CAN THE CROWD SUBSTITUTE EXPERTS IN EVALUATING CREATIVE JOBS? THE CASE OF BUSINESS MODELS

Research in Progress

Görzen, Thomas, University of Paderborn, Paderborn, Germany,
thomas.goerzen@wiwi.uni-paderborn.de

Kundisch, Dennis, University of Paderborn, Paderborn, Germany,
dennis.kundisch@wiwi.uni-paderborn.de

Abstract

Crowdvoting offers companies the opportunity to get business innovation ideas evaluated by a large number of contributors (the crowd), as an alternative to the currently prevalent practice of involving experts. In this paper we investigate whether the evaluation of a large number of complex ideas by the crowd can be conformant with that of experts. The ideas used to compare these evaluations were generated in an experiment involving students, with eighty different business models featuring as complex ideas. The results of the evaluation by experts were compared with that of an online crowd involved via a crowdvoting-platform. Our results show that an anonymous online-crowd is not as adept as experts when it comes to evaluating complex ideas such as business models. In contrast with previous studies of crowd evaluations for simple aesthetic tasks, our study provides first evidence of the limitations of crowd evaluations, and warns against the substitutions of experts for the evaluation of more complex ideas.

Keywords: Crowdvoting, Business Model, Idea Generation

1 Introduction

'Consider the following problem: You have a serious illness and want to find the best treatment to be healthy again. You employ an open mode and post your problem on the internet with the hope to get an advice. Due to this open call, you get 50 possible ideas, all looking interesting. Immediately, you face one major problem: how to evaluate the large number of ideas and identify the best idea without investing tremendous effort (visiting doctors and so on), especially in case you only have one shot at getting the right treatment' (Pisano and Verganti, 2008, p. 79)?

With the development of Internet-based tools for crowdsourcing, companies increasingly involve external partners, such as their consumers, in their idea generation processes (Forbes, 2005). Crowdsourcing can support managerial decision making, problem solving, and opportunity exploiting (Chiu et al., 2014). Previous studies (e.g. Poetz and Schreier, 2012; Bayus, 2013; Huang et al., 2014) have shown that the crowd is able to generate valuable ideas and therefore can be used to support the idea generation process. However, a successful idea generation process might easily generate hundreds of ideas which a company needs to screen in order to identify those with the largest potential (Toubia and Flores, 2007). This leads to a situation whereby "*firms simply [are] confronted with "too many" ideas and face the problem of not being able to filter and select the most promising ones, or only being able to do so with tremendous effort"* (Poetz and Schreier, 2012).

Several research studies have highlighted idea evaluation as one of the most important stages in an innovation process (Cooper, 1988; Hammedi et al., 2011; Johne and Storey, 1998). Unsurprisingly, then, idea screening previously rested solely in the hands of experts (Mollick and Nanda, 2015). Experts are by definition characterized by their experience and, in contrast to non-experts, typically consider relevant issues such as feasibility and business potential when assessing an idea (Bayus, 2013). Moreover, an analysis of the practices in the academic literature for judging idea quality illustrates the superiority of assigning trained experts to this task (Girotra, Terwiesch and Ulrich, 2010).

However, this conventional approach to engaging experts to screen ideas entails several disadvantages: First, experts are often hard to locate and therefore cannot easily be involved (Galati, 2015). Second, there can be logistical challenges as experts are often much sought after persons with very limited time availability (Galati, 2015). Finally, the cost of expert evaluations is often high, and rises as the number of ideas needing to be evaluated increases (Di Gangi et al., 2010). Companies could currently easily involve the crowd to generate ideas, but they still face the major challenge of a potentially high resource allocation for the evaluation, as long as this process continues to be assigned to experts (Riedl et al., 2011).

One approach to solving this problem is to outsource idea evaluation to the crowd, via so-called *crowdvoting* (Toubia and Flores, 2007). Recent technological developments in IS have made it significantly easier for companies to directly involve a large number of anonymous contributors - the crowd - via crowdvoting-platforms, to evaluate even a large number of ideas (Mollick and Nanda, 2015). The idea of crowdvoting differs from other concepts of crowd involvement, such as crowdfunding where contributor *give* their money to support projects, by instead involving contributors to offer their time in return for *receiving* a relatively small amount of money. By offering a cheap alternative to companies' in-house labour availability, crowdvoting-platforms can been seen as a promising approach to solving the problem of idea evaluation, especially when confronted with a large number of ideas (Howe, 2006).

Previous studies have investigated the general concept of evaluating ideas by a large number of contributors on specialized user platforms, for example, involving users of a company's products or company's employees who would have been preselected for the task. However, these approaches entail different obstacles, as customer involvement requires investment for setting up a dedicated online-forum. Additionally, it could also be expensive to engage a company´s in-house staff as the work time devoted to the evaluation task is then lost to normal tasks. An anonymous online crowdvoting-platform, however, is able to avoid such obstacles, as these platforms offer a large number of external contributors who are motivated to solve different types of tasks in exchange of a typically modest monetary reward.

Additionally, prior research mainly focuses on the evaluation of relatively simple ideas. Existing studies investigated, for example, the accuracy of the evaluation of a large number of ideas for the disposal of medical equipment (Onarheim and Christensen, 2012), household consumer products (Kornish and Ulrich, 2014) or T-Shirt designs (Piller and Walcher, 2006).

However, up to now, it has not been clear whether a real online crowdvoting-platform with an anonymous crowd can easily be used to assess a large number of complex ideas as a substitute for experts' evaluations, at the same or a comparable level. Thus, we aim to shed light on this question as we investigate whether, for the evaluation of a large number of complex ideas, experts can be substituted with a real online crowd. These considerations lead us to the following research question:

Is an evaluation made by a crowd on an anonymous online-crowd conformant with expert´s assessment for the evaluation of a large number of complex ideas?

To address this research question, we use 80 business model ideas generated by students in a class-room experiment. Numerous scholars regard a business model as a key to companies' performance (Demil & Lecocq, 2010; Ireland et al., 2001; Johnson et al., 2008; Sosna et al., 2010) which reinforces our choice for using business models as valid ideas for our experiment. In contrast with previous

studies, the evaluation of business model ideas differs from simple aesthetic tasks insofar as a larger number of components need to be evaluated, jointly, due to the degree of interdependencies between them. Therefore, the complexity of the evaluation of a business model with different components and strong interdependencies between these components represents a complex construct (Galati, 2015).

We first evaluate the ideas assessed by experts, then use a real-life online crowdvoting-platform to evaluate the same business model ideas, and finally compare the results of experts' assessments with the results of the crowd evaluation. With this experiment we make new contributions to research in two ways: (1) We contribute to an enhanced understanding of crowd evaluation compared to experts by investigating the conformance between crowd and experts' assessments by using an anonymous crowd on an online crowdvoting-platform. (2) We show whether crowd evaluation can be used to assess complex ideas such as business models. The organization of this paper is as follows: Section 2 presents the related literature on business model innovation and crowd evaluation. Section 3 describes the methods deployed in our experiment. Section 4 presents our results. Section 5 outlines the contribution of the paper and plans for the next steps for our research.

2 Related Literature

2.1 Business model innovation

In recent years, attention on business models has steadily increased in practice as well as in academic research, emanating from different disciplines such as innovation and technology management, strategy and information systems (Zott et al., 2011). As companies commercialize their innovative ideas and technologies through their business models (Zott et al., 2011), business models themselves need to be innovated to unlock the value potential of new technologies. For this reason, and because business model innovation is a key to companies' performance, business models themselves represent a new topic for innovation (Mitchell and Coles, 2003) and increasing scholarly interest. Indeed, business model innovation is considered to be key to companies' performance and a key driver for corporate transformation (Demil & Lecocq, 2010; Ireland, Hitt, Camp, & Sexton 2001; Johnson, Christensen, & Kagermann, 2008; Sosna, Trevinyo-Rodríguez, & Velamuri, 2010). Business model innovation, however, is not a routine activity but rather a creative process (Teece, 2010) which can be broken down into *analysis*, *generation*, *evaluation* and *communication/implementation* phases (Howard et al., 2008). The creative process of business model innovation can also be seen as a generic design process with different feedback loops between the different phases (Basadur et al., 2000). Within and between these different phases, both convergent and divergent thinking is a necessary skill to make this process work (Basadur et al., 2000)

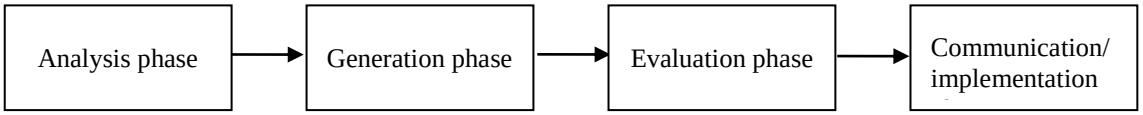

Figure 1. Phases of business model innovation process

The *analysis phase* mainly consists of steps such as problem finding and the recognition of a need for business model innovation and is therefore mainly undertaken by internal resources (Mitchell and Coles, 2004). Internet-based tools for crowdsourcing have previously mainly been used in the *generation phase*, especially by involving a large number of customers or user communities to generate a large number of ideas. Previous studies (e.g., Poetz and Schreier, 2012; Bayus, 2013; Huang et al., 2014) suggest that the crowd is able to generate valuable ideas and therefore can be used to support the idea generation phase. The potentially large numbers of ideas thus generated may, however, lead to a bottleneck in the *evaluation phase*, and companies face time and resource

constraints, while mainly relying on evaluation undertaken by suitable experts (Riedl et al., 2011). Whilst the results of recent studies indicate that crowdvoting can be used for simple aesthetic issues to support the *evaluation phase* (Magnusson et al., 2014, Kornish and Ulrich, 2014), there has not been any study that examined whether crowdvoting can be used to support the *evaluation phase* for the screening of a complex idea such as a business model.

2.2 Crowd evaluation

Related literature was found that explored the concept of idea evaluation by a large number of independent judgements (Table 1). Previous research showed the conformance of idea evaluation between community selection, such as open innovation contest or design competitions and expert ratings in different contexts and focus areas. One field experiment showed conformance between a group of executives, on the one hand, and of employees, on the other, who were both given the task of screening the design for medical equipment. As could be expected, the more experienced employees showed a higher level of conformance than inexperienced employees (Onarheim and Christensen, 2012). In line with these results, the simulation of different algorithms for adaptive idea screening showed that a large number of consumers could be compared with the evaluation of a small number of experts when making binary holistic decisions such as whether something "is a good idea or not" (Toubia and Flores, 2007). These simulation results were also confirmed in a field experiment (Toubia and Flores, 2007). Another study analyzed the appropriateness of users screening ideas for future mobile phone services. The study compared technically naïve and skilled users with the ratings of experts. The results showed no conformance between the two user panels and the expert ratings when comparing their absolute score, however, in a relative comparison (ranking of ideas) both user panels were conformant with the professional experts (Magnusson et al., 2014). A comparison of funding success of theatre projects launched on a crowdfunding platform and ex post evaluation by experts showed significant agreement between funding success and expert ratings (Mollick and Nanda, 2015). Finally, a comparison of assessments from online consumer panellists and experts of 249 user-generated ideas for household consumer products showed that consumers turned out better at identifying a good idea than experts (Kornish and Ulrich, 2014). Related literature showed that different types of contributors, such as companies' own employees or selected users/customers on an online forum, can assess a large number of relatively simple ideas as adequately as experts. Contrary to companies' own employees or selected users/customers on an online forum, members of a crowd on a real online crowdvoting-platform do not have the same incentives, such as reputation or appreciation by the company involved (Onarheim and Christensen, 2012). Moreover, the evaluation of a business model cannot be split into bite-sized, independent chunks because the different parts of a business model, such as revenue streams, cost structure and costumer have to be considered in relation to each other (John and Kundisch, 2015).

Nevertheless, a comparison between the results of experts' ratings with that of an anonymous crowd evaluation, using a real crowdvoting-platform, has not yet been undertaken as part of a research study. Our aim, therefore, is to investigate the ability of crowd evaluation for a large number of complex ideas, namely business model ideas.

Article	Aim of study	Context of ideas	Judges	Result[1]
Mollick and Nanda (2015)	Examine whether funding decision of the crowd and experts evaluation agree	120 theater projects from a crowdfunding-platform	Crowd on a crowdfunding platform and experts for arts	Significant agreement between crowds' funding decision and experts rating for theater projects
Onarheim and Christensen (2012)	Comparison of employees´ rating with executives´ rating	99 ideas for disposable medical equipment	Employees of the company and small team of executives	Employee voting significantly correlates with executives' choices
Tuobia and Flores, (2007)	Analysis of an idea-screening algorithm and comparison with rating of members of an online panel	99 consumer-generated ideas on new cell phone features	195 members of crmmetrix´s online panel	Developed algorithm to show convergence pattern for the results observed in the field experiment
Magnusson, Wästlund and Netz (2014)	Investigation of conformance of users' assessment and experts´ assessment	83 user generated ideas for telecommunication services	30 users (19 technically skilled; 11 technically naïve) and 4 experts	No conformance between users and experts on absolute scores, but conformance on relative comparison (ranking of ideas)
Kornish and Ulrich (2014)	Exploration whether online consumer panelists or experts can better identify a raw ideas´ potential	249 consumer generated ideas for household consumer products	Online consumer panelist and experts from an android user platform	Online consumer panels are a better way to determine a "good" idea than rating by experts

Table 1. Studies on accuracy on large number of judgements in idea screening

3 Study Method

3.1 Idea generation

The ideas used in the study were taken from previous research (John and Kundisch, 2015) as part of an experiment for business model idea generation. The aim was to generate a large number of business model ideas and involved 89 non-expert participants (students). From the perspective of creativity research, it is a well-accepted practice to employ student participants (e.g., Baer et al., 2010; Girotra et al., 2010). The idea generation experiment consisted of two parts: a training part in which each participant was given basic knowledge of business models and of the idea generation task. Students generated the ideas within the well-established framework of the Business Model Canvas (Osterwalder and Pigneur, 2010). In the second part of the experiment, participants generated business model ideas for the product *perfume.* This product was chosen for two reasons: First, perfume is an everyday product, which reduces the risk of prior knowledge being heterogeneously distributed among participants (John and Kundisch, 2015). Second, perfume is used as a *'standalone product'*, which steers creative thinking away from product innovation and towards business model innovation (John and Kundisch, 2015). Additionally, regarding the evaluation task, it is necessary to have at least a minimum of domain knowledge to be able to judge an idea; this is a relevant criterion especially for the crowds' evaluation task (Galati, 2015). At the end of the idea generation task, each participant was

[1] Further analyses that were not directly related to our investigation are not listed.

asked to select his or her favourite idea, and to answer task-related and background questions. In sum, participants generated 89 different business model ideas. After a first idea screening, two ideas were excluded as they had nothing to do with perfume and therefore were deemed not useful.

3.2 Expert rating

To determine the quality of the ideas, we recruited two experts - one from our university´s entrepreneurship center and the other, a senior in-house consultant, from a large corporation. Both experts were well-qualified to evaluate business model ideas in our experiment, their qualifications matching the qualifications reported in prior research (e.g., Baer et al., 2010; Scopelliti et al., 2013): the expert from the university regularly advises academic sponsors in terms of developing viable business models, and the corporate expert works as a consultant on corporate innovation projects. Before starting on the task, the experts were given the evaluation criteria and the rating scales to make sure that each had a comparable understanding of the task at hand (John and Kundisch, 2015). Both experts were asked to rate the creativity of each idea in terms of whether they deemed it to be novel and useful (similarly to Amabile, 1996; Baer et al., 2010), on a scale ranging from (1), "not at all", to (7), "very creative" (John and Kundisch, 2015). The experts were instructed to disregard any language-related weaknesses they might notice in the business models.

3.3 Crowd rating

In this paper we propose a practical approach to investigating a crowd's ability to evaluate ideas. Because the number of ideas generated by our students was large (87 ideas), it would be unreasonable to ask each member of the crowd to rate all ideas, not least because the evaluation was performed online in a non-controlled environment (Dahan and Hauser, 2001). Hence we first divided the ideas into eight blocks of ten ideas each. Because we needed eight equally sized blocks to compare the evaluations, we excluded seven randomly chosen ideas. We then randomly assigned each idea to one block and afterwards controlled whether the randomization worked by analysing descriptive statistics as well as the distribution of the experts' rankings in each block. Descriptive analysis indicated that there were no differences between the blocks in terms of average idea quality, nor in the distribution of idea quality. To evaluate these eight blocks of ideas, we used a commercial crowdworking-platform, *Crowdflower,* which has millions of potential contributors worldwide. We designed eight jobs, one job for each block of ten distinct business model ideas (following Toubia and Flores, 2007). We formulated our instructions following Magnusson et al. (2014) and tested these instructions, the payment for the contributors and the design of the evaluation task, in nine pretests. Because all business ideas were formulated in German, we limited access to the jobs to German speaking contributors, mainly from Germany, Switzerland and Austria. Finally, we launched the eight evaluation jobs on the platform on two different days (four blocks on each day) to prevent the same user from rating all eight blocks in succession. The crowd evaluated each business model idea using the same seven-point scale as the experts. Before the contributors started on their evaluation, they were asked to answer a few demographical questions as well as questions on their prior knowledge of business models, retail and perfume. Every contributor earned 0.50€ for the evaluation of one block, which took less than an hour, on average. Every business model idea was evaluated by 20 different members of the crowd, so we received 1,600 different evaluations in total. The data analysis was carried out as follows: first we pre-screened the judgements of the crowd to identify any invalid judgements. Next we compared the descriptive statistics of both experts and crowd evaluation and analyzed the distribution of both ratings. Then we conducted non-parametric tests to analyze the extent to which the crowd's judgements were aligned to those of the experts. Finally we compared the top five ideas of both ratings.

4 Results

Ideally, with the ultimate view of keeping down the cost of crowd evaluations, the results of crowdvoting should not require any additional pre-screening. However, our pretests showed that there will always be some contributors in the crowd who do not adhere to task instructions and, instead, click through the task without giving the ideas due consideration. Therefore, we assessed the time needed to finish the task as well as the standard deviation of each contributor. We identified the crowd workers who clicked through the task too quickly for a proper evaluation on the basis of a standard deviation below 0.5 and an average process time of less than two minutes[2]. As a result we had to exclude ten judgements distributed over all eight business model blocks. To compare the ratings of both the experts and the crowd, we used their respective averages ratings and median for each business model. In a first step, we compared the overall descriptive statistics (Table 2) of each evaluation, which showed no significant differences, except a higher standard deviation for the experts due to the fact that their average was formed out of a lower number of ratings.

	Experts	Crowd
Mean	4.331	4.584
Median	4.500	4.583
Std. Dev.	1.214	0.412

Table 2. Descriptive Statistics of experts and crowd rating

Next, we tested both distributions - for 80 average ratings from experts and 80 average ratings from the crowd – to see whether they were normally distributed. The Kolmogorov-Smirnov test showed that the experts' ratings deviate significantly from normal ($D(80) = 0.205$, $p < 0.01$). Hence, we used non-parametric rank correlation tests to analyze a potential coherence between evaluation by experts and by the crowd. Because of the relatively large number of split ranks, we also considered Kendall´s tau in addition to Spearman´s rho. We calculated the values based on the average rating of each of the 80 ideas and, to reduce the influence of possible outliers, we considered the correlation based on the median rating. All values show only very little positive correlation between experts and crowd rating and non-significant p-values (Table 3).

	Average rating of ideas	Sig._Average (two-tailed)	Median rating of ideas	Sig._Median (two-tailed)
Spearman´s-Rho	.065	.566	.023	.838
Kendall´s-Tau	.053	.515	.021	.827

Table 3. Correlation coefficients for average and median rating from the crowd and the experts

Finally, we compared the relative ratings of experts and crowd evaluation because companies in real-life situations would only be interested in selecting the best from a large pool of ideas and would eliminate the least promising ones. Therefore, we compared the top five ideas identified by the experts with those identified by the crowd, based on average and median ratings. Experts´ ratings and crowd

[2] Five of them even below one minute. The overall average time to solve the task was 4:32 minutes

evaluation only share one idea between them, within a comparison based on average ratings, and none based on medians.

5 Conclusion

In the context of the growing importance of crowd based decision making and the currently limited body of literature in the context of crowdvoting, we consider this field and investigate whether a crowd on a real online crowdvoting-platform can evaluate business model ideas as well as - or conformant with - experts. To the best of our knowledge, this research in progress is the first to investigate this question. Unlike previous studies that mainly investigated non-expert idea evaluations based on simple aesthetic judgements, our task differs in terms of the complexity of the ideas, in our case, a large number of business model ideas. Another key difference between previous studies, which mainly used companies' own employees (Onarheim and Christensen, 2012) or selected online users from specified online forums with some investment in the idea to be evaluated (Magnusson et al., 2014), and our experiment is that our crowd is anonymous, and involved via a crowdvoting-platform. Considering different measures, we find first evidence that an anonymous crowd on a crowdvoting-platform cannot be used to evaluate business model ideas and therefore cannot be seen as a potential substitute for experts for the evaluation of complex ideas. The results show no correlation between the ratings by experts and by the crowd. Even a relative comparison of the top five ideas of both evaluations, which can be seen as a real-life procedure for companies to decide which ideas need to be considered in more detail, showed that the two groups only agreed on one of the five ideas as their top idea. Further research could include different aspects of crowds and task characteristics to gain more insights into the potential use of crowdvoting. This would make both a theoretical as well as a practical contribution. From a theoretical perspective, one could consider the reasons why a crowd is more adept at evaluating relatively simple aesthetic rather than complex tasks. Investigating different complexity levels based on the complexity dimensions of task design could add further insights on suitable task-levels for crowdvoting and explain potential limitations inherent in a crowd's cognitive ability. From a practical perspective, task designs such as context framing, or the influence of splitting complex evaluation tasks into a greater number of simple tasks, could also be considered. To gain additional insights into the ability of a crowd to evaluate ideas, we will perform additional analyses. The aim is to investigate whether previous knowledge of business models or of the product itself could affect the ability to evaluate the business model ideas. Furthermore, we will analyze the influence of experience gained from evaluating one block of ideas as well as the aggregated experience across ideas, due to the fact that some members of the crowd assigned themselves to several tasks. Beyond this study we also plan to investigate additional evaluation mechanisms for idea evaluation. As we only used a holistic rating based on a seven-point scale, a pairwise comparison of ideas, for example, could help the crowd to identify the best and the worst out of a large pool of ideas.

References

Amabile, T. M. (1996). "Creativity in context: Update to The social psychology of creativity." Boulder, CO: Westview Press.

Baer, M., Leenders, Roger, T. A., Oldham, G. R. and A. K. Vadera (2010). "Win or lose the battle for creativity: The power and perils of intergroup competition." *Academy of Management Journal* 53 (4), 827–845.

Basadur, M., Pringle, P., Speranzini, G. and M. Bacot. (2000). "Collaborative problem solving through creativity in problem definition: expanding the pie." *Creativity and Innovation Management* 9 (1), 54-76.

Bayus, B. L. (2013). "Crowdsourcing New Product Ideas over Time: An Analysis of the Dell IdeaStorm Community." *Management Science* 59 (1), 226-244.

Chiu, C. M., Liang, T. P. and E. Turban (2014). "What can crowdsourcing do for decision support?" *Decision Support Systems* 65, 40-49.

Cooper, R. G. (1988). "Predevelopment activities determine new product success." *Industrial Marketing Management*, 17 (3), 237-247.

Dahan, E. and J.R. Hauser (2001). "The virtual customer." *Journal of Product Innovation Management* 19 (5), 332-354.

Demil, B., and X. Lecocq, (2010). "Business model evolution: In search of dynamic consistency." *Long Range Planning* 43, 227-246.

Di Gangi, P. M., Wasko, M. and R. E. Hooker (2010). "Getting customers' ideas to work for you: Learning from Dell how to succeed with online user innovation communities." *MIS Quarterly Executive* 9 (4), 213–228.

Forbes (2005). "Why companies need your ideas – How they´ve tapping costumers to develop new products." February (14), 78-86.

Galati, F. (2015). "Complexity of Judgement: What makes possible the convergence of expert and nonexpert ratings in assessing creativity." *Creativity Research Journal*, 27 (1), 24-30.

Girotra, K., Terwiesch, C. and K. T. Ulrich (2010). "Idea Generation and the Quality of the Best Idea." *Management Science* 56 (4), 591-605.

Hammedi, W., Van Riel, A. C., and Z. Sasovova (2011). "Antecedents and Consequences of Reflexivity in New Product Idea Screening." *Journal of Product Innovation Management*, 28 (5), 662-679.

Howard, T.J., Culley, S.J., and E. Dekoninck (2008). "Describing the creative design process by the integration of engineering design and cognitive psychology literature." *Design studies* 29 (2), 160-180.

Howe, J. (2006). "The rise of crowdsourcing." *Wired* 14 (6).

Huang, Y., Singh, P. V. and K. Srinivasan (2014). "Crowdsourcing New Product Ideas Under Consumer Learning." *Management Science* 60 (9), 2138-2159.

Ireland, R. D., Hitt, M. A., Camp, M., and D. L. Sexton (2001). "Integrating entrepreneurship and strategic management actions to create firm wealth." *Academy of Management Executive* 15, 49-63.

John, T. and D. Kundisch (2015) Why Use the Canvas for Ida Generation? A Design Theory and First Evidence. Working Paper, Paderborn University.

Johne, A., and C. Storey (1998). "New Service Development: a review of the literature and annotated bibliography." *European Journal of Marketing*, 32 (3/4).

Johnson, M. W., Christensen, C. C., and H. Kagermann (2008). "Reinventing your business model." *Harvard Business Review*, 86 (12), 50-59.

Kornish, L. J. and K. T. Ulrich (2014). "The importance of the raw idea in innovation: Testing the sow's ear hypothesis." *Journal of Marketing Research*, 51 (1), 14-26.

Magnusson, P. R., Wästlund, E. and J. Netz (2014). "Exploring User´Appropriateness as a Proxy for Experts When Screening New Product/Service Ideas." *Journal of Product Innovation Management.*

Mitchell, D. W. and C. Coles (2003). "The ultimate competitive advantage of continuing business model innovation." *Journal of Business Strategy*, 24 (5), 15-21.

Mitchell, D. W. and C. Coles (2004). "Establishing a continuing business model innovation process." *Journal of Business Strategy*, 25 (3), 39-49.

Mollick, E. and R. Nanda (2015). "Wisdom or Madness? Comparing Crowds with Expert Evaluation in Funding the Arts." *Management Science* (Article in advance), 1-21.

Onarheim, B. and Christensen, B.T. (2012). "Distributed idea screening in stage-gate development processes." *Journal of Engineering Design* 23 (9), 660-673.

Osterwalder, A. and Y. Pigneur (2010). "Business model generation: A handbook for visionaries, game changers, and challengers." Hoboken, NJ: Wiley.

Piller, F.T. and D. Walcher (2006). "Toolkits for idea competitions: a novel method to integrate users in new product development." *R&D Management* 36 (3), 307-318.

Pisano, G. P., and Verganti, R. (2008). "Which kind of collaboration is right for you?" *Harvard Business Review* 86 (12), 78-86.

Poetz, M. K., and M. Schreier (2012). "The Value of Crowdsourcing: Can Users Really Compete with Professionals in Generating New Product Ideas?" *Journal of Product Innovation Management* 2 (29), 245-256.

Riedl, C., Blohm, I., Leimeister, J.M. and H. Krcmar (2011). "Idea evaluation mechanisms for collective intelligence in open innovation communities: do traders outperform raters?" *Proceedings of 32nd International Conference on Information Systems*, 2011, 1-24.

Scopelliti, I., Cillo, P., Busacca, B., and D. Mazursky (2013). "How Do Financial Constraints Affect Creativity?" *Journal of Product Innovation Management* 31 (5), 880-893.

Sosna, M., Trevinyo-Rodríguez, R. N., and S. R. Velamuri (2010). "Business models innovation through trial-and-error learning: The Naturhouse case." *Long Range Planning* 43, 383-407.

Teece, D.J. (2010). "Business Models, Business Strategy and Innovation." *Long Range Planning* 43, 172-194.

Toubia, O., and L. Flores (2007). "Adaptive idea screening using consumers." *Marketing Science* 26 (3), 342-360.

Velamuri, V. K., Schneckenberg, D., Haller, J. B. A. and K. M. Moeslein (2015). "Open evaluation of new product concepts at the front end of innovation: objectives and contingency." *R&D Management* (forthcoming).

Zott, C., Amit, R., and Massa, L. 2011. "The business model: Recent developments and future research." *Journal of Management* 37 (4), 1019-1042.