Rethinking the Design of Innovation Crowdsourcing Competitions: Strategies for Shaping Participation Structure and Maximizing Value Creation

Emmanuel Peterle *[1], **Juan Pablo Reyes Ochoa***[2]

[1]University of Franche Comté
[2]University of Burgundy Franche Comté
30 Av. de l'Observatoire
25030 Besançon, France
juan_pablo.reyes_ochoa@univ-fcomte.fr

Abstract

Crowdsourcing through virtual innovation contests have emerged as a prominent option for addressing creative and complex R&D problems. This approach is flexible, encourages voluntary participation, and combines competition and collaboration. Organizations that turn to innovation crowdsourcing (ICS) are looking for transformative solutions, not incremental improvements. However, little is known about how to shape teams and optimize their capacity to create value in CS competitions. To address this knowledge gap, our study delves into analyzing how contest design influences participation structure and project value creation. Our approach extends the conclusions of previous research that emphasized the importance of motivation and active engagement as fundamental factors for effective problem-solving. Our research goes further by examining the role played by other dimensions of participation structure, such as team size, as well as solvers' social and intellectual capital.

Introduction

In the process of problem-solving, organizations often face constraints of time, information, resources, and cognitive biases that make it difficult to find optimal solutions (Afuah and Tucci 2012). This restriction leads decision-makers to seek alternatives that achieve a minimum performance level. Typically, companies rely on their familiar environment, based on routines, knowledge, and past experiences to make secure decisions. However, in innovative situations, they need to explore new fields of knowledge. This is where crowdsourcing (CS) comes into play, transforming distant searches into local ones by delegating tasks to dispersed groups over the Internet, who seek solutions and share them for rewards.

Virtual innovation contests have emerged as a prominent option for addressing creative and complex R&D problems. This approach is flexible, encourages voluntary participation, and combines competition and collaboration. Its main advantage is maximizing the value of the highest-performing result (Terwiesch and Xu 2008). In innovation, it is not quantity but identifying the best opportunity that matters (Girotra, Terwiesch, and Ulrich 2010). Organizations that turn to innovation crowdsourcing (ICS) are looking for transformative solutions, not incremental improvements. Despite this, the literature has focused on predicting factors that determine individuals' performance or their ability to reach higher rank positions and earn rewards, overlooking the exploration of knowledge generation and distribution dynamics in teams and their influence on solution quality. Little is known about shaping teams and optimizing their capacity to create value in CS contests.

To address this knowledge gap, our study delves into analyzing how contest design parameters influence participation structure and project value creation. Our approach extends the conclusions of previous research that emphasized the importance of motivation and active engagement as fundamental factors for effective problem-solving (Garcia Martinez 2015; Archak 2010). However, our research goes further by examining the role played by other elements of participation structure, such as team size, as well as solvers' social and intellectual capital. As our primary focus lies within ICS, our approach transcends mere assessments of individual or team-based value creation. Rather, we delve into the examination of relative performance. To illustrate, we employ the metric of Top Normalized Scores to ascertain whether a given competition has engendered solutions that exhibit notable deviations from the overarching mean of contributions. Furthermore, we employ Performance Progression analysis to elucidate the dynamics of knowledge accumulation throughout the progression of the competitions.

In particular, our research aims to address the following questions:

RQ1: How do contest design affect participation structure?

RQ2: What role does cooperation play in the effective design of ICS projects?

RQ3: How social and intellectual capital distribution within teams influence contest relative performance?

To answer these questions, we work with a dataset of 3,169,640 teams participating in 204 contests organized between 2016 and 2023 on the CS platform Kaggle, specialized in data science competitions. We will use structural

*These authors contributed equally.

equation modeling (SEM) as data analysis method to evaluate and model the magnitude and direction of relationships between our variables following the recommendations of Gana and Broc (2019) (see variable summary in Appendix).

We propose that distinct contest configurations not only lead to significant changes in team composition but also shape their inherent capabilities. Individuals contribute two types of capital to teams: intellectual capital, which refers to task-related skills including experience, education, and knowledge, and social capital, which represents the advantage a person gains from their position in the social structure (Burt 2007). From a social network theory perspective, we propose that increasing monetary rewards contributes to build teams with higher intellectual capital but negatively influences their size and social capital. The number of rewards positively influences team size and social capital, but negatively affects intellectual capital. Non-monetary incentives related to status and hierarchy only have a positive effect on team size and social capital. We posit that contests with a higher team participation ratio produce better results. The combination of cooperative teams with higher social and intellectual capital has significant positive effects on their capacity to create value.

To our knowledge, this is one of the first works that focus on analyzing the influence of different contest design alternatives on team composition and mechanisms to maximize value creation. Our comparative analysis of competitions with varied characteristics ranging from code building to detecting ink within logs dating back over 2,000 years (Alex Lourenco et al. 2023), has the potential to provide valuable insights into the cooperative and competitive strategies participants adopt based on contest characteristics and their influence on the quality of their contributions. These insights can be used by decision-makers from the requesting organizations or CS platforms to design more efficient innovation contests.

This is a work in progress, so in this report, we present our research strategy. The document is structured as follows: Section 2 introduces the context and hypotheses, Section 3 provides an overview of our empirical context, data preprocessing, and methodology, and Section 4 presents our conclusion.

Context and Hypotheses

Recent research has examined participation and performance in ICS contests. Studies such as DiPalantino and Vojnović (2009) find a relationship between reward size and participation, particularly notable among expert users. Liu et al. (2014) find positive links between reward variations, contributions, and quality. Contest duration, according to Liu et al. (2014), forms an inverted ”U” curve, suggesting that motivation and participation have an optimal peak. The compensation-effort relationship, explored by Horton and Chilton (2010), varies with task duration and complexity. While there is little exploration of motivations for cooperation, Huang, Zhou, and Chen (2022) notes the positive influence of task complexity on team formation.

Recent research has improved our understanding of factors affecting participation and performance in ICS contests, demonstrating that the relationship between rewards and participation is complex and context-dependent. While more exploration is needed regarding motivations for cooperation and the impact of non-monetary rewards on team structure, our RQ1 focuses on the influence of contest parameters on participation structure, so we propose the following:

H1. Monetary and non-monetary incentives have a positive effect on the volume of active participation.

H2. Reward size and status and hierarchy rewards have a positive influence on the proportion of teams in a competition, whereas this relationship is negative concerning the number of rewards.

H3. Contests with higher monetary rewards attract competitors with higher intellectual capital, and contests with more non monetary rewards attract competitors with higher social capital.

Recent scholarly investigations have delved into the intricate interplay between cooperation and competition, with a primary focus on gauging both individual and team performance within the context of ICS contests. The findings of these studies have underscored the pivotal role of teamwork in enhancing the likelihood of victory (Dissanayake, Zhang, and Gu 2015). Furthermore, the degree of collaboration, encompassing proactive engagement within forums and the exchange of solutions, has demonstrated a significant correlation with elevated individual performance (Javadi Khasraghi and Hirschheim 2022). In parallel, insights gleaned from an examination of Kaggle data have underscored the profound impact of social and intellectual capital on team performance (Zhang et al. 2020). This investigation notably emphasizes the influence of knowledge distribution and the cohesiveness of the community in shaping the trajectory of success.

Thus, the literature provides solid clues about the influence of cooperative structure and distribution of knowledge and networks on participants' intrinsic motivation. However, examining how these characteristics shape contest outcomes is required; following RQ2 and RQ3, we present the following hypotheses:

H4. An increase in the proportion of teams with at least two members has a significant positive effect on contest value creation.

H5. The proportion of teams with at least two members is dependent on competitors' social and intellectual capital.

H6. The probability of attaining superior relative performance is heightened in contests that aggregate teams characterized by elevated levels of both social and intellectual capital.

Data and Methodology

This section describes the empirical context and data preprocessing. We also describe the data analysis strategy and provide a description of our metrics.

In our research, we use the Meta Kaggle database (updated to July 2023) (Megan Risdal and Timo Bozsolik 2023), which includes 32 data tables on solutions from over 13 million users, distributed among more than 6 million teams participating in 5,586 competitions organized between 2010 and 2023. We chose this empirical context considering the variety and size of data it offers for contests that occur in coopetitive environments, addressing the theoretical dimensions of interest in this work without introducing systematic biases (Seawright and Gerring 2008). Kaggle is a crowdsourcing platform specialized in the field of data science. With a global presence across 194 countries, this dynamic community consists of over 536,000 active members who together generate an average of 150,000 contributions per month. This highlights the engaged and active nature of this extensive network.

Our research delves into competitions that intertwine both monetary and non-monetary incentives, incorporating cooperative and competitive dynamics. Our sample acquisition involved the following criteria: BanTeamMergers being 'False', more than one in TotalTeams, and RewardType denoted as 'USD' or 'EUR'.

The Meta Kaggle platform has been gathering user performance data (evaluated through metrics encompassing points, medals, and expertise levels) since 2016. Consequently, our study comprises contests from the year 2016 up to 2023, culminating in a total of 204 contests. Subsequently, we filtered the Teams, Team Memberships, Submissions, and Users tables, focusing solely on users and teams participating in the sample competitions.

Data Analysis Strategy

To analyze the intricate relationships among our variables, we will use R data analysis software in conjunction with the Lavaan package to conduct a Partial Least Squares Structural Equation Modeling (PLS-SEM) analysis with a 95% confidence level(Gana and Broc 2019). Our analysis will encompass three distinct structural equations. The first equation will consider the control variables, namely, Time (number of years after 2016) and Competition Intensity. In the second equation, we will solely focus on the independent variables. In the third equation, we will delve into the moderation effects we anticipate for Distribution Effort in relation to Team Ratio, Distribution of Intellectual and Social Capital, and dependent variables (i.e. Top Normalized Scores and Performance Progression).

To ensure the robustness of our findings, we will apply the minimum coefficient of determination (R^2) criterion to estimate appropriate levels of statistical power, as well as significant coefficients at each step ($p < 0.05$) (Kock and Hadaya 2018).

Furthermore, it is important to highlight that our model posits that the participation structure plays a mediating role in the connection between competition design and contest relative performance. To accurately assess the direct influence of our independent variables on the dependent ones, in comparison to indirect effects through mediating variables, we will employ the Sobel Test (Sobel 1982).

For a detailed description of the variables we will use to test our hypotheses, see Appendix .

Conclusions

Based on the results of our study, we aim to provide a series of practical applications targeted towards professionals and leaders within organizations. These applications are intended to strengthen strategic decision-making in the conception of ICS competitions. Understanding how team composition and skills impact the quality of final solutions provides a solid foundation for making informed decisions when forming and fostering collaboration among teams, which in turn translates into achieving higher relative performance.

Our conclusions regarding how incentives influence team formation are invaluable for optimizing engagement strategies. For instance, if the goal is to enhance the presence of knowledge within teams, it is possible to adjust incentive designs accordingly. In the event that we identify that cooperation among teams with high intellectual and social capital leads to superior results, this could inspire the development of platforms facilitating interaction between such teams.

The conclusions related to the relevance of intellectual and social capital within teams can influence how platforms shape and empower their community members, promoting the acquisition of technical skills and the building of effective social networks. Furthermore, these conclusions can lay the groundwork for future research and guide approaches in team management and process optimization within ICS projects.

References

Afuah, A.; and Tucci, C. L. 2012. Crowdsourcing As a Solution to Distant Search. *Academy of Management Review*, 37(3): 355–375.

Alex Lourenco; Brent Seales; Christy Chapman; Daniel Havir; Ian Janicki; JP Posma; Nat Friedman; Ryan Holbrook; Seth P.; Stephen Parsons; and Will Cukierski. 2023. Vesuvius Challenge - Ink Detection. https://www.kaggle.com/competitions/vesuvius-challenge-ink-detection.

Archak, N. 2010. Money, Glory and Cheap Talk: Analyzing Strategic Behavior of Contestants in Simultaneous Crowdsourcing Contests on TopCoder.Com. In *Proceedings of the 19th International Conference on World Wide Web, WWW '10*, 21–30. Raleigh, NC.

Burt, R. S. 2007. *Brokerage and Closure: An Introduction to Social Capital*. Oxford: Oxford Univ. Press, 1. publ. in paperback edition. ISBN 978-0-19-924915-2 978-0-19-924914-5.

DiPalantino, D.; and Vojnović, M. 2009. Crowdsourcing and All-Pay Auctions. In *Proceedings of the ACM Conference on Electronic Commerce*, 119–128. Stanford, CA.

Dissanayake, I.; Zhang, J.; and Gu, B. 2015. Task Division for Team Success in Crowdsourcing Contests: Resource Allocation and Alignment Effects. *Journal of Management Information Systems*, 32(2): 8–39.

Gana, K.; and Broc, G. 2019. *Structural Equation Modeling with Lavaan*. Mathematics and Statistics. London : Hoboken, NJ: ISTE Ltd ; John Wilery & Sons, Inc. ISBN 978-1-78630-369-1.

Garcia Martinez, M. 2015. Solver Engagement in Knowledge Sharing in Crowdsourcing Communities: Exploring the Link to Creativity. *Research Policy*, 44(8): 1419–1430.

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

Horton, J.; and Chilton, L. 2010. The Labor Economics of Paid Crowdsourcing. In *Proceedings of the ACM Conference on Electronic Commerce*, 209–218. Cambridge, MA.

Huang, K.; Zhou, J.; and Chen, S. 2022. Being a Solo Endeavor or Team Worker in Crowdsourcing Contests? It Is a Long-term Decision You Need to Make. *Proceedings of the ACM on Human-Computer Interaction*, 6(CSCW2): 1–32.

Javadi Khasraghi, H.; and Hirschheim, R. 2022. Collaboration in Crowdsourcing Contests: How Different Levels of Collaboration Affect Team Performance. *Behaviour & Information Technology*, 41(7): 1566–1582.

Kock, N.; and Hadaya, P. 2018. Minimum Sample Size Estimation in PLS-SEM: The Inverse Square Root and Gamma-Exponential Methods: Sample Size in PLS-based SEM. *Information Systems Journal*, 28(1): 227–261.

Liu, T. X.; Yang, J.; Adamic, L. A.; and Chen, Y. 2014. Crowdsourcing with All-Pay Auctions: A Field Experiment on Taskcn. *Management Science*, 60(8): 2020–2037.

Megan Risdal; and Timo Bozsolik. 2023. Meta Kaggle.

Seawright, J.; and Gerring, J. 2008. Case Selection Techniques in Case Study Research: A Menu of Qualitative and Quantitative Options. *Political Research Quarterly*, 61(2): 294–308.

Sobel, M. E. 1982. Asymptotic Confidence Intervals for Indirect Effects in Structural Equation Models. *Sociological Methodology*, 13: 290.

Terwiesch, C.; and Xu, Y. 2008. Innovation Contests, Open Innovation, and Multiagent Problem Solving. *Management Science*, 54(9): 1529–1543.

Zhang, X.; Duan, K.; Zhao, H.; Zhao, Y.; Wang, X.; and de Pablos, P. O. 2020. Can Cooperation Drive the Success of Suppliers in B2B Crowdsourcing Innovation Projects? A Large Scale Data Perspective. *Industrial Marketing Management*, 90: 570–580.

Appendix: Metrics Used in the Study

Table 1: Metrics Used in the Study

Variable	Description
Incentives	
Reward Size	Amount in US dollars to be distributed among the winners.
Number of Prizes	Number of winning teams among which the reward amount will be divided.
Performance Medals	Non-monetary incentive (status and hierarchy) that awards permanent gold, silver, and bronze medals and influences platform expertise categories (e.g. Expert, Master, and GrandMaster). The number of medals to be distributed depends on the number of active teams in the competition. This variable is in boolean format.
Performance Tiers	Unlike performance medals that are permanent representations, Kaggle's progression system makes user points decay over time to keep global rankings updated and competitive. Kaggle uses the following formula to calculate points, where t is the number of days since the point was awarded: $e^{-t/500}$
Task Complexity	
Competition Length	Number of days between the start date and submission deadline.
Kernel Submission	Boolean variable indicating if the contest only allows kernel-style solutions. Kernels are scripts or notebooks with data that allow running programming language libraries like R or Python. Requesting kernel-style submissions reduces the cost related to team coordination, indicating low task complexity.
Data Density	Size of the data train file made available to participants.
Other Contest Parameters	
Maximum Team Size	Maximum number of members within a team.
Leaderboard Percentage	Percentage of data used in evaluating the scores displayed on the contest page leaderboard. A lower percentage increases competitors' uncertainty about final results.
Submission Limit	Daily submission limit per team.
Control Variables	
Time	Variable to control the impact of time passing on the number of registered users, past experiences, and accumulated knowledge of competitors.
Competition Intensity	Variable used to control the impact of competition intensity on participants' self-selection process. We use the Herfindahl Index (HHI) to measure this variable. $$CompetitionIntensityj = \frac{\sum_{i=1}^{n} TeamIC_{ij}^2}{(\sum_{i=1}^{n} TeamIC_{ij})^2}$$
Participation Structure	
Team Ratio	Proportion of teams with at least two members participating in the contest, taking all teams (including teams of one member).
Social Capital Distribution	Measure of the team's internal social capital based on past connections between team members before the contest. Average of all participant i's past connections with other team members, where aij represents the connection between nodes i and j. If there is a connection between i and j, aij is 1; otherwise, it is 0. $$DegreeCentralityi = \sum{}_j a_{ij}$$
Intellectual Capital Distribution	Average sum of points earned by team members in past competitions.
Effort Distribution	Normalized number of submissions made by teams in each contest. $$e' = \frac{e}{Teams\mu_{ej}}$$
Contest Performance	
Top Normalized Scores	Calculation of the average z-score of the top 10% teams' scores in each contest.
Performance Progression	The rate of variation between the average normalized scores of the initial 10 days and the final 10 days of the competition.