

**“OPTIMAL” FEEDBACK USE IN CROWDSOURCING CONTESTS:
SOURCE EFFECT AND PRIMING INTERVENTION**

Tat Koon Koh

Hong Kong University of Science and Technology

koh@ust.hk

Online Appendix

Appendix A. Field Survey in Study A

<p>Scenario: You are taking part in a logo design contest, in which the client will award a US\$200 prize to the designer who submits the winning design. During this contest, you can submit multiple designs, and the client and/or other competing designers can give you feedback for the designs that you submit. Suppose you receive a feedback with suggestions of specific changes that you can make to a logo design that you have submitted.</p>
<p>[Intention to use seeker feedback] If the feedback was given by the client, do you agree with the following statements? #</p> <ul style="list-style-type: none"> • I will use the client's suggestions in the feedback to create designs for the contest. • I will apply the suggestions from the client's feedback in my designs for the contest. • I will generate designs based on the suggestions from the client's feedback for the contest.
<p>[Intention to use peer feedback] If the feedback was given by a competing designer in the contest, do you agree with the following statements? #</p> <ul style="list-style-type: none"> • I will use the competing designer's suggestions in the feedback to create designs for the contest. • I will apply the suggestions from the competing designer's feedback in my designs for the contest. • I will generate designs based on the suggestions from the competing designer's feedback for the contest.
<p>[Perceptions about seekers] Do you agree with the following statements about the clients using design contests? #</p> <ul style="list-style-type: none"> • Using clients' feedback increases my chance of winning contests. • Clients can be trusted in contests.
<p>[Perceptions about peers] Do you agree with the following statements about the competing designers participating in design contests? #</p> <ul style="list-style-type: none"> • Using competing designers' feedback increases my chance of winning contests. • Competing designers can be trusted in contests.
<p>[Contest experience]</p> <ul style="list-style-type: none"> • How many years of design experience do you have? • Have you ever participated in design contests? * • Based on your estimate, how many contests have you won on 99designs? (Note: leave blank if you do not use 99designs) • Based on your estimate, how many contests have you won on Crowdspring? (Note: leave blank if you do not use Crowdspring) • Based on your estimate, how many contests have you won on DesignContest? (Note: leave blank if you do not use DesignContest)
<p>[Demographic]</p> <ul style="list-style-type: none"> • In which continent do you currently reside? • What is your year of birth (YYYY)? • What is your gender? • (Optional) Please provide your email address if you wish to receive a summary report of the survey.
<p>Notes: Text in brackets are not shown in the survey. # 7-point Likert scale (<i>strongly disagree</i>... <i>strongly agree</i>) * Only asked for respondents whom I invited on Behance</p>

Appendix B. Robustness Checks for Study A

I conducted various analyses to check the robustness of the results. First, I split the sample into two groups based on whether respondents had previously won on 99designs; this was to address the possibility that the feedback use intention of 99designs winners was systematically different. According to a one-way ANOVA for the winners, feedback use differed by source ($F(1, 39) = 54.21, p < .001$). A contrast analysis showed that these respondents had a stronger intention to use seeker feedback ($M = 5.83, SD = 1.29$) than peer feedback ($M = 3.50, SD = 1.87$) ($t(39) = 7.36, p < .001$; Cohen's $d = 1.16$). In an ANOVA for respondents without winning experience on 99designs, feedback use intention also varied by source ($F(1, 11) = 11.33, p = .006$). A contrast analysis indicated a stronger intention to use seeker feedback ($M = 5.67, SD = 1.68$) than peer feedback ($M = 4.44, SD = 1.67$) ($t(11) = 3.37, p = .006$; Cohen's $d = .97$).

Second, I grouped respondents based on whether they provided their email address during the survey; those who did so could have taken the survey more seriously because they were keen to receive the report. In an ANOVA for respondents who provided their email address, feedback use differed by source ($F(1, 36) = 37.44, p < .001$). A contrast analysis indicated a stronger feedback use intention for seeker feedback ($M = 5.77, SD = 1.53$) than for peer feedback ($M = 3.79, SD = 1.97$) ($t(36) = 6.12, p < .001$; Cohen's $d = 1.01$). In an ANOVA for respondents who did not provide an email address, feedback use intention also differed by source ($F(1, 14) = 25.18, p < .001$). A contrast analysis showed that respondents had a stronger intention to use seeker feedback ($M = 5.84, SD = .95$) than peer feedback ($M = 3.53, SD = 1.57$) ($t(14) = 5.02, p < .001$; Cohen's $d = 1.30$).

Third, I controlled for gender, age, and graphic design experience in an ANOVA. The results showed that respondents' feedback use differed by source ($F(1, 14) = 10.37, p = .006$), even after accounting for their demographics.

Appendix C. Contest Procedure in Study B

Figure C1. Project Brief (Step 1)

Step 1. Submit Preliminary Idea

Contest:	Product name for software
Client:	Syfin Tech
Type of Business:	Financial Software and Technology
Detailed Brief:	We are looking for a product name for our new software that tracks real-time quotes and charting for stocks, bonds, and futures. Users can submit buy/sell orders and manage their investment portfolio through the software.
Winning Prize:	US\$15.00 bonus prize for the winning participant
Restrictions:	Do not use "Syfin" in the product names
Style of Product Name:	Open to all ideas
Maximum Number of Words in Product Names:	No Limit

Submit ONE preliminary idea; entries with multiple preliminary ideas will not be considered. You can give an explanation of your idea if you wish to do so.

Preliminary Idea: Explanation of Idea:

All solvers were shown the same project brief. I used "client" instead of "seeker" in the experiment because "client" was more intuitive to workers on microtask platforms such as MTurk.

Figure C2. Filler Task (Step 2)

Step 2. Provide Feedback

Thank you for submitting your preliminary idea. Kindly wait a moment to receive the feedback for your idea. In the meantime, please give your feedback to a preliminary idea by another competing worker.

Preliminary Idea:	Financial Trading Software
--------------------------	----------------------------

Your feedback:

I used "worker" instead of "peer" or "solver" in the experiment because "worker" was the terminology on MTurk.

Figure C3. System Prompt After Completing Filler Task (Step 2)

Step 2. Provide Feedback

Thank you for submitting your preliminary idea. While we wait for the client's feedback, please give your feedback to a preliminary idea by another competing worker.

Preliminary Idea: I think this name is too common, nothing stands out about it being real-time or unique.

Submit your feedback

This figure shows the system prompt for solvers who were assigned to receive seeker feedback after they submitted their feedback for the competing worker. For solvers who were assigned to receive peer feedback, the first sentence of the system prompt was “A competing worker is still providing a feedback for your preliminary idea.”

Figure C4. Feedback for Solvers' Preliminary Idea (Step 3)

Step 3. Submit More Name Ideas

Below is a feedback for your preliminary idea from the client.

Your Product Name Idea: StockTick

Client: Syfin Tech


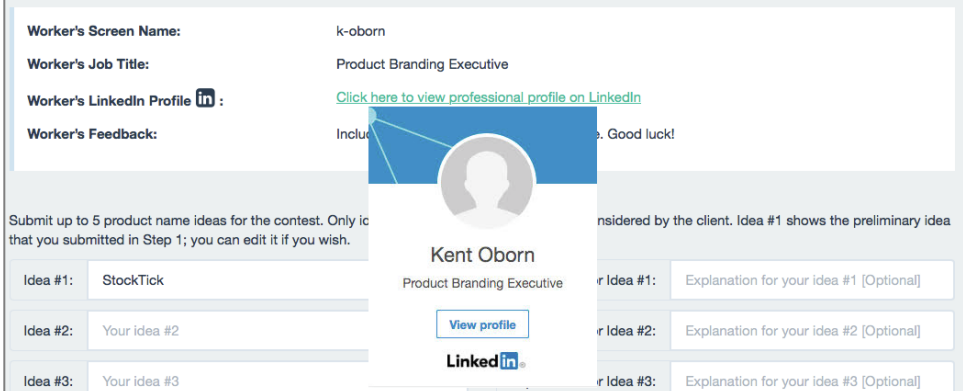
Client's Feedback: Include "Professional" in the product name. Good luck!

Submit up to 5 product name ideas for the contest. Only ideas in this final submission will be considered by the client. Idea #1 shows the preliminary idea that you submitted in Step 1; you can edit it if you wish.

Idea #1:	StockTick	Explanation for Idea #1:	Explanation for your idea #1 [Optional]
Idea #2:	Your idea #2	Explanation for Idea #2:	Explanation for your idea #2 [Optional]
Idea #3:	Your idea #3	Explanation for Idea #3:	Explanation for your idea #3 [Optional]
Idea #4:	Your idea #4	Explanation for Idea #4:	Explanation for your idea #4 [Optional]
Idea #5:	Your idea #5	Explanation for Idea #5:	Explanation for your idea #5 [Optional]

Submit your ideas

This figure shows the constructive seeker feedback.

Figure C5. Manipulation of Peer Subtypes	
Peer-1	<p>Worker's Screen Name: k-oborn</p> <p>Worker's Feedback: Include "Professional" in the product name. Good luck!</p>
Peer-2	<p>Worker's Screen Name: k-oborn</p> <p>Worker's Job Title: Product Branding Executive</p> <p>Worker's Feedback: Include "Professional" in the product name. Good luck!</p>
Peer-3	<p>Worker's Screen Name: k-oborn</p> <p>Worker's Job Title: Product Branding Executive</p> <p>Worker's LinkedIn Profile  : Click here to view professional profile on LinkedIn</p> <p>Worker's Feedback: Include "Professional" in the product name. Good luck!</p> <p>I created a LinkedIn account that had 10 connections to other LinkedIn users. A hovering box appeared when solvers clicked on "Click here to view professional profile on LinkedIn" (see below). Solvers could click on "View profile" in the box to see the peer's actual LinkedIn profile, where the name, job title, and number of connections were shown. However, profile information for the peer's connections was not displayed.</p> 

Appendix D. Analyses in Study B

Preliminary Analyses. Solvers spent an average of 55.79 seconds on the filler task, and Table D1 shows some feedback solvers gave therein. By and large, the feedback was relevant to the context, and several comments were detailed; giving such feedback was consistent with the cooperative actions among solvers reported in field studies (Bullinger et al. 2010; Füller et al. 2014). On average, the solvers used 166.94 seconds to finalize their ideas and submitted 3.64 ideas in the contest; Table D2 shows some ideas.

Table D1. Selected Feedback That Solvers Submitted in Filler Task	
1.	"This is a very simple name. Not too catchy. I do like that it tells the buyer exactly what it is for. It could be called FTS for short. Maybe make it a little more catchy."
2.	"Boring"
3.	"Straightforward, but a little bland. Not really eye catching. It may be appealing to straight and narrow stock traders, but may be overlooked by younger, fast paced, tech savvy users who are used to software and applications with catchy title."
4.	"This is a simple and to the point name which is good. People will understand right away what the software is."
5.	"I think it is a good name as it talks about something Financial. It could have a better name as far as what it does and refer to something of how its [sic] up to date."
6.	"I think this idea has great potential. Users can feel like they are right on Wall Street, but can sell/buy from their computer."
7.	"I do like this idea but i [sic] feel that it is a little too basic. I cannot say that my idea is any better, but there should be a little more of a kick to the name of a new product to get the name out there and get companies interested in it."
8.	"Flat and uninteresting. Sounds like a description rather than a product. Maybe FTS—Financial Trading Software. Nah, sounds too much like FTD for flowers."
9.	"A bit too simplistic/generic for a product name. Perhaps something a bit more 'sexy' would be more marketable."
10.	"This is too generalized. Where's the marketability?"
11.	"It explains what the software does well but I do not think it has trendy, futuristic sounding name which might make it more appealing."
12.	"It doesn't really give a clever name to the app; it is just telling you what it is. You need something cleverer in order to win! Think of something that reflects your preliminary idea and then create a catchy title for it."
13.	"Hate to be negative but too simplistic"
14.	"too generic, unless you want to make it the mysterious and unexplained FTS, or better yet FTX"
15.	"I think this is a great idea! Maybe you can add some more describing, flowery words."
<p>Note: Solvers submitted the feedback above in the filler task in Step 2, where they commented on a product name idea by a (hypothetical) competing solver. Some feedback was generic (e.g., #3, #4, and #13), whereas other feedback was developmental by suggesting how to improve the ideas (e.g., #1, #9, and #15).</p>	

Table D2. Selected Product Name Ideas Submitted by Solvers	
Initial Idea	Ideas(s) After Receiving Feedback
InvestExpress	InvestExpress; BreakevenExpress; Breakeven Now; BreakEven RealTime; RealTime BreakEven
INSTAVEST	INSTAVEST BREAKEVEN; QUICKINVEST BREAKEVEN
MAXX	MAXX; MAXX Pro; MAX2 - MAX Squared; MAXX Professional; MAXIMUS Pro
PortfolioSoft	PortfolioSoft; PortfolioSoft Pro; InvestmentFolio
StockMonster	StockMonster; Better Than Breakeven; Breakeven, Bust, or Beautiful Bonus
StockRight	StockRight Breakeven
TradeWare	TradeWare Pro; ProTrade Software

Based on solvers' responses to seven-point Likert scale survey items in Step 4, those who considered the prize to be more attractive (*"The US\$15 bonus prize for the winning participant in this contest is attractive"*; $M = 6.49, SD = .88$) attached greater importance to winning (*"Winning this contest is important to me"*; $M = 5.61, SD = 1.44$) ($r = .39, p < .001$) and reported more significant effort (*"I put a lot of effort into this contest"*; $M = 6.06, SD = 1.03$) ($r = .38, p < .001$). Also, the number of submitted ideas correlated

positively with self-reported prize attractiveness ($r = .13, p = .014$), winning importance ($r = .11, p = .030$), and effort ($r = .16, p = .002$). These were consistent with expectations based on actual contests, such as attractive prizes could increase submissions in design contests (Bockstedt et al. 2016).

Robustness Checks. I compared the use of feedback from the seeker and the respective peer subtypes using a two-way ANOVA. I found a significant interaction between source and constructiveness ($F(3, 355) = 4.91, p = .002$). Simple main effects analysis indicated that *Feedback Use* differed by source ($F(3, 355) = 37.75, p < .001$) and constructiveness ($F(1, 355) = 55.42, p < .001$). Tukey-Kramer pairwise comparison tests showed that *Feedback Use* was higher when the feedback was from the seeker ($M = .60, SD = .31$) than from *Peer-1* ($M = .24, SD = .30$) (TK test: $p < .001$; Cohen's $d = 1.18$), *Peer-2* ($M = .26, SD = .26$) (TK test: $p < .001$; Cohen's $d = 1.16$), or *Peer-3* ($M = .27, SD = .27$) (TK test: $p < .001$; Cohen's $d = 1.12$). These were consistent with H1. As reported in the main analyses, *Feedback Use* was higher for constructive feedback than for less constructive feedback, supporting H2. However, for seeker feedback, *Feedback Use* did not vary significantly by constructiveness (Constructive: $M = .61, SD = .33$; Less Constructive: $M = .57, SD = .30$; TK test: $p = .935$; Cohen's $d = .13$). By contrast, for *Peer-1* feedback, *Feedback Use* was higher for constructive feedback ($M = .39, SD = .32$) than for less constructive feedback ($M = .08, SD = .16$) (TK test: $p < .001$; Cohen's $d = 1.20$); this was also the case for *Peer-2* feedback (Constructive: $M = .40, SD = .25$; Less Constructive: $M = .12, SD = .19$; TK test: $p < .001$; Cohen's $d = 1.30$) and *Peer-3* feedback (Constructive: $M = .35, SD = .29$; Less Constructive: $M = .18, SD = .22$; TK test: $p = .042$; Cohen's $d = .68$). Thus, across peer subtypes, the positive relationship between feedback use and constructiveness was stronger for peer feedback than for seeker feedback, supporting H3.

I considered other measures of *Feedback Use*. As solvers generated an initial idea before receiving feedback, their final ideas could be regarded as being based on the former. Thus, I examined the final ideas that did not include any terms in the corresponding initial idea to obtain a more conservative estimate of the effect of feedback. I excluded 166 (45.7%) solvers whose final ideas all contained terms in their initial idea; the number of their submitted ideas being considered was zero, rendering *Feedback Use* undefined. A two-way ANOVA indicated a significant interaction of source and constructiveness ($F(1, 193) =$

3.97, $p = .048$). Simple main effects analysis showed that *Feedback Use* differed by source ($F(1, 193) = 75.32, p < .001$) and constructiveness ($F(1, 193) = 13.61, p < .001$). Tukey-Kramer tests showed that *Feedback Use* was higher for seeker feedback ($M = .77, SD = .39$) than for peer feedback ($M = .23, SD = .38$) (TK test: $p < .001$; Cohen's $d = 1.40$). *Feedback Use* was also higher for constructive feedback ($M = .49, SD = .46$) than for less constructive feedback ($M = .26, SD = .40$) (TK test: $p < .001$; Cohen's $d = .52$). While *Feedback Use* for constructive seeker feedback ($M = .77, SD = .40$) and less constructive seeker feedback ($M = .76, SD = .39$) did not differ significantly (TK test: $p = .993$; Cohen's $d = .02$), it was higher for constructive peer feedback ($M = .38, SD = .44$) than for less constructive peer feedback ($M = .13, SD = .29$) (TK test: $p < .001$; Cohen's $d = .68$). Solvers' feedback use was thus more sensitive to feedback constructiveness for peer feedback than for seeker feedback. These results supported H1 to H3.

Table D1. Robustness Analyses	
DV: Feedback Use (Number of Ideas that Used Feedback)	
Constant	-0.73*** (0.14)
Seeker Feedback	1.56*** (0.17)
Constructive Feedback	1.00*** (0.16)
Seeker Feedback x Constructive Feedback	-0.92*** (0.21)
Wald χ^2	113.03***
Pseudo R ²	0.14
Observations	363
Robustness standard error in parentheses. *** $p < 0.001$	

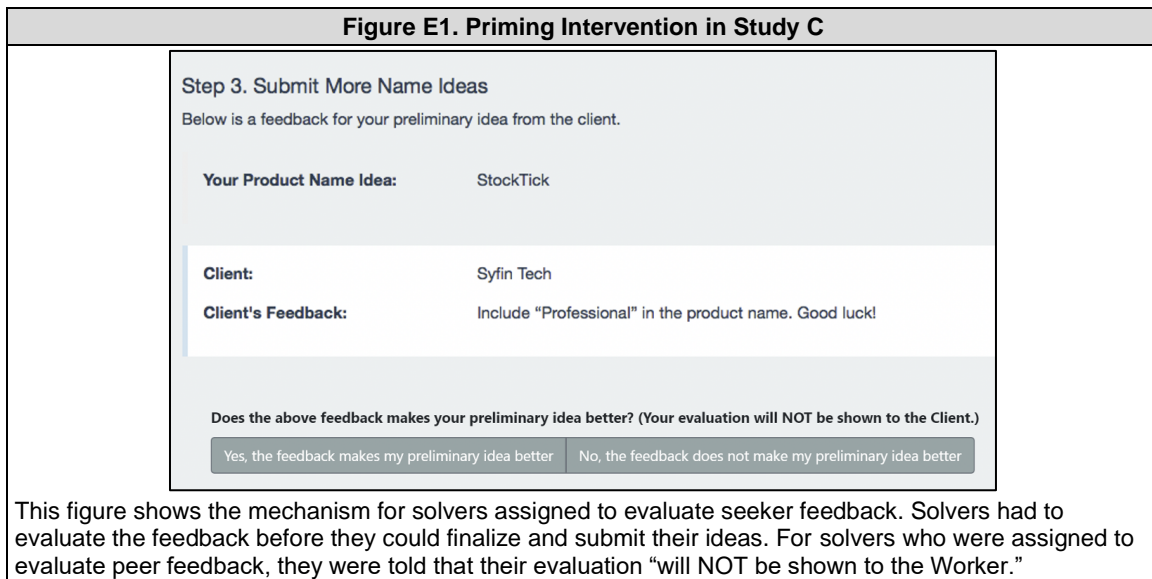
I then used the count of ideas incorporating the feedback as the dependent variable because this measure might also reflect the feedback's influence. Table D1 shows the results using Poisson regression. The estimate for *Seeker Feedback* was positive ($\beta = 1.56, p < .001$), supporting H1. The estimate for *Constructive Feedback* was also positive ($\beta = 1.00, p < .001$), supporting H2. The interaction of *Seeker Feedback* and *Constructive Feedback* was negative ($\beta = -.92, p < .001$). Thus, the positive impact of feedback constructiveness on feedback use was more substantial for peer feedback than for seeker feedback, supporting H3. These results were consistent with the main results.

Additional Analyses. A one-way ANOVA showed that solvers' evaluation of feedback constructiveness varied across manipulations of constructiveness ($F(1, 360) = 101.34, p < .001$) but not of the source ($F(1,$

360) = 2.60, $p = .108$). They perceived constructive seeker feedback ($M = 5.55$, $SD = 1.35$) as more appropriate and helpful than less constructive seeker feedback ($M = 4.36$, $SD = 1.83$) ($t(87) = 3.47$, $p < .001$). In addition, they neither perceived constructive seeker feedback as better than constructive peer feedback ($M = 5.54$, $SD = 1.33$) ($t(184) = .04$, $p = .965$) nor less constructive seeker feedback as substantially better than less constructive peer feedback ($M = 3.81$, $SD = 1.62$) ($t(175) = 1.72$, $p = .090$).

Considering solvers' evaluation and use of feedback together underscores the source effect in their behaviors despite their discerning feedback assessments. Although solvers better perceived constructive feedback than less constructive feedback, they used both to a similar extent when given by the seeker but the former more strongly when provided by a peer. By contrast, despite similar perceptions of constructive feedback from both sources, they used it more strongly when it was from the seeker; the same applied to less constructive feedback. Also, they perceived less constructive seeker feedback as less appropriate and helpful than constructive peer feedback ($t(179) = 3.90$, $p < .001$) but used the former more strongly.

Appendix E. Contest Procedure in Study C



Appendix F. Analyses in Study C

As H1 to H3 involve only feedback source and/or constructiveness, I pooled the different levels for priming intervention to test these hypotheses. For the experiment on MTurk, a two-way ANOVA indicated a significant interaction of feedback source and constructiveness ($F(1, 605) = 3.99$, $p = .046$). Simple main

effects analysis showed that *Feedback Use* differed by source ($F(1, 605) = 142.20, p < .001$), constructiveness ($F(1, 605) = 75.64, p < .001$), and priming intervention ($F(1, 605) = 17.92, p < .001$). Tukey-Kramer tests showed that *Feedback Use* was higher for seeker feedback ($M = .48, SD = .34$) than for peer feedback ($M = .20, SD = .28$) (TK test: $p < .001$; Cohen's $d = .89$), consistent with H1. *Feedback Use* was higher for constructive feedback ($M = .42, SD = .35$) than for less constructive feedback ($M = .21, SD = .29$) (TK test: $p < .001$; Cohen's $d = .68$), supporting H2. The difference in *Feedback Use* between constructive and less constructive feedback from the seeker (Constructive: $M = .55, SD = .34$; Less Constructive: $M = .39, SD = .32$; TK test: $p < .001$; Cohen's $d = .47$) was smaller than that from the peer (Constructive: $M = .32, SD = .32$; Less Constructive: $M = .08, SD = .17$; TK test: $p < .001$; Cohen's $d = .95$), supporting H3.

For the replication experiment on Prolific, a two-way ANOVA showed a significant feedback source and constructiveness interaction ($F(1, 1,386) = 4.05, p = .044$). Simple main effects analysis indicated that *Feedback Use* differed by source ($F(1, 1,386) = 522.29, p < .001$), constructiveness ($F(1, 1,386) = 165.96, p < .001$), and priming intervention ($F(1, 1,386) = 90.33, p < .001$). *Feedback Use* was higher for seeker feedback ($M = .53, SD = .33$) than for peer feedback ($M = .20, SD = .26$) (TK test: $p < .001$; Cohen's $d = 1.12$). *Feedback Use* was also higher for constructive feedback ($M = .44, SD = .32$) than for less constructive feedback ($M = .26, SD = .32$) (TK test: $p < .001$; Cohen's $d = .57$). The difference in *Feedback Use* between constructive and less constructive feedback from the seeker (Constructive: $M = .61, SD = .30$; Less Constructive: $M = .46, SD = .34$; TK test: $p < .001$; Cohen's $d = .48$) was smaller than that from the peer (Constructive: $M = .31, SD = .28$; Less Constructive: $M = .10, SD = .19$; TK test: $p < .001$; Cohen's $d = .88$). These results supported H1 to H3.

A three-way ANOVA showed that the three-way interaction was weakly significant ($F(1, 1,383) = 3.30, p = .069$). Similar to the MTurk experiment, the interactions of (i) feedback source and constructiveness ($F(1, 1,383) = 4.15, p = .042$) and (ii) feedback source and priming intervention ($F(1, 1,383) = 30.65, p < .001$) were significant, but the interaction of feedback constructiveness and priming intervention was not ($F(1, 1,383) = 1.84, p = .175$).

REFERENCES

- Bockstedt J., Druehl C., Mishra A. (2016) Heterogeneous Submission Behavior and its Implications for Success in Innovation Contests with Public Submissions. *Production & Operations Management* **25**(7) 1157-1176.
- Bullinger A.C., Neyer A.-K., Rass M., Moeslein K.M. (2010) Community-Based Innovation Contests: Where Competition Meets Cooperation. *Creativity and Innovation Management* **19**(3) 290-303.
- Füller J., Hutter K., Hautz J., Matzler K. (2014) User Roles and Contributions in Innovation-Contest Communities. *Journal of Management Information Systems* **31**(1) 273-308.