Should We Reach for the Stars? Examining the Convergence between Online Product Ratings and Objective Product Quality and Their Impacts on Sales Performance

Sarah Köcher*

Doctoral Candidate in Marketing
Department of Marketing
TU Dortmund University
sarah.koecher@tu-dortmund.de

Sören Köcher

Assistant Professor of Marketing
Department of Marketing
TU Dortmund University
soeren.koecher@tu-dortmund.de

Suggested citation:
Köcher, Sarah and Sören Köcher (2018), "Should We Reach for the Stars? Examining the Convergence Between Online Product Ratings and Objective Product Quality and Their Impacts on Sales Performance," *Journal of Marketing Behavior*, 3 (2), 167–183. http://dx.doi.org/10.1561/107.00000050

* Corresponding author

The two authors contributed equally to this paper. The authors thank the replication editor and the reviewers for their valuable feedback.

Should We Reach for the Stars? Examining the Convergence between Online Product Ratings and Objective Product Quality and Their Impacts on Sales Performance

Abstract

By documenting that online ratings poorly correlate with quality scores provided by Consumer Reports—presumably a measure of 'objective' product quality—de Langhe *et al.* (2016a) found that consumers rely more heavily on such ratings when making quality inferences than they should. Aside from replicating this finding, we examine the moderating effect of product age on the convergence between objective and rated quality and investigate which quality indicator is a better predictor of sales performance.

Keywords: Online product ratings, Objective quality, Sales performance, Product age

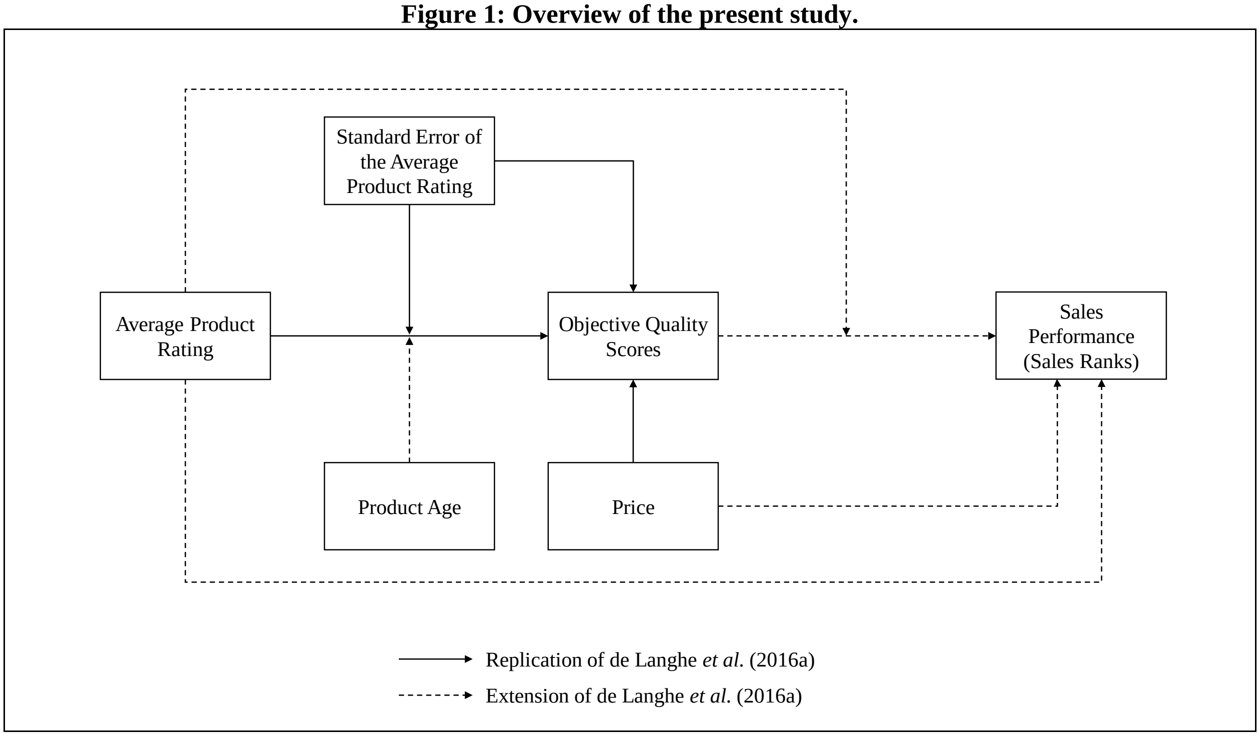
Should We Reach for the Stars? Examining the Convergence between Online Product Ratings and Objective Product Quality and Their Impacts on Sales Performance

Introduction

De Langhe, Fernbach, and Lichtenstein (2016a) recently discovered a substantial gap between the extent to which consumers trust in online product ratings when making inferences about the quality of a product and the actual validity of such ratings as an indicator of a product's 'objective' performance. More precisely, across a series of consumer studies the authors found that people place enormous weight on average product ratings when assessing the quality of a product, while the convergence between average ratings and the quality scores provided by *Consumer Reports* (CR)—presumably a measure of objective quality—and, thus, their validity as a quality indicator, is evidentially weak. These findings have caused a lively discussion among several eminent researchers (De Langhe *et al.* 2016b; Kozinets 2016; Simonson 2016; Winer and Fader 2016) primarily questioning the actual relevance of the reported results, the reliability of CR scores as a measure of objective quality, as well as the simplicity of analysis neglecting consumer heterogeneity and dynamic changes in product ratings over time. Some of the stated assertions, however, lack empirical evidence.

The purpose of our study is to contribute to this debate by empirically testing three of the critics' annotations: First, Simonson (2016) doubted that CR scores actually capture objective product quality; inter alia referring to an occasion where CR's methodology has come under severe criticism. We tested whether the claimed inaccuracy and distortedness of CR ratings per se are the ultimate source of the low convergence between rated and objective quality by replicating de Langhe *et al.*'s (2016a) findings using a database very similar to the one they have used. However, instead of CR scores, we collected the quality scores provided by *Stiftung Warentest*—the German equivalent of CR—and inspected their convergence with ratings provided on Amazon's German website. Second, Winer and Fader (2016) asserted that the correlation between rated and objective quality is determined by the dynamics of reviews and criticize the lack of dynamic aspects in the original study. Among other considerations, they suggested that the correlation between objective and rated quality may change over a product's life cycle. Inspired by this assumption we examined potential differences in the convergence between objective and rated quality scores across older and newer products and, thereby, extend the original work. Finally, motivated by a further suggestion of the discussants (Winer and Fader 2016; Simonson 2016) we investigated the extent to which different pieces of quality information influence purchase behavior by examining the impact of rated and objective quality on sales performance. Figure 1 illustrates the relationships put under scrutiny in the present study.

Figure 1: Overview of the present study.



Data

The German consumer organization *Stiftung Warentest* publishes a monthly magazine with tests of a variety of consumer products. We downloaded all tests of consumer electronic products that had been published between 2014 and 2017 from the organization's website and extracted quality scores[1] for each of the tested items. This resulted in quality ratings for 2,473 products across 352 categories. As in de Langhe *et al.* (2016a), we defined product categories at the lowest level of abstraction (e.g., we considered over-ear and on-ear Bluetooth headphones as separate categories). In addition, if a category had been tested multiple times, we treated each test as an individual subcategory (e.g., smartphones tested in February 2014 and smartphones tested in November 2017 were considered as separate subcategories) such that items within a subcategory were relatively homogeneous and quality ratings were comparable. We then searched the Amazon.de website for each product for which we had a *Stiftung Warentest* score and recorded all product ratings, selling prices, bestseller ranks, and launch dates. We restricted the database to items that have been rated five or more times, and categories comprising at least three products. Our final database consisted of 1,322 products across 224 categories[2].

Replications

Simple Correlations

We first calculated the Pearson correlation between average ratings and *Stiftung Warentest* scores for each of the 224 product categories. Similar to the original findings, the average correlation was only 0.18, and 36.3% of the correlations were negative.

---

[1] In contrast to the quality scores provided by CR, *Stiftung Warentest* scores range from 1 to 6 (with one decimal place); with lower values indicating better quality. To facilitate the comparison of our results with those reported by de Langhe *et al.* (2016a), we reversed these scores in our data set.

[2] See Appendix for a list of all product categories.

Regression Analyses

We then examined the convergence between average ratings and *Stiftung Warentest* scores using regression analyses. In a first step, we regressed *Stiftung Warentest* scores on the average rating, the standard error (SE) of the average rating—as a measure of the accuracy of the average rating—and the interaction between the average rating and the SE. As in de Langhe *et al.* (2016a), we z-standardized all variables by subcategory. A comparison of parameter estimates and confidence intervals (CIs) with de Langhe *et al.*'s (2016a) results appears in Table 1.

Table 1: Parameter estimates (and confidence intervals) for the original and present study.

	De Langhe *et al.* (2016a)		Present study	
	Model A	Model B	Model A	Model B
Dependent variable	*Consumer Reports* quality scores		*Stiftung Warentest* quality scores	
Independent variables				
Average rating	0.16 (0.10 to 0.22)	0.09 (0.03 to 0.15)	0.13 (0.07 to 0.19)	0.08 (0.02 to 0.13)
Price		0.34 (0.28 to 0.39)		0.31 (0.26 to 0.36)
Standard error	–0.13 (–0.20 to –0.07)	–0.15 (–0.21 to –0.09)	–0.15 (–0.21 to –0.09)	–0.18 (–0.23 to –0.12)
Average rating × standard error	–0.06 (–0.12 to –0.01)	–0.07 (–0.12 to –0.02)	–0.01 (–0.06 to 0.04)	–0.01 (–0.05 to 0.04)
Average correlation between average rating and objective quality scores	0.18		0.18	
Percentage of negative correlations	34 %		36.3 %	
Data source of independent variables	Amazon.com		Amazon.de	
Number of observations	N = 1,272 products across 120 categories		N = 1,322 consumer electronic products across 224 categories	

Basically, our results support the findings of de Langhe *et al.* (2016a) regarding the weak relationship between average ratings and objective quality scores (b = 0.13, CI_{95}: 0.07 to 0.19; see Model A). However, although we also found a negative main effect of SE (b = –0.15, CI_{95}: –0.21 to –0.09) on *Stiftung Warentest* scores, we unexpectedly could not support the reported interaction between average rating and SE (b = –0.01, CI_{95}: –0.06 to 0.04); i.e., the correspondence between average rating and *Stiftung Warentest* scores in our sample was independent of the SE[3]. We assume that the lack of interaction might be traced back to at least one of two differences between our database and the one used by de Langhe *et al.* (2016a). First, although both the SE (M = 0.20 vs. $M_{\text{de Langhe et al.}}$ = 0.22, Median = 0.16 vs. $\text{Median}_{\text{de Langhe et al.}}$ = 0.17) and the SD (M = 1.32 vs. $M_{\text{de Langhe et al.}}$ = 1.31, Median = 1.35 vs. $\text{Median}_{\text{de Langhe et al.}}$ = 1.36) were similarly distributed, the distribution of the number of ratings was somewhat different; while the average number of ratings in our data set was considerably smaller (M = 181 vs. $M_{\text{de Langhe et al.}}$ = 271), the median number of ratings was larger (Median = 63 vs. $\text{Median}_{\text{de Langhe et al.}}$ = 50). Hence, the SEs in our database were based on different sample sizes. Second, the average number of products per category was remarkably larger in the original study ($M_{\text{de Langhe et al.}}$ = 10.6; 1,272 products across 120 product categories) than in ours (M = 5.9; 1,322 products across 224 product categories).

In a next step, we benchmarked the effect of average ratings on *Stiftung Warentest* scores against that of price (see Model B). This analysis revealed significant effects of average rating (b = 0.08, CI_{95}: 0.02 to 0.13), SE (b = –0.18, CI_{95}: –0.23 to –0.12), and price (b = 0.31, CI_{95}: 0.26 to 0.36) on *Stiftung Warentest* scores. Again, the interaction between average rating and SE was not significant (b = –0.01, CI_{95}: –0.05 to 0.04). To evaluate the relative amount of unique variance in

[3] We also estimated another regression model including, in addition to the average rating, the standard deviation (SD) and the number of ratings (i.e., the two components of the SE) as well as their interactions with average rating. This analysis revealed that neither the SD (b = –0.02, CI_{95}: –0.08 to 0.03) nor sample size (b = –0.04, CI_{95}: –0.10 to 0.01) moderated the relationship between average ratings and *Stiftung Warentest* scores.

Stiftung Warentest scores explained by price and average rating, we computed squared semipartial correlations. Price uniquely explained 9.61% of the variance, 21 times more than average rating ($sr^2_{average\ rating}$ = 0.46%); implying that the price of a product is a much better indicator of its quality than its average rating. In sum, these results basically match up with the original findings; indicating that the CR scores used by de Langhe *et al.* (2016a) cannot be held responsible for the low convergence between rated and objective quality.

Extensions

The Moderating Effect of Product Age on the Relationship between Rated and Objective Quality

In their commentary on the original work, Winer and Fader (2016) speculated that the correlation between online ratings and objective performance may be different early in a product's life cycle versus later. They argue that since early adopters tend to be more knowledgeable than later adopters, the correlation between early adopters' ratings and objective quality should be higher than that of later adopters. If this is so, then the correspondence between the average ratings of products that have already been on the market for a relatively long period of time and objective quality scores should be weaker than that of newer products.

Data

To explore this conjecture, we screened our database for product categories that have been evaluated by *Stiftung Warentest* more than once during the last four years. This resulted in a database of 546 products across 29 categories that have been tested, on average, 2.8 times. As a measure of product age, we calculated for how long each product had already been available on Amazon.de using its launch date. The average age of the products in this data set was 32.6 months (Median = 34). The average range of product age within the 29 categories was 35.2 months (Median

= 32). Items that had been evaluated in the earliest test of a category were significantly older (M = 41.2 months, SD = 14.4) than those that had been assessed in the most recent test of the same product category (M = 24.3 months, SD = 14.9; t = 10.51, p < .01)[4]. The average number of products within a category was 18.8 (Median = 12) and, on average, these products have received 191 ratings (Median = 74).

Results

We regressed *Stiftung Warentest* scores on the average rating, the age of the product, and the interaction between average rating and product age. Before running the analysis, we z-standardized *Stiftung Warentest* scores and average ratings by each test of each product category (e.g., smartphones tested in February 2014 vs. smartphones tested in November 2017) and standardized product age by category (e.g., smartphones). This analysis revealed a positive effect of average rating on *Stiftung Warentest* scores (b = 0.17, CI_{95}: 0.09 to 0.25) as well as an average rating × product age interaction (b = –0.11, CI_{95}: –0.20 to –0.03) such that average ratings of newer products were stronger related to *Stiftung Warentest* scores (–1 SD from the average age: b = 0.28, CI_{95}: 0.16 to 0.40) than average ratings of older products (+1 SD from the average age: b = 0.10, CI_{95}: –0.06 to 0.18); the main effect of product age was not significant (b = –0.01, CI_{95}: –0.09 to 0.07)[5].

[4] *Stiftung Warentest* quality scores did not differ between products that have been tested earlier (M = 4.38, SD = 0.63) versus later (M = 4.46, SD = 0.57; t = 1.33; p = 0.18); precluding that only older products with high quality assessments maintained on the market. In addition, *Stiftung Warentest* scores between products that were available on Amazon (M = 4.42, SD = 0.59) and those that were not available (M = 4.34, SD = 0.62) were not significantly different (t = 1.36, p = 0.17).

[5] We also examined whether the relationship between product age and *Stiftung Warentest* scores might have been non-linear. However, none of the tested alternative model specifications (i.e., logarithmic, quadratic, inverse, and exponential) yielded a significant effect of product age.

The Impacts of Product Ratings and Objective Quality Scores on Sales Performance

Winer and Fader (2016) and Simonson (2016) likewise expressed their interest in a better understanding regarding the extent to which different pieces of available quality information affect decision making and, consequently, sales. Thus, we examined which predicts a product's sales performance better, average ratings or objective quality scores.

Data

As in previous studies (e.g., Chevalier and Mayzlin 2006; Sun 2012; see also Floyd et al. 2014), we used Amazon's bestseller ranks as an inverse proxy for sales performance. Hence, aside from the initially described restriction criteria, we had to further limit our database to product categories comprising three or more products that have been ranked within a common category on Amazon.de (e.g., 'Camera & Photo'). This resulted in a database of 1,220 products across 213 categories.

Results

We stepwise investigated the impacts of rated and objective quality on sales ranks. We first regressed the bestseller rank on the average rating and *Stiftung Warentest* score (Model C). Then, we added the interaction between rated and objective quality to the regression model (Model D). Finally, we incorporated selling prices as a covariate (Model E). In sum, across the three estimated models, we found that both objective and rated quality affected sales ranks (see Table 2).

Table 2: The effects of rated and objective product quality on sales performance.

	Model C	Model D	Model E
Dependent variable		*Amazon Bestseller Ranks*	
Independent variables			
Average rating	–0.15 (–0.20 to –0.09)	–0.14 (–0.20 to –0.09)	–0.15 (–0.21 to –0.101)
Stiftung Warentest quality scores	–0.22 (–0.28 to –0.17)	–0.22 (–0.27 to –0.16)	–0.28 (–0.33 to –0.22)
Average rating × *Stiftung Warentest* quality scores		0.08 (0.02 to 0.14)	0.07 (0.01 to 0.13)
Price			0.19 (0.14 to 0.25)

Note: All variables were z-standardized by subcategory before analysis.

Next, we inspected the relative amount of unique variance in bestseller ranks explained by each predictor using squared semipartial correlations. Interestingly, objective quality scores uniquely explained the highest proportion of variance ($sr^2_{SW\ scores}$ = 6.60%; sr^2_{price} = 3.34%; $sr^2_{average\ rating}$ = 2.22%; Model E). Please note, this finding does not necessarily imply that consumers actually consulted the quality judgments provided by *Stiftung Warentest*. Instead, consumers might be at least partially able to infer a product's objective performance on their own using the provided product information (e.g., they may conclude that a vacuum cleaner with 800 watts performs better than a vacuum cleaner with 600 watts). In other words, although we cannot make a statement about whether or not consumers care about the quality scores provided by *Stiftung Warentest* per se, our findings indicate that consumers seem to care about what they convey. In fact, the information these scores represent can explain three times more variance in sales ranks than average ratings. However, in addition to the described findings, we also found a significant average rating × *Stiftung Warentest* score interaction (see Model D and E); indicating that the influence of each of the two pieces of quality information decreases with the favorability of the other. Pessimistically stated, this finding implies that purchase behavior is less reliant on the objective performance of a product as its rated quality increases.

General Discussion

This study replicates and extends de Langhe *et al.*'s (2016a) seminal work on the limited convergence between online product ratings and measures of objective product performance which has been controversially discussed among several eminent marketing researchers. With this paper, we contribute to this discussion in three important ways. First, by replicating the original findings using a different data source for objective quality information (i.e., product assessments published by *Stiftung Warentest*), we rule out that the detected low convergence has to be merely ascribed to methodological defects in the product evaluations provided by CR. Second, inspired by Winer and Fader (2016), we examined potential differences in the correlation between rated and objective quality across products that have already been on the market for a relatively long period of time and newer products. Our findings reveal that the relationship between average ratings and objective quality scores is negatively moderated by product age. Hence, in particular, when trying to get an impression of the quality of older products, people should be careful not to be misled by average ratings. Please note that, although this result is in line with Winer and Fader's (2016) posit that the correlation between online ratings and objective performance may decrease over a product's life cycle, a conservative and explicit test of this postulate would investigate the relationship between objective quality scores and periodical average ratings using the dates on which online ratings have been posted while keeping the products under consideration constant. Finally, our investigation of the degree to which both rated and objective quality influence sales performance reveals that the information conveyed by objective quality scores is three times more influential in driving sales ranks—a commonly used proxy for a product's sales performance—than average ratings.

Furthermore, we found that the relationship between objective quality and bestseller ranks decreases as the favorability of rated quality increases. Thus, high customer ratings may even be able to disguise a product’s objective quality to some degree.

References

Chevalier, Judith A. and Dina Mayzlin (2006), "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research*, 43(3), 345–354.

De Langhe, Bart, Philip M. Fernbach, and Donald R. Lichtenstein (2016a), "Navigating by the Stars: Investigating the Actual and Perceived Validity of Online User Ratings," *Journal of Consumer Research*, 42(6), 817–833.

De Langhe, Bart, Philip M. Fernbach, and Donald R. Lichtenstein (2016b), "Star Wars: Response to Simonson, Winer/Fader, and Kozinets," *Journal of Consumer Research*, 42(6), 850–857.

Floyd, Kristopher, Ryan Freling, Saad Alhoqail, Hyun Young Cho, and Traci Freling (2014), "How Online Product Reviews Affect Retail Sales: A Meta-Analysis," *Journal of Retailing*, 90(2), 217–232.

Kozinets, Robert V. (2016), "Amazonian Forests and Trees: Multiplicity and Objectivity in Studies of Online Consumer-Generated Ratings and Reviews, A commentary on de Langhe, Fernbach, and Lichtenstein. *Journal of Consumer Research*, 42(6), 834–839.

Simonson, Itamar (2016), "Imperfect Progress: An Objective Quality Assessment of the Role of User Reviews in Consumer Decision Making, A Commentary on de Langhe, Fernbach, and Lichtenstein," *Journal of Consumer Research*, 42(6), 840–845.

Sun, Monic (2012), "How Does the Variance of Product Ratings Matter?" *Management Science*, 58(4), 696–707.

Winer, Russell S. and Peter S. Fader (2016), "Objective vs. Online Ratings: Are Low Correlations Unexpected and Does It Matter? A Commentary on de Langhe, Fernbach, and Lichtenstein," *Journal of Consumer Research*, 42(6), 846–849.

Appendix: Product categories and Pearson correlations between average ratings from Amazon.de and *Stiftung Warentest* quality scores.

Category	Issue	Correlation	Category	Issue	Correlation
All-in-one inkjet printers (✓)			Camcorders (✓)	6/17	0.718
*with fax**	4/15	-0.711	*action camcorders**	8/14	-0.262
	4/17	0.041		7/16	0.619
with fax and automatic document feeder	4/16	-0.392	*HD camcorders with hard drive*	10/14	0.791
*without fax**	4/16	-0.667	Cameras (✓)		
	4/17	0.756	*small models with large zoom**	9/14	-0.016
All-in-one laser printers (✓)				9/15	0.230
*black and white laser (with fax)**	4/14	-0.226	*large models with extra large zoom**	9/14	-[a]
	9/17	0.966		9/15	-0.058
*black and white laser (without fax)**	10/14	0.081	Simple compact models		
	9/16	0.881	*super zoom*	9/16	0.410
color laser (with fax)	9/17	0.345	*standard zoom*	9/16	0.475
color laser (without fax)	9/16	-0.662	Premium compact models	12/14	-0.471
Baby monitors (✓)			*standard zoom*	9/16	-0.347
audio models	5/15	0.505	*compact models*	12/15	0.476
video models	5/15	0.904	*with zoom lens*	1/14	0.160
Blood pressure monitors (✓)			*robust cameras*	7/14	-0.817
wrist models	5/16	0.656	Camera travel lenses (✓)		
arm models	5/16	0.226	*for Canon*	3/16	-0.829
Bluetooth headphones			*for Nikon*	3/16	-0.213
over-ear	6/17	-0.388	Coffee makers (✓)		
on-ear	6/17	0.344	*coffee makers cup brew and dispense models*	11/15	-0.059
Bluetooth receivers	8/17	0.263			
Bluetooth speakers*	6/15	0.206	*espresso machines*	12/14	0.643
	4/16	0.255	*espresso makers*	12/16	0.396
	9/17	-0.287	Coffeemaker combos	12/16	0.081
Blu-ray players	1/16	-0.785	*with automatic milk frother*	12/17	-0.099
Built-in refrigerators (*small size models)* (✓)	5/17	0.500			

Appendix: ***(continued)***

Category	Issue	Correlation
Computer monitors (✓)		
widescreen (16:9 ratio)	5/15	0.333
ultrawide (21:9 ratio)	5/15	0.786
Computer tablets (✓)		
large computer tablets	12/14	0.719
small computer tablets	12/14	0.660
computer tablets with	8/16	-0.085
*keyboard**	1/17	0.546
	7/17	0.972
*6.8 – 8.4 inch models**	6/14	-0.381
	1/15	0.673
	12/15	0.532
	1/16	0.229
6.9 – 8 inch models	8/16	0.920
7 – 8 inch models	12/16	0.945
8.7 – 9.8 inch models	6/14	0.614
8.7 – 10.9 inch models	12/15	0.933
8.9 – 10 inch models	1/15	0.081
8.9 – 10.9 inch models	7/15	0.409
9.4 – 10 inch models	7/17	0.588
9.6 – 10.1 inch models	12/16	0.870
10 inch models	8/16	1.000
Conventional dishwashers *(60*	5/15	-0.506
cm) (✓)*	6/16	-0.121
Cooktops *(ceramic cooktops)*	2/15	-0.028
Cordless hedge trimmers		
hedge trimmers	8/17	0.056
pole hedge trimmers	8/17	0.168
Cordless phones (✓)		
simple models	1/14	-0.871
comfort models	1/14	0.712
with base station	9/15	0.612
without base station	9/15	-0.533
with touchscreen	1/14	-0.747
Digital radios		
DAB+	7/15	0.664
DAB+ and Internet radio	7/15	-0.738
Digital scales (✓)	1/14	0.232
Drones *(with GPS)*	12/17	-0.111
Drilling machines (✓)		
Cordless drills/drivers		
light use cordless use drills/drivers	3/15	-0.720
cordless impact drills/drivers	3/15	0.995
impact drills/drivers	3/15	0.134
rotary hammer drills	3/15	-0.758
DVB-T2 HD receivers *(with decoder)*	2/17	0.602
DVB-T2 outdoor antennas	3/17	0.299
E-book readers *(Black and White)* (✓)	2/14	0.083
Electric grills (✓)		
contact grills	6/15	-0.251
electric griddles	6/15	0.158
Electric toothbrushes*	3/16	0.586
	1/17	0.685
	11/17	-0.533

Appendix: ***(continued)***

Category	Issue	Correlation	Category	Issue	Correlation
Electric toothbrushes for kids	1/15	0.502	Jig saws		
Electric mixers (✓)			*corded barrel-grip*	3/16	0.150
up to 1000 watts	10/16	0.873	*cordless top handle*	3/16	0.958
more than 1000 watts	10/16	0.686	Kitchen machines (*with heating mode)*	12/15	-0.546
Electric razors (✓)	5/17	0.389			
Exercise bikes *(upright ergometer)*	1/15	-0.463	Laptop computers and ultrabooks		
Fitness trackers (✓)	12/17	-0.581	*ultrabook PCs with Windows*	4/17	-0.674
with heart rate monitor	1/16	0.947	*convertibles with Windows*	4/17	-0.278
without heart rate monitor	1/16	0.893	Laser printers (✓)		
Fitness watches	12/17	0.258	*black and white laser*	10/14	-0.129
Freezers			*printers**	9/17	0.937
small size freezers	8/15	0.762	*color laser printers*	9/15	-0.111
large size freezers	8/15	-0.961	Lawn mowers		
freestanding Freezers (large size)	8/17	0.721	*cordless lawn mower*	4/17	-0.616
			corded lawn mower	4/14	0.567
GPS navigators (✓)			Microwaves (✓)		
*5 inch screen size**	2/14	0.742	*with grill and oven*	8/16	0.085
	2/15	0.540	*with grill*	8/16	0.061
*6 – 7 inch screen size**	2/14	-0.716	Mini Hi-Fi systems	12/15	-0.138
	2/15	0.035	Mini PCs	10/16	-0.917
Hair dryers (*ionic*) (✓)	1/15	0.470	Network receivers (*AV receivers*)	8/17	0.865
Headphones (✓)					
in-ear headphones	8/15	-0.305	PC sticks	10/16	0.985
wired headphones	5/14	0.091	Personal clouds		
Sports headphones			*single drive*	2/16	0.250
wired sports headphones	8/16	0.362	*dual drive*	2/16	0.139
with bluetooth	8/16	0.151	Power banks		
High pressure washers	4/14	-0.175	*2200 – 3000 mAh capacity*	6/16	-0.448
Indoor antennas for DVB-T2	2/17	0.780	*5200 – 6000 mAh capacity*	6/16	-0.214
Inkjet printers (✓)	4/15	-0.348			

Appendix: *(continued)*

Category	Issue	Correlation	Category	Issue	Correlation
Projectors			Smartphones for seniors (*simple models*)	1/17	0.525
long throw	6/16	0.263			
short throw	6/16	0.196	Smartwatches*	10/15	0.712
full HD	6/14	-0.691		12/17	0.997
Refrigerators (✓)			Smoke alarms		
large size models	5/17	0.780	*battery operated smoke alarm*	1/16	0.442
compact models	8/14	-0.135	*interconnected battery operated smoke alarm*	1/16	-0.781
Refrigerator freezer combos (✓) *(without chill compartment)*	7/16	0.190			
			Smoothie mixers (✓)	10/16	0.873
Routers			Soundbars and soundplates (✓)	12/14	-0.480
DSL	11/17	0.551	*soundbars**	11/15	0.272
with ADSL modem	8/14	-0.936		11/17	0.702
with VDSL and ADSL modem	8/14	0.932	*soundbar bundles with wireless bass module*	11/17	0.104
Satellite TV receivers	4/14	-0.094	*soundplates*	11/15	0.383
single tuner	6/15	0.245	Steam irons (✓)		
twin tuner	6/15	0.327	*conventional*	12/16	-0.053
Security cameras			*steam ironing systems*	12/16	0.994
outdoor	10/17	0.174	System cameras (✓)		
indoor	10/17	-0.430	*with viewfinder*	3/14	0.352
Small water filters (✓)	5/15	-0.376	*with electronic viewfinder**	3/15	-0.106
Smartphones*	2/14	-0.203		3/16	-0.885
	7/14	0.194		4/17	0.165
	11/14	0.274	*with optical viewfinder**	3/15	0.713
	3/15	-0.032		3/16	-0.135
	8/15	-0.146		4/17	0.159
	1/16	0.341	*without viewfinder**	3/15	0.175
	5/16	0.554		3/16	0.610
	11/16	0.325		4/17	0.611
	5/17	0.426	Tankless water heaters (*electric models*)	1/15	0.958
	11/17	0.517			

Appendix: ***(continued)***

Category	**Issue**	**Correlation**	**Category**	**Issue**	**Correlation**
Telephoto lenses (✓)			Vacuum cleaners (✓)		
for Canon cameras			*bagged**	6/15	-0.088
large maximum aperture	7/17	0.706		5/16	0.944
small maximum aperture	7/17	0.744		7/17	-0.495
for Nikon cameras			*bagless**	5/16	-0.725
large maximum aperture	7/17	0.967		7/17	0.810
Thermostats (✓)			*cord-free vacuum*	2/16	0.025
programmable thermostats	1/17	-0.792	*robotic vacuum cleaners*	2/15	0.590
thermostats with Wi-Fi	1/17	-0.086	Washing machines (*front load*	11/15	-0.357
Toasters (✓)	4/16	0.697	*washer*)*	11/16	0.072
Tumble dryers	10/17	-0.597	Wi-Fi receivers		
*with heat pump**	9/14	0.301	*network audio players*	8/17	-0.731
	9/15	-0.120	*connectors*	8/17	-0.011
	9/16	0.755	Wi-Fi speakers	12/16	0.457
without heat pump	9/16	0.689	Wireless speakers	11/14	0.032
TV's (✓)					
*32 inches**	10/16	0.663			
	2/17	-0.094			
	10/17	-0.719			
*40-43 inches**	10/15	0.896			
	12/15	-0.078			
	2/16	-0.287			
	10/16	0.556			
	10/17	0.993			
*48-50 inches**	2/16	0.619			
	6/16	-0.988			
	10/17	0.229			
49 inches	10/16	0.893			
49-50 inches (LCD models)	12/17	0.175			
55 inches (OLED models)	12/17	0.621			
*55-58 inches**	12/16	-0.680			
	12/17	-0.645			

Note: The check mark indicates that the same or a similar product category was also in de Langhe *et al.*'s (2016a) database; [a]no variation in *Stiftung Warentest* quality scores; asterisked categories have been used to explore the moderating effect of product age on the relationship between average ratings and *Stiftung Warentest* scores.