expected for writing quality review that everyone *should* express anger. Thus, emotional expressiveness is not included.

Convincingness. We only found one recent paper (Hernandez-Bocanegra and Ziegler, 2019), in which convincingness is found to partially mediate the relationship between review objectivity and helpfulness. That is, review objectivity increases helpfulness partly because a review that contains objective content is perceived to be more convincing. Please note that we do include review objectivity (Ghose and Ipeirotis 2007, 2011; Otterbacher 2009) as one of the indicators, thereby allowing us to capture convincingness to some extent. Seeing others provide more objective and factual descriptions about the focal product or service can set up the expectation that subsequent writers *should* also provide more objective account in writing their reviews.

References

- Ghose, A. and Ipeirotis, P. G. 2007. "Designing Novel Review Ranking Systems: Predicting the Usefulness and Impact of Reviews," in *Proceedings of the Ninth International Conference on Electronic commerce (ICEC 2007)*. New York, NY, USA: ACM.
- Ghose, A. and Ipeirotis, P. G. 2011. "Estimating the Helpfulness and Economic Impact of Product Reviews: Mining Text and Reviewer Characteristics," *IEEE Transactions on Knowledge and Data Engineering* (23:10), 1498-1512.
- Hernandez-Bocanegra, D. C. and Ziegler, J. 2019. "Assessing the Helpfulness of Review Content for Explaining Recommendations," in *Proceedings of SIGIR 2019 Workshop on Explainable Recommendation and Search (EARS'19)*. ACM, New York, NY, USA.
- Kwon, O. and Sung, Y. 2015. "The Consumer-Generated Product Review: Its Effect on Consumers and Marketers," in *Cracking the Code: Leveraging Consumer Psychology to Drive Profitability*, Posavac, S. S. (ed.), New York: Routledge, 200-218.
- Otterbacher, J. 2009. "'Helpfulness' in Online Communities: A Measure of Message Quality," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ACM, 955-964.
- Shoham, M., Moldovan, S., and Steinhart, Y. 2017. "Positively Useless: Irrelevant Negative Information Enhances Positive Impressions," *Journal of Consumer Psychology* (27:2), 147-159.
- Wang, Y. and Chaudhry, A. 2018. "When and How Managers' Responses to Online Reviews Affect Subsequent Reviews," *Journal of Marketing Research* (55:2), 163-177.
- Yin, D., Bond, S., and Zhang, H. 2014. "Anxious or Angry Effects of Discrete Emotions on the Perceived Helpfulness of Online Reviews," *MIS Quarterly* (38:2), 539-560.

Online Appendix H. Machine-Learning Methods and Programming Packages Used, and Evaluation Results in Quality-1 Measurement

For each of the indices included in the four dimensions of the Quality-1 measure of quality review writing behavior, the following reports their definitions, specific programming packages or resources used, and the reliability evaluations of the supervised-learning classifiers trained where applicable.

- **Length:** length measures the length of a review text
 - #syllable: Python package textstat (<https://pypi.org/project/textstat/>)
 - #words: Python package nltk (<https://www.nltk.org/>)
 - #nonstop words: #words excluding all stop words. Stop words in English are those that do not add much meaning to the sentence, e.g., “this,” “that,” “been,” etc.; thus, excessive use of stop words increases word count but does not necessarily add much more information. Python package nltk (<https://www.nltk.org/>)
 - #sentences: custom program using Python native string methods and functions

- **Readability:** readability measures a review text’s level of ease with which a reader can understand the written text. The following are the most commonly used English readability indices in natural language processing (NLP):
 - Automated Readability Index (ARI)
 - Coleman–Liau Index (CLI)
 - Flesch–Kincaid Grade Level (FKGL)
 - Flesch Reading Ease (FRE)
 - Gunning–Fog Index (GFI)
 - Simple Measure of Gobbledygook (SMOG)
 - Dale-Chall Readability (DCR)
 - Linsear Write Readability Formula (LWRF)

Python package textstat (<https://pypi.org/project/textstat/>) is used to calculate these readability indices

- **Objectivity:** objectivity measures how closely related the review text is to the product or service descriptions publicly available, presumably on the merchant site. Objective reviews tend to include more factual descriptions about products or services regarding their functions and features, and how well they perform, whereas subjective reviews tend to include more emotional expressions as a result of using the products or services. Three different methods are considered and evaluated as follows:

- bigram overlap: proportion of consecutive two-word sequences in the review text that also appear in the product description texts found on the merchant site. Custom program using nltk (<https://www.nltk.org/>) and other Python functions
- probability of objective sentence: method adapted from Ghose and Ipeiritos 2011, it is calculated based on the output from a custom supervised machine-learning classifier built in Python using package textstat (<https://pypi.org/project/textstat/>). The specific model used is the naïve Bayes classifier. After training, reliability of the classifier is evaluated. Using ten-fold cross-validation, the classifier had acceptable Area Under the Receiver Operating Characteristic Curve (AUC-ROC) of .862 (calculator) and .846 (headphone), according to Mandrekar (2010) and Hand & Till (2001)
- normalized longest common subsequence (NLCS): it first finds the longest common subsequence (LCS) of words (i.e., longest common phrase) that the review text and the product description texts found on the merchant site have in common, then normalizes it by the length of the review text
- **#product-relevant points**: number of different product features and/or functions mentioned and discussed in a review text. It is calculated as follows:
 - a custom supervised text-mining classifier first detects what product feature(s) each sentence of a review text discusses (or no feature is discussed), then compute the number of *unique* product features detected among all the processed sentences of a review. The end result is the number of unique product features detected in the discussion of a review text. After training, reliability of the classifier is evaluated. Using ten-fold cross-validation, the classifier achieved acceptable AUC-ROC of .83 and .837 for calculators and headphones, respectively (Hand & Till 2001, Mandrekar 2010)

References

- Hand, D. J. and Till, R. J. 2001. "A Simple Generalisation of the Area Under the ROC Curve for Multiple Class Classification Problems," *Machine Learning* (45), 171-186.
- Mandrekar, J. N. 2010. "Receiver Operating Characteristic Curve in Diagnostic Test Assessment," *Journal of Thoracic Oncology* (5:9), 1315-1316.

Online Appendix I. Reliability of the Machine-Learning Algorithm for Calculating the Number of Product (Service)-Relevant Points

We trained a supervised-learning algorithm (i.e., naïve Bayes classifier) to automatically compute the number of product-relevant points conveyed in each review collected from the experiment. This result is included in the indices for the Quality-1 measure of the outcome variable. Two coauthors pre-labeled a sizable portion of the collected data under each product to feed into the machine-learning algorithm; inter-coder reliability was acceptable.

Also recall that we had previously collected human-coded data on the number of product-relevant points. Because human intelligence can differentiate nuanced topic space in both the main features and their sub-dimensions (i.e., subfeature), we instructed the human raters to count both and add them together. In other words, the “instructions” and flexibility provided to the machine-learning algorithms and to the human coders are different; therefore, the measures produced by both approaches can be used to evaluate their correlation and reliability, as a means to assessing the robustness of the machine-learning method.

The correlations between the two approaches were acceptable (Calculator: $\rho = .60$; Headphone: $\rho = .59$). If the measures produced by the two approaches were to form a scale under each product, their reliability index were also acceptable (Calculator: Cronbach’s $\alpha = .73$; Headphone: Cronbach’s $\alpha = .70$). We conclude that the number of product-relevant points included in the indices for the Quality-1 measure is reliable and robust, as benchmarked by the human-coded counterpart judging the same dimension of the review texts.

Online Appendix J. Exploratory Factor Analysis (EFA) Results: Quality-1 Indicators

Calculator: all indicators

	Factor 1	Factor 2	Factor 3	Factor 4
Automated Readability Index (ARI)	.849			
Coleman–Liau Index (CLI)	.844			
Flesch–Kincaid Grade Level (FKGL)	.867			
Flesch Reading Ease (FRE) – reversed	.897			
Gunning–Fog Index (GFI)	.736			
Simple Measure of Gobbledygook (SMOG)	.930			
Dale–Chall Readability (DCR)	.914			
Linsear Write Readability Formula (LWRF)	.887			
Number of Syllables		.946		
Number of Nonstop Words		.943		
Number of All Words		.935		
Number of Sentences		.876		
Mean Probability of Objective Sentences			.763	
Number of Bigram Overlaps			.824	
Normalized Longest Common Subsequence (NLCS) *		-.956		
Number of Relevant Points				.785

Notes: Loadings with absolute values below 0.5 omitted. Varimax rotation with Kaiser Normalization. N = 40.

* Indicator dropped because of unexpected conflicted loading.

Calculator: final indicators

	Factor 1	Factor 2	Factor 3	Factor 4
Automated Readability Index (ARI)	.851			
Coleman–Liau Index (CLI)	.844			
Flesch–Kincaid Grade Level (FKGL)	.869			
Flesch Reading Ease (FRE) – reversed	.899			
Gunning–Fog Index (GFI)	.745			
Simple Measure of Gobbledygook (SMOG)	.931			
Dale–Chall Readability (DCR)	.915			
Linsear Write Readability Formula (LWRF)	.891			
Number of Syllables		.938		
Number of Nonstop Words		.934		
Number of All Words		.924		
Number of Sentences		.877		
Mean Probability of Objective Sentences			.760	
Number of Bigram Overlaps			.827	
Number of Relevant Points				.798
Variance explained	34%	24%	18%	10%

Notes: Loadings with absolute values below 0.5 omitted. Varimax rotation with Kaiser Normalization. N = 40.

Headphone: all indicators

	Factor 1	Factor 2	Factor 3	Factor 4
Automated Readability Index (ARI)	.767			.507
Coleman–Liau Index (CLI)	.779			
Flesch–Kincaid Grade Level (FKGL)	.828			
Flesch Reading Ease (FRE) – reversed	.861			
Gunning–Fog Index (GFI)	.618			
Simple Measure of Gobbledygook (SMOG)	.915			
Dale–Chall Readability (DCR)	.910			
Linsear Write Readability Formula (LWRF)	.807			
Number of Syllables		.967		
Number of Nonstop Words		.969		
Number of All Words		.959		
Number of Sentences		.905		
Mean Probability of Objective Sentences			.764	
Number of Bigram Overlaps			.826	
Normalized Longest Common Subsequence (NLCS) *		-.969		
Number of Relevant Points				.562

Notes: Loadings with absolute values below 0.5 omitted. Varimax rotation with Kaiser Normalization. N = 40.

* Indicator dropped because of unexpected conflicted loading.

Headphone: final indicators

	Factor 1	Factor 2	Factor 3	Factor 4
Automated Readability Index (ARI)	.770			
Coleman–Liau Index (CLI)	.779			
Flesch–Kincaid Grade Level (FKGL)	.832			
Flesch Reading Ease (FRE) – reversed	.863			
Gunning–Fog Index (GFI)	.626			
Simple Measure of Gobbledygook (SMOG)	.913			
Dale–Chall Readability (DCR)	.909			
Linsear Write Readability Formula (LWRF)	.810			
Number of Syllables		.962		
Number of Nonstop Words		.964		
Number of All Words		.952		
Number of Sentences		.907		
Mean Probability of Objective Sentences			.766	
Number of Bigram Overlaps			.825	
Number of Relevant Points				.577
Variance explained	33%	24%	17%	9%

Notes: Loadings with absolute values below 0.5 omitted. Varimax rotation with Kaiser Normalization. N = 40.

Results Summary

In conclusion, four factors emerged from the above EFA results, and this is consistent with our anticipation about four factors. We name the four factors as follows:

- Factor 1: Readability
- Factor 2: Length
- Factor 3: Objectivity
- Factor 4: Number of product (service)-relevant points

In the subsequent Online Appendix, we examine the factor structure by pooling these four textual quality indices with the four human coded indices. We present the detailed factor analysis and results.

Online Appendix K. Confirmatory Factor Analysis (CFA) Results: The Behavior of Writing High-Quality Reviews

In the previous Online Appendix, we show that four factors consistently emerge from the textual quality indicators captured by various machine-learning algorithms in a random sample of 40 participants' reviews of the calculator and the headphone. To ensure that these factors (indices) meaningfully measure the behavior of writing high-quality reviews, we also collected human-rated data on a scale with four indices pertaining to our definition of the behavior of writing quality reviews: relevance, thoughtfulness, informative-ness, and helpfulness. The rationale for selecting these four dimensions are provided in the main manuscript (i.e., Chapter *Study 1*, Sec. 3.4. *Measures for Writing Quality Reviews*).

In this Appendix, we examine whether the four textual-analysis based indices are distinct from those of the human-rated items, and if so, whether two factors would emerge from the data and whether each is reliable and both are consistent (highly correlated) with each other.

We used the main study dataset for the CFA ($n = 168$). We tested three alternative measurement models in each product, and the below Table presents the results of model comparison. The results indicated that the two-factor measurement model—i.e., Model (2): readability, length, objectivity, and number of product (service)-relevant points load onto factor Quality-1 ($\alpha > .76$ across two products); relevance, thoughtfulness, informative-ness, and helpfulness load onto factor Quality-2 ($\alpha > .99$)—showed acceptable fit indices (Calculator: $\chi^2(19) = 37.95$, $p = .006$, comparative fit index (CFI) = .99, Tucker-Lewis index (TLI) = .99, root mean square error of approximation (RMSEA) = .08. Headphone: $\chi^2(19) = 34.16$, $p = .02$, CFI = .99, TLI = .99, RMSEA = .07). Model comparison results showed that, whereas the two-factor model had superior fit to the more parsimonious one-factor model ($p < .01$ across two products),

thus favoring Model (2), its fit was similar to that of the less parsimonious three-factor model ($p > .05$). Therefore, Model (2) had the best measurement model specification. For this model, all items loaded adequately to their respective factors (factor loadings $> .60$). Average variance extracted (AVE) of each factor was adequate (AVE $> .60$). Lastly, the square root of the AVE of each factor exceeded its correlation with the other factor. In sum, the results confirmed that the scales for the two DVs were reliable and valid as expected. The Figure (further below) summarizes the results of the best CFA models. Quality-1 and Quality-2 both had high reliability ($\alpha > .76$). The Figure also shows that Quality-1 and Quality-2 had high correlations.

Table. CFA Results Comparison: Competing Measurement Models

Model	Chi2	df	Δ Chi2	Δ df	CFI	TLI	RMSEA
Calculator							
(1) three-factor model	37.10**	17	--	--	.99	.98	.08
(2) two-factor model	37.95**	19	.85 (1 vs. 2)	2	.99	.99	.08
(3) one-factor model	52.40***	20	14.45 (2 vs. 3)***	1	.98	.98	.10
Headphone							
(1) three-factor model	31.28*	17	--	--	.99	.98	.07
(2) two-factor model	34.16*	19	2.88 (1 vs. 2)	2	.99	.99	.07
(3) one-factor model	42.01**	20	7.85 (2 vs. 3)**	1	.99	.98	.08

Notes.

Results for both Calculator and Headphone indicate that Model (2) with two factors is the best fitted specification.

Model (1) is the least parsimonious, Model (3) the most parsimonious loading all indicators onto 1 factor.

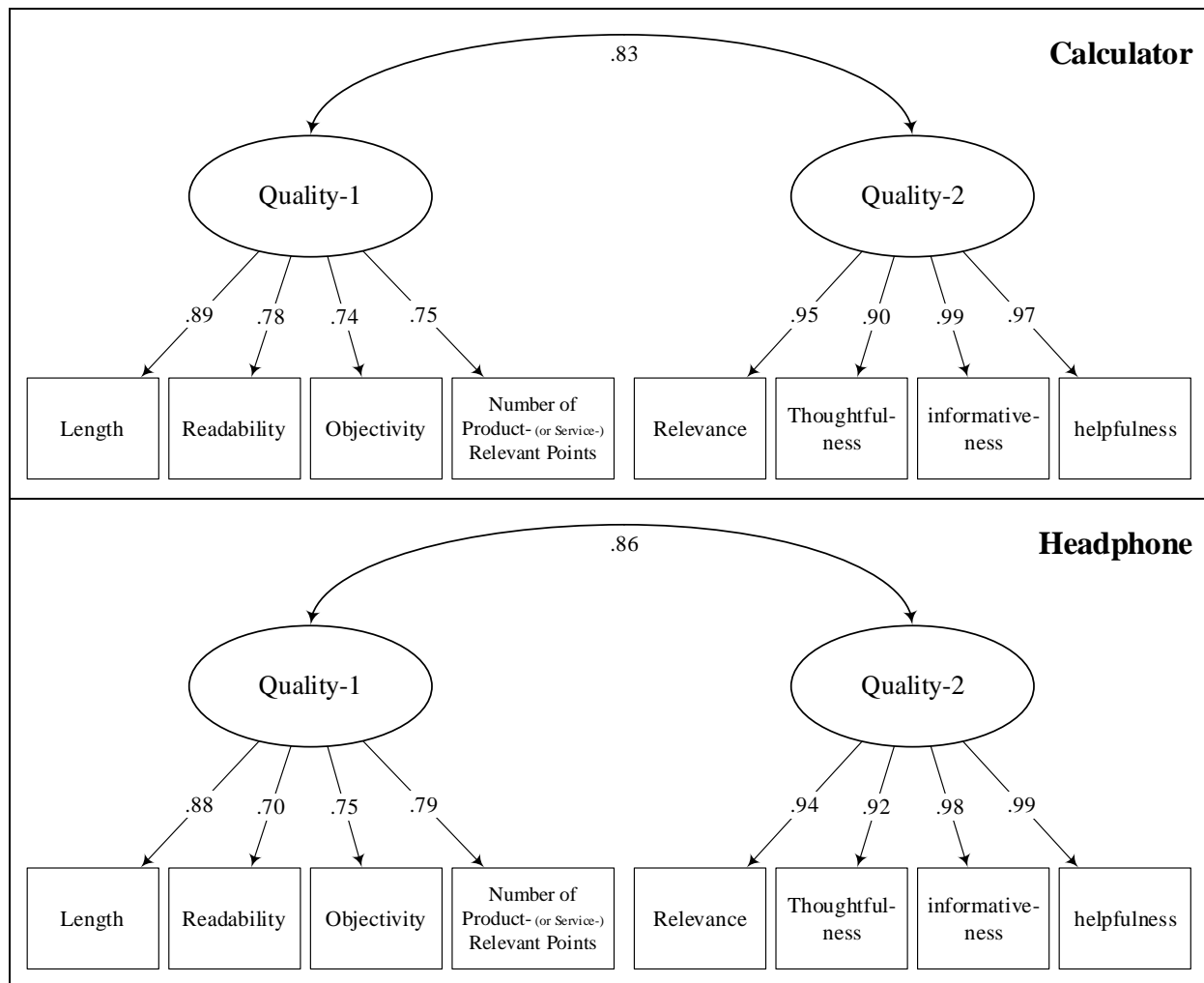
Model (1) specification: readability and length load onto factor 1; objectivity and number of product-relevant points load onto factor 2; relevance, thoughtfulness, informative-ness, and helpfulness load onto factor 3.

Model (2) specification: readability, length, objectivity, and number of product-relevant points load onto factor 1 (Quality-1); relevance, thoughtfulness, informative-ness, and helpfulness load onto factor 2 (Quality-2).

CFI = Comparative Fit Index, TLI = Tucker-Lewis Index, RMSEA = Root Mean Square Error of Approximation.

* $p < .05$, ** $p < .01$, *** $p < .001$

Figure. Results of CFA Measurement Model with the Best Fit



Online Appendix L. Means (SDs) of the Dependent Measure Indicators and Manipulation Checks By Conditions

Measures	<i>Calculator</i>			<i>Headphone</i>		
	High-norm (n = 53)	Low-norm (n = 58)	Control (n = 57)	High-norm (n = 53)	Low-norm (n = 58)	Control (n = 57)
<i>Quality-1 Composite Measure</i>						
Length						
#syllable	133.91 (33.54)	51.78 (19.51)	88.40 (35.15)	140.92 (34.38)	52.86 (23.3)	101.70 (47.35)
#nonstop words	44.13 (10.73)	17.05 (6.04)	29.84 (12.15)	50.40 (12.42)	19.98 (8.36)	36.42 (17.44)
#words	94.06 (23.84)	36.19 (14.24)	62.07 (25.33)	107.53 (27.13)	40.12 (18.67)	75.16 (33.22)
#sentences	6.96 (1.97)	3.26 (1.36)	5.58 (2.56)	9.51 (3.02)	3.81 (1.81)	6.95 (3.60)
Objectivity						
classifier result	.42 (.21)	.26 (.23)	.42 (.24)	.48 (.21)	.39 (.26)	.45 (.24)
bigram overlap	.20 (.04)	.10 (.05)	.12 (.05)	.19 (.04)	.10 (.06)	.14 (.05)
Readability						
ARI	10.03 (3.2)	8.73 (3.4)	8.22 (3.12)	9.68 (3.17)	8.21 (3.6)	8.38 (3.17)
CLI	8.23 (1.66)	8.09 (2.76)	7.90 (1.96)	8.24 (1.54)	8.06 (2.74)	8.29 (2.17)
FKGL	9.72 (2.96)	8.34 (3.17)	7.99 (2.83)	8.13 (3.26)	6.58 (3.46)	7.04 (3.09)
FRE (reversed)	41.33 (4.44)	30.89 (5.55)	36.22 (5.78)	40.79 (5.89)	29.72 (4.31)	33.22 (4.61)
GFI	11.26 (2.78)	10.05 (3.05)	9.81 (2.52)	10.05 (3.17)	8.99 (3.29)	9.02 (2.88)
SMOG	10.38 (3.33)	3.67 (4.81)	8.67 (3.95)	7.99 (3.1)	4.63 (4.32)	7.47 (3.51)
DCR	7.38 (.8)	7.36 (1.36)	7.42 (1.04)	7.15 (1.24)	7.28 (1.37)	7.24 (1.36)
LWRF	10.53 (2.99)	8.82 (3.87)	8.58 (2.95)	8.70 (2.8)	7.13 (3.13)	7.95 (2.60)
#Points	3.96 (1.09)	2.44 (.84)	3.28 (1.1)	5.06 (1.98)	3.41 (1.5)	4.14 (1.60)
<i>Quality-2 Composite Measure</i>						
Relevance	6.58 (.57)	2.31 (.68)	4.06 (1.1)	6.51 (.57)	2.56 (.77)	4.57 (1.3)
Thoughtfulness	6.29 (.67)	1.86 (.8)	3.79 (1.16)	6.39 (.65)	2.25 (.77)	4.34 (1.39)
Informative-ness	6.27 (.67)	1.84 (.63)	3.75 (1.28)	6.37 (.64)	2.16 (.71)	4.36 (1.4)
Helpfulness	6.25 (.77)	2.12 (.75)	3.96 (1.25)	6.42 (.58)	2.5 (.82)	4.63 (1.3)
<i>Perceived Social Presence</i>	4.30 (1.40)	3.66 (1.40)	3.27 (1.39)	4.30 (1.40)	3.66 (1.40)	3.27 (1.39)
Manipulation Checks						
Others' quality review behavior	5.82 (0.62)	3.66 (0.75)	n.a.	5.82 (0.62)	3.66 (0.75)	n.a.
Valence	5.34 (0.90)	5.26 (1.05)	n.a.	5.34 (0.90)	5.26 (1.05)	n.a.

Notes: Each cell reports the corresponding Mean (Standard Deviation).

ARI: Automated Readability Index, CLI: Coleman–Liau Index, FKGL: Flesch–Kincaid Grade Level, FRE (reversed): Flesch Reading Ease, in reversed scale, GFI: Gunning–Fog Index, SMOG: Simple Measure of Gobbledygook, DCR: Dale-Chall Readability, LWRF: Linsear Write Readability Formula. Unlike all the other readability indices, Flesch Reading Ease (FRE) records the extent to which the text is easy; in other words, higher FRE scores reflect less sophisticated texts, whereas lower FRE scores indicate more sophisticated texts. To aggregate the indices more easily and keep their scales consistent, we reverse-coded the FRE calculations by subtracting the original FRE scores from the maximum possible value of 100. Refer to the *Readability* article on [www.Wikipedia.org \(https://en.wikipedia.org/wiki/Readability\)](https://en.wikipedia.org/wiki/Readability) for more details about these formulas.

Online Appendix M. Main Study 1 Model Results

As with the control model of the calculator reviews discussed in the main paper, the control variables here explained minimal variance of the DVs. In the headphone reviews (Table below), some of the control variables had significant coefficients: gender had a significant effect on Quality-1 implying that females tended to write lower-quality reviews judged by textual analyses; age had a significant negative effect on Quality-2 implying that older consumers tended to write lower-quality reviews judged by human raters.

The middle-column results show that relative to the control condition (i.e., no norms), the participants' behaviors of writing quality reviews in the high-norm condition were significantly higher (Quality-1 M1-2, $\beta_{\text{headphone, High}} = .69$, robust $t = 4.54$, $p < .001$, Cohen's $d = 0.91$; Quality-2 M2-2, $\beta_{\text{headphone, High}} = 1.04$, robust $t = 10.74$, $p < .001$, Cohen's $d = 2.15$). On the other hand, this behavior in the low-norm condition were significantly lower (Quality-1, $\beta_{\text{calculator, Low}} = -.78$, robust $t = -5.16$, $p < .001$, Cohen's $d = -1.03$; Quality-2, $\beta_{\text{calculator, Low}} = -1.05$, robust $t = -10.96$, $p < .001$, Cohen's $d = -2.119$).

Tables M-B and M-C present the results of additional analyses which make hypotheses testing more straightforward, especially H1. Instead of using the control group as the baseline and adding the high- and low-norm treatment dummies, these analyses used the low-norm group as baseline and added dummy terms for the high-norm and control groups.

Table M-A. Main Study Model Results, Control Condition as Baseline (Headphone)

	DV = Quality-1 Textual Analysis Composite Measure			DV = Quality-2 Human-Coded Composite Measure		
	M1-1	M1-2	M1-3	M2-1	M2-2	M2-3
Baseline: Intercept	2.63*(1.26)	1.68(1.02)	1.47(1.01)	3.02*(1.26)	1.61*(.65)	1.41*(.63)
Control Variables						
Past review experience	-.06(.05)	-.01(.03)	.00(.04)	-.09(.05)	-.02(.02)	-.02(.02)
Age	-.11(.06)	-.07(.05)	-.06(.05)	-.13*(.06)	-.07*(.03)	-.06*(.03)
Gender: M = 0, F = 1	-.31*(.14)	-.27*(.12)	-.25*(.12)	-.14(.14)	-.08(.07)	-.06(.07)
Main Effect						
PSP: Perceived social presence			-.05(.06)			-.05(.04)
Treatment Effect: (β s) Space Norm						
High		.69***(.15)	.64***(.16)		1.04***(.10)	1.0***(.10)
Low		-.78***(.15)	-.81***(.15)		-1.05***(.10)	-1.1***(.09)
Interaction Term (δ s)						
High * PSP			.03(.16)			-.07(.10)
Low * PSP			-.33*(.15)			-.35***(.10)
Model Fit						
F (df1, df2)	2.97*(3, 164)	21.24***(5, 162)	14.57***(8, 159)	3.22*(3, 164)	99.55***(5, 162)	69.46***(8, 159)
R^2	.052	.396	.423	.056	.754	.778

Notes: Maximum VIF of all models 1.52. Heteroscedasticity-robust standard error in parentheses.

* $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed test).

Table M-B. Main Study Model Results, Low-Norm Condition as Baseline (Calculator)

	DV = Quality-1 Textual Analysis Composite Measure		DV = Quality-2 Human-Coded Composite Measure	
	M3-1	M3-2	M4-1	M4-2
Baseline: Intercept	-.93(.93)	-1.12(.89)	-.95(.62)	-1.14*(.6)
Control Variables				
Past review experience	.00(.03)	.01(.03)	-.01(.02)	.00(.02)
Age	.00(.04)	.01(.04)	-.00(.03)	.00(.03)
Gender: M = 0, F = 1	-.04(.11)	-.04(.10)	.09(.07)	.10(.07)
Main Effect				
PSP: Perceived social presence		.10(.06)		.04(.04)
Treatment Effect: (β s) Space Norm				
High ^{H1}	1.74***(.14)	1.63***(.14)	2.16***(.09)	2.1***(.09)
Control	.87***(.14)	.94***(.13)	.97***(.09)	1.03***(.09)
Interaction Term (δ s)				
High * PSP ^{H2}		.55***(.14)		.4***(.09)
Control * PSP		.39**(.13)		.33***(.09)
Model Fit				
F (df1, df2)	32.45***(5, 162)	24.9***(8, 159)	111.0***(5, 162)	81.04***(8, 159)
R ²	.500	.556	.774	.803

Notes: H1: Results used to test H1; H2: Results used to test H2. Maximum VIF of all models 1.52.

Heteroscedasticity-robust standard error in parentheses. * $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed test).

Table M-C. Main Study Model Results, Low-Norm Condition as Baseline (Headphone)

	DV = Quality-1 Textual Analysis Composite Measure		DV = Quality-2 Human-Coded Composite Measure	
	M5-1	M5-2	M6-1	M6-2
Baseline: Intercept	.90(1.02)	.66(1.02)	-.56(.65)	.31(.63)
Control Variables				
Past review experience	-.01(.03)	.00(.04)	-.02(.02)	-.02(.02)
Age	-.07(.05)	-.06(.05)	-.07*(.03)	-.06*(.03)
Gender: M = 0, F = 1	-.27*(.11)	-.25*(.12)	-.08(.07)	-.06(.07)
Main Effect				
PSP: Perceived social presence		-.05(.06)		-.05(.04)
Treatment Effect: (β s) Space Norm				
High ^{H1}	1.46***(.15)	1.45***(.16)	2.09***(.10)	2.1***(.1)
Control	.78***(.15)	.81***(.15)	1.05***(.10)	1.1***(.09)
Interaction Term (δ s)				
High * PSP ^{H2}		.36*(.16)		.28**(.1)
Control * PSP		.33*(.15)		.35***(.1)
Model Fit				
F (df1, df2)	21.24***(5, 162)	14.57***(8, 159)	99.55***(5, 162)	69.46***(8, 159)
R ²	.396	.423	.754	.778

Notes: H1: Results used to test H1; H2: Results used to test H2. Maximum VIF of all models 1.52.

Heteroscedasticity-robust standard error in parentheses. * $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed test).

Online Appendix N. Moderated Regression Analyses

To clearly show the interaction effect between space norm and social presence (PSP), we followed the process prescribed by the moderated regression analysis (MRA; Sharma et al. 1981; Hayes 2013). First, we created two multiplicative interaction terms by multiplying the mean-centered (standardized) values of PSP and the two dummies for the space norm treatment (high- and low-norm). The control condition served as the baseline group.

Second, adapting from the formulas in Sharma et al. (1981), we ran the following three regressions (eq. 1-3). We denote y as the DV, a as the intercept, x_H and x_L as the treatment dummies (b_1 and b_2 their regression coefficients), z as PSP (coefficient b_3), and $z * x_H$ and $z * x_L$ as the two interaction terms (b_4 and b_5 their coefficients). In all these regressions, control variables were included but not shown.

$$y = a + b_1x_H + b_2x_L \quad \text{eq. 1}$$

$$y = a + b_1x_H + b_2x_L + b_3z \quad \text{eq. 2}$$

$$y = a + b_1x_H + b_2x_L + b_3z + b_4z * x_H + b_5z * x_L \quad \text{eq. 3}$$

Third, we present the treatment effects on writing quality reviews (i.e., group mean differences) that varies across different levels of the continuous moderator values (i.e., PSP) (Cohen 1988). Rearranging the terms in eq. 3, we have the following eq. 4. We use eq. 4 to draw graphs of the moderated effect and show the significant interaction. They depict the interaction effect where the low-norm treatment effect on writing quality reviews (i.e., heights of the points on y-axis) is evaluated at one standard deviation (SD) above and below the mean of PSP (i.e., x-axis). Please refer to Figure 1 in the main manuscript.

$$y = (a + b_3z) + (b_1 + b_4z) * x_H + (b_2 + b_5z) * x_L \quad \text{eq. 4}$$

Calculator and Quality-1

In the first two equations (eq. 1 and 2), the results in Table N-1 showed that space norm treatment significantly affects the review writing behavior, explaining 50 percent of the variance (Eq. 1). Both PSP and space norm treatment explain 50.8 percent of the variance of the review writing behavior (Eq. 2); moreover, the result indicated that PSP alone was not significantly associated with the review writing behavior (Eq. 2: $\Delta R^2 = .008$, $b_3 = .091$, robust $t = 1.582$, $p = .116$). In the third equation (eq. 3), we regressed the behavior of writing quality reviews on space norm treatment, PSP, and their interactions, in addition to the controls. The percent increase in the variance explained by the addition of the interaction terms was statistically significant (Eq. 3: $\Delta R^2 = .048$, $F(2, 159) = 8.613$, $p < .001$). Thus, H2 was supported for Quality-1 and Calculator data. The interpretation of the Figure 1a in the main manuscript was already provided there.

Table N-1. Moderated Regression Analysis (Calculator, DV = Quantity-1)

Eq. 1	Estimate	SE	Robust t	p	95% CI
Intercept (a)	-.063	.93	-0.068	.946	-1.900, 1.774
$x_H (b_1)$.869	.138	6.302	<.001	.597, 1.142
$x_L (b_2)$	-.872	.137	-6.39	<.001	-1.142, -.603
Model $R^2 = .500$, $F(5, 162) = 32.45$, $p < .001$.					
Eq. 2	Estimate	SE	Robust t	p	95% CI
Intercept (a)	.018	.927	.02	.984	-1.813, 1.85
$x_H (b_1)$.802	.144	5.592	<.001	.519, 1.086
$x_L (b_2)$	-.894	.137	-6.55	<.001	-1.165, -.625
$z (b_3)$.091	.058	1.582	.116	-.023, .207
Model $R^2 = .508$, $F(6, 161) = 27.71$, $p < .001$. $\Delta R^2 = .008$, $p = .116$.					
Eq. 3	Estimate	SE	Robust t	p	95% CI
Intercept (a)	-.174	.888	-.196	.845	-1.929, 1.58
$x_H (b_1)$.681	.14	4.855	<.001	.404, .958
$x_L (b_2)$	-.947	.133	-7.103	<.001	-1.211, -.684
$z (b_3)$.098	.056	1.766	.079	-.012, .208
$z * x_H (b_4)$.164	.138	1.198	.233	-.107, .437
$z * x_L (b_5)$	-.385	.134	-2.868	.005	-.651, -.12
Model $R^2 = .556$, $F(8, 159) = 24.9$, $p < .001$. $\Delta R^2 = .048$, $F(2, 159) = 8.613$, $p < .001$.					

Notes. All models included control variables but do not show their results. Table 3 in the main paper include those results.

Calculator and Quality-2

In the first two equations, the results in Table N-2 showed that space norm treatment significantly affects the review writing behavior, explaining 77.4 percent of the variance (Eq. 1). Both PSP and space norm treatment explain 77.5 percent of the variance of the review writing behavior (Eq. 2). PSP alone was not significantly associated with the review writing behavior (Eq. 2: $\Delta R^2 = .001$, $b_3 = .032$, robust $t = .814$, $p = .417$). In the third equation, we regressed the behavior of writing quality reviews on space norm treatment, PSP, and their interactions. The percent increase in the variance explained by the addition of the interaction terms was statistically significant (Eq. 3: $\Delta R^2 = .028$, $F(2, 159) = 11.295$, $p < .001$). Thus, H2 was supported for Quality-2 and Calculator data. A significant interaction between low-norm treatment and PSP was found ($b_5 = -.334$, robust $t = -3.736$, $p < .001$); the magnitude of this interaction effect might be considered “medium” (Green 1991). Relative to the control condition, subjects in the low-norm condition wrote significantly lower-quality reviews (Eq. 3: $b_2 = -1.028$, robust $t = -11.574$, $p < .001$). As PSP increased, subjects in the low-norm condition wrote even lower-quality reviews, above and beyond the main effect.

In Figure 1b of the main manuscript (upper-right chart), at one SD below the mean of PSP, i.e., when social presence was low, participants on average wrote relatively less-lower-quality reviews in the low-norm condition than in the control condition (i.e., $b_2 - b_5 = -1.028 - (-.334) = -.694$). When social presence was high (i.e., one SD above the mean), the group mean difference in review writing quality between the low-norm and the control conditions was much larger (i.e., $b_2 + b_5 = -1.028 + (-.334) = -1.362$).

Table N-2. Moderated Regression Analysis (Calculator, DV = Quantity-2)

Eq. 1	Estimate	SE	Robust <i>t</i>	<i>p</i>	95% CI
Intercept (a)	.011	.625	.019	.985	-1.223, 1.247
$x_H (b_1)$	1.189	.093	12.828	<.001	1.007, 1.373
$x_L (b_2)$	-.97	.092	-10.57	<.001	-1.151, -.789
Model $R^2 = .774$, $F(5, 162) = 111$, $p < .001$.					
Eq. 2	Estimate	SE	Robust <i>t</i>	<i>p</i>	95% CI
Intercept (a)	.04	.627	.064	.949	-1.198, 1.279
$x_H (b_1)$	1.166	.097	12.018	<.001	.975, 1.358
$x_L (b_2)$	-.977	.092	-10.586	<.001	-1.16, -.796
$z (b_3)$.032	.039	.814	.417	-.046, .11
Model $R^2 = .775$, $F(6, 161) = 92.46$, $p < .001$. $\Delta R^2 = .001$, $p = .417$.					
Eq. 3	Estimate	SE	Robust <i>t</i>	<i>p</i>	95% CI
Intercept (a)	-.121	.592	-.205	.838	-1.29, 1.047
$x_H (b_1)$	1.074	.093	11.496	<.001	.89, 1.259
$x_L (b_2)$	-1.028	.089	-11.574	<.001	-1.204, -.853
$z (b_3)$.036	.037	.992	.323	-.036, .11
$z * x_H (b_4)$.065	.092	.717	.475	-.115, .247
$z * x_L (b_5)$	-.334	.089	-3.736	<.001	-.511, -.158
Model $R^2 = .803$, $F(8, 159) = 81.04$, $p < .001$. $\Delta R^2 = .028$, $F(2, 159) = 11.295$, $p < .001$.					

Notes. All models included control variables but do not shown their results. Online Appendix M includes those results.

Headphone and Quality-1

In the first two equations, the results in Table N-3 showed that space norm treatment significantly affects the review writing behavior, explaining 39.6 percent of the variance (Eq. 1). Both PSP and space norm treatment explain 39.9 percent of the variance of the review writing behavior (Eq. 2). PSP alone was not significantly associated with the review writing behavior (Eq. 2: $\Delta R^2 = .003$, $b_3 = -.054$, robust $t = -.852$, $p = .396$). In the third equation, we regressed the behavior of writing quality reviews on space norm treatment, PSP, and their interactions. The percent increase in the variance explained by the addition of the interaction terms was statistically significant (Eq. 3: $\Delta R^2 = .024$, $F(2, 159) = 3.339$, $p = .037$). Thus, H2 was supported for Quality-1 and Headphone data. A significant interaction between low-norm treatment and

PSP was found ($b_5 = -.329$, robust $t = -2.149$, $p = .033$); the magnitude of this interaction effect might be considered “medium” (Green 1991). Relative to the control condition, subjects in the low-norm condition wrote significantly lower-quality reviews (Eq. 3: $b_2 = -.813$, robust $t = -5.35$, $p < .001$). As PSP increased, subjects in the low-norm condition wrote even lower-quality reviews, above and beyond the main effect.

Table N-3. Moderated Regression Analysis (Headphone, DV = Quality-1)

Eq. 1	Estimate	SE	Robust t	p	95% CI
Intercept (a)	1.675	1.023	1.638	.103	-.344, 3.696
$x_H (b_1)$.689	.152	4.543	<.001	.39, .988
$x_L (b_2)$	-.775	.15	-5.166	<.001	-1.072, -.479
Model $R^2 = .396$, $F(5, 162) = 21.24$, $p < .001$.					
Eq. 2	Estimate	SE	Robust t	p	95% CI
Intercept (a)	1.627	1.025	1.587	.115	-.398, 3.652
$x_H (b_1)$.728	.159	4.59	<.001	.415, 1.042
$x_L (b_2)$	-.761	.151	-5.044	<.001	-1.06, -.464
$z (b_3)$	-.054	.064	-.852	.396	-.182, .072
Model $R^2 = .399$, $F(6, 161) = 17.79$, $p < .001$. $\Delta R^2 = .003$, $p = .396$.					
Eq. 3	Estimate	SE	Robust t	p	95% CI
Intercept (a)	1.47	1.013	1.452	.149	-.53, 3.471
$x_H (b_1)$.643	.16	4.024	<.001	.328, .96
$x_L (b_2)$	-.813	.152	-5.35	<.001	-1.114, -.513
$z (b_3)$	-.05	.063	-.795	.428	-.175, .075
$z * x_H (b_4)$.029	.157	.19	.849	-.28, .34
$z * x_L (b_5)$	-.329	.153	-2.149	.033	-.632, -.027
Model $R^2 = .423$, $F(8, 159) = 14.57$, $p < .001$. $\Delta R^2 = .024$, $F(2, 159) = 3.339$, $p = .037$.					

Notes. All models included control variables but do not shown their results. Online Appendix M includes those results.

In Figure 1c of the main manuscript (lower-left chart), at one SD below the mean of PSP, i.e., when social presence was low, participants on average wrote relatively less-lower-quality reviews in the low-norm condition than in the control condition (i.e., $b_2 - b_5 = -.813 - (-.329) = -.484$). When social presence was high (i.e., one SD above the mean), the group mean difference

in review writing quality between the low-norm and the control conditions was much larger (i.e., $b_2 + b_5 = -.813 + (-.329) = -1.142$).

Headphone and Quality-2

In the first two equations, the results in Table N-4 showed that space norm treatment significantly affects the review writing behavior, explaining 75.4 percent of the variance (Eq. 1). Both PSP and space norm treatment explain 75.7 percent of the variance of the review writing behavior (Eq. 2). PSP alone was not significantly associated with the review writing behavior (Eq. 2: $\Delta R^2 = .003$, $b_3 = -.049$, robust $t = -1.207$, $p = .229$). In the third equation, we regressed the behavior of writing quality reviews on space norm treatment, PSP, and their interactions. The percent increase in the variance explained by the addition of the interaction terms was statistically significant (Eq. 3: $\Delta R^2 = .021$, $F(2, 159) = 7.462$, $p < .001$). Thus, H2 was supported for Quality-2 and Headphone data. A significant interaction between low-norm treatment and PSP was found ($b_5 = -.347$, robust $t = -3.654$, $p < .001$); the magnitude of this interaction effect might be considered “medium” (Green 1991). Relative to the control condition, subjects in the low-norm condition wrote significantly lower-quality reviews (Eq. 3: $b_2 = -1.097$, robust $t = -11.628$, $p < .001$). As PSP increased, subjects in the low-norm condition wrote even lower-quality reviews, above and beyond the main effect.

In Figure 1d of the main manuscript (lower-right chart), at one SD below the mean of PSP, i.e., when social presence was low, participants on average wrote relatively less-lower-quality reviews in the low-norm condition than in the control condition (i.e., $b_2 - b_5 = -1.097 - (-.347) = -.75$). When social presence was high (i.e., one SD above the mean), the group mean difference in review writing quality between the low-norm and the control conditions was much larger (i.e., $b_2 + b_5 = -1.097 + (-.347) = -1.444$).

Table N-4. Moderated Regression Analysis (Headphone, DV = Quantity-2)

Eq. 1	Estimate	SE	Robust <i>t</i>	<i>p</i>	95% CI
Intercept (a)	1.61	.652	2.469	.015	.322, 2.898
$x_H (b_1)$	1.039	.097	10.745	<.001	.848, 1.23
$x_L (b_2)$	-1.049	.096	-10.964	<.001	-1.238, -.86
Model $R^2 = .754$, $F(5, 162) = 99.55$, $p < .001$.					
Eq. 2	Estimate	SE	Robust <i>t</i>	<i>p</i>	95% CI
Intercept (a)	1.566	.652	2.401	.017	.278, 2.854
$x_H (b_1)$	1.074	.101	10.643	<.001	.875, 1.274
$x_L (b_2)$	-1.037	.096	-10.793	<.001	-1.227, -.847
$z (b_3)$	-.049	.041	-1.207	.229	-.13, .031
Model $R^2 = .757$, $F(6, 161) = 83.43$, $p < .001$. $\Delta R^2 = .003$, $p = .229$.					
Eq. 3	Estimate	SE	Robust <i>t</i>	<i>p</i>	95% CI
Intercept (a)	1.408	.629	2.239	.027	.166, 2.65
$x_H (b_1)$	1.002	.099	10.092	<.001	.806, 1.199
$x_L (b_2)$	-1.097	.094	-11.628	<.001	-1.284, -.911
$z (b_3)$	-.045	.039	-1.159	.248	-.123, .032
$z * x_H (b_4)$	-.069	.097	-.709	.479	-.261, .123
$z * x_L (b_5)$	-.347	.095	-3.654	<.001	-.535, -.16
Model $R^2 = .778$, $F(8, 159) = 69.46$, $p < .001$. $\Delta R^2 = .021$, $F(2, 159) = 7.462$, $p < .001$.					

Notes. All models included control variables but do not shown their results. Online Appendix M includes those results.

References

- Cohen, J. (1988). *Statistical Power Analysis for the Behavioral Sciences* (2nd Ed). Hillsdale, NJ: Lawrence Erlbaum.
- Green, S. 1991. "How Many Subjects Does It Take to Do a Regression Analysis?" *Multivariate Behavioral Research* (26), 499-510.
- Hayes, A. F. 2013. *Introduction to Mediation, Moderation, and Conditional Process Analysis: A Regression-based Approach*. New York: Guilford Press.
- Sharma, S., Durand, R. M., & Gur-Arie, O. 1981. "Identification and Analysis of Moderator Variables," *Journal of Marketing Research* (18:3), 291-300.

Online Appendix O. Post-Hoc Analysis: Ruling Out the Imitation Explanation

Assuming participants imitated (i.e., copied the words and the writing style of) what others say about the products (i.e., the treatment reviews), it follows that certain attributes of participant reviews would mimic those of the treatment reviews. We analyzed this possibility by using two approaches: (1) an established framework of natural language processing (NLP) for identifying authorship of online content (Zheng et al. 2006), (2) an unsupervised deep-learning method (latent Dirichlet allocation). We used both approaches because the former is more theory-driven and the latter is more data-driven; therefore, consistent results across the methods, if obtained, would bolster the conclusion.

Following Zheng et al. (2006), we focused on two key attributes. The first one is the *lexical* attribute of writing texts (Zheng et al. 2006). Specifically, we analyzed to what degree the high frequency words in the participant reviews overlapped with those in the treatment reviews. Table O-1 below shows the results of matched words between the participant reviews and the treatment reviews. More importantly, a baseline for the degree of word overlap due to random chance is needed for a meaningful comparison, because as long as the reviews are written for the same product, it is still highly likely that reviews would use the same word(s) to describe salient features of a given product (e.g., many review writers use the word “clear” to describe the sound feature of a quality headphone even though they have never seen any review from other writers). Thus, as a baseline level of random word overlap, we also calculated the degree to which the top-frequency words of the participant reviews matched those of the treatment reviews *unseen* to the focal participants — i.e., such treatment reviews were instead displayed to the participants in the other experimental condition — in which case any matched top-frequency words were only due to chance. The results below provided evidence that imitation (i.e., copying of words) was not

substantial. For the calculator reviews, in no participant review did we find over three matched words that also appeared in the *seen* treatment reviews; this same result was found in both the high- and low-norm conditions. Notice that in the baseline cross-condition result, none of the reviews had more than three matched words with the opposite-condition treatment reviews, indicating that the degree of overlap by chance was also minimal. Furthermore, only 5.6% of the high-norm participants wrote reviews containing more than one matched word that also appeared in the seen treatment reviews. This percentage was 27.8% in the low-norm condition. Although it seems like a high percent of overlap, this result has to be viewed with reference to the baseline result of 7.1%. We found a similar results pattern in the headphone reviews. Overall, none of the results indicated that the majority of the participants used more than one same word from the treatment reviews. Therefore, imitation of the treatment reviews by means of copying words was not serious.

Table O-1. Overlap of Top-Frequency Words of Participant and Treatment Reviews

	Threshold Number of Matched Words			
	> 1	> 2	> 3	> 4
Product: calculator				
Participants in high-norm condition	5.6%	0%	0%	0%
<i>Baseline: cross-condition word match</i>	<i>8.7%</i>	<i>3.2%</i>	<i>0%</i>	<i>0%</i>
Participants in low-norm condition	27.8%	3.9%	0%	0%
<i>Baseline: cross-condition word match</i>	<i>7.1%</i>	<i>0%</i>	<i>0%</i>	<i>0%</i>
Product: headphone				
Participants in high-norm condition	32.5%	12.7%	4.8%	0.8%
<i>Baseline: cross-condition word match</i>	<i>6.3%</i>	<i>0</i>	<i>0</i>	<i>0%</i>
Participants in low-norm condition	12.7%	2.3%	0%	0%
<i>Baseline: cross-condition word match</i>	<i>38.1%</i>	<i>14.3%</i>	<i>2.4%</i>	<i>0%</i>

Notes:

1. Each non-italic cell is the percent of the data sample with more than a threshold number of overlapped top-frequency words. The threshold is indicated in the respective column header. The percentages are calculated based on raw data following this procedure: after removing from the corpuses English stop words and words in the product names (see note below), we calculated the 10 most frequent words of each treatment review and

participant review; subsequently, we recorded the number of the top-frequency words in the participant review that also appeared in the list of the treatment reviews that the participant saw. Based on these raw comparison reviews, the above table summarizes the percentages of the dataset with more than a threshold number of matched top-frequency words, calculated by each product, by the high- and low-norm conditions, and by four different threshold numbers, respectively.

2. The results in the italic-typeface calculate the baseline percentages of top-frequency word overlap in the participant reviews of that condition with the *opposite-condition* treatment reviews; in other words, we compared the top-frequency-word list of participant reviews with that of the treatment reviews *unseen* to these participants, thus, the results indicate the *by-random-chance* percentages of matched top-frequency words by specific threshold numbers.
3. As for the list of stop words, we used the same R utility included in the “Text Mining Package” mentioned in the main manuscript.
4. As for the removed words in the product names, they are: “Texas instruments,” “TI,” “pro,” “engineering,” “scientific,” and “calculator” for the calculator; “Bohm,” “Bluetooth,” “headphone,” “microphone,” and “wireless” for the headphone.

The second attribute of writing texts is the sentence length of reviews, which is denoted as a *structural* attribute in Zheng et al. (2006)’s framework and is used by earlier work in similar domains (Yule 1938). If participants imitated the writing style of the treatment reviews, we would expect that the average sentence length of reviews written by participants in the high-norm condition should differ from that of reviews written by the low-norm participants (because the sentence lengths of the treatment reviews were different between the two experimental conditions; calculator treatment reviews: average sentence length 17.4 and 10, respectively, in high and low conditions; headphone: 11.2 and 9.1, respectively). The results are in Table O-2. For both products, participants in the high- and low-norm conditions wrote reviews with statistically the same sentence length despite the different sentence lengths in the treatment reviews.

In sum, the above analyses as inspired by Zheng et al. (2006)’s framework led us to conclude that the participants did not imitate the treatment reviews that they had seen.

Table O-2. T-Test of Equal Average Median Sentence Lengths of Participant Reviews between Experimental Conditions

	Experimental Conditions		T-test [†]	Assumption of equal variances [‡]
	High-Norm Condition	Low-Norm Condition		
Product: calculator				
Average Median Sentence Lengths of Participant Reviews	11.06	9.78	1.57 (.12)	Verified valid
Product: headphone				
Average Median Sentence Lengths of Participant Reviews	10	9.39	0.92 (.35)	Verified valid

Notes:

[†] p value of statistical test in parenthesis.

[‡] Test of homogeneous variance of the median sentence lengths of participant reviews between the high- and low-norm conditions used a robust parametric Levene's test (Levene 1960) and a robust nonparametric Fligner-Killeen test (Conover et al. 1981).

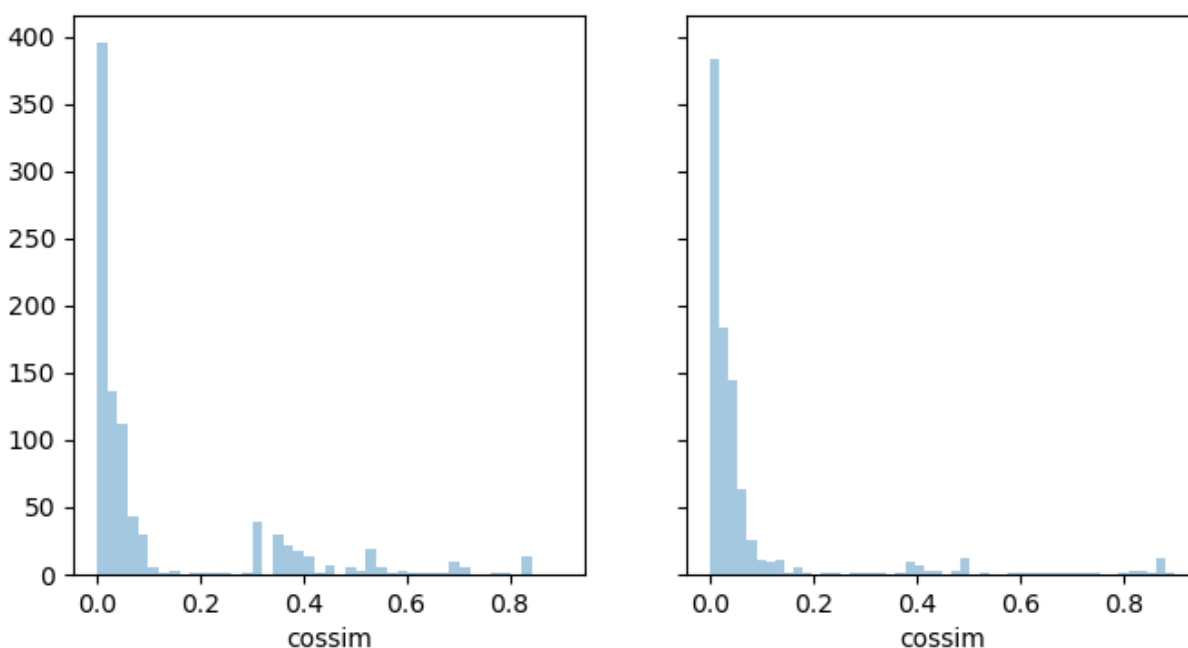
Next, we used an unsupervised machine-learning model to quantify similarity between each pair of reviews. The deep-learning model latent Dirichlet allocation (LDA) is a powerful NLP algorithm widely used in information science and business research (Wang and Chaudhry 2018). LDA automatically detects different topic dimensions emerging from a large body of texts. LDA is helpful in our context because product reviews may span several topic dimensions such as different features and functions (Wang and Chaudhry 2018). A participant could imitate others' reviews by not only copying words and sentences (as tested earlier), but also summarizing their opinions about the same set of product features (or service features). As such, a participant review with highly similar topic spaces (i.e., product features) may also be considered imitation. In this case, the earlier approach based on word overlap and sentence length might not work well, and we needed models such as LDA to detect topic space overlap.

We applied LDA to the entire texts of product reviews in the experiment. LDA requires to specify *ax-ante* the number of topic dimensions (denoted as *k*), which is not a mechanical

process but rather often requires subjective judgment involving the context, data size, and feedback from the actual results (Blei et al. 2003, Wang and Choudhry 2006). We ultimately chose $k = 10$ topic dimensions by following a rigorous process. First, we argue that the minimum k should not be lower than 5. The reason is that our specific purpose of conducting LDA is not the same as that of the original model. We adopted LDA to detect or rule out imitation behavior. If k was smaller than 5, the chance of “detecting” imitation behavior which in fact is not—a false positive error—would be high, i.e., over 20%. Second, by the same token, a value for k that is too large, e.g., over 50, may artificially inflate the chance of false negative errors. Therefore, we empirically analyzed numerous LDA by varying k between 5 and 50. Finally, we chose an intuitive number 10 by which we sought to balance “between having too many topics to be interpretable and too few to allow meaningfully variation” (Haans 2019, p.13). LDA assigned optimal weights to associate each word with the topic dimensions (Blei et al. 2003). Given these weight matrices and the emerged topic dimensions (see Wang and Chaudhry 2018), we obtained a 10-item weight vector for each review, and each weight represents the likelihood that the review is associated with one of the 10 topics. Using this setup and the weight vectors, we derived the cosine similarity score (CSS) for each pair of participant-treatment reviews. For more technical details about LDA model and cosine similarity score, refer to the Web Appendix of Wang and Choudhry (2006). In this study, we used gensim (version 3.8.2), a mature and popular open-source software (OSS) package in Python, for analyses.

CSS is normalized (bounded) in 0 and 1 and larger value means more similar, but there lacks an agreeable threshold for judging similarity (Wang & Chaudhry 2018). Figure below shows the distributions of CSS in our sample including each pair of a participant review with four treatment reviews seen by the focal participant ($n = 444$ in each product). Using an arbitrary but intuitive

threshold of 0.5, it is quite clear that the vast majority of participant reviews were dissimilar to their seen treatment reviews (percent CSS below 0.5: calculator 93%, headphone 95%).



Lacking an agreeable threshold, even these percentages may not be conclusive. We sought to establish a sensible benchmark for comparison. As discussed earlier, a reasonable baseline is the *by-random-chance* similarity with the *unseen* treatment reviews — these were not seen by the focal participants; instead, they were displayed in the other condition. If imitation took place, participant reviews are expected to be more similar to the treatment reviews displayed in the assigned condition than random. We therefore tested the following alternative hypothesis:

***Ha*: CSS of the pairs of reviews in the same condition (i.e., between participant reviews and seen treatment reviews) will be higher than that in the opposite conditions.**

We tested this hypothesis in the high-norm and low-norm subsamples. Because the experimental site displayed four treatment reviews simultaneously in each condition, we tested hypothesis using two versions of pooled sample: one pool with all possible pairs between each participant and four treatment reviews; another pool with only one pair, where we retained the highest CSS

among the four treatment reviews. The rationale for the highest CSS is that a participant may not necessarily imitate all four reviews; if most participants imitated only one of the seen reviews, keeping the participant-specific highest CSS would include the most likely imitation target.

We used the paired-sample t-test because the participant reviews used in each test are the same whereas the comparison reviews differ. We verified the equal variance assumption in all tests. Table O-3 below shows all results. Each thick box is a t-test comparing the mean CSS of the review pairs in the same-condition sample (shaded cell) against the opposite-condition sample (unshaded italic). The above hypothesis expected the shaded statistic to be higher than the unshaded italic statistic; moreover, it expected the difference to be statistically significant. As Table O-3 shows, all eight tests *failed to reject the null hypothesis of equal means* at a level of significance of at least 0.05; thus, these results rejected the alternative hypothesis stated above. Therefore, all participant reviews were as similar to the seen reviews as to those unseen treatment reviews. Given an overall average CSS of 0.16, these results clearly show that imitation in writing participant reviews was unlikely.

Table O-3. T-Test of Equal Similarity Means with Seen vs. Unseen Treatment Reviews

Participant Reviews	Samples of Treatment Reviews					
	1 Sample Includes All Treatment Reviews per Condition			2 Sample Includes Only the Most Similar Treatment Review Among All Four per Condition		
	High Norm	Low Norm	T-Test [†]	High Norm	Low Norm	T-Test [†]
Product: Calculator						
High-Norm Condition	.133	.174	-2.723 (.997)	.513	.508	.12 (.452)
Low-Norm Condition	.202	.220	.945 (.172)	.532	.595	1.378 (.086)
Product: Headphone						
High-Norm Condition	.141	.128	1.044 (.149)	.431	.407	.769 (.222)
Low-Norm Condition	.17	.159	-.904 (.817)	.461	.481	.595 (.277)

Notes: [†] *p* value of one-tailed t-test in parenthesis.

H0: shaded sample mean of similarity is equal to or lower than unshaded italic sample mean of similarity.

Ha: shaded sample mean of similarity is greater than the unshaded italic sample mean of similarity.

Similarity was measured by cosine similarity score (CSS). Test of homogeneous variance of similarity used a robust parametric Levene's test (Levene 1960) and a robust nonparametric Fligner-Killeen test (Conover et al. 1981).

All tests failed to reject the equal variance assumption.

Each thick box is a paired-sample t-test of mean similarity. Under a specific product and treatment assignment of participants (see row name), similarity is calculated between each participant review and each seen and unseen treatment review. Shaded cells denote mean similarity with the seen treatment reviews as the row and column headers match; unshaded and italicized cells denote mean similarity with the unseen treatment reviews (i.e., cross-condition). The unshaded mean similarity serves as a comparison baseline because participants did not see those treatment reviews and the observed similarity is generated by chance. Tests under column 1 pool all similarity scores of the condition, whereas tests under column 2 only retain one out of four treatment reviews that has the highest similarity with a participant review. In each t-test, because mean similarity in the shaded cell (seen reviews) is expected to be higher than the unshaded cell (unseen reviews), one-tailed tests were used.

Summary

We adopted two different NLP approaches to conduct extensive tests on the possibility of imitation behavior in participants' writing of product reviews during the experiment. Converging results from these analyses, as shown above in this Appendix, led us to conclude that imitation behavior could not have explained the main results of the experimental study.

References

- Blei, D. M., Ng, A. Y., & Jordan, M. I. 2003. "Latent Dirichlet Allocation," *Journal of Machine Learning Research* (3), 993-1022.
- Conover, W. J., Johnson, M. E., and Johnson, M. M. 1981. "A Comparative Study of Tests for Homogeneity of Variances, with Applications to the Outer Continental Shelf Bidding Data," *Technometrics* (23), pp. 351-361.
- Haans, R. F. J. 2019. "What's the Value of Being Different When Everyone Is? The Effects of Distinctiveness on Performance in Homogeneous Versus Heterogeneous Categories," *Strategic Management Journal* (40:1), 3-27.
- Levene, H. 1960. "Robust Tests for Equality of Variances," in *Contributions to Probability and Statistics: Essays in Honor of Harold Hotelling*, Olkin, I. (ed.), Palo Alto: Stanford University Press, pp. 278-292.
- Wang, Y. and Chaudhry, A. 2018. "When and How Managers' Responses to Online Reviews Affect Subsequent Reviews," *Journal of Marketing Research* (55:2), 163-177.
- Yule, G. U. 1938. "On Sentence Length as a Statistical Characteristic of Style in Prose," *Biometrika* (30), pp. 363-390.
- Zheng, R., Li, J., Chen, H., and Huang, Z. 2006. "A Framework for Authorship Identification of Online Messages: Writing-Style Features and Classification Techniques," *Journal of the American Society for Information Science and Technology*, (57:3), pp. 378-393.

Online Appendix P. Sample Product Review Pages


Sample Product Review Page of Aritzia.com (US)

198 styles marked down EVEN DEEPER. In stores and online. [Details](#)


USA (USD) | [STORE LOCATOR](#) | [CAREERS](#)
ARITZIA
SEARCH | [SIGN IN](#)

[Sale](#) [New](#) [Clothing](#) [Outerwear](#) [Brands](#) [Accessories](#)

Clothing / Sweaters / Salette Sweater



Styled with [Anise Pant](#).



Styled with [Anise Pant](#).

Reviews ★★★★★ 7

7 Customers Say It Fits True to Size

Smaller
Larger

True to Size

Sorted by [Most Recent](#)

Size Purchased:

★★★★★ THANK YOU FOR MAKING A SWEATER WITHOUT WOOL!!!

It's so hard to find nice, quality sweaters that don't contain wool. Thank you for making sweater season so much better! Drapes nicely, super soft and... [read more](#)

Megan C – 10 days ago

Size Purchased:

★★★★★ Salette sweater

This sweater is great with jeans or dressed up Love and fits great

Lisa C – 15 days ago

Fits: True to Size ↔

Usual Size: XS

Size Purchased: XS

Height: 5ft 4in

★★★★★ Color is beautiful very soft

Like the purple color very much, feel very comfortable and soft, but very easy be snagged.

Yan Z – 26 days ago

Size Purchased:

★★★★★ Perfect knit!

I love how this sweater slouches in all the right ways. It's super versatile and not too warm. The fabric is so soft as well! I purchased a size xs be... [read more](#)

Tracy R – 27 days ago

Size Purchased:

★★★★★ Love, love, love

I usually don't like sweaters that are cuffed at the bottom but this is an exception because it's so soft and stylish and can be worn with basically e... [read more](#)

Cynthia L – 1 month ago

Fits: True to Size ↔

Usual Size: XS

Size Purchased: XS

Height: 5ft 0in

★★★★★ Very pretty

Very soft and pretty sweater. Love the color (bought in pink)

Munn G – 1 month ago


Reviews 1 – 6 of 7
[Prev](#) [Next](#)

Questions? We have the answers.

Sample Product Review Page of USA.philips.com/Shop


PHILIPS Products
1-866-309-3263

[Home](#) > [Personal care](#) > [FACE Stylers and grooming kits](#) > [Multigroomers](#) > [Norelco Multigroom 3000 multipurpose trimmer](#)



360°

[Images](#) | [360°](#) | [Videos](#)



Norelco Multigroom 3000 multipurpose trimmer

MG3750/60

★★★★★ Ⓟ
[Read all 43 reviews](#) | [Write a review](#)

\$ 20.99

- ✔ 13 pieces
- ✔ tempered steel blades
- ✔ lithium power: 60 min runtime
- ✔ impact-resistant guards

Quantity:

Add to cart

Product Details
Specification
Reviews
Support

★★★★★

Overall rating 4.3 / 5

5 Stars	<div style="width: 27%; background-color: #0070C0;"></div>	27
4 Stars	<div style="width: 9%; background-color: #0070C0;"></div>	9
3 Stars	<div style="width: 3%; background-color: #0070C0;"></div>	3
2 Stars	<div style="width: 0%; background-color: #0070C0;"></div>	0
1 Star	<div style="width: 0%; background-color: #0070C0;"></div>	4

Performance 85% reviewers would recommend this product.

★★★★★

Ease of use ★★★★★

Design ★★★★★

Write a review

What is relevancy sort?

CrappieTim

From: East Texas

Gender: Male

Age: Over 50

I Have Used This Product For: **Less than a month**

★★★★★

5 Star

26 January 2019

I have used Norelco products for a very long time and have never had any problems. I'll never use anything else since my last try of another shaver. Nothing works as well at any cost. I'm very confident that this product will have a long life.

+ great products, dependable

✔ Yes, I would recommend this product.

Written by a customer while visiting philips.com

PHILIPS

Dagent

From: Miami, Florida

Gender: Male

Age: Over 50

I Have Used This Product For: **7-12 months**

★★★★★

I've used it for less than a year and the battery shorted on me.

14 January 2019

I love the design and performance but the battery life surprised me (less than a year).

+ design and performance

- battery life

PHILIPS

By Consumer Care

16 January 2019

Hello there,

Thank you for sharing your feedback, and we're very sorry to hear about your experience. Your utmost satisfaction is extremely important to us, as we pride ourselves on our quality and performance. Our Philips Norelco Multigroomers come with a two-year Warranty and an additional six months if you register your product. We also offer a 45-Day Money Back Guarantee on this product. At your earliest convenience, please get in touch

Online Appendix Q. Specification Tests for Study 2's Econometric Model

We carefully checked the assumption of using fixed effects (FE) model as opposed to random effects (RE) model. An RE model may also control for unobserved product-level effects; yet an underlying assumption of the RE model is that the unobserved factors are uncorrelated with any other included variables; violating this assumption would cause inconsistent model estimates (Wooldridge 2010). To test this key assumption (the null), we conducted Hausman tests (Hausman 1978) separately for the two datasets. The test rejects the null if the RE model produces inconsistent estimates, making the FE model the only option; or it fails to reject the null if both models produce consistent estimates (Hausman 1978). Our results rejected the null in both cases, thereby leading us to use the FE model (Philips data: Hausman $\chi^2(3) = 29, p < .001$. Aritzia data: Hausman $\chi^2(3) = 59.94, p < .001$).

References

- Hausman, J. A. 1978. "Specification Tests in Econometrics," *Econometrica* (46:6), 1251-1271.
Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data* (2nd Ed).
Cambridge, MA: The MIT Press.

Online Appendix R. Study 2 Robustness Checks

Robustness Checks: Alternative Model Results

One potential concern about panel-data analysis is the serial correlation in longitudinal observations. Due to the small sample size of repeated observations for each product (maximum 3), modeling the serial correlation was infeasible (Wooldridge 2010). Thus, for robustness, we ran several OLS models separately for each review rank (4th, 5th, or 6th review) on each site. The results in Table R-1 are similar to the FE models discussed in the main manuscript. These results indicated a positive and significant effect of Mean Quality across the review ranks and study sites (all $\beta_{\text{Mean Quality}} \geq .44$, $t \geq 2.142$, $p \leq .036$). These consistent results may relieve any potential concern about the panel-data FE model results.

Table R-1. Alternative Model Results (OLS)

Variable	Philips Data			Aritzia Data		
	4 th Review O1-1	5 th Review O1-2	6 th Review O1-3	4 th Review O2-1	5 th Review O2-2	6 th Review O2-3
Intercept	-.19(.26)	-.11(.66)	.05(.56)	-.36(.73)	.72(.85)	2.18 [†] (1.17)
Rating	.05(.06)	.14(.15)	-.00(.13)	.08(.15)	-.13(.17)	-.41 [†] (.24)
Quality Variance	.01(.08)	-.17(.19)	-.03(.14)	-.09(.09)	-.11(.15)	-.47 [†] (.24)
Mean Quality	.44 ^{**} (.16)	1.55 ^{***} (.33)	1.32 ^{***} (.28)	.67 [*] (.31)	.57 [*] (.26)	.60 [*] (.27)
Model Fit						
<i>F</i> (df1, df2)	3.40 [*] (3, 64)	7.945 ^{***} (3, 64)	12.48 ^{***} (3, 63)	2.985 [*] (3, 91)	2.303 [†] (3, 63)	2.89 [*] (3, 48)
<i>R</i> ²	.137	.271	.373	.122	.099	.137
Largest VIF	1.21	1.26	1.79	1.52	1.78	2.09

Notes: [†] $p < .10$, * $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed test). Heteroscedasticity-robust standard error in parentheses. Largest VIF well below 5.

Robustness Checks: More Control Variables and Effects

Besides the controls such as Rating and Quality Variance in the main analyses, we conducted additional robustness checks to further evaluate the robustness of the main result (i.e., Mean Quality). It may be argued that the effect of space norms could be influenced by the ratings

already displayed on the product review page (“Mean ratings”), or by their variance (“Variance of prior ratings”); further, space norms could be moderated by Quality Variance or Mean ratings. We tested these various possibilities, and the results are shown below, separate for the Philips and Aritzia data. It is quite clear that the explanatory power of the fixed effects (FE) models did not improve significantly after including any of the possibilities mentioned above; further, none of their regression coefficients tested significantly. We conclude that despite their plausibility, none of these possible effects was confirmed in the two field sites that we examined.

Table R-2. Panel-Data Model Results (Philips data)

Variable	O3-1	O3-2	O3-3	O3-4	O3-5
Mean Quality	1.19**(.45)	1.43**(.56)	1.19**(.46)	1.14**(.45)	1.2**(.48)
Rating (star rating of focal review)	.03(.07)	.01(.07)	-.01(.09)	-.01(.09)	.03(.08)
Quality Variance	-.05(.18)	.17(.31)	-.04(.18)	-.09(.18)	-.05(.18)
Quality Variance × Mean Quality		.11(.24)			
Mean ratings			-.36(.45)	-.03(.48)	
Mean ratings × Mean Quality				-.79(.54)	
Variance of prior ratings					-.04(.14)
Product fixed effects (FE)	included	included	included	included	included
Model Fit					
R^2	.627	.631	.629	.639	.628
Sample Size	203	203	203	203	203

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed test). Panel-data fixed effects (FE) models. Intercept omitted due to fixed effects. Heteroscedasticity-robust standard error in parentheses. Largest VIF of all models well below 5.

Table R-3. Panel-Data Model Results (Aritzia data)

Variable	O4-1	O4-2	O4-3	O4-4	O4-5
Mean Quality	2.74***(.34)	2.81***(.41)	2.77***(.34)	2.77***(.34)	2.66***(.34)
Rating (star rating of focal review)	-.08(.10)	-.08(.10)	-.04(.13)	-.03(.12)	-.15(.12)
Quality Variance	-.08(.17)	-.11(.22)	-.07(.18)	-.07(.18)	-.1 (.18)
Quality Variance × Mean Quality		.06(.2)			
Mean ratings			.32(.55)	.31(.55)	
Mean ratings × Mean Quality				.02(.73)	
Variance of prior ratings					.35(.3)
Product fixed effects (FE)	included	Included	included	Included	included
Model Fit					
R^2	.693	.693	.694	.694	.696
Sample Size	214	214	214	214	214

Notes: * $p < .05$, ** $p < .01$, *** $p < .001$ (two-tailed test). Panel-data fixed effects (FE) models. Intercept omitted due to fixed effects. Heteroscedasticity-robust standard error in parentheses. Largest VIF of all models well below 5.

References

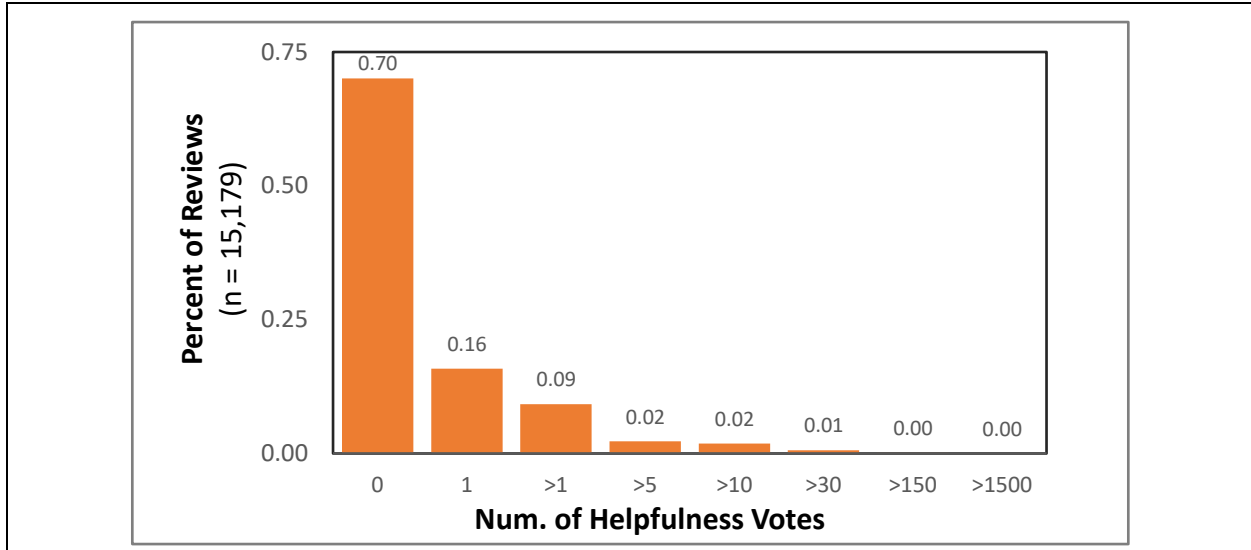
Wooldridge, J. M. 2010. *Econometric Analysis of Cross Section and Panel Data* (2nd Ed). Cambridge, MA: The MIT Press.

Online Appendix S. Product Review Quality in the Marketplace

The goal of this analysis is to examine the prevalence of low-quality customer product reviews in the marketplace. Utilizing a customized crawler, we collected 15,179 reviews from Amazon.com for 54 highly popular products in two product types, i.e., 34 consumer electronics (search products) and 20 video games (experience products).¹ We used the number of helpfulness votes a review received as a proxy for the quality review content. The below Figure displays the percent of them that have had a certain number or range of helpfulness votes. We found that 70% of reviews had zero helpfulness vote, and only less than 15% of reviews had more than one helpfulness vote. While we acknowledge that user-assessed helpfulness vote is not the sole determiner of quality reviews, it nevertheless presents a commonly-applied indicator in literature. This analysis indicates that low-quality reviews make up the majority of customer product reviews in the marketplace.

¹ We think the included products constitute a representative sample. The selection of products under each product type covered a broad range of popular brands, product lines, and their features. For instance, under the search products, popular laptop choices such as Apple MacBook Air, Microsoft Surface, and ASUS Chromebook Flip were included and they varied widely across the pricing levels, i.e., from almost \$2,000 to below \$300. In terms of product lines, the search products included laptops, smartphones, smartwatches, and other electronics including headphones, wireless keyboards, and wireless mouse. The selection of 20 video games covered a wide variety of consoles and titles, for instance, FIFA 18' for PlayStation 3, Lego Batman for Wii, etc.

Figure. Percent of Reviews by the Number of Helpfulness Votes



Notes: These popular products include: (1) laptops in various price levels, e.g., Apple’s MacBook Pro, Microsoft Surface, ASUS Chromebook, etc.; (2) smartphones, e.g., Apple’s iPhone 7, Samsung’s Galaxy S8, Microsoft Lumia 640, etc.; (3) headphones, e.g., Bose QuietComfort, Jelly Comb, etc.; (4) smartwatches, e.g., Fitbit; (5) miscellaneous consumer electronics, e.g., keyboard and mouse, wireless laser presenter, etc.; (6) video games, e.g., FIFA 18’, Final Fantasy XIII, Wii Sports Resort, Mario Party 9, etc. For each product, all reviews with a posted date no later than November 14, 2017 were collected and included in analysis.