2021

Online Reviews Generated Through Product Testing: Can More Favorable Reviews Be Enticed with Free Products?

Ina Garnefeld

Tabea Krah

Eva Böhm

Dwayne D. Gremler
Bowling Green State University, gremler@bgsu.edu

ORIGINAL EMPIRICAL RESEARCH

Online reviews generated through product testing: can more favorable reviews be enticed with free products?

Ina Garnefeld[1] · Tabea Krah[1] · Eva Böhm[2] · Dwayne D. Gremler[3]

Abstract

Online reviews have profound impacts on firm success in terms of sales volume and how much customers are willing to pay, yet firms remain highly dependent on customers' voluntary contributions. A popular way to increase the number of online reviews is to use product testing programs, which offer participants free products in exchange for writing reviews. Firms that employ this practice generally hope to increase review quality and secure higher product rating scores. However, a qualitative study, experimental study, and multilevel analysis of a field study dataset of more than 200,000 online reviews by product testers combine to reveal that product testing programs do not necessarily generate higher quality reviews, nor better product ratings. Only in certain circumstances (e.g., higher priced products) does offering a product testing program generate these benefits for the firm. Therefore, companies should consider carefully if and when they want to offer product testing programs.

Keywords Product testing · Online product reviews · Equity theory · Reactance · Multilevel analysis · Experimental study

Online product reviews—defined as "peer-generated product evaluations posted on company or third party websites" (Mudambi and Schuff 2010, p. 186)—can substantially increase product acceptance. Recent research suggests that 91% of consumers read online product reviews regularly, and 84% trust online reviews as much as personal recommendations (Brightlocal 2018). Positive reviews increase readers' intentions to buy a product (Marchand et al. 2017) and their

willingness to pay a higher price (Kübler et al. 2018). For example, a one-star improvement in a review rating[1] reportedly results in a 9% increase in sales (Luca 2011) and a willingness to pay nearly 50 euros more for an ebook reader (Kostyra et al. 2016). Readers of online reviews also perceive high-quality reviews as more helpful than low-quality reviews and thus are more likely to follow these recommendations (Lu et al. 2018). Such benefits of online product reviews can directly affect a firm's performance.

Although the importance of online reviews is generally well-known, companies often suffer the challenges of few reviews, low product ratings, and poor quality reviews—especially for new products just introduced to the market (Cui et al. 2012). In general, customers tend to write reviews only if they are extremely satisfied or dissatisfied, leading to a J-shaped distribution of star ratings (Chevalier and Mayzlin 2006; Hu et al. 2009; Liu 2006). Moreover, the majority of online reviews are brief and do not provide sufficiently useful information (Cao et al. 2011; Mudambi and Schuff 2010). Companies thus seek ways to improve both product ratings and the quality of reviews, but no simple or specific tactic exists for them to do so. After a purchase, companies often ask customers to provide product reviews, but only a minority

---

Alina Sorescu served as Guest Editor for this article.

✉ Ina Garnefeld
garnefeld@wiwi.uni-wuppertal.de

Tabea Krah
krah@wiwi.uni-wuppertal.de

Eva Böhm
eva.boehm@tu-dortmund.de

Dwayne D. Gremler
gremler@bgsu.edu

[1] Department of Marketing, University of Wuppertal, Gaussstr. 20, 42119 Wuppertal, Germany

[2] Department of Marketing, Technical University of Dortmund, Otto-Hahn-Str. 6, 44227 Dortmund, Germany

[3] Department of Marketing, College of Business, Bowling Green State University, Bowling Green, OH 43403, USA

---

[1] We define *review rating* as the number of stars given by a reviewer in rating a product. In most situations, it ranges from one to five stars, with five stars representing the highest rating for a product.

704

J. of the Acad. Mark. Sci. (2021) 49:703–722

of customers comply (Magno et al. 2018). Other firms offer financial incentives to increase customers' motivation to provide online product reviews. However, these incentives have been found to have mixed effects on the rating, leading Garnefeld et al. (2020) to advise companies to only carefully apply incentives for online reviews.

Product testing offers another, relatively recent approach to encourage customers to provide online reviews (Kim et al. 2016). A typical product test consists of five steps:

1. The product testing provider contacts selected customers or customers contact the provider (after it openly posts a testing opportunity) and apply to participate.
2. Selected participants receive the test product for free.
3. They test it.
4. They write an online review within a mandated timeframe.
5. Once they have done so, they are allowed to keep the product.

However, the details of the programs differ by provider. Some manufacturers offer their own products for testing (e.g., Team Clean by Henkel, Vocalpoint by P&G, Philips), but in other cases, retailers (e.g., Best Buy, Amazon) or marketing agencies (e.g., SheSpeaks, Toluna, Trnd) offer a product test on behalf of a manufacturer. Typically, product tests occur at the beginning of a product's lifecycle, when it has few online reviews and the company would benefit from more reviews (see exemplary program descriptions of Amazon 2020, Philips 2020, Super savvy me 2020). Hence, product tests can be regarded as a type of seeding program that aims at increasing the diffusion of products early in the adoption process (Haenlein and Libai 2017).

The programs result in more online reviews for test products, but in addition to volume, manufacturers seek favorable product ratings and high-quality reviews. Two psychological theories—equity theory (Adams 1963; Ajzen 1982) and the theory of psychological reactance (Brehm 1966, 1972, 1989; Clee and Wicklund 1980)—suggest though that product testing might exert contrasting effects related to these two goals. On the one hand, participants may perceive their outcome (free test product) as more favorable than the company's outcome (publication of an online review). To give something back and restore equity, these reviewers may feel compelled to increase their rating or effort when writing the review, leading to more positive reviews of higher quality, relative to reviews written outside the realm of a product testing program. On the other hand, product testers might feel pressured to perform the review writing task. That is, after signing up for the product testing opportunity and receiving the test product for free, they cannot choose if and when to write a review about it. This obligation may lead to reactance (Brehm 1989), such that they write the reviews as required, but they do so with less effort and offer lower ratings and poorer quality. To predict how product testing programs affect reviewing behavior, we thus aim to shed light on which psychological effects occur.

Relying on a qualitative study, an experimental study, and a field study of 207,254 reviews by participants in one of the world's largest product testing programs—Amazon Vine—we derive three main findings that contribute pertinent insights. First, we challenge the common assumption that product testing increases review ratings. Firms might anticipate that product testing–induced reviews will be positively biased, because participants are "bribed" with a free test product (Wu 2019). Our comparison of reviews written by customers enrolled in the Amazon Vine program who receive a free product in exchange for writing a review versus Vine participants who purchased a product shows instead that product testing does not always increase review ratings. Second, we identify price and the number of previously published reviews as important contextual variables that influence review ratings. Compared with other reviewers, product testers assign higher review ratings to high-priced products. Furthermore, if many reviews have been published about a product, testers give it higher ratings and evaluate the product more favorably. This finding suggests a way that companies can successfully increase review ratings. Third, we show that product testing is effective for increasing review quality if the product is high priced or complex. However, review quality decreases if product testers are asked to write a review after many reviews have already been published about the product. That is, in certain circumstances, product testing can be an effective means to manage review quality.

The remainder of this article is structured as follows: First, as we regard product testing as a type of seeding program, we provide a review of relevant literature on seeding programs. Second, we develop hypotheses to predict how product testers might behave differently when they receive a product for free rather than purchase the product. Third, we describe insights gained from our qualitative study outlining positive as well as negative effects of product testing programs and shed light on the contextual factors that influence product testers' reviewing behavior. Fourth, we report the results of an experimental study in which we analyze the predicted psychological mechanisms of perceived inequity and perceived pressure. Fifth, we test our hypotheses with a multilevel analysis of field study data from the Amazon Vine program. Sixth, we discuss our results and offer advice for companies: They should carefully consider if, when, and in what conditions to offer product testing. We also present some limitations and further research opportunities.

Product testing literature

Product testing programs mostly aim to increase online reviews of new products, so they constitute a type of seeding program or a plan "to get a (typically new) product into the hands of some individuals, in the hope that this early social influence will help to accelerate and expand the growth process" (Haenlein and Libai 2017, p. 71). Seeding programs offer discounts, samples, or trial periods, and sometimes products, to selected people (i.e., "seeds") mostly early in the adoption process (Haenlein and Libai 2013). In what follows, we provide an overview of the literature that yields insights on the influencing factors and consequences of seeding programs and contrast them with our study.

Influencing factors of seeding programs

Literature on seeding programs identifies different design factors that influence their effectiveness. In particular, seed selection is critical to the success of seeding campaigns. However, no consensus exists regarding which seeds are most effective to target (Chen et al. 2017; Libai et al. 2013). In their review, Hinz et al. (2011) identify debates about whether it is optimal to target hubs (most connected people), fringes (less connected people), or bridges (people who connect parts of the network that otherwise would not be connected). The size of the seed sample (Jain et al. 1995; Nejad et al. 2015) and geographical spread (Haenlein and Libai 2017) also influence program success.

Outcomes of seeding programs

Seeding programs can affect sales through two routes: directly through the behavior of the seed (direct value) and indirectly through other customers acquired due to the social influence of the seed (social value) (Haenlein and Libai 2017, p. 74). Several studies focus on direct value and find a positive effect of seeding programs on seeds' own purchase behavior and, in turn, on sales (Bawa and Shoemaker 2004; Gedenk and Neslin 1999; Li et al. 2019). However, seeds acquired with a free trial period ultimately tend to be less profitable than customers acquired through other marketing instruments (Datta et al. 2015).

Several studies also confirm the social value created by seeding programs (Table 1). For example, Haenlein and Libai (2013) find that profitable seeds (in terms of high customer lifetime value) create high social value, because profitable customers tend to engage in networks with other potentially profitable customers. Seeding affects the communication behavior of seeds and also can lead to spillover effects for nonseeds (Chae et al. 2017), such that after they read posts written by seeds, nonseeds often increase their communication about the product too, whether for self-presentation reasons or to signal their own expertise. Seeding programs can be seen as "double-edged swords" (Foubert and Gijbrechts 2016) in that offers of good quality products can increase positive word of mouth (WOM) and accelerated adoption, but trials of lower quality products can lead to opposite effects. Hence, trials might "alienate consumers and trigger adverse WOM effects, thus driving away customers who would have adopted now or later" (Foubert and Gijsbrechts 2016, p. 825).

Contributions of this study relative to seeding program literature

Our study of product testing extends research into the social value of seeding programs in three important ways. First, prior research predominately has focused on seeding programs in which customers receive products for free for a limited time (Foubert and Gijsbrechts 2016) or with a limited size (Chae et al. 2017; Kim et al. 2014). Thus, the companies anticipate future revenues from the seeds themselves, once they adopt the product (direct value). In contrast, product testing programs primarily aim to generate online reviews by product testers, not acquire them as customers. Of course, gaining participants as ongoing customers could be a positive side effect, but it rarely is the primary focus, and it tends to be unlikely, because product testers receive free products that typically do not require repurchases in the near future (e.g., book, electronics equipment). Therefore, the program providers' primary interest is social value, not direct value.

Second, seeds generally receive a sample version or may use the service for limited time, but product testing programs provide more substantial offerings, beyond a sample or trial period. Receiving this larger "gift" might evoke distinct and stronger psychological effects. In particular, it likely prompts stronger perceptions of inequity in the relationship and a motivation to give something back (Larsen and Watson 2001).

Third, seeding programs hope to encourage social influence (Haenlein and Libai 2017, p. 71). In a product testing program, participants are required to write online reviews in exchange for the free products. They lack full freedom of choice to decide whether to write a review, so they likely perceive the task differently than participants in conventional seeding programs who can freely decide whether, how, and when to use their social influence, such as by writing a review.

Conceptual framework and hypotheses

To predict the effects of product testing on review rating and review quality, we rely on equity theory (Adams 1963; Ajzen 1982) and the theory of psychological reactance (Brehm 1966, 1972, 1989; Clee and Wicklund 1980). Figure 1 provides an overview of the hypothesized effects.

706

J. of the Acad. Mark. Sci. (2021) 49:703–722

Table 1 Selected literature on the social value of seeding programs

| Authors	Seeding Program Type	Consequences Investigated	Study Type	Key Findings
Chae et al. (2017)	Out-of-store sampling (cosmetics)	- Spillover effects of WOM (i.e., WOM by nonseeds) on product, category, or brand level	Empirical	Seeding increases nonseed WOM about the focal product and decreases WOM about other products from the brand or the same category.
Chen et al. (2017)	Stimulation of an initial group of leaders to spread WOM about the microfinance program among their contacts	- Product diffusion	Bayesian modeling	Development of a new methodology to show how relationship characteristics influence the diffusion process.
Foubert and Gijsbrechts (2016)	Free trial for an interactive television service	- Adoption - Reference price - WOM	Mathematical modeling, simulation	For high/low product quality, free trial increases/decreases the number of users through WOM and accelerates/slows down adoption.
Haenlein and Libai (2013)	Program to encourage seeds to adopt a new product early, with the goal of jump-starting contagion	- Adoption - Social influence	Agent based simulation	Both revenue leader seeding and opinion leader seeding can create greater value than random customer seeding.
Haenlein and Libai (2017)	Seeding approaches include discounts, samples, or free products	- Direct value - Social influence	–	Development of a new method to measure the impact of seeding on customer equity.
Kim et al. (2014)	In-store sampling (razors) and free product (print of a photo portrait)	- Repeat purchase - Word-of-mouth	Empirical	Sampling leads to lower repeat purchases and WOM compared with pay-what-you-want.
Libai et al. (2013)	Program to get a group of target customers to adopt the product early, in an effort to enhance the contagion process for other customers	- Social influence	Agent based simulation	Seeding creates social value (expansion and acceleration), influenced by competition, program targeting, profit decline, and retention.
Our study	**Product testing program**	**- Online review ratings** **- Online review quality**	**Empirical**	**Product testing does not increase review ratings nor review quality in general, but only in certain contexts.**

Positive effects of product testing programs on reviewing behavior

Equity theory (Adams 1963; Ajzen 1982) suggests that people seek to restore equity if they sense they are under- or over-rewarded in an exchange. Inequity creates tension, which people attempt to decrease (Adams 1965). The action a person takes to restore equity is proportional to the perceived inequity (Adams 1963). Equity theory receives empirical support in diverse research fields, such as sociology (e.g., Austin and Walster 1975), psychology (Greenberg 1982), management (Carrell and Dittrich 1978), marketing (Fang et al. 2008; Homburg et al. 2007; Scheer et al. 2003), and incentivized online reviews (Petrescu et al. 2018).

In product testing programs, a product tester and a company enter into a reciprocal exchange. The company offers a

J. of the Acad. Mark. Sci. (2021) 49:703–722

707

Fig. 1 Overview of empirical studies and hypotheses
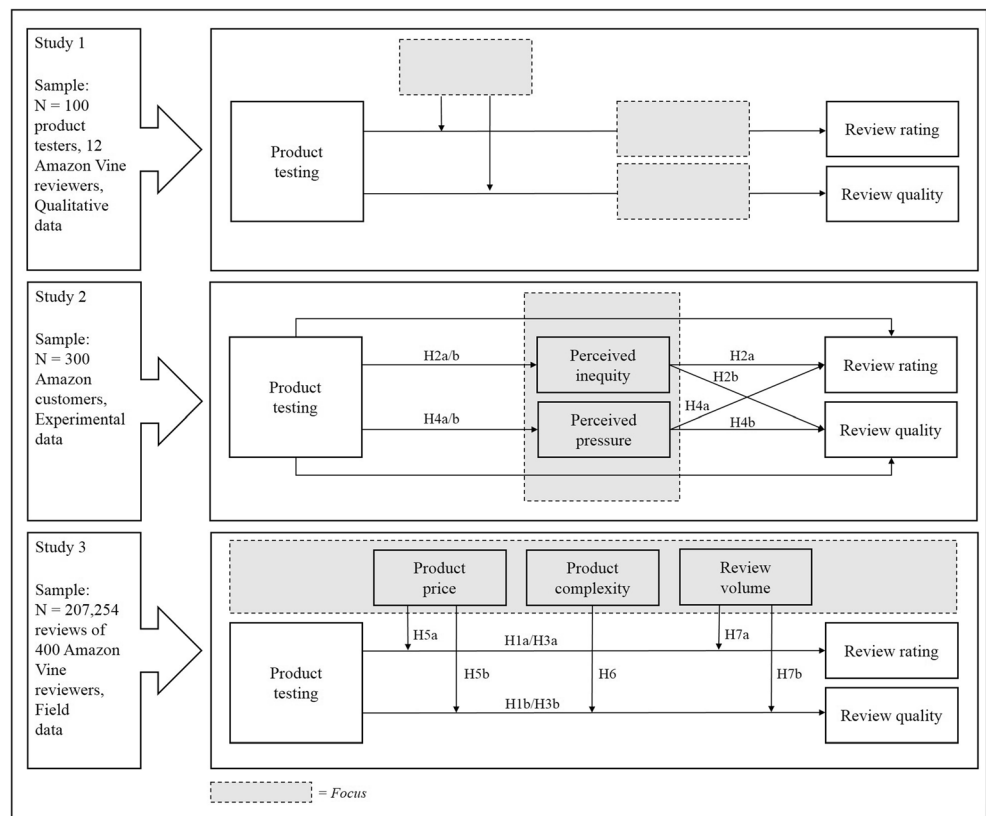
product at no cost, and the product tester writes a review in exchange. According to equity theory, assessments of the outcome-to-input ratio influence each party's behavior in the exchange. If the relationship is perceived as inequitable, product testers should adjust their behavior to create balance. In assessing their outcome-to-input ratio, they likely consider what they gain from the negotiated exchange (outcome) and what they provide (input), relative to the company's outcome-to-input ratio (Homburg et al. 2007). Thus, they assess what they receive from the company (test product as their outcome) and what they must do (writing a review as their input), and that assessment affects their reviewing behavior (i.e., rating or quality of the review). Simultaneously, they consider the company's outcome-to-input ratio: It gives a product away for free (input) and receives an online review from the product tester (outcome). If these ratios seem equal, the relationship appears equitable. However, receiving a free product might seem like a more considerable outcome, relative to the less effortful behavior of writing an online review. If product testers thus perceive their own benefit as more favorable than what the company attains (i.e., perceived inequity), they may strive to rebalance the relationship in various ways. First, they might increase the company's outcome by offering a better review rating. Second, they can increase both their input and the company's outcome by putting more effort into the review, which should lead to a higher quality review. Therefore, we posit that product testing leads to more

favorable review ratings and reviews of higher quality, due to the effects of inequity perceived by product testers. In the following, we regard perceived inequity as positive inequity which occurs if one's own input-to-outcome ratio is higher than that of the other party (Scheer et al. 2003). Consequently, we predict:

H1: Product testing reviews (vs. other types of reviews) increase (a) review ratings and (b) review quality.

H2: Customers' perceived inequity mediates the positive effects of product testing reviews (vs. other types of reviews) on (a) review ratings and (b) review quality.

Negative effects of product testing programs on reviewing behavior

The theory of psychological reactance predicts opposite effects. It suggests that people respond negatively when others attempt to influence them (Brehm 1989). According to Brehm (1966), people believe that they should be able to do what they want, when they want, and in the way they want. If their freedom is threatened or eliminated, it leads to a psychological state of reactance—an unpleasant motivational arousal to redress a perceived threat or constraint on a specific behavioral freedom—that induces attitudinal and behavioral reactions (Brehm 1972). A person may form a negative or hostile

708

J. of the Acad. Mark. Sci. (2021) 49:703–722

attitude toward the source of the influence and engage in defensive behaviors to reestablish behavioral freedom (Clee and Wicklund 1980). Individual perceptions are pivotal too. That is, a person does not need proof that freedom is being constrained but can exhibit reactance following a mere perception of such influences.

Applying the theory of psychological reactance to product testing programs, we predict that participants might perceive their freedom as threatened by an obligation to review a product. If testers, obliged to write a public online review in a given timeframe, regard their behavioral freedom as constrained by this duty, they may respond negatively and act against the interests of the company. The perceived pressure to write a product review in a mandated timeframe also may induce a negative attitudinal shift toward the company or its products, along with negative behaviors, such as giving a low rating or putting less effort into writing the review. Therefore, we offer competing hypotheses, relative to H1 and H2:

H3: Product testing reviews (vs. other types of reviews) lower (a) review ratings and (b) review quality.

H4: Customers' perceived pressure mediates the negative effects of product testing reviews (vs. other types of reviews) on (a) review ratings and (b) review quality.

Moderators

Product price In practice, large discrepancies exist in the prices of test products, ranging from ebooks that sell for a few dollars to vacuum cleaners priced at $500 for example. Equity theory (Adams 1963; Ajzen 1982) suggests that the exchange relationship becomes especially inequitable if the test product costs more. Testers of expensive products, rather than less expensive products, thus may feel more obligated to increase the company's outcomes by giving more positive ratings or to increase their own input by putting more effort into writing the review. Therefore, we hypothesize:

H5: Product price moderates the effects of product testing reviews (vs. other types of reviews) on (a) review ratings and (b) review quality, such that the effects become more positive as the price increases.

Product complexity Test products also vary in their complexity; some products contain a multitude of attributes to evaluate (e.g., smart television), but others are simpler to test and evaluate (e.g., pencil). For more complex products, companies strongly prefer detailed, high-quality reviews, because other potential customers need extensive information and may turn to online reviews to obtain it. Hence, product testers may feel that for complex products a review of high quality is even more valuable to the company (compared to less complex products). In turn, product testers are more likely to rebalance perceived inequity by writing a review of high quality as product complexity increases. We accordingly predict:

H6: Product complexity moderates the effects of product testing reviews (vs. other types of reviews) on review quality, such that the effect becomes more positive as product complexity increases.

However, we do not anticipate that product complexity will affect the relationship between product testing and review rating. Product testers have no particular reason to evaluate a complex product more positively or negatively than a less complex one.

Previously published review volume Although product tests usually are offered at the beginning of a product's lifecycle, the number of already available reviews of a test product can vary. Some product testers will be among the first to review a product; others might be asked to write a review after many other reviews have been published (e.g., later in the product lifecycle, for products of great interest to customers). A review writer's exposure to previous reviews influences his or her own review (e.g., Askalidis et al. 2017; Sridhar and Srinivasan 2012; Sunder et al. 2019), because previous reviews offer insights into how others have perceived the product. According to equity theory, product testers usually try to treat the company fairly. Even if they discover potential drawbacks, the existing reviews give them a means to determine if their opinion is universal or if others perceive the same drawbacks. In turn, they might reconsider their rating, to determine if it is legitimate or should be adjusted. Therefore, we hypothesize:

H7a: Previously published review volume moderates the effects of product testing reviews (vs. other types of reviews) on review ratings, such that the effect becomes more positive when more reviews have been published previously.

Alternatively though, product testers may believe the company does not need more high-quality reviews if a multitude already are available. Thus, they might devote less effort to writing a review of high quality when many reviews already have been published. We accordingly predict:

H7b: Previously published review volume moderates the effects of product testing reviews (vs. other types of reviews) on review quality, such that the effect becomes less positive when more reviews have been published previously.

To explore these predictions, we adopt a mixed method approach that combines qualitative and quantitative research methods. Figure 1 details the three studies we conducted. In Study 1, we explore the different feelings, perceptions, and behaviors evoked by product testing programs with a content analysis of responses to open-ended questions, gathered in surveys of 100 product testers from various testing programs and 12 in-depth interviews with actual Amazon Vine product testers. This study also reveals some contextual influences on the effects of product testing programs on reviewing behavior. In Study 2, we experimentally test the positive and negative psychological mechanisms of our theoretically derived hypotheses, as shown in Study 1. Building on Studies 1 and 2, Study 3 tests the hypothesized moderating effects of product price, product complexity, and previously published review volume on product testers' actual behavior with field data obtained from more than 200,000 product reviews.

Study 1: Questionnaire and expert interviews

To gain qualitative insights into the feelings and perceptions of product testers, we administered an open-ended questionnaire to 100 participants from various product testing programs in the United Kingdom (Sample A), and we conducted in-depth interviews with 12 experts, namely, Amazon Vine product testers from Germany (Sample B). In collecting and analyzing these data, we pursue two goals. First, noting the paucity of literature on the effects of product testing programs on reviewing behavior, we use the data to shed light on both positive and negative effects of product testing programs on participants. Second, they provide insights into the thoughts and feelings of a wide range of product testers, representing various test programs, including Amazon Vine. We present detailed information about the participants and data collection procedures for both samples in Appendix 1. Briefly though, Sample A data were collected online, and participants provided written responses; Sample B data were gathered through video conferences, recorded, and transcribed.

Method

Following a research strategy suggested by Miles and Huberman (1994), we combined within- and cross-case analyses, starting with line-by-line open coding to assign descriptive codes to the answers obtained in the questionnaire and interview transcripts. Next, we subjected the data to cross-case comparisons to identify similarities among product testers and systematic associations among the focal variables. The cross-text analysis then sought to ascertain if respondents could be grouped into categories, according to their responses to a particular topic. We grouped statements into (1) positive reactions to participating in the product testing programs that led to more positive reviews of higher quality and (2) negative reactions that lead to more negative reviews of lower quality. We also categorized statements according to the factors that appear to influence the effects of the product testing program on reviewing behavior, including product price, product complexity, and previously published review volume.

Findings

Positive consequences of product testing for companies
Overall, the product testers expressed gratitude for the opportunity to test products and often mentioned their sensed need to return the favor by giving a more positive rating and/or putting more effort into their review than they would into a routine review for a purchased produc:

> I believe I gave it a fair and honest review which was perhaps leaning to a more positive review. Again, this was out of gratitude for getting a free gift. [*Sample A, female, 32, product tester for 2 years*]
> I was probably slightly more positive, as I felt it was the least I could do. [*Sample A, female, 31, product tester for 3 years*]
> Since Amazon gives me this opportunity, I would like to return the favor. [*Sample B, male, 43, product tester for 1 year*]
> It is an expensive program at the end of the day. If you look at the whole thing, the products for free, the shipping, the logistics behind it. Then you feel guilty towards the program. [*Sample B, male, 53, product tester for 1 year*]

These comments are consistent with H1 and H2.

Negative consequences of product testing for companies
The participants also mentioned negative perceptions and feelings, including feeling pressured to write a review and do so within a certain timeframe. They also cited a sense of exhaustion brought on by their participation in the program sometimes and noted that the requirement to publish a review within a particular timeframe could prevent them from writing a thoughtful review.

> You get the product for free, therefore you feel compelled to post a review. [*Sample A, female, 45, product tester for 8 years*]
> I felt more pressured by my ability to complete the task at all. [*Sample A, female, 32, product tester for 2 years*]
> So there is, let's say, a little bit of trouble. It's like this—you get selected and that's maybe, I would say, a certain honor, but on the other hand you have to

evaluate the products very rapidly and timely. What is actually contradictory—such a pressure to evaluate—is actually contradictory to a sincere type of evaluation, I think. [*Sample B, male, 47, product tester for 1 year*] Well, it does put you under pressure somehow, … you sit there and you know you have to make something out of thin air, but somehow it's like, well, it's a bit of a pressure feeling. [*Sample B, female, 24, product tester 1 year*]

These comments are consistent with H3 and H4.

Product price We find more expressions of gratitude in response to receiving a high-priced test product compared with a low-priced one. The participants want to return this favor by writing more positive reviews and devoting more effort to craft high-quality reviews:

> If I received a more expensive product for free, I would feel more gratitude towards it as I'd feel more grateful having reviewed it for free. [*Sample A, female, 24, product tester for 4 years*]
> I would feel much more valued by the company and spend a lot more time reviewing and testing the product. [*Sample A, male, 35, product tester for 5 years*]
> If you liked it, but you didn't like little certain things about it, would you write that very negatively or would you rather say, "Come on, I got this as a gift. It costs 1,800 euros"? I think you are a little bit more willing [to give it a favorable review]. [*Sample B, male, 53, product tester for 1 year*]
> When it comes to high-priced products you can't allow yourself to only write two lines about it. [*Sample B, male, 47, product tester for 1 year*]

Thus, our data suggest that product testers pay attention to the product price, even though they do not have to pay for the product. In turn, they feel the need to rate high-priced test products more favorably and put more effort into their reviews, in line with H5.

Product complexity Many product testers believe that companies expect more thorough online reviews from them when giving away complex test products compared to less complex test products:

> [If the product was complex], I would feel I should give a more measured and intense review. I would also feel more was expected of me [in terms of review quality]. [*Sample A, female, 60, product tester for 5 years*]

If I received a more complex product for free, I would make sure that I thoroughly understood all of the features so that I could review the product comprehensively. [*Sample A, female, 64, product tester for 10 years*]
I also had a shower gel to test, there is not much to write about it, the complexity is limited. The scope of the review is not very large, because you don't necessarily want to reinvent the wheel now and you don't have to write a seven hundred page homage to a shower gel. If a few points can simply be listed, then I think it's legitimate. However, as I said, if it is about a saw or simply about certain IT devices, then it simply needs a certain differentiation, also in the writing/text. [*Sample B, male, 29, product tester for 10 years*]

As these comments reveal, product testers put more effort into reviewing complex test products compared to other products. Generally, the testers believe that companies value high-quality reviews more for complex products compared to less complex products, in line with H6.

Previously published review volume Finally, many product testers asserted that if many reviews already were available, the company did not need another extensive review. This belief affected how much effort they put into their own reviews. They also used these previously published reviews as points of reference, which in some cases prompted them to second-guess their own review or adjust their ratings.

> If I were writing a review after many other reviews have been published about the test product, I would provide an honest report but wouldn't go into as much detail as other reviews as I feel that some people could have already provided honest and helpful feedback that can benefit other people so my review may not be as recognizable as others and appreciated. [*Sample A, female, 28, product tester for 5 years*]
> If I were posting a review after many others I think it would be harder to post a very different review, for instance if everyone else gave it four stars and I hated it, I might be reluctant to give it a really poor review, I might end up second guessing my own opinion. [*Sample A, female, 59, product tester for 4 years*]
> If I were among the first to evaluate a test product, I would definitely re-read my review before sending and try to address as many points as I could think of, so as not to miss anything. [*Sample A, female, 38, product tester for 2 years*]

These comments are in line with H7.

Overall then, our Study 1 findings indicate that product testing programs can evoke both positive and negative

outcomes for the firm. The contextual variables also appear to influence product testers' reviewing behavior, consistent with our hypotheses.

Study 2: Experimental study

In Study 2, we test the theoretically derived psychological mechanisms and the Study 1 findings related to product testers' reviewing behavior with an experimental approach. More specifically, we examine how perceived inequity and perceived pressure mediate the relationship between product testing and review rating and the relationship between product testing and review quality. The conceptual model is in Fig. 1.

Method

Research design and participants We used a posttest-only control group design (Campbell and Stanley 1963) and manipulated the type of review (product testing versus non-product testing) by randomly assigning participants to one of two groups. Three hundred participants were recruited from Prolific, a well-established platform for online research (Paharia 2020). We restricted the sample to participants who had an Amazon account, shopped online on average at least once a month, were from Great Britain, and spoke English as their first language. The participants were prescreened by applying Prolific's filter options, so participation was available only to those who met the criteria.

Of the recruited participants, 69.7% were women. Most of the respondents are in the 18–30 (29.5%) or 31–40 (31.2%) age groups, followed by the 41–50 (20.5%) and 50+ (18.8%) age groups. Overall, the sample has a mean age of 39 years. Regarding their educational background, 10.7% have achieved a secondary school degree as their highest level of education, 16.7% a high school degree, 17.3% a vocational training or technical school degree, 38% a bachelor's degree, and 16.3% a post-graduate degree; 1% of the participants indicated they have not achieved any of the aforementioned degrees. The experimental and control groups comprised 150 participants each.

Procedure The participants completed the survey by accessing a link that directed them to a website with the scenario descriptions and questionnaire. Participants from both groups were asked to recall the last product they purchased on Amazon that they had already used. In the product testing (experimental) group, they had to imagine they received this specific product for free, as part of Amazon's product testing program. Similar to the actual Amazon Vine procedure, they read that they had been requested to be a product tester for this specific product, received the product free of charge, were given some time to test it, and

after a trial period were required to write an online review. The other (control) group of participants had to imagine that after purchasing and using the product (their most recent purchase on Amazon), they decided to write a review of it. Both groups then wrote an actual review, suitable for posting on Amazon.co.uk, and rated the product on a typical scale with one to five stars. They also indicated how much effort they put into writing the review. Finally, all participants completed a short questionnaire with items related to the hypothesized psychological mechanisms (i.e. , perceived inequity and perceived pressure), followed by manipulation and realism checks.

Measures All of the scales are in Table 2. The *review rating* measure ranges from one star ("very poor") to five stars ("very good"). For *review quality*, we use a seven-point Likert-type scale with three items, adapted from Yin et al. (2017). To measure *perceived inequity*, we adapt a three-item scale from Brady et al. (2012) to our study context. *Perceived pressure* is measured on a four-item scale adapted from Unger and Kernan (1983). Both psychological mechanisms are measured on seven-point Likert-type scales, ranging from 1 = "I strongly disagree" to 7 = "I strongly agree."

Manipulation and realism checks The manipulation was successful. Respondents assigned to the product tester condition agreed that they had been instructed to imagine they had received the product free of charge, whereas customers from the control group did not ($M_{product\ tester}$ = 6.75, $SD_{product\ tester}$ = .94; $M_{regular\ customer}$ = 1.31, $SD_{regular\ customer}$ = .98; t = −49.014, p < .001; 1 = "I strongly disagree", 7 = "I strongly agree"). The realism check also indicated that participants perceived the scenarios as realistic (M = 6.08, SD = 1.19; 1 = "The situation was very difficult to imagine," 7 = "The situation was very easy to imagine"), which confirms that they could put themselves in the described situation.

Validity assessment A confirmatory factor analysis of our three multi-item scales (review quality, perceived inequity, and perceived pressure) provides support for convergent validity, according to the factor loadings (>.82), average variance extracted (>.76), and Cronbach's alphas (>.84). All values exceed the common thresholds (see Table 2).

Results

We employ the PROCESS procedure (Hayes 2018) and examine the direct and indirect effects of product testing programs on review ratings and review quality. With a bootstrapping mediation analysis, with 5000 bootstrapped samples in Model 4, we find support for H2 and H4, because the effects of product testing programs on review

712

J. of the Acad. Mark. Sci. (2021) 49:703–722

Table 2 Reliability assessment and construct measurements, Study 2

Construct	Items	Factor Loadings	AVE	CA
Review rating	Star rating (one to five stars)	–	–	–
Review quality (adapted from Yin et al. 2017)	I put a lot of effort into writing the review.	.93	.76	.84
	I gave a lot of thought to this review.	.90		
	I spent a lot of time writing the review.	.79		
Perceived inequity (adapted from Brady et al. 2012)	I got more out of this transaction than the product provider.	.87	.80	.87
	I got more than I deserved.	.88		
	I got more benefits than the product provider.	.93		
Perceived pressure (adapted from Unger and Kernan 1983)	I felt forced by the product provider to write the review.	.91	.78	.91
	I felt that I did not write this review voluntarily.	.82		
	I felt obligated by the product provider to write the review.	.89		
	In my opinion, the product provider talked me into writing the review.	.92		

Notes: AVE = average variance extracted; CA = Cronbach's alpha. Items were measured on 7-point scales (1 = "I totally disagree," 7 = "I totally agree"), except for star rating (one to five stars)

ratings and review quality are fully mediated by perceived inequity and perceived pressure. Specifically, both indirect effects of product testing programs on review ratings are significant: perceived inequity (H2a: b = .0856, SE = .0363, confidence interval [CI$_{90\%}$] = [.0301, .1481]) and perceived pressure (H4a: b = −.0669, SE = .0322, CI$_{90\%}$ = [−.1193, −.0125]). The direct effect of product testing programs on review ratings is not significant though (b = −.0587, SE = .0798, t = −.7348, p = .4630), indicating that the effect is fully mediated by perceived inequity and perceived pressure. The non-significant total effect on review ratings (b = −.0400, SE = .0699, t = −.5721; p = .5677) further suggests a cancelling out result, due to the combination of a positive effect through perceived inequity and a negative effect through perceived pressure.

Then, in line with H2b and H4b, both indirect effects of product testing programs on review quality are significant, through perceived inequity (b = .2178, SE = .0675, CI$_{90\%}$ = [.1123, .3355]) and perceived pressure (b = −.2732, SE = .0868, CI$_{90\%}$ = [−.4239, −.1382]). The effect on review quality is fully mediated by perceived inequity and perceived pressure; the direct effect of product testing programs on

review quality is not significant (b = .1710, SE = .1626, t = 1.0514, p = .2939). The total effect also is not significant (b = .1156, SE = .1455, t = .7941, p = .4278), so we again find evidence of cancelling out, such that the negative effect through perceived pressure balances the positive effect of perceived inequity.

In summary, consistent with our hypotheses and Study 1, product testing programs can stimulate feelings of inequity and pressure, each of which leads to distinct behavioral responses, manifested in participants' review ratings and review quality. The product testing program appears to signal restrictions on customers' sense of freedom, such as whether and when they want to write an online review. This perceived pressure in turn decreases review ratings and review quality. However, participating in the product testing program increases perceptions of inequity, because the product testers regard receiving the product for free as a more beneficial outcome than what they input to the exchange (i.e., writing an online review). These two contradictory effects cancel out each other, and the simultaneous presence of positive and negative mechanisms suggests that the effect of product testing programs on review ratings and review quality hinges on contextual factors.

Study 3: Field study

With this study, we expand the Study 2 findings with field data from 207,254 online reviews written by 400 Amazon Vine reviewers and investigate the potential moderating effects of product price, product complexity, and previously published review volume. Relying on a multilevel analysis, we examine reviews written by customers enrolled in the Vine program (who received products free of charge) and compare them with reviews of other products that they purchased. The conceptual model for Study 3 is in Fig. 2.

Method

Design and data collection We rely on data from Amazon Vine, one of the world's largest product testing programs, to compare reviews by consumers when they receive products for free, as part of the Vine program, versus when they purchase products. The Amazon Vine program offers products, provided by manufacturers, publishers, or music labels, to selected customers who are required to write an online review in exchange. However, Amazon does not specify the valence of the reviews and rather explicitly "welcome[s] honest opinion[s] about the product—positive or negative" (for more detailed information, see Amazon 2020).

Using a customized Python-based web data crawler (version 3.6.2), we retrieved publicly available review data on the German version of Amazon's web site (Amazon.de). (See

Fig. 2 Conceptual model of Study 3
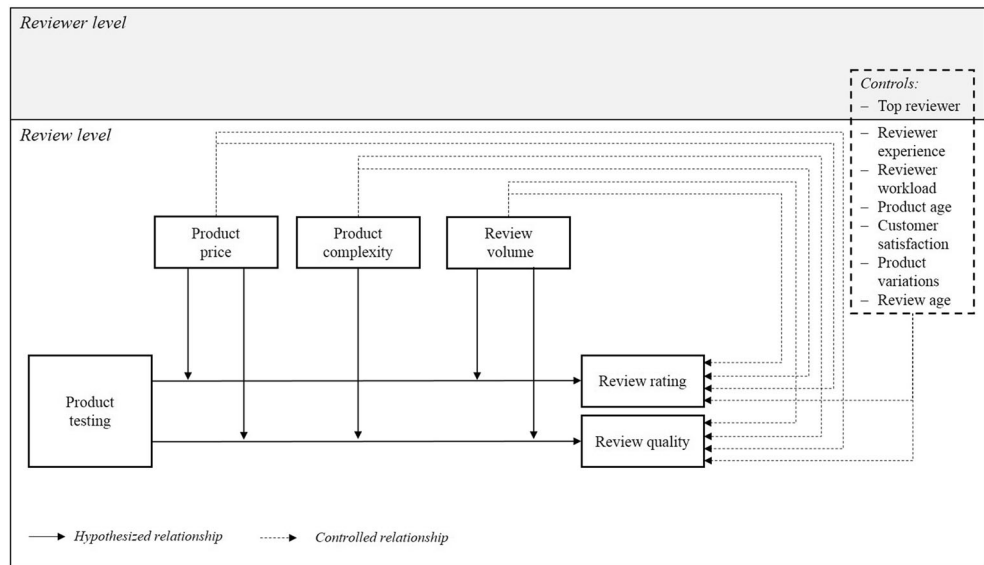
Appendix 2 for a detailed description of the data collection procedure.) All reviews were published in German between September 2000 and March 2020. The featured products spanned 13 categories, such as books, electronics, and toys (see Fig. 3). More than half (59%) of the test products were offered within 180 days of their introductions, and 72% of the Vine reviews were among the first 20 reviews of the product. That is, most product tests took place early in a product's lifecycle, when relatively few reviews were available.

Measures Tables 3, 4 and 5 detail the study measures and their descriptive statistics. We operationalized *product testing*, the independent variable, using the assigned badge "Vine Customer Review of Free Product," which automatically identifies reviews written by Amazon product testers. For the dependent variable *review rating*, we used the star rating (one to five stars) included in each review. Then we measured the dependent variable *review quality* with the Dickes-Steiwer index, an extended version of the commonly applied English-language Flesch reading ease index (Berger et al. 2020), adapted to German (Dickes and Steiwer 1977). The Flesch reading ease index has been applied previously to assess review quality (Agnihotri and Bhattacharya 2016; Gao et al. 2017; Sridhar and Srinivasan 2012). The Dickes-Steiwer index is represented by the following formula:

$$\text{Dickes}-\text{Steiwer index} = 235.96 - 73.02 \times \ln\left(\left(\frac{\text{total number of characters}}{\text{total number of words}}\right) + 1\right) - 12.56$$
$$\times \ln\left(\left(\frac{\text{total number of words}}{\text{total number of sentences}}\right) + 1\right) - 50.03 \left(\frac{\text{unique words}}{\text{total number of words}}\right),$$

such that a higher score indicates better readability or ease with which a reader can understand the written text.
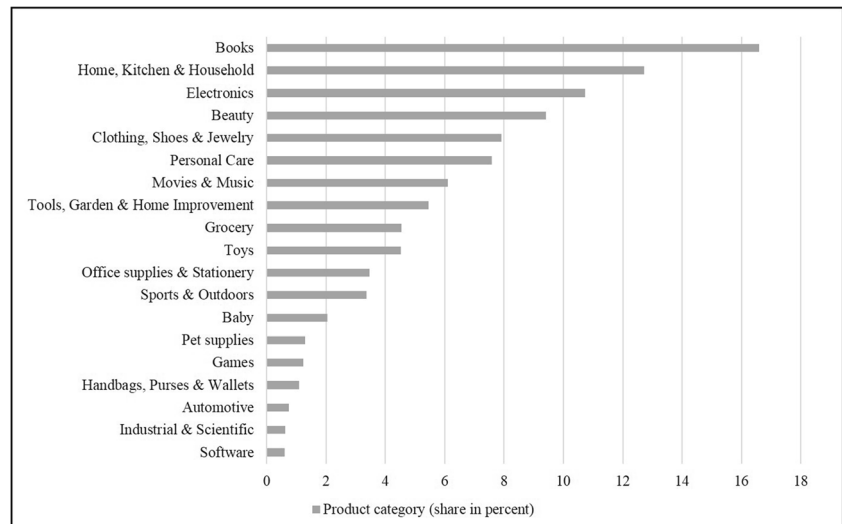
The moderating variable *product price* is the euro amount, indicated on the Amazon product page. To determine the moderator *product complexity*, we retrieve the number of questions posed by Amazon customers on the product page. *Previously published review volume* is operationalized as the number of reviews published before a given review.

In addition to the variables in our theoretical framework, we include several controls. To account for reviewer characteristics, we include the Amazon reviewer ranking, such that reviewers assigned to the Top 1000 reviewers list or the Amazon Hall of Fame are classified as top reviewers, and all others are regular reviewers. According to Amazon's guidelines, reviewers are honored as members of the Top 1000 reviewers list depending on how many reviews they write and how helpful their reviews are to other customers, and it also assigns more weight to recent reviews.[2] The Hall of Fame contains reviewers who have been successful contributors for multiple years and anyone ever identified among the Top 10. Any reviewer with at least one of these Amazon honorifics is coded as a "top reviewer."

---

[2] https://www.amazon.com/gp/customer-reviews/guidelines/top-reviewers.html/

714

J. of the Acad. Mark. Sci. (2021) 49:703–722

Fig. 3 Product category percentages in Study 3

On the review level, we control for *reviewer experience*, measured by the total number of reviews written by the reviewer at the time of the focal review. Furthermore, we include *reviewer workload* as a control variable, by retrieving the number of reviews written by the reviewer on the same day as the focal review. The control variable *product age* pertains to the time of the review, operationalized as the number of days between when the product first became available and the review date. To control for *customer satisfaction*, we obtain the average number of stars assigned to a product across all reviews. Moreover, we control for the *availability of product variations* (e.g., different colors, editions, sizes), to acknowledge that some products are standardized and the same for everyone, whereas others might be adapted to personal preferences. Finally, we note the *review age* by recording how many days had passed since it was written at the moment of the data collection.

Multilevel approach Because online reviews are nested within reviewers, we apply multilevel modeling. In contrast with an ordinary least squares approach, multilevel modeling acknowledges that online reviews written by the same person are not independent. The way a review is written likely varies from reviewer to reviewer (e.g., some product testers evaluate products more positively in general and write higher quality reviews than others), so we simultaneously analyze the data on two levels (i.e., review level and reviewer level).

The intraclass correlations (ICCs) confirm the need for multilevel modeling. The results show that up to 10.5% of the total variance of review rating and up to 31.7% of the total variance of review quality may be attributed to differences among reviewers, indicating significant variations. The ICC value thus signals a high proportion of between-group variance relative to total variance for both constructs. Values of 1–5% can already lead to significant distortions (Cohen et al.

2003), such that our ICCs indicate the need for multilevel modeling.

To analyze our multilevel data, we calculate random intercept and slopes models, with within-level interactions, using MPlus 8.3 for the two dependent variables, review rating and review quality (Muthén and Muthén 2020). The binary variables were zero-centered; all other variables were grand-mean-centered, in line with recommendations in multilevel methodology literature (Luke 2019). The review-level (level 1) equation is as follows:

$$Y_{ij} = \beta_{0j} + \beta_{1j} \times \text{Product testing}_{ij} + \beta_2$$
$$\times \text{Product price}_{ij} + \beta_3 \times \text{Product testing}_{ij}$$
$$\times \text{Product price}_{ij} + \beta_4 \times \text{Product complexity}_{ij}$$
$$+ \beta_5 \times \text{Product testing}_{ij} \times \text{Product complexity}_{ij}$$
$$+ \beta_6 \times \text{Review volume}_{ij} + \beta_7 \times \text{Product testing}_{ij}$$
$$\times \text{Review volume}_{ij} + \beta_8 \times \text{Reviewer experience}_{ij}$$
$$+ \beta_9 \times \text{Reviewer workload}_{ij} + \beta_{10}$$
$$\times \text{Product age}_{ij} + \beta_{11} \times \text{Customer satisfaction}_{ij}$$
$$+ \beta_{12} \times \text{Product variations}_{ij} + \beta_{13} \times \text{Review age}_{ij}$$
$$+ e_{ij}, \tag{1}$$

where i denotes review, j indicates the reviewer, and Y refers to the dependent variable (either review ratings or review quality). The reviewer-level model (level 2) then captures the differences between reviewers and explains the regression intercept and the slope of product testing, respectively:

$$\beta_{0j} = \gamma_{00} + \gamma_{01} \times \text{Top reviewer}_j + \mu_{0j} \tag{2}$$

Table 3 Measures for Study 3

Construct	Measure	Description
Independent variables		
Product testing	Amazon Vine badge	
Product price	Amazon price	Price indicated on the Amazon product page
Product complexity	Number of questions	Number of questions posed by other customers on the Amazon product page
Review volume	Number of previously published reviews	How many reviews of the product were published before the focal review on Amazon
Dependent variables		
Review rating	Number of stars	Rating (one to five stars)
Review quality	Dickes-Steiwer index	Index of the readability level of the text
Controls		
Top reviewer	Top reviewer (yes/no)	Top 1000 reviewer or Amazon Hall of Fame (HOF)*
Reviewer experience	Total number of reviews per reviewer at the time of the review	Gained experience in writing online reviews at the time of the review
Reviewer workload	Number of reviews per reviewer written on the same day	Workload in terms of the number of reviews written on the same day
Product age	Product age in days	Days between first product availability on Amazon and review date
Customer satisfaction	Average product rating	Average number of stars of all online reviews on the product
Product variations	Product variations (yes/no)	Variations of the product available (e.g., color, editions, sizes)
Review age	Review age in days	Review age in days at the time of data collection

* The Amazon Hall of Fame honors reviewers that have been highly ranked. Any reviewer who has risen to the Top 10, even if for a day only, will receive a permanent Hall of Fame badge

$$\beta_{1j} = \gamma_{10} + \mu_{1j} \qquad\qquad (3)$$

where β represents the regression coefficients on level 1; γ refers to the regression coefficients on level 2; e is the residual value on level 1; and μ indicates residual values on level 2.

Results

Table 6 summarizes the results of our multilevel analysis, which reveal no significant direct effects of product testing on review ratings (b = .002, p = .958) or review quality (b = .060, p = .829). In contrast with H1 and H3, but in line with our results from Study 2, product testing does not lead to more positive product ratings nor to higher quality reviews.

In line with H5, we find support for the moderating effect of product price, including positive interaction effects of product testing and product price on review ratings (b = .025, p = .006) and review quality (b = .317, p = .001). As the price of the product increases, product testers tend to give more positive ratings and reviews of higher quality. Product complexity influences the relationship between product testing and review quality (b = .002, p = .046), as we predicted in H6. However, the moderating effect of product complexity on the link between

product testing and review ratings is not significant (b = .001, p = .122). When they receive more complex products, testers do not necessarily give higher or lower ratings, but they write higher quality reviews. In line with H7a, we find a positive interaction effect between product testing and previously published review volume on review ratings (b = .015, p = .017). Consistent with H7b, previously published review volume also negatively moderates the effect of product testing on review quality (b = −.169, p = .030). As the volume of previously published reviews increases, product testing positively affects review ratings, but it negatively influences review quality.

Discussion

Key findings

Marketing managers widely acknowledge that online product reviews can shape readers' attitudes and behaviors (Minnema et al. 2016) and thereby influence the firm's performance (Chintagunta et al. 2010). Thus it is not surprising that firms continually search for ways to increase the number of positive reviews of their products (Haenlein

Table 4 Correlations, means, and standard deviations for Study 3

Variables	1	2	3	4	5	6	7	8	9	10	11	12	13
Reviewer level													
1. Top reviewer	1												
Review level													
2. Product testing	−.031**	1											
3. Product price[a]	.058**	.267**	1										
4. Product complexity	.027**	.011**	.226**	1									
5. Review volume[a]	−.034**	.140**	−.003	.242**	1								
6. Reviewer experience[a]	.214**	.404**	.162**	−.017**	−.005*	1							
7. Reviewer workload[a]	.089**	.163**	.006*	−.030**	.125**	.196**	1						
8. Product age[a]	.036**	−.180**	−.054**	.004	.215**	−.102**	.100**	1					
9. Customer satisfaction	.021**	−.071**	.007*	.022**	.067**	.058**	.123**	.042**	1				
10. Product variations	−.042**	−.155**	−.118**	.076**	.072**	−.014**	−.042**	−.040**	.138**	1			
11. Review age[a]	−.038**	−.321**	−.163**	.032**	−.137**	−.283**	−.357**	.197**	−.224**	.047**	1		
12. Review rating	.006*	.006*	.051**	.018**	−.022**	.042**	.080**	−.011**	.332**	.022**	−.079**	1	
13. Review quality	.061**	.031**	.041**	.019**	.026**	.112**	.048**	−.091**	−.010**	.035**	−.058**	−.019**	1
Mean	.21	.64	2.88	9.80	2.45	5.77	1.01	5.41	4.16	.60	6.71	4.28	30.26
Standard deviation	.41	.48	1.15	45.19	1.48	1.34	.91	1.47	.53	.49	1.30	.97	7.91

* $p < .05$. ** $p < .01$

Notes: Significance is based on two-tailed tests

[a] The variable is non-normally distributed and therefore ln-transformed

and Libai 2017; Kim et al. 2016), such as by offering financial incentives (Burtch et al. 2018; Khern-am-nuai et al. 2018) or potentially even engaging in illegal practices, such as paying for fake reviews (Anderson and Simester 2014; Moon et al. 2019; Wu et al. 2020). However, these approaches are neither consistently effective nor advisable for firms.

With three studies, we suggest some alternative methods, and we contribute to marketing literature by assessing the effectiveness of product testing programs. As we find, product testing programs do not necessarily prompt better quality or higher review ratings; rather, the effect depends on the context, as established by product prices, product complexity, and previously published review volume. Customers taking part in product testing programs offer more positive ratings for higher priced and more extensively reviewed products. Moreover, when the test product is high priced or complex, participants in the program offer higher quality reviews. In contrast, if many reviews already are available, product testers tend to devote less effort and thus produce a review of lower quality. These findings offer important implications for marketing theory and practice.

Theoretical implications

We contribute to marketing theory in two ways. First, we extend previous work on seeding programs by conceptualizing and assessing product testing programs as a special format of seeding programs. But previously studied seeding programs differ from product testing programs, in terms of the products provided and the associated pressures on participants. That is, free samples only potentially evoke social influence from seeds, who typically do not face any obligations in exchange for receiving a test product, so prior seeding program research cannot account for the specific effects of product testing programs on reviewing behavior. This contribution is particularly notable, in that we find some positive effects of product testing programs on review ratings and quality, as desired by companies, but we also identify some contextual factors that lead to negative effects.

Second, this study highlights the importance of accounting for different, potentially contradictory theoretical mechanisms when anticipating the outcomes of product testing programs. We theoretically derive and experimentally confirm two

Table 5 Descriptive statistics, two-level dataset, Study 3

Variable	Min	Max	N	Mean	SD	Mean of Product Testing Reviews	Mean of Non-Product Testing Reviews
Product testing	0	1	207,254	.64	.48	–	–
Product price[a]	−4.61	9.39	144,346	2.88	1.15	3.11	2.47
Product complexity	0	977	204,874	9.80	45.19	10.16	9.16
Review volume[a]	0	8.48	161,450	2.45	1.48	2.60	2.17
Top reviewer	0	1	207,254	.21	.41	.20	.23
Reviewer experience[a]	0	8.15	206,853	5.77	1.34	6.18	5.05
Reviewer workload[a]	0	4.84	207,254	1.00	.91	1.12	.81
Product age[a]	0	10.68	164,612	5.41	1.47	5.24	5.82
Customer satisfaction	1	5	195,206	4.16	.53	4.13	4.21
Product variations	0	1	207,254	.60	.49	.54	.70
Review age[a]	0	10.64	194,113	6.71	1.30	6.42	7.30
Review rating	1	5	207,254	4.28	.97	4.29	4.28
Review quality	−35.96	79.75	207,254	30.26	7.91	30.44	29.93

[a] The variable is non-normally distributed and therefore ln-transformed

opposing mechanisms that explain the effects of product testing programs on reviewing behavior: perceived inequity and perceived pressure. On the one hand, product testing program participants assess what they receive, compared with what they must do in return. Their assessment of this outcome-to-input ratio typically is in their favor—a free product seems like a good outcome for writing a review—so product testers sense inequity, which they try to resolve by giving more to the exchange, in the form of a better rating or higher quality review. On the other hand, the pressure to complete the review can make product testing program participants feel restricted in their behavioral freedom, in terms of whether and when to write the online review. This sentiment can reduce review ratings and review quality, potentially cancelling out the positive effects of product testing programs via perceived inequity.

Managerial implications

When designing product testing programs, companies might seek three distinct goals: increase the number of reviews (volume), increase product ratings, or increase review quality. Our study offers guidelines for how managers can leverage product testing programs to achieve their specific goals (Fig. 4).

First, companies often want to increase the number of reviews their products receive, because greater review volume has positive consequences for purchase behavior (Dellarocas et al. 2007; Liu 2006). Companies offering a new product are especially likely to suffer from insufficient online reviews. Product testing program participants are obliged to write an online review, so such programs result in more online reviews. For firms interested mainly in increasing the number of online reviews, a product testing program is a viable tactic.

Second, beyond the number of reviews, companies likely care also about the ratings their products receive, with the recognition that ratings exert strong influences on customer purchasing behavior and willingness to pay a higher price (Kübler et al. 2018; Marchand et al. 2017). In our analysis of more than 200,000 online reviews in Study 3, we do not find an overall effect of product testing programs on review ratings; that is, review ratings offered within a product testing program context are not generally better than routine reviews for purchased products. This finding might surprise many companies that hope to prompt better ratings by giving away free products. Notably, product ratings improve if the product testing program offers high-priced products or those that already have accumulated many reviews. However, even in these cases, the improvement to product ratings is only modest. Thus, companies that primarily seek to increase product ratings should avoid product testing programs. Because the programs do not reduce review ratings though, companies that seek greater review volume still can successfully offer product testing programs without risking a decline in their product ratings.

Third, if a company's goal is to attract higher quality reviews, product testing programs can be effective in specific circumstances, namely, if the program includes high-priced or more complex products. These findings are especially worthwhile for companies that sell such products. Customers perceive purchases of expensive, complex products as risky and tend to rely heavily on product reviews to gather additional, company-independent, and (theoretically) unbiased information (Liu et al. 2019). High-quality reviews effectively reduce uncertainty in the prepurchase phase (Kostyra et al. 2016). Consequently, these companies should offer their high-priced,

718

J. of the Acad. Mark. Sci. (2021) 49:703–722

Table 6 Results, two-level dataset, Study 3

Independent Variables	Dependent Variables			
	Review rating		Review quality	
	b	SE	b	SE
Intercept	4.226***	0.032	29.026***	0.327
Simple effects				
Product testing (yes=1)	0.002	0.032	0.060	0.276
Product price[a]	0.023**	0.008	0.359***	0.107
Product complexity	< 0.001*	< 0.001	−0.001	0.001
Review volume[a]	−0.038***	0.007	0.076	0.087
Interactions				
Product testing x Product price[a]	0.025**	0.010	0.317***	0.103
Product testing x Product complexity	< 0.001	< 0.001	0.002*	0.001
Product testing x Review volume[a]	0.015*	0.007	−0.169*	0.089
Controls				
Top reviewer (yes=1)	0.072	0.052	0.991	0.863
Reviewer experience[a]	−0.007	0.010	0.533***	0.115
Reviewer workload[a]	0.029***	0.007	−0.478***	0.076
Product age[a]	−0.001	0.004	−0.363***	0.035
Customer satisfaction	0.618***	0.023	−0.683***	0.062
Product variations (yes=1)	0.019*	0.009	0.571***	0.080
Review age[a]	0.016*	0.007	0.038	0.075
−2 log-likelihood	241,886.388 (d.f. = 17)		628,463.358 (d.f. = 17)	

*$p < .05$; **$p < .01$; ***$p < .001$. The p-values for directional hypotheses are one-sided

[a] The variable is non-normally distributed and therefore ln-transformed

complex products through product testing programs. Complex products should only be offered for a limited period though. That is, once the product has attracted a multitude of reviews, product testing programs tend to lead to lower quality reviews, even if the ratings in these reviews might be better. Overall, product testing programs should be used strategically by companies interested in increasing review quality.

Regardless of their primary goal, companies should recognize that the effects of product testing programs on reviewing behavior are complex and context-dependent and thus calculate the returns of the program for their specific products. A product testing program could enhance sales by increasing the number of reviews; in certain circumstances, the program can increase product ratings and the quality of online reviews too,
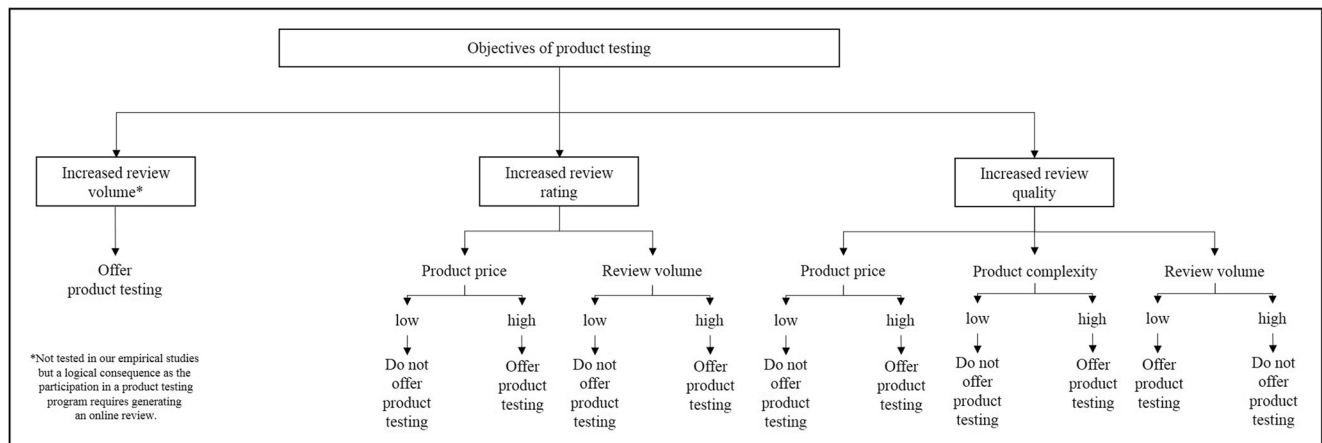

Fig. 4 Guidelines for managers

which heighten purchase intentions and willingness to pay (Kostyra et al. 2016). However, product testing programs also have substantial costs; the manufacturer has to provide the test products for free and also might pay fees to a retailer or agency that manages the program. Therefore, profitability must be calculated on an individual product basis.

Our theoretically derived and experimentally confirmed psychological mechanisms offer further managerial insights. Both perceived inequity and perceived pressure can be leveraged with appropriate program designs. For example, if product testers view the outcome-to-input ratio as favorable, they are likely to strive to reestablish an equitable relationship by providing more positive product ratings and higher quality reviews. Companies might aim to increase participants' perceived inequity, such as by presenting their test product as especially worthy or emphasizing its popularity, as well as by making partcipants' inputs less effortful, whether by eliminating minimum length requirements or facilitating the process for submitting online reviews. Increasing perceived outcome while minimizing perceived input can increase product testers' sense of inequity, which should increase their review ratings and quality.

Managers also might seek to decrease perceived pressures on product testers, to prevent the potential negative effects of product testing programs. For example, companies that care less about review volume might simply state that they would appreciate, but do not require, the product tester's feedback in the form of a review. If companies are interested in review volume and do not face stringent time constraints, they also could give product testers more time to write their reviews and thus grant them freedom with regard to when to accomplish the task. Such extensions also would grant product testers more time to experience the product, which may reduce their perceptions of pressure further.

Limitations and further research

The study has limitations that suggest research opportunities. First, our investigation focuses on product testers who write online reviews, not the effects of product testing programs on review readers. It would be interesting to examine how recipients perceive online reviews clearly marked as written by a product tester. They might see these reviews as less trustworthy, because they interpret the free product provision as a sort of "bribe" to the reviewer. But they also might regard product testers as experts, which could increase their review credibility. It also would be interesting to analyze recipients' own communication behaviors, after they read product testers' reviews. Chae et al. (2017) find spillover effects in other types of seeding programs, which lead us to posit that readers of reviews written by product testers might be motivated to write their own reviews and add their opinions to the set of available

reviews. Continued research can address these potential effects on review recipients' perceptions and behaviors.

Second, we analyze reviewer behavior within the Amazon Vine program, offered by a retailer rather than by manufacturers, which is a common practice. However, in such product testing programs, participants actually interact with two parties: the manufacturer of the test product and the retailer or agency that conducts the product test. In our three studies, our focus was on the relationship between the participant and the manufacturer, but additional effects might pertain to the relationship between the participant and the retailer or agency. For example, a sense of inequity might prompt a felt need to give something back to the retailer or agency, as well as other potential attitudinal and behavioral shifts.

Third, we consider three contextual moderating effects (product price, product complexity, and previously published review volume) that reveal important implications with regard to whether and when to offer product testing programs. Continued studies can build on these results to test other contextual factors. For example, some product testing programs require participants to apply actively to test a particular product, but others do not. In some cases, the product testing programs offer products at a reduced price, rather than for free. Each of these program design elements might exert distinct effects on participants' reviewing behaviors. According to Shampanier et al.'s (2007) finding that customers overreact to free products, participants paying a reduced price might devote less effort to reviewing a product than participants who receive it for free. By comparing different program characteristics, researchers could generate additional advice for appropriate designs of product testing programs.

Appendix 1: Data collection and analyses (Study 1: Qualitative studies)

Sample A

We needed a sample of product testers, a specific requirement not included in Prolific's general filter options, so we ran an individualized, prescreening question, designed to identify participants of interest for Study 1. Some of the prescreening effort relied on Prolific's existing filter options. That is, we restricted the sample to participants who had an Amazon account, shopped online at least about once a month, were from Great Britain, and spoke English as their first language. Then the individualized prescreening question was posed to 1000 potential participants who met these criteria. Namely, we asked if they had ever participated in a product testing program, and 120 people who indicated they had were then invited to participate in the survey, with open-ended questions. The first 100 individuals who responded to the invitation were sent the survey; this sample of product testers ranged in age between 21

and 78 years, with an average age of 43 years, and their average experience as product testers was 5 years.

The survey first asked them to describe the product testing program in which they had participated most recently in detail, including who offered it and the brand of the test product. We encouraged these participants to describe their thoughts and feelings during each part of the product test process (e.g., application, product testing, reviewing). Sample questions included, "What were your first thoughts/feelings after having been selected as a product tester/when you received the test product?" "Did the product testing experience create negative/positive feelings about the company or the product?" and "Do you think that product testing reviews differ in any way from other reviews?" Finally, we asked them to provide some sociodemographic information (e.g., age, education, occupation).

Sample B

With a purposive sampling method (Miles and Huberman 1994), we identified 40 Amazon Vine product testers with varying levels of experience (long and short reviewing history, top-ranked and average reviewers), distinct product interests, and different ages and occupational backgrounds, all based in Germany. Using the contact details in their Amazon profile, we sent them personalized invitation letters, via e-mail, outlining our research in terms of the general topic, estimated duration of the interview, and intention to record the interview. One author then contacted testers, one at a time, and interviewed each until information redundancy occurred—which happened after 12 product testers had been interviewed. This number has been found to be sufficient for reaching theoretical saturation (Miles and Huberman 1994). These product testers ranged between 19 and 62 years of age, with an average of 40 years, and their average experience as product testers was 4 years.

By gathering these data through in-depth interviews, we could gain insights into the respondents' own interpretations of their environments and understand their underlying thoughts and feelings better (Miles and Huberman 1994). We used a semi-structured interview guide, such that after we provided a brief description of the research project and some introductory questions about their shopping and reviewing behaviors, we asked the participants to talk freely about their most recent product testing experience and describe all its steps, from the first to the last contact with Amazon, as well as how they behaved and felt during each step. They also were asked to indicate if these behaviors and feelings were typical for the product tests in which they had participated and, if not, to identify what factors might have led them to behave or feel differently.

The 12 interviews were carried out in German, via video conferencing, during September–November 2020. One author conducted all of the interviews, which varied in length from 24 to 86 min (average of 42 min). With the consent of the participants, the interviews were audio recorded and transcribed verbatim (152 single-spaced pages).

Appendix 2: Amazon Vine data collection procedure (Study 3)

To acquire a random selection of Amazon Vine product testers, we developed an algorithm, starting with 10 Vine product testers chosen by our customized Python program. The Amazon reviewer IDs (with which links to their reviewer profiles can automatically be created) were saved in a separate .TXT file. From this list, we reviewed randomly selected Amazon Vine testers' reviewer profiles and checked all reviews written by this specific reviewer, to determine if they took place within the Amazon Vine program (i.e., if the review featured the badge "Vine Customer Review of Free Product"). If a review was written within the product testing program, the product ID (with which a link to a given product page can automatically be created) was saved in a second, separate .TXT file.

After we read all reviews by a specific Vine tester, whether written within the confines of the Amazon Vine program or not, we chose a product from the Vine product list randomly, to find other Vine testers. If a review writer for a specific product was part of the Amazon Vine program, the reviewer ID was added to the Vine tester list. Thus, the Vine tester list was continually being expanded. After all reviews of a product were checked, we restarted the process by opening the Vine tester list and checking the next randomly chosen Vine tester from that list for products in the Amazon Vine program.

Thus, we created a list of 400 randomly collected Amazon Vine tester IDs, from which we retrieved data about their Amazon profiles, including all the reviews each tester had written. Amazon's website is based on HTML and JavaScript code. Our Python script removed the HTML and JavaScript formatting from the text and extracted the data of interest. It then saved this information to Excel files. For each review, we captured whether it was written as part of the Vine product testing program, the review rating, the review text, previously published review volume (i.e., how many reviews of the product were published before the focal review), and the review age at the time of the data collection. Moreover, we retrieved the product price, the average rating the product received, the product age at the time of the review, if product variations were available (e.g., color, sizes, editions), and the number of questions posed by Amazon customers about the product. Finally, we captured the reviewer's rank, the total number of reviews by each reviewer at the time of each review, and the Amazon reviewer ID, to identify to which reviewer a review belonged.

References

Adams, J. S. (1963). Towards an understanding of inequity. *Journal of Abnormal and Social Psychology, 67*(5), 422–436.

Adams, J. S. (1965). Inequity in social exchange. In L. Berkowitz (Ed.), *Advances in experimental social psychology* (Vol. 2, pp. 267–299). New York: Academic Press.

Agnihotri, A., & Bhattacharya, S. (2016). Online review helpfulness: Role of qualitative factors. *Psychology & Marketing, 33*(11), 1006–1017.

Ajzen, I. (1982). On behaving in accordance with one's attitudes. In M. P. Zanna, E. T. Higgins, & C. P. Herman (Eds.), *Consistency in social behavior: The Ontario symposium* (Vol. 2, pp. 3–15). Hillsdale: Erlbaum.

Amazon (2020). What is Amazon Vine? https://www.amazon.com/gp/vine/help?language=en_US. Accessed 8 December 2020.

Anderson, E. T., & Simester, D. I. (2014). Reviews without a purchase: Low ratings, loyal customers, and deception. *Journal of Marketing Research, 51*(3), 249–269.

Askalidis, G., Kim, S. J., & Malthouse, E. C. (2017). Understanding and overcoming biases in online review systems. *Decision Support Systems, 97*, 23–30.

Austin, W., & Walster, E. (1975). Equity with the world: The trans-relational effects of equity and inequity. *Sociometry, 38*(4), 474–496.

Bawa, K., & Shoemaker, R. (2004). The effects of free sample promotions on incremental brand sales. *Marketing Science, 23*(3), 345–363.

Berger, J., Humphreys, A., Ludwig, S., Moe, W. W., Netzer, O., & Schweidel, D. A. (2020). Uniting the tribes: Using text for marketing insight. *Journal of Marketing, 84*(1), 1–25.

Brady, M. K., Voorhees, C. M., & Brusco, M. J. (2012). Service sweethearting: Its antecedents and customer consequences. *Journal of Marketing, 76*(2), 81–98.

Brehm, J. W. (1966). *A theory of psychological reactance*. Oxford: Academic Press.

Brehm, J. W. (1972). *Responses to loss of freedom: A theory of psychological reactance*. Morristown: General Learning Press.

Brehm, J. W. (1989). Psychological reactance: Theory and applications. *Advances in Consumer Research, 16*, 72–75.

Brightlocal (2018). Local consumer review survey, https://www.brightlocal.com/research/local-consumer-review-survey/?SSAID=314743&SSCID=81k3_bvosq. Accessed 8 December 2020.

Burtch, G., Hong, Y., Bapna, R., & Griskevicius, V. (2018). Stimulating online reviews by combining financial incentives and social norms. *Management Science, 64*(5), 2065–2082.

Campbell, D. T., & Stanley, J. C. (1963). *Experimental and quasi-experimental designs for research*. Boston: Houghton Mifflin.

Cao, Q., Duan, W., & Gan, Q. (2011). Exploring determinants of voting for the "helpfulness" of online user reviews: A text mining approach. *Decision Support Systems, 50*(2), 511–521.

Carrell, M. R., & Dittrich, J. E. (1978). Equity theory: The recent literature, methodological considerations, and new directions. *Academy of Management Review, 3*(2), 202–210.

Chae, I., Stephen, A. T., Bart, Y., & Yao, D. (2017). Spillover effects in seeded word-of-mouth marketing campaigns. *Marketing Science, 36*(1), 89–104.

Chen, X., Van Der Lans, R., & Phan, T. Q. (2017). Uncovering the importance of relationship characteristics in social networks: Implications for seeding strategies. *Journal of Marketing Research, 54*(2), 187–201.

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

Chintagunta, P. K., Gopinath, S., & Venkataraman, S. (2010). The effects of online user reviews on movie box office performance: Accounting for sequential rollout and aggregation across local markets. *Marketing Science, 29*(5), 944–957.

Clee, M. A., & Wicklund, R. A. (1980). Consumer behavior and psychological reactance. *Journal of Consumer Research, 6*(4), 389–405.

Cohen, J., Cohen, P., West, S. G., & Aiken, L. S. (2003). *Applied multiple regression/correlation analysis for the behavioral sciences*. Mahwah: Lawrence Erlbaum Associates.

Cui, G., Lui, H. K., & Guo, X. (2012). The effect of online consumer reviews on new product sales. *International Journal of Electronic Commerce, 17*(1), 39–57.

Datta, H., Foubert, B., & Van Heerde, H. J. (2015). The challenge of retaining customers acquired with free trials. *Journal of Marketing Research, 52*(2), 217–234.

Dellarocas, C., Zhang, X. M., & Awad, N. F. (2007). Exploring the value of online product reviews in forecasting sales: The case of motion pictures. *Journal of Interactive Marketing, 21*(4), 23–45.

Dickes, P., & Steiwer, L. (1977). Ausarbeitung von Lesbarkeitsformeln für die deutsche Sprache. *Zeitschrift für Entwicklungspsychologie und Pädagogische Psychologie, 9*(1), 20–28.

Fang, E., Palmatier, R. W., & Evans, K. R. (2008). Influence of customer participation on creating and sharing of new product value. *Journal of the Academy of Marketing Science, 36*(3), 322–336.

Foubert, B., & Gijsbrechts, E. (2016). Try it, you'll like it—Or will you? The perils of early free-trial promotions for high-tech service adoption. *Marketing Science, 35*(5), 810–826.

Gao, B., Hu, N., & Bose, I. (2017). Follow the herd or be myself? An analysis of consistency in behavior of reviewers and helpfulness of their reviews. *Decision Support Systems, 95*, 1–11.

Garnefeld, I., Helm, S., & Grötschel, A. (2020). May we buy your love? Psychological effects of monetary incentives on writing likelihood and valence of online product reviews. *Electronic Markets, 30*, 805–820.

Gedenk, K., & Neslin, S. A. (1999). The role of retail promotion in determining future brand loyalty: Its effect on purchase event feedback. *Journal of Retailing, 75*(4), 433–459.

Greenberg, J. (1982). Approaching equity and avoiding inequity in groups and organizations. In J. Greenberg & R. L. Cohen (Eds.), *Equity and justice in social behavior* (pp. 389–435). New York: Academic Press.

Haenlein, M., & Libai, B. (2013). Targeting revenue leaders for a new product. *Journal of Marketing, 77*(3), 65–80.

Haenlein, M., & Libai, B. (2017). Seeding, referral, and recommendation: Creating profitable word-of-mouth programs. *California Management Review, 59*(2), 68–91.

Hayes, A. F. (2018). *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach* (2nd ed.). New York: Guilford Publications.

Hinz, O., Skiera, B., Barrot, C., & Becker, J. U. (2011). Seeding strategies for viral marketing: An empirical comparison. *Journal of Marketing, 75*(6), 55–71.

Homburg, C., Hoyer, W. D., & Stock, R. M. (2007). How to get lost customers back? *Journal of the Academy of Marketing Science, 35*(4), 461–474.

Hu, N., Zhang, J., & Pavlou, P. A. (2009). Overcoming the J-shaped distribution of product reviews. *Communications of the ACM, 52*(10), 144–147.

Jain, D., Mahajan, V., & Muller, E. (1995). An approach for determining optimal product sampling for the diffusion of a new product. *Journal of Product Innovation Management, 12*(2), 124–135.

Khern-am-nuai, W., Kannan, K., & Ghasemkhani, H. (2018). Extrinsic versus intrinsic rewards for contributing reviews in an online plat-form. *Information Systems Research, 29*(4), 871–892.

Kim, J., Naylor, G., Sivadas, E., & Sugumaran, V. (2016). The unrealized value of incentivized eWOM recommendations. *Marketing Letters, 27*(3), 411–421.

Kim, J. Y., Natter, M., & Spann, M. (2014). Sampling, discounts or pay-what-you-want: Two field experiments. *International Journal of Research in Marketing, 31*(3), 327–334.

Kostyra, D. S., Reiner, J., Natter, M., & Klapper, D. (2016). Decomposing the effects of online customer reviews on brand, price, and product attributes. *International Journal of Research in Marketing, 33*(1), 11–26.

Kübler, R., Pauwels, K., Yildirim, G., & Fandrich, T. (2018). App pop-ularity: Where in the world are consumers most sensitive to price and user ratings? *Journal of Marketing, 82*(5), 20–44.

Larsen, D., & Watson, J. J. (2001). A guide map to the terrain of gift value. *Psychology & Marketing, 18*(8), 889–906.

Li, H., Jain, S., & Kannan, P. K. (2019). Optimal design of free samples for digital products and services. *Journal of Marketing Research, 56*(3), 419–438.

Libai, B., Muller, E., & Peres, R. (2013). Decomposing the value of word-of-mouth seeding programs: Acceleration versus expansion. *Journal of Marketing Research, 50*(2), 161–176.

Liu, X., Lee, D., & Srinivasan, K. (2019). Large-scale cross-category analysis of consumer review content on sales conversion leveraging deep learning. *Journal of Marketing Research, 56*(6), 918–943.

Liu, Y. (2006). Word of mouth for movies: Its dynamics and impact on box office revenue. *Journal of Marketing, 70*(3), 74–89.

Lu, S., Wu, J., & Tseng, S. L. A. (2018). How online reviews become helpful: A dynamic perspective. *Journal of Interactive Marketing, 44*, 17–28.

Luca, M. (2011). Reviews, reputation, and revenue: The case of Yelp.com. *Harvard Business School NOM Unit Working Paper, 12-016*, 1–39.

Luke, D. A. (2019). *Multilevel modeling* (Vol. 143, 2nd ed.). Thousand Oaks: Sage.

Magno, F., Cassia, F., & Bruni, A. (2018). "Please write a (great) online review for my hotel!" guests' reactions to solicited reviews. *Journal of Vacation Marketing, 24*(2), 148–158.

Marchand, A., Hennig-Thurau, T., & Wiertz, C. (2017). Not all digital word of mouth is created equal: Understanding the respective impact of consumer reviews and microblogs on new product success. *International Journal of Research in Marketing, 34*(2), 336–354.

Miles, M. B., & Huberman, A. M. (1994). *Qualitative data analysis: An expanded sourcebook*. Thousand Oaks: Sage.

Minnema, A., Bijmolt, T. H. A., Gensler, S., & Wiesel, T. (2016). To keep or not to keep: Effects of online customer reviews on product returns. *Journal of Retailing, 92*(3), 253–267.

Moon, S., Kim, M. Y., & Bergey, P. K. (2019). Estimating deception in consumer reviews based on extreme terms: Comparison analysis of open vs. closed hotel reservation platforms. *Journal of Business Research, 102*, 83–96.

Mudambi, S. M., & Schuff, D. (2010). What makes a helpful review? A study of customer reviews on Amazon.com. *MIS Quarterly, 34*(1), 185–200.

Muthén, L. K., & Muthén, B. (2020). *Mplus. The comprehensive model-ing program for applied researchers: User's guide* (8th ed.). Los Angeles: Muthén & Muthén.

Nejad, M. G., Amini, M., & Babakus, E. (2015). Success factors in product seeding: The role of homophily. *Journal of Retailing, 91*(1), 68–88.

Paharia, N. (2020). Who receives credit or blame? The effects of made-to-order production on responses to unethical and ethical company production practices. *Journal of Marketing, 84*(1), 88–104.

Petrescu, M., O'Leary, K., Goldring, D., & Mrad, S. B. (2018). Incentivized reviews: Promising the moon for a few stars. *Journal of Retailing and Consumer Services, 41*, 288–295.

Philips (2020). About the Philips product tester program. https://www.producttester.philips.com/s/?language=en_US&locale=en_US. Accessed 8 December 2020.

Scheer, L. K., Kumar, N., & Steenkamp, J. B. E. (2003). Reactions to perceived inequity in U.S. and Dutch interorganizational relation-ships. *Academy of Management Journal, 46*(3), 303–316.

Shampanier, K., Mazar, N., & Ariely, D. (2007). Zero as a special price: The true value of free products. *Marketing Science, 26*(6), 742–757.

Sridhar, S., & Srinivasan, R. (2012). Social influence effects in online product ratings. *Journal of Marketing, 76*(5), 70–88.

Sunder, S., Kim, K. H., & Yorkston, E. A. (2019). What drives herding behavior in online ratings? The role of rater experience, product portfolio, and diverging opinions. *Journal of Marketing, 83*(6), 93–112.

Super savvy me (2020). Welcome to the savvy circle! https://circle.supersavvyme.co.uk/uk/. Accessed 8 December 2020.

Unger, L. S., & Kernan, J. B. (1983). On the meaning of leisure: An investigation of some determinants of the subjective experience. *Journal of Consumer Research, 9*(4), 381–392.

Wu, P. F. (2019). Motivation crowding in online product reviewing: A qualitative study of amazon reviewers. *Information & Management, 56*(8), 103–163.

Wu, Y., Ngai, E. W. T., Wu, P., & Wu, C. (2020). Fake online reviews: Literature review, synthesis, and directions for future research. *Decision Support Systems, 132*(5), 113280.

Yin, D., Bond, S. D., & Zhang, H. (2017). Keep your cool or let it out: Nonlinear effects of expressed arousal on perceptions of consumer reviews. *Journal of Marketing Research, 54*(3), 447–463.