OPEN INNOVATION CONTESTS IN ONLINE MARKETS: IDEA GENERATION AND IDEA EVALUATION WITH COLLECTIVE INTELLIGENCE

A Dissertation
Submitted to
the Temple University Graduate Board

---

in Partial Fulfillment
of the Requirements for the Degree of
DOCTOR OF PHILOSOPHY

By
Yang Yang
January, 2012

Examining Committee Members:

Pei-yu Chen, Advisory Chair, Management Information Systems
Paul A. Pavlou, Management Information Systems
Jose M. Plehn-Dujowich, Accounting Department
Yuexiao Dong, External Member, Statistics Department

ABSTRACT

Open Innovation Contests in Online Markets: Idea Generation and Idea Evaluation with Collective Intelligence
Yang Yang
Doctor of Philosophy
Temple University, 2012
Doctoral Advisory Committee Chair: Dr. Pei-yu Chen

To overcome constrained resources, firms can actively seek innovative opportunities from the external world. This innovation approach, called open innovation (Chesbrough 2003; Hippel 2005; Terwiesch and Ulrich 2009; Terwiesch and Xu 2008), is receiving more and more attention. Facilitated by the global Internet and emerging forms of information technology, it has become very easy for companies to generate large numbers of innovative solutions through the use of online *open innovation contests* or *crowdsourcing contests* (Archak and Sundararajan 2009; Terwiesch and Ulrich 2009; Terwiesch and Xu 2008; Yang et al. 2009).

For an innovation project to succeed, it is necessary to generate not only a large number of good ideas or solutions, but also to identify those that are "exceptional" (Terwiesch and Ulrich 2009). This dissertation contains three studies that aim to improve our understanding of how best to use contests as a tool to aggregate external resources (collective intelligence) in the generation and evaluation of solutions.

The first study views an innovation contest from the innovation seeker's perspective and provides insights on how to improve contest performance. The second study views an innovation contest from the innovation solver's perspective examining the characteristics and strategies of winners and solvers. Finally, in the third study, a new approach to the solution evaluation process is introduced, which is referred to as open

evaluation. In this approach, a prediction market is used as an aggregation mechanism to coordinate the crowd in the evaluation of proposed solutions.

These three studies make a number of contributions to the literature, addressing core issues in the area of online innovation contests. The analyses, which leverage large-scale empirical data, produce a number of profound results, which can help people to understand how best to use and design innovation contests in an online environment, for idea generation. Further, these studies present a variety of managerial implications associated with the aggregation of individual effort (collective intelligence) to evaluate the ideas that are generated by an innovation contest. We hope that our studies can help open innovation pioneers, such as Google, to systematically generate and identify exceptionally good ideas at much lower costs. By utilizing our findings, we expect that more firms will be able to adopt an open innovation strategy, both systematically and easily.

ACKNOWLEDGEMENTS

I would like to start by thanking those directly related to my successful completion of this dissertation. First and foremost, I would like to thank Pei-yu Chen, my committee chair and doctoral advisor, for her support and guidance throughout this journey. She allowed me to choose a research topic that I am truly interested in and she believed in my ability to complete this dissertation study. Without her consistent encouragement and affirmation, I would not have survived this Ph.D. program. I would also like to express my appreciation to Paul A. Pavlou, Jose M. Plehn-Dujowich, and Yuexiao Dong for their help and advice as members of my committee. Their suggestions and comments at both the proposal and the dissertation stages were indispensable to the successful completion of my dissertation. I would also like to thank Rajiv Banker for his help in collaborating on one of the studies herein.

Beyond the committee, I would like to thank the faculty members of our MIS department. First, to Munir Mandviwalla, Youngjin Yoo, and Daniel Fesenmaier, thank you for providing me with this precious opportunity to pursue my Ph.D. at Temple University. These individuals guided me to the world of MIS during my first year of the Ph.D. program. I made lots of mistakes in my first year and they helped me a lot to find the right way to be successful in MIS. In particular, when I worked for Daniel, I gained much knowledge on information search, which helped me a lot in my dissertation study. Thanks also to Steven Johnson and David Schuff for sharing their ideas with me in seminars and talks. Thanks as well Sunil Wattal, Than Lam and Jinyoung Min, with whom I created and taught a new course. Thanks very much for all the help; it was my great pleasure and honor to work with them. I would also want to thank two professors in the Marketing Department, Richard A. Lancioni and James Hunt. They allowed me to sit in their seminars in three semesters and nearly all of my studies contain some interesting marketing theories.

I would also want to thank Gordon Burtch and Kevin Hong, both of whom are probably the best PhD students in our department. Thanks for their very helpful feedback to my dissertation. My appreciation further goes to my colleagues in the doctoral program at the Fox School of Business. In this regard, I would like to specially acknowledge Douglas Schutz, Yuhei Inoue, Angela Zenteno-Hidalgo, Ke Li, Shih-Ching Wang, and William Kline for your friendship and constant support. I would also want to thank all the people that I have worked with when I was working in the National Laboratory for Tourism & E-Commerce (NLTeC).

Finally, and most importantly, I would like to show my deepest appreciation to my wife, Wei Qi, and my parents, Shifei Yang and Zhijun Yang, for their patience, genuine care, and continuous encouragement. Without my wife’s tireless support, it would have been impossible to complete my doctoral study. Thanks to my 6-month old daughter Beiling Qi; she gives me lots of pressure to work hard and brings me lots of fun to enjoy a better life.

TABLE OF CONTENT

ABSTRACT .... **iv**

ACKNOWLEDGEMENTS .... **vi**

LIST OF TABLES .... **xii**

LIST OF FIGURES .... **xiv**

CHAPTER 1 INTRODUCTION **1**

1.1 Background on Open Innovation Contests 1

1.2 Literature Streams 3

1.3 An Innovation Contest in an Online Market 5

1.3.1 Step 1: The Seeker Launches a Contest 6

1.3.2 Step 2: Solvers Submit Solutions 7

1.3.3 Step 3: Communication 8

1.3.4 Step 4: Evaluation 8

1.4 Chapter Summaries 9

CHAPTER 2 ENHANCING OPEN INNOVATION CONTESTS **15**

2.1 Introduction 15

2.2 Literature Review and Context 20

2.3 Performance Model of an Online Open Innovation Contest 24

2.3.1 Prior Performance Model 25

2.3.2 Adapted Performance Model 26

2.4 Investigation of Factors Influencing Contest Performance 32
2.4.1 Contest Design Parameters 34
2.4.2 Project Intrinsic Characteristic 40
2.4.3 Market Environment Factors 41
2.4.4 Between Performance Proxies 44
2.5 Data and Methodology 45
2.5.1 Data Collection and Descriptive Analysis 45
2.5.2 Estimation Model 51
2.6 Result and Analysis 52
2.7 Robustness Check 60
2.8 Contributions, Implications and Future Research 61
2.8.1 Contributions 61
2.8.2 Managerial Implications 64
2.8.3 Implications for the Design of Contest Structure for Complex Projects 65
2.8.4 Limitations and Future Research 67
2.9 Conclusion 70

CHAPTER 3 WINNER DETERMINATION 71
3.1 Introduction 71
3.2 Theory and Hypotheses 75
3.2.1 Impact of Expertise/Past Experience 76
3.2.2 Temporal Strategy 77

3.2.3 Strategic Choice of Projects 82

3.3 Data and Estimation Model 83

3.3.1 Research Site and Sample Selection 83

3.3.2 Sample Data and Measurement.. 84

3.4 Estimation Model... 90

3.5 Results and Analysis 91

3.6 Discussion and Conclusion 99

CHAPTER 4 OPEN EVALUATION FOR OPEN INNOVATION......................... 102

4.1 Introduction... 102

4.2 ZBJ Open Evaluation System 106

4.3 Elements of Open Evaluation 108

4.3.1 Aggregation of Collective Intelligence 109

4.3.2 Evaluation Method 111

4.3.3 Solution/Evaluation Interactions 115

4.4 Performance Measurement Theory... 118

4.4.1 The Precision Metric 119

4.4.2 The Hit-or-Miss Metric 121

4.5 Data Analysis 122

4.5.1 Measurement 122

4.5.2 Methodology 125

4.6 Results... 126

4.6.1 Overall Analysis 129

4.6.2 Sectional Analysis 130

4.6.3 Robustness Check 134

4.7 Discussion and Conclusion 134

4.8 Limitations and Future Work 136

CHAPTER 5 CONCLUSION 138

5.1 Theoretical Contribution 138

5.1.1 NPD 139

5.1.2 Collective Intelligence 141

5.2 Overall Managerial Implications 142

5.2.1 Innovation Seekers 142

5.2.2 Problem Solvers 143

5.2.3 Market Operators 144

5.3 Future Research Suggestions 148

5.3.1 Reverse Auction vs. Innovation Contest 148

5.3.2 Multiple-stage Contest 149

5.3.3 Payment Policy 149

5.3.4 Open Evaluation 150

REFERENCES CITED 152

APPENDIX: FEEDBACK IMPACT 164

LIST OF TABLES

Table 2.1 Traditional Contests vs. Online Contests 23

Table 2.2 Variable Definitions and Measurements 46

Table 2.3 Category Means and Features 49

Table 2.4 General Descriptive Statistics 50

Table 2.5 General Regression (SUR) Analysis 53

Table 2.6 Sectional Results by Dimension 59

Table 3.1 Temporal Strategy Analysis of Solvers 80

Table 3.2 Expertise Variables Definition 85

Table 3.3 Descriptive Statistics of Expertise Variables 85

Table 3.4 Strategy Variables Definition 87

Table 3.5 Descriptive Statistics of Strategy Variables 87

Table 3.6 Correlation Coefficients of Participation Variables 88

Table 3.7 Descriptive Statistics of Fixed Project Variables 88

Table 3.8 Category Means and Features of Projects 89

Table 3.9 Conditional Logit Model and Results 91

Table 3.10 Sectional Results by Category 94

Table 3.11 Sectional Results by Dimension 94

Table 3.12 Sectional Results by Prize or Duration 96

Table 3.13 Correlations between Contest Designs and Expertise Signals 98

Table 4.1 Contest A vs. Contest B 116

Table 4.2 Original Result Vs. Grouped Result 121

Table 4.3 Descriptive Statistics 124

Table 4.4 Pearson Correlation Matrix 125

Table 4.5 Hit rate Descriptive Summary 127

Table 4.6 Overall Regression Result 128

Table 4.7 Sectional Descriptive Analysis 131

Table 4.8 Sectional Regression Results (Metric: Precision) 132

LIST OF FIGURES

Figure 1.1 Daily Submission Frequency 7

Figure 1.2 Step 3: Communication 8

Figure 1.3 NPD Framework (Terwiesch and Ulrich 2009) 10

Figure 2.1 Timeline for an Online Contest Vs. Timeline for a Traditional Contest 22

Figure 2.2 Feedback Impact on Extreme Value Distribution of v_i 30

Figure 2.3 Research Model 33

Figure 2.4 Result Model 56

Figure 3.1 Dynamic Timeline of Online Contest 74

Figure 3.2 Estimated Utility – U shape 92

Figure 4.1 Workflow of ZBJ Open Evaluation System 107

Figure 4.2 Framework of an Open Evaluation System 108

Figure 4.3 Role of Criteria Information 113

CHAPTER 1

INTRODUCTION

1.1 Background on Open Innovation Contests

Innovation is key for a firm to compete and survive in a rivalrous market. The identification of the best approaches to enhancing innovation performance is at the core of both practice and academic research. The traditional innovation paradigm focuses on optimal management and resource allocation within firms. However, due to the resource constraints, it is often difficult to significantly increase innovation performance using this approach (Chesbrough 2003). To overcome the resource constraints, firms can actively seek innovative opportunities from the external world. This innovation approach, referred to as open innovation, is receiving more and more attention of late (Chesbrough 2003; Hippel 2005; Terwiesch and Ulrich 2009; Terwiesch and Xu 2008). Facilitated by the global Internet and emerging forms of information technology, it has become very easy for companies to generate large numbers of innovative solutions through the use of online *open innovation contests* or *crowdsourcing contests* (Archak and Sundararajan 2009; Terwiesch and Ulrich 2009; Terwiesch and Xu 2008; Yang et al. 2009).

An open innovation contest, in which an innovation seeker (e.g. a firm, an organization or an individual) holds a contest to seek innovative ideas or solutions from external solvers for a specific problem, is an important approach to open innovation. The rapid development of the Internet has made it possible to leverage online open innovation contests as an efficient tool in aggregating external intelligence, which can reduce costs dramatically (Williams 2006). Further, Bonabeau (2009) notes that these contests often perform "*better than theorists can explain*," especially in terms of idea generation. Because of these obvious benefits, open innovation contests have been adopted by many firms for problem solving and new product development (NPD). However, these contests are applicable to nearly all sizes of firms. Large IT firms, such as Google and Netflix, can easily launch self-hosted innovation contests. For firms that lack IT deployment capacity or channels to reach huge volumes of potential solvers, launching innovation contests through an established online market is a wise choice. In practice, open innovation contest can employ the collective intelligence of a large pool of external solvers, which can help facilitate faster, more diversified and potentially better ideas or solutions, compared to internal innovation efforts (Bonabeau 2009; Terwiesch and Ulrich 2009). Besides, in a contest, the innovation seeker only needs to pay for the winning solution; that which outperforms all others. Thus the return on investment from innovation contests is higher than the traditional innovation approach, in which firms have to cover the risk of failure. Further, with the extremely low cost of online contest launch, participation, and communication, facilitated by the global Internet, online contests are becoming a popular method of open innovation.

The contest can also be viewed as an aggregation tool to achieve collective intelligence in problem solving. In practice, the power of collective intelligence is best applied to idea generation and evaluation (Bonabeau 2009), which is important, given that a successful innovation project needs to first generate numerous good ideas or solutions, and to then evaluate those solutions, to identify those that are "*exceptional*" (Terwiesch and Ulrich 2009).

1.2 Literature Streams

Five streams of literature have contributed most to the study of open innovation contests.

The first stream of literature is in regard to new product development (NPD) (Dahan and Mendelson 2001; Gassmann 2006; Girotra et al. 2010; Lakhani et al. 2007; Laursen and Salter 2006; Loch et al. 2006; Terwiesch and Ulrich 2009; Terwiesch and Xu 2008; VanGundy 1988). VanGundy (1988) and colleagues initiated this research stream by investigating the core issues in traditional NPD, including basic solution evaluation and management. More recently, Terwiesch and Xu (2008) provided a linear model of innovation contests, based on the extreme value model (Dahan and Mendelson 2001), incorporating multiple projects dimensions (Loch et al. 2006). Chao and Kavadias (2001) provided a comprehensive review of idea evaluation for all types of problems in the NPD literature. Terwiesch and Ulrich (2009) formulated a NPD framework, which divides the process into three parts: idea generation, idea evaluation, and phase-gate production (Figure 1.3). This framework suggests an integrated roadmap, which can be leveraged to guide the use of contests as tools to accelerate innovation. Given that open innovation contests are intended to address the same fundamental issue as traditional NPD projects, the rich knowledge from the NPD research stream can be borrowed or extended to study open innovation contests.

The second stream of literature is rooted in economics. Lazear and Rosen (1981) proposed the first contest model, in a linear format, to identify the optimal design for a contest prize. For example, if there are multiple winners, their model suggests how to set the optimal prize amount for different winners in order to help contest owners obtain the highest output from contestants (Modovanu and Sela 2001). Many studies have also investigated the design of an optimal contest prize structure, conditional on other aspects of the contest, such as project type (Archak and Sundararajan 2009; Che and Gale 2003; Dasgupta and Nti 1998; Glazer and Hassin 1988; Runkel 2006; Runkel et al. 2003; Sheremeta 2009). Similarly, Moldovanu and Sela (2006) have studied the optimal design of contest structure, noting, for instance, that in some cases it would be better to have contestants compete in a playoff tournament structure, where winners of sub-contests subsequently compete against one another in later stages.

The third stream of literature comes from marketing research. Sales contests have been studied extensively in marketing because they are an important promotional technique (Hart et al. 1989; Kalra and Shi 2001; Liu et al. 2007; Murphy et al. 2004; Murthy and Mantrala 2005). However, sales contests are a little different from innovation contests, since sales managers are concerned with the total sales generated by all contestants, while innovation managers are only interested in the quality of the best solutions (Terwiesch and Xu 2008).

The fourth stream of literature is that pertaining to the notion of collective (collected) intelligence. This stream views a contest as a powerful tool to aggregate the efforts of the crowd (Bonabeau 2009; Bothos et al. 2009; Lévy 1997; Mataric 1993; Watkins 2007). Collective intelligence is a shared or group intelligence that emerges from aggregated individuals through some coordinating mechanism, such as collaboration or competition (Bothos et al. 2009; Surowiecki 2004; Watkins 2007). Bonabeau (2009) views collective intelligence as an external human brain that can help in making decisions. He suggests that the applications of collective intelligence can be divided into two categories: idea generation and idea evaluation. This is highly consistent with Terwiesch and Ulrich (2009)'s innovation framework; the process is simply considered from a different perspective. This stream of literature is thus unique in that it compares contests to other aggregation mechanisms that can be used to achieve collective intelligence, thereby extending our perspective on problem solving.

The fifth and final stream of literature draws jointly from behavior economics and the psychology of reward discounting. This field of work pays unique attention to the irrational human response that is often observed when a problem solver must contend with multiple competitors, or when an individual must pay or invest to some degree before a reward can be received (Akerlof 1991; Edwards 1956; Green and Myerson 2004). These studies provide findings that are similar to those in the aforementioned stream of economic work, however, in contrast, they provide better explanations of certain empirical findings, such as the fact that individuals tend to procrastinate when the contest duration is longer, and that individuals invest a lower equilibrium effort when there are more competitors.

The five streams of literature described above comprise the most important theories related to open innovation contests. As such, the three studies in this dissertation draw on different aspects from each. Additionally, the three studies presented herein are informed by the literature on information retrieval and prediction markets.

1.3 An Innovation Contest in an Online Market

Before the results of prior studies can be used to produce something new, it is necessary to first introduce the full workflow of a typical online innovation contest, from the very beginning to the end. Based on our initial examination of several popular online contest markets, e.g., Zhubajie.com, TaskCN.com, TopCoder.com, 99Designs.com, CrowdDesgin.com, and Freelancer.com, we have formulated the below description a typical contest's timeline:

1.3.1 Step 1: The Seeker Launches a Contest

To launch a contest, an innovation seeker first needs to decide several things.

- Winner Prize. The total amount of the prize to be awarded, how many winners there should be and how much of the prize should be awarded to each winner. In most markets, the innovation seeker must pay the prize in advance, before the problem solvers can access the project.
- Project Description. The seeker needs to input their project details, which usually include background information, the project objective, and the criteria for determining winners.
- Contest duration. This indicates how many days the contest will be open for solution submission.

After inputting the above information, the contest can be launched successfully and displayed in the open contest list. In most markets, the latest contests will be displayed first. The exception to this is 99Designs, which prioritizes the display of contests based on how soon they are ending.

A common question faced by the innovation seeker at this stage is how to decide the optimal contest configuration.

1.3.2 Step 2: Solvers Submit Solutions

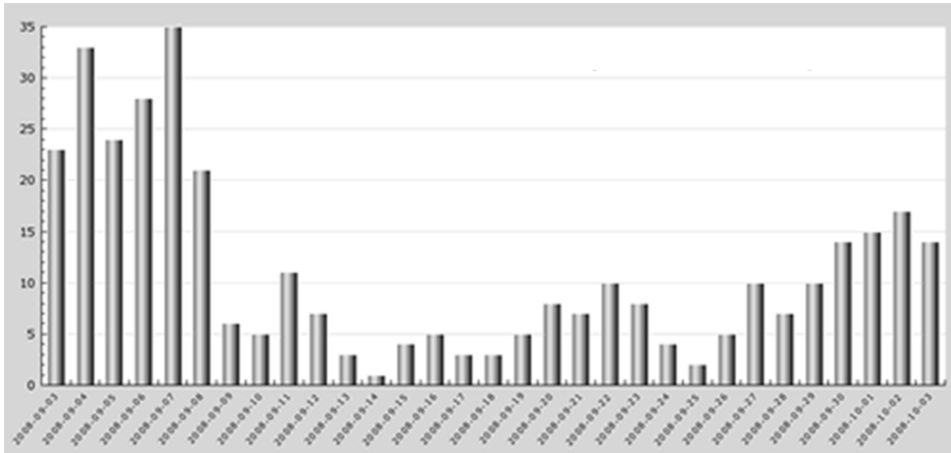


Figure 1.1 Daily Submission Frequency

Solvers next view and evaluate the contest. Each solver needs to decide whether or not to participate. It is interesting to note that most markets require a public announcement from each participant, although an announced solver is not required to submit a solution in the end. Solvers can submit a solution at any time before the contest is closed. Hence, no one knows how many solutions will be received before the contest is ended (**Figure 1.1**). It is also interesting to note that, by default, in most of these markets, submitted solutions are publicly viewable. One reason why market operators do this is that it makes the market more attractive to new visitors, who may be potential innovation seekers.

A common decision that innovation solvers may face at this stage is when to should submit their solutions. E.g., submit early vs. submit later.

1.3.3 Step 3: Communication

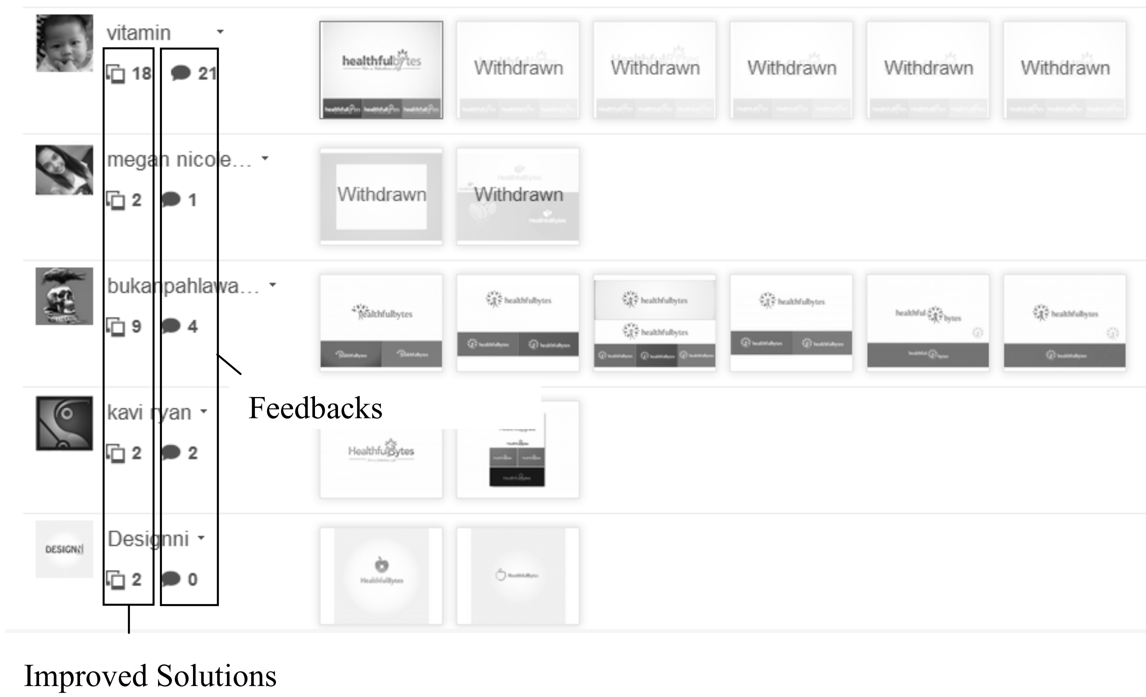


Figure 1.2 Step 3: Communication

After problem solvers begin submitting solutions, the seeker can begin to examine them. In most markets, the seeker is encouraged to send feedback to solvers about their submitted solutions. We have observed that solvers usually prefer to submit improved solutions after they receive feedback with suggestions for improvement, as shown in **Figure 1.2**. An interesting question that might arise at this stage, which we attempt to address, is how to measure contest performance, when submission feedback is allowed.

1.3.4 Step 4: Evaluation

Ultimately, the seeker needs to evaluate all solutions and announce the winner(s). If there are multiple winners, the seeker also needs to decide their order. In most markets, the seeker is responsible for all evaluation costs and is required to indicate the best solution(s). For each solution, TopCoder, a contest market mainly for software projects, provides a detailed evaluation score, which is supplied by a team of experts for the innovation seeker's reference. ZBJ Network provides a unique open evaluation system, which aggregates external evaluators to predict the solution that the innovation seeker will choose. It would therefore be interesting to examine whether this type of open evaluation approach can improve upon traditional evaluation.

This is the final step of a contest. Usually, the innovation seeker obtains a conceptual solution as the output of a contest. After obtaining this conceptual solution, innovation seekers need to convert it into a final product, which is another area of research in the field of NPD.

1.4 Chapter Summaries

The remainder of this dissertation contains three studies, which aim to improve our understanding of how contests can best be used as a tool to generate high quality ideas and to evaluate those ideas efficiently. Terwiesch and Ulrich (2009) divided all innovation problems into three parts: idea generation, idea evaluation and the phase-gate development process (Figure 1.3). They argue that the first two parts constitute the dual-core of the entire NPD framework, because they largely decide the quality of the final product.

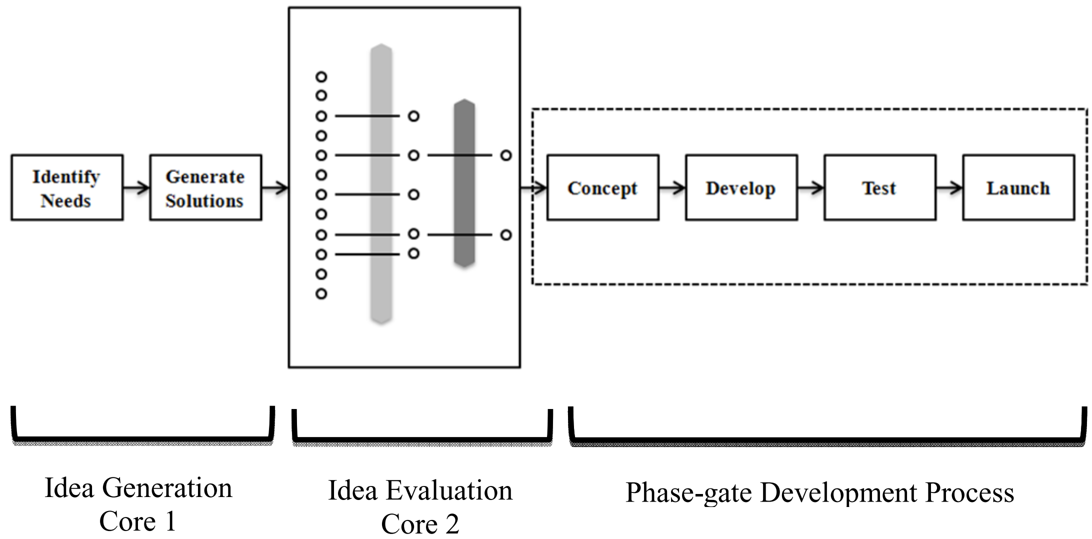


Figure 1.3 NPD Framework (Terwiesch and Ulrich 2009)

According to this framework, the first study (Chapter 2) and the second study (Chapter 3) fall within core 1: idea generation, while the third study (Chapter 4) falls within core 2: idea evaluation. The abstracts for each chapter follow.

In chapter 2, we consider innovation contests from the innovation seeker's perspective, in an effort to understand and improve contest performance. While there is a long stream of theoretical work on contests and tournaments, we identify several features of real-world open innovation contests in an online setting that make them vastly different from traditional contests. In particular, the presence of a feedback system makes these contests quite different. A feedback system is a software agent via which seekers can send feedback to solvers, based on the evaluation of their submitted solutions. These systems are widely used in online contests. We show that these features have profound impacts on the performance of online open innovation contests. We show that with the effective use of a feedback system, an increase can be achieved in the number of solvers and the submission speed, leading to higher contest performance, all else being equal. Using the number of solvers and submission speed as key measures, we empirically examine the factors that influence contest performance. Based on a large-scale dataset pertaining to open innovation contests in an online market, we identify three groups of factors that influence performance: contest design parameters (i.e. prize, description length and duration), project intrinsic characteristics (i.e. complexity) and market environment factors (i.e. competition intensity and market price). In general, we find that contests experience a higher submission speed when they have shorter descriptions, shorter durations and involve less complex projects. Further, we find that contests entertain a greater number of solvers when they offer larger prizes, involve less complex projects, have longer durations, lower levels of competition intensity, lower market prices, and faster submission speeds. Dimensionally, as defined by Terwiesch and Xu (2007), we find that ideation-based contests are most sensitive to variations in prize, whereas prize has relatively little influence on expertise-based contests. Surprisingly, longer project descriptions draw fewer solvers to ideation-based contests, yet they draw more solvers to expertise-based contests. In terms of their ability to attract large numbers of solvers, innovation contests exhibit the best aggregation power when they pertain to ideation-based projects. For a contest that involves a complex project, to avoid a situation in which no solutions are obtained, it seems it would be best to first aggregate solvers to work on the ideation-based portion of the project, prior to tackling the expertise-based portion.

In Chapter 3, we study innovation contests from the innovation solvers' perspective, in order to examine the characteristics of winners and solvers. Most studies of contests have taken the perspective of innovation seekers, thus little is known about solvers' strategies and responses. However, evaluating contest performance is also dependent on our understanding solvers' responses. This paper provides insights on these questions. Specifically, we show that the past experience of a solver is a good predictor of his future winning probability and, further, that winners are more likely to be those who submit early or late in the submission period, rather than those who submit in the middle. We also find that "strategic waiting" (for solution submissions) is associated with a higher winning probability. Furthermore, we show that different contests appear to attract solvers with different distributions of expertise, which invalidates a common assumption made in numerous previous studies: the presence of a fixed solver expertise distribution across projects. This finding also has strategic implications for the design of contest parameters.

In Chapter 4, we suggest a new evaluation approach for open innovation contests - open evaluation. By employing a prediction market as an aggregation mechanism, seekers can leverage the crowd to evaluate ideas or products. We first introduce the open evaluation system used by a large online contest market. A framework for open evaluation systems is then provided and several typical problems are discussed. We develop a performance model based on two different metrics: hit-or-miss and precision. By examining a large-scale empirical dataset, we identify several interesting findings. Notably, the usefulness of criteria information is found to be dependent on the criteria format and the project type. For graphic design projects, visual criteria appear to be very helpful while textual criteria appear to be only useful for highly expertise based projects such as software development. Further, and surprisingly, background information serving as implicating criteria appear to be misleading to external evaluators. The prediction market aggregation mechanism is extraordinarily efficient; we find that seekers can draw more evaluators by setting a higher evaluation prize. On average, US $2.00 can aggregate over 100 evaluators. Our results also show that the involvement of more evaluators results in better evaluation performance (i.e., a greater collective intelligence is employed). Although a public voting policy is commonly used for evaluation, we find that this approach introduces a herding effect, especially when there are numerous solutions. Furthermore, the herding effect results in a higher evaluation disparity among evaluators, which is associated with lower performance in the open evaluation process. If open innovation contest and open evaluation can be employed jointly and systematically, open innovation may be a useful complement to Google's internal innovation model such as 80/20 innovation time-off model. Further, open evaluation can be used in many other domains. For instance, leveraging open evaluation, a firm could systematically identify exceptional candidates for job positions, with lower recruitment costs, which would be very beneficial to large firms.

Lastly, in the concluding chapter, we provide a review and summary of three studies. Drawing on a large-scale empirical dataset, we obtain a series of findings. These findings improve our understanding of online innovation contests, allowing us to infer how best to leverage them for idea generation. In addition, these findings suggest a number of managerial implications regarding how firms can aggregate the effort of the crowd to achieve collective intelligence for the evaluation of ideas generated by an innovation contest. We hope that these studies can help open innovation pioneers, such as Google, to systematically generate and identify exceptionally good ideas with much less effort and at lower cost. By utilizing our findings, we expect more firms can adopt open innovation in a systematic manner.

CHAPTER 2

ENHANCING OPEN INNOVATION CONTESTS[1]

2.1 Introduction

Innovation, which refers to the introduction of something new; a new idea, method, or device (Webster's dictionary) or "new stuff that is made useful" (Barras 1986), is key to an organization's performance and growth. Traditionally, firms have performed innovation internally, pursuing it by way of individual projects or via ongoing R&D effort. In recent years, a new approach, called *open innovation*, has emerged as an efficient avenue to innovation (Chesbrough 2003; Hippel 2005; Terwiesch and Ulrich 2009; Terwiesch and Xu 2008). This approach relies on the undefined public, outside of the firm, to achieve innovation. This approach of using outside solvers to address questions and problems internal to the firm was initially employed to resolve research problems in the natural sciences, such as chemistry and physics (Lakhani et al. 2007). Later, this approach began to see wide application to resolve a variety of problems, such as graphic design, and algorithm development. An appealing feature of the open innovation approach is that *innovation seekers* only pay for the success of innovation projects, and not their failures. In addition, the potential for a larger pool of *innovation solvers* from outside the firm extends the scope of available knowledge resources and may offer faster and better innovation outcomes at lower cost. A popular practice in open innovation is to launch an open contest, seeking ideas and solutions to a pre-defined

---

[1] I co-authored this article with Professor Pei-yu Chen and Professor Paul A. Pavlou from the Management Information Systems Department, Fox School of Business, Temple University.

problem, an approach known as an *open innovation contest* (Terwiesch and Ulrich 2009; Terwiesch and Xu 2008). When this type of contest is conducted online, it is called an online open innovation contest or a *crowdsourcing contest* (Archak and Sundararajan 2009; Howe 2006; Yang et al. 2008; Yoo and Hill 2010). Many large firms are adopting open innovation contests with the hope of capturing more and potentially better external ideas or solutions to a specific problem. For instance, in September 2008, Google funded the $10M launch of Project 10^100, which called for outside ideas to change the world. This was done in an attempt to solicit as many innovative ideas as possible from individuals outside of the firm. Since 2006, Netflix has offered a yearly $1M prize, with the goal of substantially improving the accuracy of predictions of viewers' enjoyment from watching different movies, based on movie preferences. Beyond these sorts of firm-induced open innovation contests, several online markets exist that facilitate open innovation contests. *InnoCentive,* founded in 2001, was the first online market to host open innovation projects in the form of contests (Allio 2004). A variety of project types are posted on InnoCentive, including logo, website, algorithm, and construction design, and the innovation seekers who post these projects could be either individuals or firms. There are a number of other online contest markets, such as TopCoder, DesignCrowd and TaskCN, each with a different project or geographic focus. Many individuals and organizations are now using these online markets for open innovation contests. Google Trends shows that both the search and news reference volumes for the terms *"open innovation"* and *"crowdsourcing"* have been increasing dramatically since 2008. The volume of Google searches for the term "crowdsourcing" is projected to reach 4 times what it was in 2008, by the end of 2011.

While there is a long stream of research on contests and tournaments in the economics and new product development (NPD) literatures, most previous studies have been purely theoretical and have mainly focused on the optimal design of prize structures, e.g. the best allocation of a prize between the first and second place winners (Archak and Sundararajan 2009; Lazear and Rosen 1981; Liu et al. 2007; Moldovanu and Sela 2001; Moldovanu and Sela 2006; Terwiesch and Xu 2008). However, the derivations have been based on several assumptions, such as the innovation seeker having foreknowledge about the number of solvers before the contest is initiated, or the presence of simultaneous competition among solvers. However, many of these assumptions do not hold in real world online contests, due to the dynamism of the participation process (e.g. solvers can join a contest any time before the contest ends). In general, very little empirical research has been conducted on contests.

The purpose of this research is to unveil open innovation contests in online markets and to empirically examine factors that enhance the performance of these contests from an innovation seeker's perspective. While contest performance is a variable that is not observed directly, by employing an analytical model in which contest performance is measured by the best solution (closest to the seeker's goal or ideal solution), we show that the number of solvers and the solvers' submission speed can be good proxies for contest performance when a feedback system is used effectively. Feedback systems, which often accompany online contests in practice, are commonly used between seekers and solvers to communicate questions and answers, as well as for seekers to provide feedback to solvers regarding their solutions. The advantage of using a feedback system of this sort is that it allows solvers to direct their efforts appropriately, in line with the seeker's preference.

Using the number of solvers and submission speed as proxies for contest performance, we empirically examine the influence of a number of factors. Based on a large-scale dataset pertaining to a number of open innovation contests from a real online market, we identify three groups of factors*:* contest design parameters (i.e. prize, description length, and duration), project intrinsic characteristics (i.e. complexity) and market environment factors (i.e. competition intensity and market price) that influence the two key performance metrics: number of solvers and submission speed. In general, we find that contests draw higher submission speeds (i.e., solvers submit their solutions faster) when they have shorter descriptions, have shorter durations, and involve less complex projects. Further, we find that contests draw a greater number of solvers when they have larger prizes, have longer durations, involve less complex projects, exhibit lower competition intensity, have lower market prices, and have faster submission speeds. Dimensionally, as defined by Terwiesch and Xu (2007), we find that the performance of ideation-based contests is most sensitive to variation in prize amounts, while the performance of expertise-based contests is not. Surprisingly, longer project descriptions draw fewer solvers for ideation-based contests, while drawing more solvers for expertise-based contest. With regard to their usefulness in terms of their ability to draw numerous solvers, innovation contests appear to be best suited for use with ideation-based projects.

Our study makes several unique contributions. First, we identify two features of contests in online markets that make them very different from their offline counterparts, or what has typically been assumed in previous studies (Table 2.1). Specifically, a real world online contest has an uncertain number of solvers due to the dynamism of the participation process and the competitive market environment. Second, the use of a feedback system can encourage solvers to significantly increase their effort in the appropriate direction, preferred by seekers, thereby mitigating the negative effect from the number of solvers on solver effort level. Third, we show that with the effective use of a feedback system, the number of solvers and the submission speed that emerge in a particular contest can serve as good proxies for contest performance. To the best of our knowledge, submission speed, which refers to the duration of time from when a contest initiates to when it receives a solution, has not been studied in the prior literature. Since the number of solvers and submission speed are easily observable, these values provide a simple way to measure contest performance, which has generally proven difficult for researchers to measure. Fourth, unlike previous studies, which mainly consider prize structure, we extend researchers attention to the impact of three groups of variables: contest design parameters (i.e. prize, project description length, and duration), project intrinsic characteristics (i.e. project complexity) and market environment factors (i.e. competition intensity and market price).

The remainder of this paper is organized as follows. In the next section, we review the literature and introduce the context of our study. An adapted performance model is then introduced and two performance metrics are described. Based on the new performance model, we propose a research model that consists of three groups of variables. Lastly, the research model is tested using empirical data. A variety of implications are then discussed, according to the empirical analysis.

2.2 Literature Review and Context

An innovation contest is a form of a contest, a game in which several agents expend resources to win prizes (Moldovanu and Sela 2001) [2]. Most open innovation contests are one-stage contests, in which there is just one project without any subsequent problems to solve. In most prior studies, monetary incentives have proven critical, and, accordingly, the prize structure has lain at the core of determining the overall contest performance. In prior work, performance has generally been measured according to the seeker's subjective goal and, as such, the measurement approach has varied with the seeker, even for identical contest output. For instance, early work by Lazear and Rosen (1981) proposed a simple contest model with only two competitors. This model focused on identifying the optimal prize structure, in order to stimulate the best solver contest performance. In most contest studies, information is complete and the contest performance is evaluated along one dimension, from seeker's perspective, such as quality or quantity (Archak and Sundararajan 2009; Lazear and Rosen 1981; Liu et al. 2007; Moldovanu and Sela 2001). An important finding from the literature is that the presence of only two solvers in the competition can drive each solver to exert their best effort. In contrast, having many solvers working on an innovation contest will lead to a lower level of equilibrium effort from each solver, which is undesirable from the seeker's point of view. However, Terwiesch and Xu (2008) argue that, despite the lower level of equilibrium effort, having a higher number of solvers with diversified backgrounds can contribute to the innovativeness of the solution. Another substantial contribution of

[2] Later, Moldovanu and Sela (2006) improved their one-stage model by allowing multi-stage contests.

Terwiesch and Xu's work is that they introduce three dimensions that can be used to describe contest projects: ideation, expertise, and trial-and-error based projects. *Ideation*-based projects are those that look for innovative ideas, such as the name of a new company. *Expertise*-based projects are those that require specialized expertise, such as software development. *Trial-and-error*-based projects are innovative problems with a very "rugged" solution landscape, where solvers cannot determine the solution without conducting trials.

Terweisch and Ulrich (2009) concluded that there are three approaches to improving contest performance: (1) increase the number of solvers, (2) increase the background diversity of solvers, and (3) increase the average quality of each solution. However, there exists a tradeoff between these approaches: while having more solvers may also increase the diversity of solutions, having more solvers will also decrease the solvers' equilibrium effort, leading to a lower average quality of solutions.

Overall, there are many studies in the fields of economics and NPD that shed lights on open innovation contests that are conducted in a traditional format. However, to date, very few studies have been conducted on open innovation contests in online markets. Although many online services are derived from traditional services models, there are differences due to the temporal and geographical dispersion of the online users (Gregg and Walczak 2003). This is also true when we compare online contests to traditional, offline contests. Nowadays, online open innovation contests have attracted increased interest due to the rapid development of information technology and its obvious benefits. Therefore, it is very important to redefine the contest scenario before we go further.

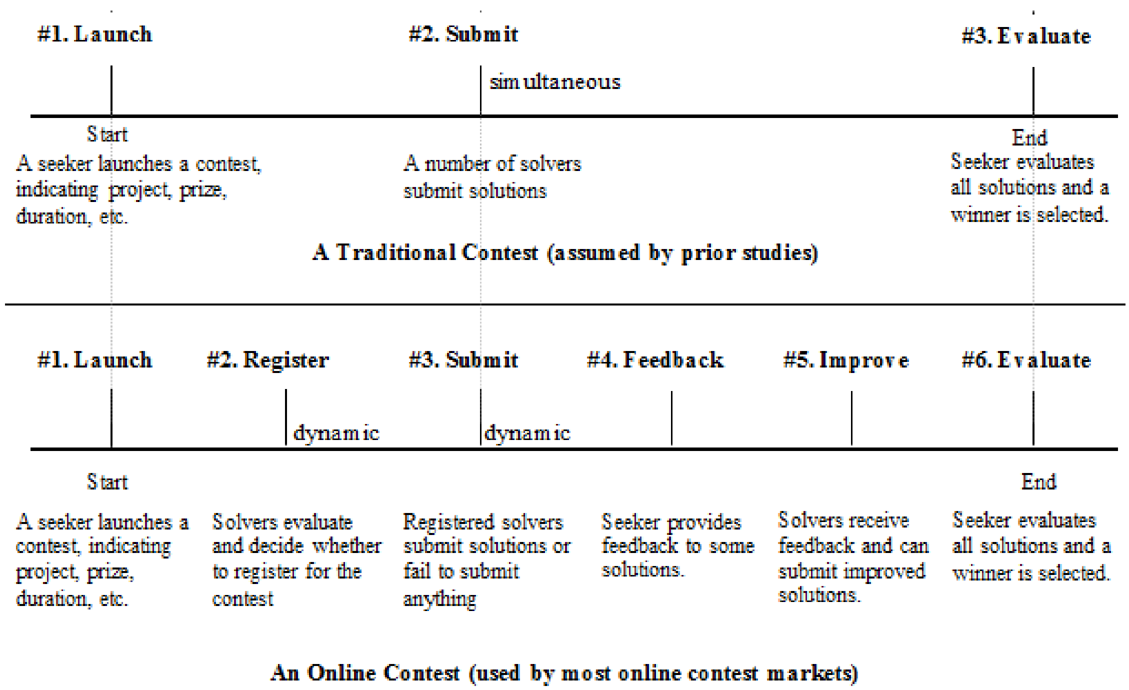


Figure 2.1 Timeline for an Online Contest Vs. Timeline for a Traditional Contest

We summarize the typical timeline of a traditional contest, which has been assumed by previous studies, in **Figure 2.1**. We also summarize the timeline of a typical online contest, which is based on our survey of many online contest markets, including TaskCN, Zhubajie, and Freelancer.com.

Table 2.1 Traditional Contests vs. Online Contests

Distinctions	Traditional Contests	Online Contests
Participation Process	A certain number of solvers compete *simultaneously* and the number of competitors is almost fixed. Further, no alternative contests are considered. **Consequence:** the number of solvers is taken as given.	Solvers have many contests to choose from, and can register and submit solutions at any time before the contest has ended. **Consequence:** The number of solvers is uncertain and is related to many preset factors.
Feedback	Seekers usually have no communication with solvers. **Consequence:** Each solver has identical information and will only execute an equilibrium effort.	Seekers commonly leave feedback on preferred solutions. Such feedback may indicate the changes desired by the seeker. **Consequence:** Solvers who received feedback perceive higher probability of winning, and would likely exert more effort to increase their solution quality, hoping to enhance their winning probability

In a typical online contest, an innovation seeker launches the contest in an online market with some pre-defined contest design parameters, i.e., a description of the project goal, a set prize amount[3], a set contest duration (how soon to end), etc. After that point, potential solvers with relevant backgrounds and experience come to evaluate the contest and decide whether to join the contest by registering. Once registered, the individual will be listed as a registered solver for this contest and the number of registered solvers will be increased by 1. It is interesting to note that most online contest markets require registration, though this process is not essential. The registration is free and is conducting

[3] We only consider single-winner contests in this study.

using a survey tool, which guides seekers and solvers through the process. However registration does not necessitate that a solver submit a solution. We observe that nearly all contests exhibit some “nonfeasance” solvers who register but fail to submit a solution. Once a solution is submitted, it becomes immediately available for the seeker to evaluate and issue feedback about, which may tell the solver how to make the solution better, from the seeker’s point of view. After receiving feedback, the solver may choose to submit an improved solution or to ignore it. Finally, the winner with the best solution is selected from all available solutions by the seeker. At that time, the contest is ended.

As summarized in Table 2.1, two features make online contests distinctive from traditional contests: (1) the dynamic participation process, and (2) the use of a feedback system. In a market with many ongoing contests, solvers have many alternatives to choose from. They can join any contest at any time before the contest ends. So, the total number of real solvers who will enter a given contest is unknown until the contest has ended. It is also in the seekers’ interest to send feedback to solvers on their preferred solution. Such feedback may inform the solver about the changes desired by the seeker. Solvers who receive feedback may perceive that they have a higher probability of winning, and may thus be more likely to exert additional effort to increase their solution’s quality. In summary, open innovation contests in online markets constitute a quite different contest scenario and need to be treated differently from traditional contests.

2.3 Performance Model of an Online Open Innovation Contest

Performance is always at the core of optimal design. The performance of online innovation contests is often a subjective term, particularly when it is measured in terms of the seeker's satisfaction level in regard to the contest result[4]. While in some cases performance can easily be measured based on output quantities or quality, more often than not, it is difficult to measure performance across different contests. We aim to identify performance measures for online contests and to those determine factors that influence them. We accomplish this by adapting the traditional contest performance model developed by Terwiesch and Xu (2008), while also taking into account the features of online contests.

2.3.1 Prior Performance Model

Following Terwiesch and Xu (2008)'s work, a one-stage innovation contest, where all solvers compete only once, can be modeled as:

$$V = \rho \max_{i=1,\ldots,n}\{v_i\} + (1-\rho)\frac{\sum_{i=1,\ldots,n} v_i}{n}, \qquad (2.1)$$

where $0 \leq \rho \leq 1$, V is overall performance, n is the number of solvers, and ρ is the weight of the best-performing solution. If a seeker only cares about the best solution, $\rho=1$; if a seeker cares about all solutions equally, such as in a contest where the goal is to maximize the cumulative performance (e.g., sales) of all 'solutions', then $\rho=0$. We consider the most common case, where the seekers are mostly concerned about providing the best solution (i.e., $\rho=1$), and where the contest performance is decided based on the quality of the best submitted solution. Thus, the quality of the solution captures the disparity between the solution and the seeker's goal or ideal quality. We denote the

[4] A common empirical approach to evaluate the performance of a specific contest is to simply ask the seeker to report a satisfaction score. Unfortunately, this approach does not work with our research site TaskCN, where 1594 out of 1621 (over 98%) of seekers are satisfied. The binary rating scores leave very little variance and it is therefore hard to differentiate the performance of different contests.

variable v_i to be the quality of the solution submitted by solver i, where i =1… n. The performance of a specific solver, i, is given in a linear format:

$$v_i(\beta_i, e_i, \xi_i) = \beta_i + r(e_i) + \xi_i \ , \tag{2.2}$$

where β_i is the expertise level of solver i. Terwiesch and Xu assumed that the distribution of expertise is known and it is fixed across all contests. $r(e_i)$ is the output of effort when solver i executes effort e_i. $r(e_i)$ is increasing in e_i. ξ_i is a random error term for each solution. This random error also captures the unobserved preferences of the seeker toward the solution and includes the diversified ideation-based output. Since β_i is fixed, the variance in performance is mainly based on the effort output r(e_i) and random error. When feedback is not available, Terwiesch and Xu (2008) found that a larger population of solvers would bring more diversified ideas, but would also lower each solver's equilibrium effort, e_i, which is also consistent with prior literature. In other words, having more solvers can increase the probability of obtaining more diverse ideas, but it does not guarantee solutions of better quality, due to the lower equilibrium effort expended by each solver. Therefore, an outstanding question remains around whether having more solvers can increase the performance of open innovation contests. In the following section, we adapt Terwiesch and Xu's model to incorporate the important features of online contests.

2.3.2 Adapted Performance Model

Terwiesch and Xu's model is applicable to an offline contest scenario where a certain number of solvers compete in one contest simultaneously, without feedback. However, for an open innovation contest in an online market, we need to modify this model to accommodate the scenario.

An obvious feature of online contests is the dynamism of the participation process, which results in an uncertain number of solvers. In an open environment, a solver can join the contest by registering at any time, as long as the contest remains open. The final number of solvers is therefore unknown, until the contest has ended. Instead of taking the number of solvers as given, we treat it as an emergent number that reflects some underlying information about a contest. **On the one hand,** the final number of solvers will never be infinitely large, as it is bounded by some factors. When more and more solvers register, the next solver to arrive will perceive a lower and lower probability of winning (Green and Myerson 2004). Eventually, the number of solvers will reach a point of saturation, where the perceived probability of winning will become insufficient to draw additional solvers. Thus, each contest will draw a certain number of solvers and this number will be bounded by a set of variables that define the *contest's launching environment,* such as the prize amount, duration, project characteristics and market information**. On the other hand**, having more solvers also means that more external knowledge can be exploited, which has the potential to result in better performance. Chesbrough (2003) insists that a central part of open innovation is the search for outside sources that have commercial potential. Similarly, open innovation performance has been argued to be dependent on how much external knowledge a firm can access (Laursen and Salter 2006). Lakhani et al. (2007) found that the success of problem solving is related to one's ability to attract specialized and diverse solvers. In a free-entry market, more solvers from the undefined, external world will bring more diversified solutions, which will increase innovativeness (Terwiesch and Xu 2008). In a contest, each solver represents a knowledge unit. If the expertise distribution of any given solver is fixed across contests, the number of solvers represents the potential performance of a contest, $V_{potential}$, and the *contest's launching environment* information. In summary, we have:

$$V_{potential} \propto Number\,of\;\;Solvers = Contest's\;Launching\,Environment\quad, \qquad (2.3)$$

where the number of solvers only includes those solvers who submit solutions. Solvers who register but fail to submit a solution contribute nothing to the contest performance and thus would not be counted. The performance potential for a project depends on exploitation of the effort that innovation seekers exert in order to make solvers contribute more, and this is related to the second feature of online contests.

The second feature of online contests is the use of a feedback system. Prior studies have assumed no communication between seekers and solvers (Lazear and Rosen 1981; Terwiesch and Xu 2008). Although this assumption greatly simplifies the contest scenario, we have observed that a feedback system is commonly used to facilitate and organize the communication process, with the aim of improving solution quality. To gain a better understanding of the impact of a feedback system on a given contest and the quality of each solution, we collected feedback related data from Zhubajie and conducted an experiment at TaskCN (Appendix). Alexa.com online traffic data shows that Zhubajie and TaskCN are the first and second largest contest markets in China. Our results show that 70.1% of seekers used feedback systems and, on average, seekers sent feedback in response to 8.9% of solutions. We found that the receipt of feedback had a significantly positive effect on solution quality ($p<0.001$), wherein each instance of feedback increased a solver's solution quality by 0.6, on average. In particular, high quality feedback (i.e. encouraging words and detailed improvement suggestion) generated even more improved solutions. For more details of feedback's influence, please see the Appendix. In summary, our result shows that sending informative feedback can highly encourage solvers to exert more effort. The feedback improves contest performance by not only increasing the effort (e_i) that solvers exert, but also by increasing the match between the solution and the seeker's unobserved preference (ξ_i), as solvers have a better understanding of the seeker's requirements. As a result, we revise equation 2.2 as follows:

$$v_i(\beta_i, e_i, \xi_i, f_i) = \beta_i + r(e_i \mid f_i) + \xi_i(f_i),$$

where f_i indicates whether feedback is received for solver i. A solver will exert higher effort when feedback is received, compared to when no feedback is received, i.e., $r(e_i \mid f_i) > r(e_i \mid 0)$. Moreover, $\xi_i(f) > \xi_i(0)$, as discussed previously.

In summary, the performance of a solution, v_i, can be improved toward the preferences of the seeker through the provision of feedback (**Figure 2.2**). A solver's response to feedback can be explained by both economics and psychology. Taking an economic perspective, without any feedback, each solver will only exert the equilibrium effort, which is a function decreasing in the number of solvers (Terwiesch and Xu 2008). From a psychological perspective, before receiving any feedback from the seeker, solvers only perceive a discounted prize due to the presence of many competitors, thus the effort he[5] will devote is determined by his expected returns, as determined by the probability of winning and the size of the prize (Ainslie 1992; Green and Myerson 2004; Kagel et al. 1995). Once feedback is given, a solver receives an indication that his solution is preferred and thus that it has a higher probability of winning, compared to when there is no feedback received. As a result of the solver perceiving a higher probability of winning,

[5] Following the tradition of economics, we call a seeker "she" and a solver "he".

they have a greater incentive to exert more effort. In summary, a solver would likely exert a higher level of effort in the presence of feedback, compared to the equilibrium effort they would exert should they receive no feedback. When the cost of making an improved solution does not exceed the expected payoff, it is in the solver's best interest to do so.

This result suggests that effective use of feedback can counteract the negative impact of having numerous solvers; the exertion of equilibrium effort, a common finding from the literature. In particular, with the use of a feedback system, the three aforementioned approaches to improving contest performance; namely, increasing the number of solvers, increasing the diversity of solver backgrounds and increasing the average quality of solutions, can potentially be pursued in tandem.

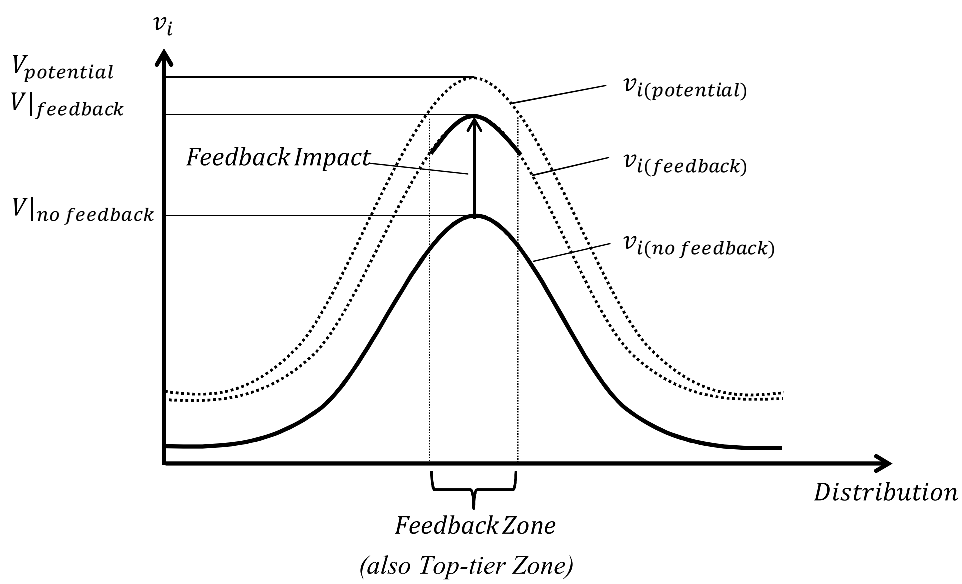


Figure 2.2 Feedback Impact on Extreme Value Distribution of v_i

From the innovation seeker's perspective, extreme value theory (Dahan and Mendelson 2001) suggests that the act of sending feedback to top-tier solutions is sufficient to improve the overall contest performance. The extreme value distributions of v_i under different circumstances are illustrated in **Figure 2.2**.

To improve the overall contest performance, V, from $V|_{no\ feedback}$ to $V|_{feedback}$ seekers only need to send feedback to top-tier solutions to improve the top-tier $v_{i\ (no\ feedback)}$ to $v_{i\ (feedback)}$. Empirical data also indicates a consistent result: in practice, seekers only send feedback to 8.9% of solutions (Appendix) and the winning solution is eventually chosen from the small set of solutions that receive feedback. Ideally, if seekers use the feedback system effectively, i.e. by sending informative feedback that provides sufficient incentive and detailed comments to solvers about how to improve their solutions to meet the seeker's goal, the negative impact of having many competitors, on the equilibrium effort, can be totally offset. As such, it becomes possible for the practical performance to reach to the maximum potential performance:

$$V\big|_{Feedback} \approx V_{potential} \propto Number\ of\ Solvers \qquad (2.4)$$ [6]

In the NPD literature, time efficiency is also a key to innovation success (Bstieler 2005; Filippini et al. 2004; Karagozoglu and Brown 1993; Swink et al. 2006). Firms not only pursue a cheaper and better innovative product, they seek to do so quickly (Swink et al. 2006). However, time efficiency has been ignored in the previous contest literature. One explanation for this is that, in the offline world, the evaluation process does not start until a contest has ended and, as such, there is not a big difference between early solutions and late solutions. However, in online contests, seekers may start to evaluate the

[6] This equation is based on the assumption that the expertise distribution is constant and that the seeker is capable of evaluating all submissions. A constant expertise distribution has been assumed by many other studies, though the validity of this assumption remains unknown. Further, if a feedback system is absent, our model is not applicable.

solutions before the contest ends. In addition, by encouraging early solutions, a seeker can also improve contest performance by providing feedback early, while also allowing sufficient time for solvers to implement the changes they desired. The benefits of feedback decrease for late solutions, as there may be insufficient time to implement the desired changes. Thus, when a feedback system is used, a seeker can increase the quality of the solutions by encouraging early submissions. Similarly, auction studies also find that bidding speed impacts the auction result and the number of bidders (Borle et al. 2006). We therefore include submission speed as a performance metric. Submission speed is measured as the duration of time between the beginning of a contest and the point at which solutions start to arrive. Altogether, the overall contest performance for seekers can therefore be proxied as follows:

$$V\,|_{feedback} \approx V_{potential} \propto Number\ of\ Solvers, Submission\ \ Speed \qquad (2.5)$$

In the remainder of this study, we will use both the number of solvers and submission speed as performance metrics for online contests.

2.4 Investigation of Factors Influencing Contest Performance

We are interested in exploring factors that influence the two contest performance metrics: the number of solvers and the submission speed. As mentioned earlier, the final number of solvers is unknown at the outset of each contest. However, this number will never be infinitely large, as it is bounded by some factors. When more and more solvers register, the next solver to arrive will perceive a lower and lower probability of winning (Green and Myerson 2004). Eventually, the number of solvers will reach a point of saturation, wherein the next potential solver's perceived probability of winning becomes insufficient to draw their interest. Thus, each contest will capture a certain number of solvers and this number is bounded by a set of variables that define the *contest's launching environment,* including the prize amount, duration, project characteristics and market information. Similarly, we would expect that the submission speed is also bounded by aspects of the contest's launching environment.

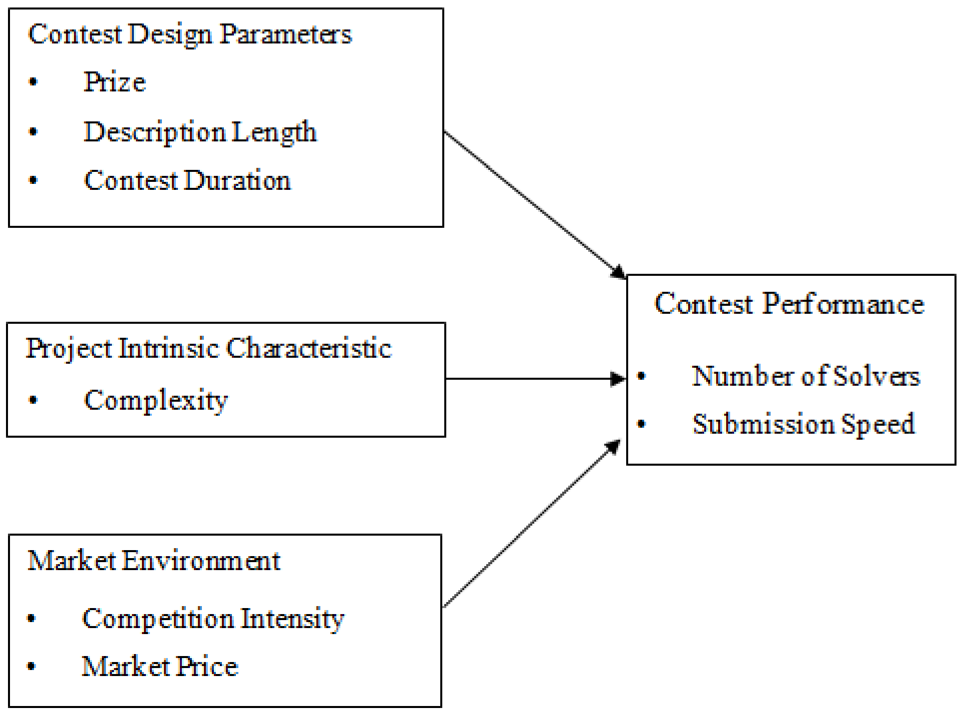


Figure 2.3 Research Model

In analyzing this launching environment, we first consider the variables that seekers can change arbitrarily. We call these variables *Contest Design Parameters* (i.e., prize, description length, and duration). For instance, the decision to set a prize of $100 or $1,000 is entirely up to the seeker. We also consider a variable that is intrinsic to the project (i.e., project complexity) and other variables related to the market environment. These variables could have an impact on contest performance, but seekers cannot change them arbitrarily. An open innovation contest basically includes a project with the objective of solving a specific problem. This problem emerges from a seeker's need or desire to achieve something. The problem that the seeker wishes to address defines the skill sets required, which is again beyond the control of seekers. We refer to this group of variables as *Project Intrinsic Characteristics*. A typical example of such variables is project complexity. The variables in the remaining group are referred to as *Market Environment Variables* (i.e., competition intensity and market price), which reflect the competitive intensity of contests relative to other, similar contests in the current marketplace. Prior literature considers each contest to be independent, however, in reality several contests are competing for talented solvers from the same pool of individuals, thus there is reason to believe that the market environment, in which the contest takes place, will influence the ultimate performance. Finally, **Figure 2.3** outlines our research model.

2.4.1 Contest Design Parameters

Contest design parameters define a contest: what is the problem to be solved (project description), how large is the prize (prize amount), and how long does the contest last (contest duration). Seekers can change contest design parameters arbitrarily. For example, for the same type of project, some seekers may provide a more detailed description of the problem, while others may provide a very vague description; some seekers may set higher prize amounts or longer contest durations than others. We are interested in how these contest design parameters impact the two performance metrics. To discover what patterns of contest design parameters can cause a contest to draw more solvers with a faster submission speed, we first need to understand solvers' information search behavior. Solvers' information search behavior has two features. **First**, each solver searches for several matched contests, instead of just participating in the first matched contest they can find (Blanchard and Diamond 1992; Mortensen 1986). Due to incomplete information, solvers will search for a number of preferred contests, in order to make a better decision. **Second**, more recent contests have a higher probability of attracting solvers' attention than contests posted earlier. In all of the online contest markets we considered, by default, the latest contests are always displayed in top-ranked positions, where more solvers can see them. This is a practice similar to most online procurement auction markets, such as Elance. Once a contest becomes 'stale', it will be pushed down the display list, where fewer solvers will access it. Information search behavior, based on the above two features, has been studied thoroughly in the information retrieval literature (Chapelle and Zhang 2009; Craswell et al. 2008; Dupret and Piwowarski 2008; Richardson et al. 2007). These studies suggest a position-based model, which implies that the probability of a solver choosing a contest depends on both his preferences and the display position of the contest. This probability can be modeled as:

$$Prob\ (Choosing\ \ a\ Contest) = \frac{1}{1+\exp\left(-\alpha_{preference\ score} + \beta_{position}\right)}, \tag{2.6}$$

where α_{pref} is positive and accounts for the likelihood that a solver chooses a contest for which they have a certain preference score. $\beta_{\mathrm{duration}}$ is positive and captures the impact of display position on solvers' likelihood of choosing a contest. A contest for which a solver has a higher preference score and which is ranked higher in terms of its display position (small ranking number) will capture more solvers. In a contest market, solvers prefer contests with a lower perceived cost and higher perceived compensation (Williamson 1998). A contest's display position is largely decided by its duration. We next explore the impacts of contest design parameters, based on the position based model (Equation 2.6) and solvers' preferences.

Prize Amount

The prize amount refers to the size of the prize that a contest winner will receive. Most previous studies have considered a monetary prize as the main incentive for solvers to compete (Lazear and Rosen 1981; Moldovanu and Sela 2001; Moldovanu and Sela 2006; Terwiesch and Xu 2008). Increasing the prize amount is believed to be an effective approach to increasing contest performance. This is because a larger prize amount increases solvers' expected payoff, and, accordingly, justifies solvers' exertion of greater effort, holding the number of solvers constant. Moreover, all else being equal, we would expect that a solver prefers a higher prize and is therefore more likely to participate in a contest that has a higher prize amount. This is because a higher prize provides a higher expected payoff, making it more likely to offset the opportunity costs a solver may face. This suggests that a contest with a higher prize should attract more solvers. This suggestion appears to be true based on our consideration of a solver's information search behavior in online contest markets. The prize amount only influences a solver's preference score; it has no bearing on the contest's display position. A contest with a higher prize amount provides higher compensation to the winner and each solver may perceive a higher benefit from participating, on average. Thus, a contest with a higher prize should be preferred by more solvers. Previous empirical research in a reverse auction setting supports this notion, as there is evidence that a higher budget indeed attracts more bids (Snir and Hitt 2003). We therefore hypothesize:

H1a: *A contest with a higher prize will attract more solvers.*

For a solver, if he participates in several contests during the same period, it is likely that he would give higher priority to the higher prized contest, due to a higher perceived payoff. As a result, all else being equal, we would expect that a higher prize would increase a solver's incentive to submit their solution earlier. We use the term submission speed to duration of time following a contest's outset before it starts to receive solution submissions. Thus, we make the following hypothesis:

H1b: *A contest with a higher prize will receive a solution sooner.*

Project Description

The project description helps solvers to formulate an idea about the specific problems of a contest project. As noted earlier, solvers have the option to propose the project using different descriptive styles. Even for identical projects, where all technical requirements and all other contest variables are the same, seekers may provide lots of details and requirements, or they may be very abstract about what needs to be done. For instance, we observe that some seekers choose to simply introduce what they need, while other seekers choose to provide more information, such as company background, culture, client features, etc. Although the project objectives and complexity are not very different, a longer description requires more time to process and possibly has a higher learning cost. As such, longer descriptions may give solvers the feeling that a project is more complicated. As a result, a contest with a longer project description may draw fewer solvers than those with similar objectives but shorter descriptions. On the other hand, one may argue that solvers get a better idea about what is needed from the seeker when there is a longer description, reducing the random error associated with guessing at what is desired by the seeker. If solvers are risk averse and try to avoid uncertainty, then longer a description may actually attract more solvers. It will be interesting to examine the relationship between description length and the number of solvers attained. We hypothesize:

H2a: *A contest with a longer description will attract fewer solvers.*

Similarly, higher learning costs prevent solvers from submitting solutions quickly, thus this can delay the solution delivery. So we also have:

H2b*: A contest with a longer description will receive solutions more slowly.*

Contest Duration

The contest duration refers to the number of days for which a contest is open and accepting new solutions. In contrast to prize and project description length, the contest duration influences a contest's display position. Immediately after launch, a contest is displayed in the top position, where solvers have the highest chance of seeing it. As time goes by, some subsequently launched contests are placed at the top of the contest list, so prior contests are moved down the contest list, where solvers have less chance of viewing them. For a contest with a duration of *d* days, the position impact can be modeled as a function of *t*:

$$\beta_{position} = f(t),\ 1 \le t \le d\ ,$$

where *t* refers to the *t*-th day after launch. The probability of a solver choosing this contest on the *t*-th day is:

$$Prob\ (Choosing\ \ a\ Contest)\,\big|_t = \frac{1}{1 + \exp\left(-\alpha_{preference\ score} + f(t)\right)},$$

For a large online market, we assume there are N_{market} visiting solvers per day. For a mature market, N_{market} is quite stable and can be treated as a constant. Then,

$$Number\ of\ Solvers\Big|_{duration=d} = \int_{t=1}^{t=d} \int_{Daily\ Visiting\ Solvers} Prob\ (Choosing\ a\ Contest)\Big|_t \cdot d(solver)dt$$

$$= N_{market} \cdot \int_{t=1}^{t=d} Prob\ (Choosing\ a\ Contest)\Big|_t \cdot dt$$

The marginal number of solvers on the *t*-th day equals:

$$\frac{d(Number\ of\ Solvers)}{dt} = N_{market} \cdot Prob\ (Choosing\ a\ Contest)\Big|_t$$

$$= \frac{N_{market}}{1+\exp\left(-\alpha_{preference\ score} + f(t)\right)} > 0 \qquad (2.7)$$

Equation 2.7 suggests that the marginal number of solvers on the *t*-th day is always positive. Thus, a contest with a longer duration can always capture more solvers. Snir and Hitt (2003) also included duration in their reverse auction study and similarly found that auctions of longer duration can capture more bidders.

However, in a highly active market, lots of new contests are launched every day and a contest launched earlier will be pushed down in its display position with each subsequent contest. Thus $f(t+1)>f(t)$, which means the marginal number of solvers is decreasing in *t*. Although a longer duration can draw more solvers, the efficiency with which new solvers can be attracted decreases as time goes by. Time is also a constrained resource. Even if H3a is supported, seekers need to consider whether it is worth waiting for too long. In summary, this leads to the following hypothesis:

H3a: *A contest with a longer duration will attract more solvers, but the marginal number of solvers is decreasing.*

All else being equal, a contest with a longer duration gives solvers more time to work on solutions and, as a result, solvers may feel less pressure to submit solutions sooner and may prioritize other, shorter-duration contests. Prior studies in behavioral economics have found that people have a lower intention to finish a task if no instant benefit or loss is perceived (Akerlof 1991; Green and Myerson 2004). In other words, longer durations make solvers more likely to procrastinate and lead to slower submission speeds. When a solver participates in several contests during the same period, he is more likely to finish the solution for a contest that has a shorter duration. So, we have the following hypothesis:

H3b: *A contest with a longer duration will receive solutions more slowly.*

2.4.2 Project Intrinsic Characteristic

Project intrinsic characteristics are factors that reflect what solvers need to do and, as such, these define the skill sets required to complete the project successfully. These characteristics are independent of the contest design or the market environment. A typical intrinsic project characteristic is its *complexity*. A higher project complexity not only increases the 'barriers to entry' for solvers, but also demands more time or effort to implement. We observe that most problem solvers in online contest markets are individuals. It is well established empirically that most individuals lack the capability and inclination to deal with complexity (Tversky and Kahneman 1974; Van de Ven 1986). When a project is more complex, the uncertainty or risk that a solver needs to bear increases, thus the project becomes less attractive. Experiments have also shown that people are less likely to choose more complex projects (Sonsino et al. 2002). This leads to the following hypotheses:

H4a: *A contest with a more complex project will attract fewer solvers.*

Less complex contests require lower setup costs and efforts and will, therefore, lead to faster submission speed due to lower time cost.

H4b: *A contest with a less complex project will receive solutions sooner.*

2.4.3 Market Environment Factors

The prior literature on contests considers each contest to be independent; however, in reality, several contests are often competing for the attention of talented solvers from the same pool of individuals. There is also reason to believe that the market environment in which a contest takes place will influence the contest's ultimate performance. *Market Environment Variables* (i.e., competition intensity and market price) reflect how competitive a contest is, relative to other, similar contests in the current marketplace. Several auction studies have indicated that market environment factors can influence the auction outcomes because bidders are inclined to search for the best deals among all available auctions with similar or substitutable products (Bapna et al. 2009; McAfee 1993; Zeithammer 2006). Similarly, when a contest is open, there are other active contests that draw solvers with related backgrounds, thus a given contest will be competing for solvers against all other similar contests being held during the same period. We refer to these contests as *overlapping contests*[7]. Given that each solver has limited capacity and energy, solvers can participate in only a small number of contests. Therefore, when the solver pool has a static size, more overlapping contests will reduce the number of solvers that each contest can receive, on average. Thus, the outcome of each contest is dependent on all the other overlapping contests. We consider two market environment factors: *competition intensity*, which indicates the number of overlapping contests of the same type that a contest has to contend with in a given period, and *market price*, which

[7] Our terminology is consistent with prior auction literature where "competing auctions" is used to denote simultaneous auctions that sell similar or substitutable products (Anwar et al. 2006). In contrast, "overlapping auctions" is used to refer to auctions that are not completely simultaneous or exactly sequential (Bapna et al. 2009), similar to the scenario in a contest market.

represents the average prize amount over all overlapping contests of the same type, within a given period. The potential influence of these two factors is discussed below.

Competition Intensity

Competition intensity reflects the degree of market competition a contest must face, in terms of quantity. Specifically, this variable measures how many overlapping contests exist, of the same type, during a period (e.g., how many overlapping graphic design contests were launched during past 30 days). In general, price information from a competitive market reflects the public and private information of all market participants (Spann and Skiera 2003; Spann and Skiera 2009). When market demands increase, the price of supply will also increase, and vice versa. For instance, having more overlapping auctions will lower bidding prices (Bapna et al. 2009). In contest markets, the prize is always fixed and the supply-demand relationship can only be captured by the number of solvers. When the number of overlapping contests increases, the market demand increases and the supply side will receive more benefits. Thus, each solver would have a higher probability of winning and each contest would attract fewer solvers. We therefore hypothesize that:

H5a: *A contest in a market with higher competition intensity will attract fewer solvers.*

When there are more overlapping contests, solvers have more contests to choose from and may feel less pressure to submit solutions early. Although, at the same time, one can also argue that facing more contests, a solver in a contest may want to submit his solution earlier so that he can move on to next contest. On average, a higher submission speed should be observed for each contest when competition intensity is high. We do not have any prior regarding the effect of competition intensity on submission speed, however, for the purposes of hypothesis testing, we have:

H5b: *A contest in a market with higher competition intensity will result in a lower submission speed.*

Market Price

Market price reflects how large of a reward solvers usually receive in contests of a similar type, across the market, which is similar to the concept of a market price for a particular job candidate. We measure market price as the average prize amount for all overlapping contests of the same type in a given period. The relative attractiveness of a contest is influenced by the relative size of its prize, among the overlapping contests. Grounded in microeconomic theory, when the market price for similar contests increases (i.e., other overlapping contests have high prizes), a solver participating in a given contest essentially faces higher opportunity costs because he loses a certain capacity to pursue other higher-prized contests. Therefore, given the same prize, it becomes more difficult for a contest to attract more solvers when it is competing in a market with higher market price. We therefore hypothesize that:

H6a*: A contest in a market with a higher market price from overlapping contests will have fewer solvers.*

We do not have any specific hypothesis regarding how market price will impact submission speed. On the one hand, an environment with a higher market price indicates it is a buyer's (i.e., solvers in our context) market. Therefore, solvers may have less pressure to submit solutions early. On the other hand, faced with a good market, a solver in a contest may want to submit his solution earlier so that he can move on to the next contest. If this were the case, a high market price would lead to a higher submission speed for each contest, on average. This remains an interesting empirical question. For the purposes of hypothesis testing, we will evaluate the following:

H6b: *A contest in a market with a higher market price from overlapping contests will result in a lower submission speed.*

2.4.4 Between Performance Proxies

We have shown that the submission speed and the number of solvers are good proxy measures for contest performance. The relationship between these two performance proxies also has important implications for our understanding of contest performance. Previous auction literature has shown that a shorter time to the receipt of a first bid will lead to a lower bidding price (Borle et al. 2006). If submission speed and the number of solvers go in opposite directions, this suggests that there may exist a tradeoff between increasing the number of solvers and increasing the submission speed. In other words, this would suggest that both dimensions of performance could not be increased simultaneously. On the other hand, if they are always moving toward the same direction, we would only need to consider one measure of contest performance. In this case, seekers might predict contest performance solely based on submission speed because submission speed information is uncovered before the number of solvers; the number of solvers remains unknown until after the contest ends. Regardless of the relationships between the two performance metrics, submission speed on its own may also contain certain useful but unobservable information that can facilitate prediction of the number of solvers. We therefore hypothesize that:

H7: *A contest with a higher submission speed will result in more solvers.*

2.5 Data and Methodology

2.5.1 Data Collection and Descriptive Analysis

Our data was collected from TaskCN.com, which was founded in China in 2005 and is one of the largest online service markets in the world. This market allows anyone to launch a contest by depositing a prize amount in advance. Solvers can participate in any contest for free. By the end of 2009, there were over 2.8 million registered solvers and over 20,000 contests in the market archive. We chose to study TaskCN for several reasons: first of all, compared to InnoCentive and Topcoder, TaskCN covers a much wider range of contests, from simple ideation-based projects to complex projects that require high levels of expertise and skills; second, compared to other contest markets, TaskCN.com is a very active online contest market – not only does it attracts a large number of contests at any given time, contests also attract 111.6 solvers on average; third, this online market provides well-organized archival data. As a result, this market provides a natural test bed with large amounts of data on real transactions from various contests, which allows us to test our hypotheses about the relative importance of different factors.

Table 2.2 Variable Definitions and Measurements

Variable	Definitions and Measurements
Performance Metrics	
Number of Solvers	This measures the number of solvers that have submitted solutions to a contest. Solvers who registered but failed to submit a solution are not counted.
Submission Speed	We first measure how soon a contest starts to receive solutions after it is launched and we refer to this as the response time. In order to make this variable consistent with the definition of speed in physics, we define the submission speed as the inverse of response time. For instance, if the response time is 0.5 hour, the submission speed is 2 submissions per hour for this contest.
Contest Design Parameters	
Prize Amount	It is the amount of money (Chinese Yuan :￥) set by the contest seeker as the prize for the winner. The market has a prize-never-refundable policy to avoid a moral hazard. In any contest, the full amount of the prize is paid to the market before the contest can be launched. The market charges 20% of the prize as a service fee for every contest, so the winner of each contest will receive 80% of the total deposited prize. Since the 20% service fee is a fixed rate, we still use the total prize as the prize amount.
Project Description	This number simply measures how many Chinese characters a seeker used in describing the contest project.
Contest Duration	We measure the duration of each contest by counting the days between the start and end time set by seekers. The start time and end time are available from TaskCN.

Table 2.2 (continued)

Variable	Definitions and Measurements
Project Intrinsic Characteristic	
Project Complexity	For each contest, solvers need to first register by announcing their participation and then work on the project. So, we know how many registered solvers there are and how many solvers submitted solutions for each contest. First, we calculate the *incompletion rate* as the percentage of registered solvers who failed to submit a solution. The incompletion rate directly measures the uncertainty around a project that solvers collectively perceive, and also acts as a measure of complexity (Tversky and Kahneman 1974; Van de Ven 1986). However, the incompletion rate for a contest is unknown before the contest ends, thus we cannot use this as an independent variable in our analysis. As such, we use an instrumental variable, the category-level incompletion rate, to capture project complexity for all contests under the category. So, our project complexity is also a categorical variable.
Category	There are 14 project categories in total: Logo design (33.25%), Graphic design (9.5%), Naming (9.09%), Creative writing (7.00%), Q&A (5.71%), House design, Packaging design, Industry design, Web design, Media, Software, Translation and Website Application. Sales contests (18.4%) were eliminated, as explained earlier.

Table 2.2 (continued)

Variable	Definitions and Measurements
Market Environment Factors	
Competition Intensity	For a specific contest, from 30 days before and after the start date, the number of overlapping contests in the same category is used to measure competition intensity. 30 days is chosen because we observed that the number of solutions rarely increases 30 days after the contest starts.
Market Price	For any given contest, the market price is the average prize amount of all the other contests in the same category, for the 30 days before and after the start date.

Table 2.3 Category Means and Features

Category	Complexity [0,1]	Prize (￥)	Duration (days)	Competition Intensity	No. Solvers	Percentage	D
Web App	0.45	285.01	20.81	21.43	13.53	3.3%	E
Translation	0.48	67.58	**13.37**	16.03	118.78	1.40%	E
Software	0.38	126.84	17.32	43.94	**8.31**	5.45%	E
Q&A	0.25	**61.70**	16.67	60.34	22.48	5.15%	Q&A
Media	0.51	316.89	16.05	22.98	15.49	3.25%	I & E
Web Design	0.53	483.57	21.66	40.51	18.19	6.65%	I & E
Industry Design	**0.64**	**545.55**	25.32	**7.812**	18.87	5.15%	I & E
Packaging Design	0.55	439.73	**30.43**	13.18	31.23	1.9%	I & E
LOGO Design	0.41	358.66	22.59	**279.8**	60.56	**40.5%**	I
Interior Design	0.53	245.17	17.24	8.46	21.46	**0.85%**	I & E
Graphic Design	0.48	288.35	19.56	94.1	38.3	13.5%	I & E
Creative Writing	0.53	159.21	19.74	68.8	71.34	7.5%	I
Naming	**0.22**	136.28	25.32	75.33	**729.22**	9.6%	I

Note: The maximum and minimum values of each column are bolded. "D" column represents the dimensions for each project category. Dimension "I" refers to projects that are purely or mostly ideation based; Dimension "E" refers to projects that require specialized expertise. These dimensions, with the exception of Q&A, were considered by Terwiesch and Xu (2008).

Table 2.4 General Descriptive Statistics

Variable	Mean	Std Dev.	Max	Min
Number of Solvers	111.67	319.94	4498.00	1.00
Submission Speed	1.81	4.46	50.00	0.01
Complexity	0.41	0.08	0.64	0.22
Prize Amount (￥)	287.15	361.95	5500.00	1.00
Description Length (char)	1022.29	966.92	9996.00	16.00
Contest Duration (days)	21.28	15.81	104.11	1.00
Competition Intensity	149.06	117.62	413	2.00
Market Price	306.04	139.45	1287.43	29.00
No. Observations		1995		

Correlation Matrix

	NS	SS	Prize	Dura	Desc	Comp	CI	MP
Number of Solvers (NS)	1.000							
Submission Speed (SS)	0.488	1.000						
Prize	0.171	-0.121	1.000					
Duration (Dura)	0.337	0.007	0.227	1.000				
Description (Desc)	0.106	-0.126	0.256	0.229	1.000			
Complexity (Comp)	-0.414	-0.391	0.207	-0.021	0.101	1.000		
Competition Intensity (CI)	0.248	-0.082	0.092	0.094	0.182	0.180	1.000	
Market Price (MP)	-0.008	-0.184	0.209	0.109	0.199	-0.340	0.392	1.000

Our data was collected during the period between September 2008 and September 2009. In total, there were 3,723 contests. We eliminated around 20% of these contests from the sample because they were multi-winner contests, and considering the impact of prize structure is not at the core of this study. Sales force contests (18.4% of total), which aim to promote product sales, were also eliminated, since they have very different performance metrics (Liu et al. 2007; Murphy et al. 2004; Terwiesch and Xu 2008). After these adjustments, 1,995 contests remained in our sample. The variable definitions, measurement methods and descriptive analysis are given in **Table 2.2**, **Table 2.3** and **Table 2.4**.

2.5.2 Estimation Model

In this study, we are mostly interested in the relative impact of the proposed antecedents of contest performance. Since our variables of interest (e.g., number of solvers, prize amount, description length, contest duration, etc.) follow long-tailed distributions, all antecedent variables are natural log-transformed and we use a log-linear model for our estimation. Thus, the coefficients in the results indicate multipliers to relative variances. For example, if a coefficient of an independent variable (IV) equals 2, then when this IV increases by 10%, the dependent variable will increase by 2 * 10%, or 20%.

Terwiesch and Xu (2007) introduce three project dimensions to describe contest projects: ideation, expertise and trial-and-error. *Ideation*-based projects are those that look for innovative ideas, such as a name for a new company. *Expertise*-based projects are those require specialized expertise, such as software development. *Trial-and-error*-based projects are innovative problems with a very “rugged” solution landscape, where solvers cannot know the result without conducting trials (Trial-and-error based projects are not allowed in TaskCN). Based on the weight of ideation and expertise, we divide all projects into four groups: (i) ideation-based, (ii) expertise-based, (iii) ideation- and expertise-based, and (iv) Q & A (**Table 2.3**). Since different types of projects (e.g., ideation-based vs. expertise-based) may have different required skill sets, the solver pool that different types of projects attract may be different. Therefore, we also control for these dimensions in our model.

Because there are potentially unobserved factors that can contribute to both the number of solvers and submission speed (i.e., errors are likely to be correlated across sub-models A and B), we use seemingly unrelated regression (SUR). Finally we have:

$$\begin{aligned} Ln(No.\,Solvers) &= \beta_a + \beta_{1a} Ln(Prize) + \beta_{2a} Ln(Description\ Length) + \beta_{3a} Ln(Duration) \\ &+ \beta_{4a} Ln(Complexity) + \beta_{5a} Ln(Competition\ Intensity) + \beta_{6a} Ln(Market\ Price) \\ &+ \beta_7 Ln(Submission\ Speed) + \beta_{8a} dimension + \xi_a \end{aligned}$$

$$\begin{aligned} Ln(Submission\ Speed) &= \beta_b + \beta_{1b} Ln(Prize) + \beta_{2b} Ln(Description\ Length) + \beta_{3b} Ln(Duration) \\ &+ \beta_{4b} Ln(Complexity) + \beta_{5b} Ln(Competition\ Intensity) + \beta_{6b} Ln(Market\ Price) + \beta_{8b} dimension + \xi_b \end{aligned}$$

In the models above, the coefficient annotations are consistent with the hypothesis annotations. For instance, β_{1a} is used to test H1a, and β_{1b} to test H1b.

2.6 Result and Analysis

The general results of the seemingly unrelated regression (SUR) are listed in **Table 2.5** and summarized in **Figure 2.4**.

Table 2.5 General Regression (SUR) Analysis

Variable	Ln(No. Solvers)	Ln(Submission Speed)
Constant	- 2.332*** (0.379)	- 3.369*** (0.711)
Contest Design Parameters:		
Ln(Prize Amount)	0.274*** (0.016)	- 0.005 (0.031)
Ln(Description Length)	- 0.001 (0.021)	- 0.128*** (0.041)
Ln(Contest Duration)	0.297*** (0.020)	- 0.183*** (0.039)
Project Intrinsic Characteristic:		
Ln(Complexity)	- 2.856*** (0.122)	- 3.292*** (0.216)
Market Environment Factors:		
Ln(Competition Intensity)	- 0.058** (0.027)	- 0.028 (0.051)
Ln(Market Price)	- 0.049* (0.028)	- 0.126 (0.120)
Between Performance Proxies:		
Ln(Submission Speed)	0.306*** (0.013)	--
Dimensions:		
Ideation	2.028*** (0.128)	1.9455*** (0.237)
Expertise	0.692*** (0.130)	1.0722*** (0.244)
*Ideation * Expertise*	- 1.214*** (0.151)	- 1.0453*** (0.284)
Number of observations	1995	
R^2	69.66%	22.44%

Note: *~p <0.1; **~p <0.05; ***~p <0.01

The result shows that contest design parameters can significantly influence both performance metrics. By increasing the prize amount, seekers can draw significantly more solvers, but the submission speed is not made significantly faster. So H1a is supported while H1b is rejected. In particular, the coefficient suggests that if the prize increases by 100%, the number of solvers would increase by 0.274*100%=27.4%. Although offering a higher prize can draw more solvers, the coefficient is less than 1, indicating that the number of solvers that each additional dollar can attract is decreasing. This result confirms that monetary motivation does exist and that it does play an important role explaining why people choose a particular contest.

The result also shows that longer descriptions do not lead to significantly fewer solvers, though they do significantly delay the submission speed. So H2a is rejected, while H2b is supported. Specifically, if description length increases by 100%, submission speed will decrease by 0.128*100%=12.8%. The result also shows that seekers can capture significantly more solvers by extending the contest duration, although the marginal number of solvers that this draws is decreasing (β_{3a}=0.297<1). Further, a longer duration also results in a slower submission speed. Therefore, both H3a and H3b are supported.

Our result also indicates that contests with less complex projects attract more solvers and that they also experience a faster submission speed. So, both H4a and H4b are supported. In addition, the coefficient shows that project complexity has a very strong impact on performance metrics. As shown in Table 2.5, the coefficients indicate that the effects of complexity on the number of solvers and on submission speed are -2.856 and -3.292, respectively. This means that if the project is 10% less complex, the number of solvers will increase by 2.856*10%=28.56% and the submission speed will increase by 3.292*10%=32.92%.

The market environment factors have a significant impact on the number of solvers, while no significant impacts on submission speed can be observed. In other words, a contest that has more overlapping contests in the presence of a higher market price will capture fewer solvers. So H5a and H6a are supported, while H5b and H6b are rejected. When the number of overlapping contests doubles, the number of solvers that each contest can attract decreases by 0.058*100%=5.8%. When the market price for a type of contest is doubled, the number of solvers will reduce by 0.049*100%=4.9%. Although both impacts are significant, the elasticity is very low. For a mature market, the turbulence of market price is usually very small, and the market environment impacts can therefore likely be neglected.

Additionally, the dimensional impacts are highly significant. In general, the result shows that the ideation dimension is more efficient in attracting solvers than the expertise dimension. This suggests that ideation-based projects have the largest pool of solvers, compared to other types of projects. This is reasonable as anyone may contribute to ideation-based projects if no specialized expertise is required. For expertise-based projects, any specific expertise narrows down the associated solver pool. If a project requires some rare expertise and the solver pool of the market is not large enough, it is possible that this project would not receive any solutions.

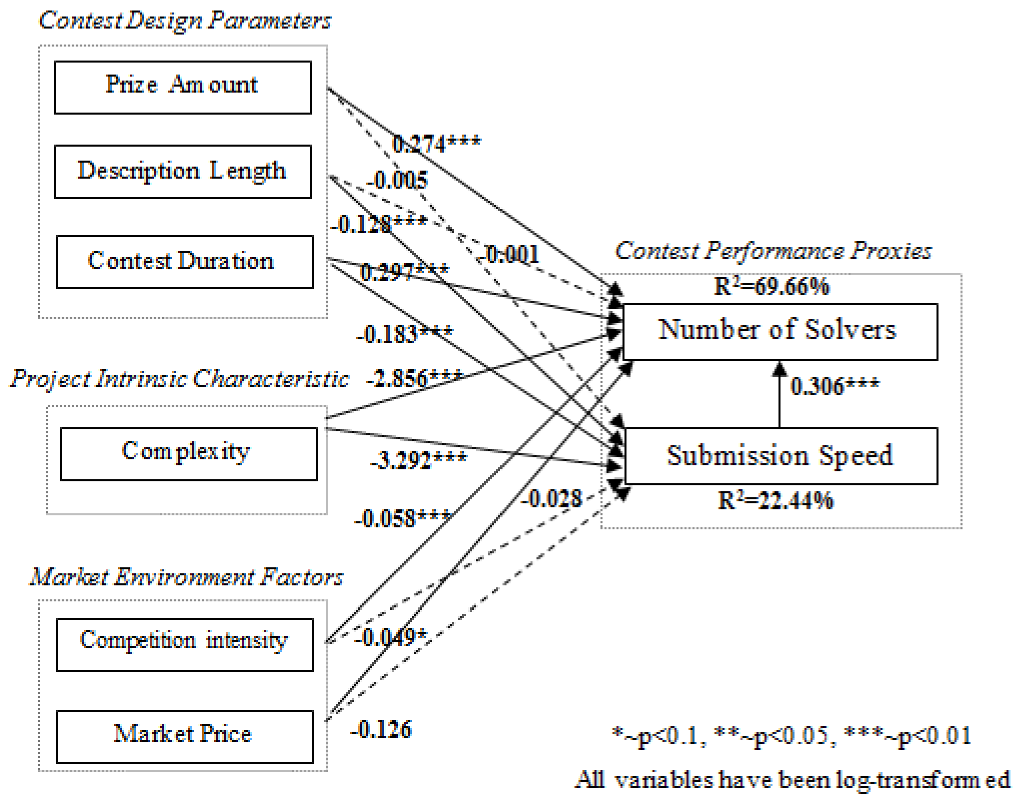


Figure 2.4 Result Model

In particular, when a project is both ideation- and expertise-based, where the interaction term between ideation and expertise equals 1, fewer solvers can be attracted. One explanation for this is that the combination of ideation and expertise increases project complexity, which prevents some solvers from participating. Combined with the impact of complexity, this result has interesting implications. In order to increase contest performance, instead of having a contest for a complex project as a whole, the seeker might wish to divide the complex project into modules, each of which has lower complexity. The seeker could then launch different contests for the different modules. In the case where this is not feasible, e.g., due to high interdependency between modules, then the seeker may implement a two-step process, as follows. Most innovation projects can be divided into a conceptual part (ideation-based) and an implementation part (expertise-based). Terwiesch and Xu (2008) argue that capturing an exceptionally good idea is the most important task in NPD. Therefore, the goal of the first step will be to obtain more diversified, exceptional ideas, which can be done by launching a contest for conceptual solutions. The second step will be to implement the chosen solution. A person with creative ideas need not be very good at implementation. Therefore, by having sequential independent contests, the seeker can ensure that appropriately qualified solvers address both contests.

Finally, the result for H7 suggests that a faster submission speed is significantly associated with a larger number of solvers, implying that the factors that increase submission speed are also likely to increase the number of solvers. Moreover, a seeker can also use the submission speed as a preliminary indication of the number of solvers that she is likely to end up with. Based on this, the seeker can plan for the required evaluation resources well in advance.

In general, our results suggest that a contest with a higher prize, longer duration, lower project complexity, lower competition intensity, lower market price and higher submission speed can attract a higher number of solvers. A contest with a shorter description, shorter duration, and lower project complexity, will lead to a higher submission speed. The R-square for estimating the number of solvers is as high as 69.66%, so our model can be used as a good forecasting tool to predict the number of solvers in practice. For example, a market owner might use this model to help seekers predict future performance before their contest is launched.

We also performed sectional regression analysis by dimension, which is presented in **Table 2.6**. This was done to identify differences between project dimensions. There are four categories, based on different combinations of dimensions: ideation-based (i.e. naming, LOGO design), expertise-based (i.e. software development, website applications), ideation-and-expertise-based (i.e. website design) and Q&A[8].

The variable impacts are mostly consistent with the general regression analysis, though in some cases the variables become insignificant. However, the following two findings are distinct and interesting.

---

[8] We retain Q&A contests because these are a common type of contest, which is similar to Yahoo! Answers. However, Q&A is not a typical innovation contest since it seeks a correct answer. Once a correct answer is given, the project is essentially ended.

Table 2.6 Sectional Results by Dimension

D.V.	I.V.	Dimension Sections			
		Ideation (1055 Obs)	Expertise (223 Obs)	I+E (624 Obs)	Q&A (93 Obs)
Ln (No. Solvers)	Constant	- 1.422***	7.037***	2.469	3.310
	Ln(Prize)	0.368***	- 0.048	0.189***	0.117*
	Ln(Description Length)	- 0.077***	0.189**	0.060	0.167
	Ln(Contest Duration)	0.300***	0.237***	0.293***	0.331***
	Ln(Complexity)	- 3.372***	- 4.940***	1.759	- 2.388
	Ln(Competition Intensity)	- 0.001	- 0.013	- 0.220***	- 0.321
	Ln(Market Price)	- 0.095*	- 0.432***	- 0.124	- 0.404
	Ln(Submission Speed)	0.284***	0.278***	0.311***	0.334***
	R^2	74.8%	58.01%	40.73%	43.38%
Ln (Submission Speed)	Constant	0.066	4.65*	- 0.042	0.720
	Ln(Prize)	- 0.025	- 0.033	- 0.261**	0.195*
	Ln(Description Length)	- 0.081*	- 0.058	- 0.164*	- 0.266
	Ln(Contest Duration)	- 0.087*	- 0.179	- 1.543	- 0.413***
	Ln(Complexity)	- 2.883***	- 4.412*	0.075	0.655
	Ln(Competition Intensity)	- 0.052	- 0.056	0.492	- 0.104
	Ln(Market Price)	- 0.465***	- 0.136	- 0.042	0.720
	R^2	30.08%	7.50%	5.71%	22.16%

Note: *~p <0.1; **~p <0.05; ***~p <0.01

First, it is very interesting to notice that for expertise-based projects, the prize shows no significant impact at all. This result confirms the argument that for contests requiring specific expertise, seekers may face the risk of receiving no solutions at all, regardless of the prize amount. Among the other three categories, ideation-based contests are most sensitive to prize amounts.

It is surprising to see that longer project descriptions draw fewer solvers for ideation-based contests, while they draw more solvers for expertise-based contest. For ideation-and-expertise-based contests, the impact is not significant, which is consistent with the general regression analysis. This result suggests that in ideation-based contests solvers do not like to see a lengthy description, as this may restrict their creativity. However, the solvers of expertise-based contests may need more details to feel confident about what seekers are looking for. In those contests based on both ideation and expertise, the significance of the description's impact is cancelled out.

Overall, if the number of solvers simply measures the potential performance (equation 2.3), the use of innovation contest shows the best application potential for ideation-based projects. .

2.7 Robustness Check

We performed the following robustness checks to ensure the validity of our results. First, scatter plots of observations and performance metrics did not show any pattern, indicating independence of observations (i.i.d.). In our regression, we constructed White (robust) standard errors to avoid issues stemming from heteroskedesticity. The effect of multicollinearity was also checked based on variation inflation factors (VIFs) for all models (Dasgupta and Nti 1998), and all VIFs were found to be below the suggested threshold (VIF<5). We also checked the correlations between variables to address the low power of VIF tests in large datasets. To avoid potentially unobserved factors that can contribute to both open evaluation performance and the number of evaluators (i.e., errors are likely to be correlated across the above models), we used the seemingly unrelated regression (SUR) method (Runkel 2006). Our result shows that the coefficients obtained with independent regressions are different, thus simultaneous regression analysis is necessary.

2.8 Contributions, Implications and Future Research

2.8.1 Contributions

Open innovation is a promising approach for innovation seekers due to the expectation of higher investment returns, faster returns and potentially better performance. By taking advantage of the global Internet, launching an innovation contest online can further enhance the performance of open innovation, due to the availability of a potentially larger pool of talented solvers and lower costs of attracting them, from all over the world. Indeed, in an online market with millions of potential solvers, a newly launched online contest can reach numerous solvers in a very short time, with controllable costs. Online contests for open innovation are becoming popular and have been adopted by many firms. However, despite increased interest in online open innovation contests, it is still unclear how innovation seekers can take advantage of online contests. Our study makes the following contributions to research on open innovation contests in online markets.

First, before conducting our analysis, we detail a more complicated scenario to reflect real-world open innovation contests. We identify that the real-world online contest process is very different from what is assumed by previous studies, because of two features. (1) Previous literature has assumed that a certain number of solvers start to compete at the same point in time, for the same duration, with symmetric information. However, the participation process of an online contest is dynamic and interdependent with other similar contests, implying that the number of solvers is uncertain and is bounded by a set of variables that define the contest's launching environment, such as its prize amount, duration, project characteristics, market information, etc. (2) Most previous studies assume that seekers simply select a winner without any interaction taking place between the seeker and solvers (Lazear and Rosen 1981; Terwiesch and Xu 2008). However, we find that feedback systems are frequently used to facilitate interactive two-way communications. Both the empirical data and our experiment show that seekers can encourage solvers to exert more effort by sending informative feedback.

More importantly, based on previous studies of contest performance and taking into consideration the above two features, we contribute to the literature by showing that the number of solvers and submission speed can be easily drawn upon as proxies for contest performance. Using feedback systems, the negative impact of having many solvers, which stems from a lower perceived probability of winning, can be eliminated. Thus, having more solvers can always bring more diverse solutions and higher contest performance. In our study, submission speed reflects how soon a contest begins to receive solutions. To our knowledge, this variable, used as a performance proxy, has not been studied before. The number of solvers and submission speed, taken together, largely simplify the performance measurement method and can be easily used in practice.

Third, instead of only considering prize, our framework considers a wider range of influential factors, which fall into three categories: (1) contest design parameters, (2) project intrinsic characteristic (i.e. project complexity), and (3) market environment factors. With this framework, we explore how contest designs, solvers, and the market interact over the contest process to achieve a high level of contest performance. In general, our result suggests that a contest with a higher prize, longer duration, lower project complexity, lower competition intensity, lower market price, and higher submission speed can draw more solvers. A contest with a shorter description, shorter duration, and lower project complexity will lead to higher submission speeds.

With the project dimensions introduced by Terwiesch and Xu (2007), we find that ideation-based contests are most sensitive to prize amounts. Further, prize appears to have no significant impact on expertise-based contests. In other words, contests requiring specific expertise or skills may face the risk of receiving no solutions at all, regardless how large the prize is. It is also surprising to find that longer project descriptions can draw fewer solvers for ideation-based contests, while they draw more solvers for expertise-based contest. This suggests that solvers of ideation-based contests do not like to see too much description, as this may limit their creativity. However, the solvers of expertise-based contests may need more details in order to become confident about what seekers are looking for. Considering the usefulness of online contests in terms of their ability to draw many solvers, these contests show the greatest potential for ideation-based projects.

Finally, although our data and model are based on contests in online markets, we believe they also offer practical insights for other types of contests (e.g., Google's 10^100 contest and the Netflix Prize contest) because contest design parameters, the external market environment, and projects' intrinsic characteristics still play important roles in determining the number of solvers that take part, which remains a useful proxy for contest performance. Netflix launched its contests using its own IT platforms, but they still need to determine the size of the prize, the contest duration and the project description. Even though it is not clear what other overlapping contests might be considered by potential solvers, there is no doubt that the external environment/conditions at the time of the contest will still impact the contest performance.

2.8.2 Managerial Implications

Our study provides several insights regarding how to help innovation seekers improve contest performance. First, seekers can and should make use of feedback systems to send high quality feedback on preferred solutions. High quality feedback refers to feedback that includes confirmatory information and detailed requirements for improvement. This action can highly encourage solvers to improve solution quality significantly. Second, when seekers set contest design parameters, they need to remain cognizant that:

- For ideation-based contests, it is better to make the description concise and give solvers more space for creativity. For expertise-based contests, it is better to provide more details about the project.

- A longer duration can help to draw more solvers; however it also lowers the submission speed, which may delay initial evaluation and feedback. Further, the marginal number of solvers per day is decreasing with time. Although setting a longer duration is free, an excessively duration may have a negative impact.
- If a project requires rare expertise or if the complexity level is high, the contest may face the risk of receiving no solutions. For this type of project, launching a procurement auction may be a better alternative.

In addition to the above, our study also provides useful suggestions to contest market operators. First, market operators can use our research model to help seekers predict future performance, before contests are launched. This function can lower seeker's adoption cost and bring more contests. Second, to increase the visibility and popularity of the market, it is more effective to increase the average prize size for the contests than it is to increase the number of overlapping contests.

2.8.3 Implications for the Design of Contest Structure for Complex Projects

Our study shows that project complexity has a substantial impact on the performance of online contests. Although a complex project is usually more valuable (**Table 2.3**), a contest with a complex project typically receives a very small number of solutions, or even no solutions. Thus, current contest mechanisms are not suitable for complex projects.

One explanation for the poor aggregation of solvers when it comes to complex projects is that most solvers are individuals, thus they prefer to avoid the risk of receiving nothing after investing a large amount of effort (Tversky and Kahneman 1974; Van de Ven 1986). So, the solutions for complex projects may come primarily from solvers who are willing to bear higher risk, instead of those with better expertise or better ideas. Complex projects are usually associated with high value. If there is no appropriate, feasible mechanism that can be employed in online contest for complex projects, the potential market for online contest is largely limited. This is likely counterintuitive for innovation seekers and market operators. To make online contests more effective for complex projects, we need a better design for the contest structure.

Terweisch and Ulrich (2009) have presented three approaches to improving contest performance: (1) increase the number of solvers, (2) increase the background diversity of solvers, and (3) increase the average quality of each solution. In an online market, solvers are presumably highly diversified. However, there is a tradeoff between drawing more solvers and simultaneously increasing the solution quality. This is because a large number of competing solvers may result in a lower perceived winning probability, causing solvers to reduce their effort. This, in turn, may result in a lower solution quality, on average. We show that by implementing an effective feedback system, this negative effect can be mitigated. Effective feedback shall give preferred solvers an incentive to further increase their effort, in anticipation of a higher winning probability. Therefore, under an effective feedback system, if seekers have the capability to evaluate all solutions and send feedback to preferred solvers, the emergent number of solvers can be used as a proxy measure of performance, or as a measure of the potential performance that a contest can ideally reach.

If we are unable to simultaneously increase the number of solvers, the diversity of solvers, and the quality in a one-stage contest, a two-stage contest structure may help to achieve all of these things, asynchronously. We therefore propose a two-stage contest model that can further increase contest performance for complex projects.

(1) In the first stage, instead of posting the "whole" project, it is suggested that seekers should lower project complexity with the goal of reaching more solvers, with more diverse ideas. A lower project complexity will lower solvers' opportunity costs and perceived risk, thus more solvers may join the contest. Also, the lowered complexity will lower seeker's evaluation cost in identifying a solution. Overall, this step allows the seeker to reach more, better solvers, compared to a one-stage contest. By evaluating solutions, a seeker can choose a certain small number of preferred solvers to proceed to the second stage of the competition.

(2) In the second stage, selected solvers will perceive a much higher probability of winning than solvers in a one-stage contest and, as such, they will have more incentive to bear the risk. Also, feedback should play a major role, increasing the quality of the preferred solutions. This two-stage structure design can mitigate the risk of missing the best idea, due to a failure to attract solvers who have good ideas but who are also unwilling to bear the high risk of taking on a complex project.

With this two-stage contest, a seeker can improve the contest performance using limited resources. In particular, this contest structure can make a complex project successful by scanning a much wider range of outside solutions.

2.8.4 Limitations and Future Research

This study has several limitations that create opportunities for future research. One limitation of our model is the assumption of a constant distribution of expertise amongst solvers, across contests, which is a common assumption in the literature.

However, it is not clear whether this assumption holds in the real world. In a reverse auction study, Snir and Hitt (2003) found that higher value projects attract a higher proportion of low skilled solvers. Boudreau and Lakhani (2011) also found that highly skilled solvers are more likely to choose competition over collaboration, thus it is very likely that contest design parameters may also affect the distribution of solver expertise. It would be interesting to explore whether a solver's expertise level impacts their choice of contest, and how the contest design parameters may influence the quality of the solvers that are attracted. For example, will a contest with a higher prize attract better quality solvers?

While our simple performance model suggests that having more solvers is better, this is true only when the number of solvers attained is within the seeker's evaluation capacity. When the costs of evaluating solutions are very high or when the seeker has capacity constraints, the seeker would not be able to evaluate all solutions adequately, and the contest performance may decrease with the number of solvers (Laursen and Salter 2006). In this case, the relationship between the number of solvers and contest performance may have an inverted-U shape and an optimal number of solvers may exist. How to calculate the optimal number of solvers is therefore also an interesting topic for future study.

We have taken into consideration the use of feedback in online contests and have shown that feedback increases the solvers' effort. However, the exact role that feedback plays in a contest requires further analysis. For example, is it better for seekers to send feedback privately or publicly? Will sending feedback bring more or fewer solvers? These are interesting research questions that can be explored further.

While our paper is one of the first in contest literature that considers submission speed, we only consider the time to submission of the first solution. It is possible that the marginal submission speed, which captures the changes of submission speed, is also interesting and informative. Later studies may expand our model to consider the speed to submission of subsequent solutions.

Our observations cover only one geographic market, China, and focus on Chinese solvers. As a result, most of our observations involve a relatively small prize amount, compared to contests on InnoCentive. Future research may consider different markets in different countries. It will also be interesting to see if there are any systematic differences across markets and solvers. The proposed model can be tested with other contest design parameters and markets.

Another interesting question worth exploring is the role of the contest market and how to formulate market policies. Some contest markets, such as TaskCN and Zhubajie, require seekers to deposit the full prize amount in advance and make this prize non-refundable. This policy avoids a moral hazard on the innovation seeker's side, who may simply take submitted ideas without awarding money to a winner. Further, this practice provides a guarantee to solvers that, if they win, they will be paid. However, under this payment policy, seekers must bear all the risk of failure, especially for complex projects that attract too few or only weak solutions. As a result, seekers may not be willing to launch cutting-edge, complex or high value projects under this 'non-refundable prize' policy. Cutting-edge and high value innovation projects are those that are most important to large firms and the most profitable ones to market owners. As such, the formulation of appropriate policies to provide enough insurance to both seekers and solvers will also be an interesting question.

2.9 Conclusion

Firms are always seeking for innovative ideas and better innovation approaches. Open innovation is capturing considerable attention due to its potential to elicit more diverse ideas from sources outside of the firm. As an important approach to open innovation, a well-designed online contest is necessary. Our paper makes a unique contribution to academic research on open innovation contests by providing a feasible performance evaluation method and a practicable framework, which covers the major of aspects that seekers need to consider. We also provide useful guidelines to help innovation seekers design online contests under different circumstances. We believe that open innovation will play an increasingly important role in the R&D process and we hope that this paper can provide academics and practitioners with a better understanding of online open innovation contests. Fundamentally, our hope is that this work aids practitioners in developing the capability to generate more and better outstanding ideas by designing and conducting optimal online contests.

CHAPTER 3

WINNER DETERMINATION[9]

3.1 Introduction

An *open innovation contest* (Terwiesch and Xu 2008; Yang et al. 2009) or *crowdsourcing contest* (Archak and Sundararajan 2009; Howe 2006; Yang et al. 2008), is a contest in which an innovation seeker, which can be a firm, organization or individual, holds a contest to seek innovative ideas or solutions to a problem. These types of contests have begun to be adopted by an increasing number of firms for problem solving and new product development (NPD). By launching online innovation contests, firms can easily access a large volume of external solvers having diversified backgrounds. A larger pool of potential solvers can help facilitate the faster generation and potentially better ideas or solutions, compared to internal innovation efforts. For instance, in September 2008, Google funded the $10M launch of an open innovation contest. The project, called Project 10^100, called for ideas to change the world, in the hopes of making the world better. Since 2006, Netflix has set a $1M prize every year, seeking to substantially improve the accuracy of predictions about the degree to which individuals will enjoy a given movie, based on their movie preferences. In addition to self-facilitated open innovation contests, several markets also exist to facilitate open innovation contests on others' behalf. InnoCentive, founded in 2001, is the first online

---

[9] I co-authored this article with Professor Pei-yu Chen, from the Management Information Systems Department, Fox School of Business, Temple University, and Professor Rajiv Banker, from the Accounting Department, Fox School of Business, Temple University.

market to host open innovation projects in the form of contests (Allio 2004). It was originally built to facilitate the search for innovative medicinal solutions. However, a variety of project types are now posted there, ranging from website LOGO design and algorithm design to complex projects, such as construction design. There also exist other contest markets that have different foci or that target different geographic regions, such as TopCoder and TaskCN. Various organizations and companies are using these markets as platforms for open innovation projects. For firms that lack an IT deployment capacity or channels via which to reach huge numbers of potential solvers, launching an innovation contest in a mature online market is an obviously better choice. In the future, we expect that more and more firms will adopt open innovation to mitigate the risks of internal project failure and to identify exceptional solutions from across the globe.

To launch an online contest, an innovation seeker needs to post project details, including a fixed prize, duration, description, etc. All potential solvers decide whether or not to join the contest and, if they join, they compete by submitting their solutions/products. After the contest ends, the seeker will assign the prize to the solver who provides the best idea/solution, according to some criteria. Several theoretical studies have been performed on contests. Dahan and Mendelson (2001) pursue an extreme value model and argue that the final performance of a product development contest is decided by the top tier distribution of solvers. Terwiesch and Xu (2008) have extended this model to a more general contest situation, where projects may have multiple dimensions, such as being expertise- or ideation-based. If the distribution of solver ability is irrelevant to the project characteristics, having more solvers will give a seeker a higher chance of obtaining better solutions. A typical assumption made in previous theoretical literature is that all solvers receive the same information and compete simultaneously.

However, in most online contest markets (e.g., Zhubajie, TaskCN, 99designs, CrowdDesign, etc.)[10], we have observed that solvers compete dynamically rather than simultaneously. As shown in the typical timeline of an online contest (**Figure 2.1**), a solver can enter the contest at any time between when the contest starts and when it ends. Since an online contest market usually makes the status of a contest public (including the number of submissions and sometimes even the content of final submissions), depending on the time at which the solver enters, they may receive different information and may employ different strategies, accordingly. For example, by waiting longer, a potential solver has more information regarding the number of competing solvers and the quality of submissions that are available. As a result, he[11] has a better understanding about his chance of winning and can take more effective action accordingly. On the other hand, by submitting a good solution early, the solver may discourage other solvers from submitting subsequent solutions. Further, submissions are usually accessible to the public and this policy makes each solver consider when to submit his solutions, in order to maximize his winning probability. Thus, the winning result is also impacted by solvers' bidding behaviors.

---

[10] We consider the scenario that applies to most popular online contest markets. Although InnoCentive receives lots of academic attention, it is not a popular online market. Thus, its business model and policies are not considered in our study.

[11] In economic research, for the convenience of discussion, the buyer is usually referred to as "she" and the seller as "he." Similarly, in this paper, we refer to a seeker as "she" and a solver as "he."

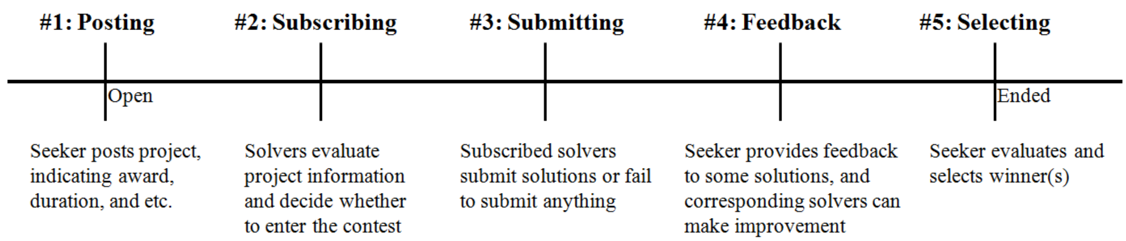


Figure 3.1 Dynamic Timeline of Online Contest

The goal of an online contest is to attract high quality solvers, to obtain good, diverse solutions (Terwiesch and Ulrich 2009; Terwiesch and Xu 2008). This depends not only on contest design parameters but also on an understanding of solvers' incentives and strategic actions. A good contest design should take into account solver behavior, and it is important to understand the strategic interactions between solvers. However, to our knowledge, scant empirical literature exists on contests and there is a lack of understanding of how solvers compete with one another strategically. This research has two goals: first, we aim to identify the factors that influence a solver's chance of winning; second, we seek to evaluate a longstanding assumption in the contest literature, that solvers are drawn from the same distribution, regardless of contest design parameters. That is, the solver distribution is the same across contests (Terwiesch and Xu 2008). In reality, however, there is often more than one contest, especially within a contest market, that a potential solver can participate in. Due to capacity constraints, solvers may self-select into different contests. That is, it is likely that some contests are better able to attract more experienced solvers than others. If this is true, then the aforementioned assumption underlying many theoretical works, which may have stemmed from a lack of scholars' understanding of how solvers compete, is invalid. We are therefore interested in testing whether this assumption is empirically true.

The remainder of this paper is organized as follows. In the next section, we discuss the features of online contest markets that may affect solvers' performance, and we develop hypotheses accordingly. We then describe our data source, variable measurement, and the empirical models used to test our hypotheses. Lastly, we present our empirical results, followed by conclusions and a discussion of implications.

3.2 Theory and Hypotheses

This study takes the perspective of contest solvers. Consider an innovation seeker posting a project on an online contest market, and the subsequent arrival of a number of solvers. To a solver, the question he is most concerned about is likely how to go about maximizing his probability of winning. The quality of the solution from solver i, v_i, where $i = 1 \ldots n$, can be modeled in a linear format (Terwiesch and Xu 2008):

$$v_i(\beta_i, e_i, \xi_i) = \beta_i + r(e_i) + \xi_i \quad , \qquad (3.1)$$

where β_i is the expertise level of solver i. It is usually assumed that the distribution of expertise is known and fixed across all contests. $r(e_i)$ is the output of effort when solver i executes effort e_i. $r(e_i)$ is increasing in e_i. ξ_i is a random error term of each solution. This random error also captures the unobserved preferences of the seekers and includes the diversified ideation-based output. Since β_i is fixed, the variance of a solver's performance is mainly based on the effort output, $r(e_i)$ and random error. As noted earlier, due to the dynamism of the competition process, solvers' bidding strategies may also impact the winning results, thus equation 3.1 can be modified to:

$$v_i(\beta_i, e_i, s_i, \xi_i) = \beta_i + r(e_i) + s_i + \xi_i \quad (3.2)$$

We next explore the impacts of a solver's expertise and bidding strategies on winning results. Effort output, r(e_i), is usually difficult to observe directly, but its impact will be discussed along with related bidding strategies.

3.2.1 Impact of Expertise/Past Experience

Previous studies indicate that an individual's performance is restricted by his expertise (Banker and Iny 2008; Snir and Hitt 2003; Terwiesch and Xu 2008). How to define expertise within an online contest scenario in a quantifiable manner is a key task for this empirical study. In the organizational literature, there are two views of expertise. The first view is based on the knowledge possessed by an individual (Rorty 1979). In this view, knowledge is a material that can be abstracted, explicitly represented, codified, and accessed (Walsh 1995). In online contest, free entry attracts solvers with diverse backgrounds. Following this view, Terwiesch and Xu (2008) suggest that expertise is usually a measure of a solver's past experience and knowledge for a particular problem. For example, the winner of a LOGO design contest is more likely to be a designer than a chemist. However, it is hard to define the diversity of expertise for a crowd of solvers, because one area of knowledge is not comparable to another. Another perspective views expertise as context dependent, emerging from patterned interactions and practices in specific scenarios (Brown and Duguid 1991; Faraj and Sproull 2000; Snir and Hitt 2003). Following this view, the expertise of a solver can be reflected by his emergent and relative contest performance results. This can be done by comparing the past behavior of one solver with that of other solvers in the same community (e.g., participation experience and past performance).

In the reverse auction literature, past experiences have proven to be good predictors of future winners (Banker and Iny 2008). In reverse auctions, agents offer bids on project proposals as signals of final solutions. If an agent is experienced, with a high propensity for winning and a lot of positive feedback, this indicates that this agent is consistently good at negotiating and maximizing the principal's utility by sending the "right signals." In online contests, solvers bid with their final solutions. In other words, solvers are bidding with their skills and capabilities; they are not negotiating. A seeker with better performance discloses his consistently better expertise. On one hand, we could conclude that this solver has a higher expertise level, and on the other hand, we might predict that this solver will have a higher probability of winning. Terwiesch and Ulrich (2009) have done an experiment with MBA students, providing evidence that the quality of submissions for a specific solver will be consistent over time. Therefore, we expect that the expertise level is positively associated with winning probability in the future. This leads us to the following hypothesis:

H1: *A solver with a higher expertise level is more likely to be the winner.*

3.2.2 Temporal Strategy

Internet enabled initiatives allow geographically distributed players to collaborate or compete with one another. Compared to traditional or offline contests, one distinguishing feature is that online contests allow players to compete dynamically, instead of simultaneously. The decision of when to enter and bid is a common concern for all solvers. A long stream of literature has sought to understand auction bidders' dynamic bidding strategies and the associated benefits. Notably, late bidding in Internet auctions has attracted a good deal of attention (Ockenfels and Roth 2001; Ockenfels and

Roth 2006; Roth and Ockenfels 2002; Vadovic 2009). The intuition behind last-minute bidding in a private value auction is that there is an incentive not to bid high when there is still time for other bidders to react. This allows one to avoid an early bidding war that will raise the expected final transaction price (Ockenfels and Roth 2001). On the opposite side of the coin, an early and high bid can lower a bidder's cost of searching for substitutions, while simultaneously making other competitors less interested in in competing (Vadovic 2009). Unfortunately, little attention has been paid to strategic bidding in online contests. In an online contest with a project search tool, a solver may strategically choose when to enter and when to submit his solution, as long as the contest has not ended. From a solver's perspective, his greatest interest is in determining the strategy that maximizes his probability of winning. To achieve this, we first need to understand all the benefits and drawbacks of early and late entry, in order to identify the dominant strategy or equilibria in this dynamic game.

The term '*benefits*' here refers to any positive impacts on the goal of winning, while '*drawbacks*' refers to the opposite impacts. Generally speaking, by entering early, the solvers have the option of choosing when to submit the initial or first solution[12], which, at least, is better than having no options. So, entering earlier is never a bad choice. Although entering early requires the solver to be more patient in waiting for the result, this strategy has no negative impact on the probability of winning. By submitting the solution early, the solver has the advantage of getting feedback from the seeker earlier and still having time to implement changes accordingly. Successfully addressing the

---

[12] In this study, we only consider each solver's first submission. The improved or sequential solutions submitted by the same solver will not be counted.

feedback communicated by the seeker will undoubtedly increase the solver's probability of winning. In addition, by submitting a good solution earlier, the solver may also discourage other solvers from competing other solvers may perceive that they have a lower winning probability and therefore may drop out. On the other hand, by purposely waiting, the solver has more information about the number of competing solvers and the quality of submissions. More information helps the solver to better evaluate their winning probability, and the solver can take action accordingly. For example, a solver will pursue the project further if he perceives that he has a high probability of winning. Conversely, the solver will drop out if he perceives that he has a low winning probability. This suggests that a solver will incur costs only when he perceives a high probability of winning, reducing the expected cost requirement. However, a solver may also submit late as a direct consequence of entering the contest late. Therefore, the time lapse between entering time and submission time may reveal some strategic information. Specifically, a solver who purposely waits will have a relatively higher time lapse compared to solvers who submit late because of entering the contest late.

Let us now take a simplified participation process as an example. Let us assume that a solver can only choose to enter or submit at one of two moments, termed *early* and *late*. After having entered, he needs to further decide when to submit. If he enters early, he can choose to submit early or late. If he enters late, he can only submit late. In total there are three strategies of participation. We summarize the benefits of each strategy in **Table 3.1**.

Table 3.1 Temporal Strategy Analysis of Solvers

		Submit Early		Submit Late
Enter Early	I	✓ More likely to receive feedback and make improvement ✓ Stop others to enter by submitting a high quality solution ✓ More time to work on solution ✓ More chances to learn from competing solutions. × More likely to have similar competing solutions	II	✓ Less likely to be copied by others ✓ More time to work on solution ✓ More chances to learn from competing solutions. × Less likely to receive feedback
Enter Late			III	✓ Less likely to be copied by others × Less likely to receive feedback × Lost benefits of entering early

Note: ✓- benefit × - drawback

By comparing strategy II to strategy III, it is obvious that strategy II has more benefits than strategy III as it gives the solver more time to work on solutions and more opportunities to learn from competing solutions. In other words, strategy II dominates strategy III.

Comparing strategy I to strategy II, after removing all the common benefits, we can see that each still has some unique benefits. A strategy I solver is more likely to receive feedback and make the improvements desired by the seeker; he may also prevent others from entering by submitting a high quality solution. A strategy II solver is more likely to submit a unique solution and make the solution more competitive. From a solver's perspective, it is not clear which strategy is dominant. Since there are no dominant strategies in a dichotomous contest, it is not surprising that no single, unique equilibrium exists.

The above analysis is based on a simplified situation where only two points of contest entry and solution submission are available. In a real contest, solvers can submit at any moment while the contest is open. At any moment, each solver has the benefit of submitting earlier than some solvers and later than others. We assume that the benefit of submitting early, $P_{submit\ early}$, at moment *t*, is convex and monotonically decreasing (i.e., $P_{submit\ early}\ |_t = C_1 + C_2 \exp(-C_3 t)$, where C_1, C_2, and C_3 are constants and C_2, C_3 >0). When a contest has just launched, $P_{submit\ early}$ is at its highest. When a contest is just ending, $P_{submit\ early}$ is at its lowest. Similarly, we also assume that the benefit of submitting late, $P_{submit\ late}$, at moment *t*, is convex and monotonically increasing (i.e., $P_{submit\ late}\ |_t = C_4 + C_5 \exp(C_6 t)$, where C_4, C_5, and C_6 are constants and C_5, C_6>0). When a contest has just launched, $P_{submit\ late}$ is at its lowest. When a contest is just ending, $P_{submit\ late}$ is at its highest. The overall benefit of submitting at moment *t* is $P|_t = P_{submit\ early} + P_{submit\ late}$, which is convex and non-monotonic. In other words, either submitting very early or very late may increase the chance of winning, over those solutions submitted in the intervening period. Thus we have:

H2a: *the probability of winning follows a U shape in terms of solvers' relative submission order.*

In a dichotomous contest, entering early is always better than entering late, and this result should still hold in a real contest, where, after a submission is observed, we are also concerned about the interaction between the entering time and submitting time. To capture the entering strategy and its interaction with the submitting strategy, the lapse in time between the entering and submission is more interesting. This time lapse reveals not only how early a solver entered, but also the strategic waiting that took place before the final submission. On the other hand, it may also be a noisy measure of effort output r(e_i) in equation 3.2. Regardless of whether this value captures strategic waiting or higher effort, we expect that a solver's chance of winning is increasing in this time lapse. We therefore have:

H2b: *A longer time lapse between entering and submission will be associated with a higher chance of winning.*

3.2.3 Strategic Choice of Projects

As noted earlier, another question of interest to us is whether solvers with different expertise levels systematically choose different projects. For example, do solvers with higher expertise tend to choose higher-prize projects? In order to simplify the theoretical model, many previous studies assume that expertise has no impact on choice. However Terweisch and Xu (2008) doubt the validity of this common assumption. In an online reverse auction market, Snir and Hitt (2003) find that high value projects attract pools of solvers with lower quality, on average. Does a contest market exhibit this same phenomenon? An experienced solver should have learned how to increase his profit by choosing the "right" projects. If so, this would suggest some self-selection bias and invalidate the common assumption held by previous studies that expertise distributions are constant across different projects. To test the validity of the above common assumption, we present the following null hypothesis:

H3: *A solver's expertise does not affect his choice of projects.*

3.3 Data and Estimation Model

In this section, we introduce our research site, sample selection, sample data, measurement of constructs, and estimation model.

3.3.1 Research Site and Sample Selection

We collect data from TaskCN.com, which is one of the largest online service markets in China, founded in 2005. By the end of 2009, TaskCN had over 2.7 million registered solvers. This market allows anyone to launch a contest with an advance prize deposit. All contests are free to enter. It is important to note that TaskCN employs the same contest model that has been adopted by most other online contest markets, such as Zhubajie, TopCoder, DesignCrowd, etc.

We collected data on all contests launched between June 2006 and October 2008. In total, there are 7,728 contest projects. Around 20% of the projects are multi-winner projects. We eliminated these projects since the role of prize structure design is not well understood and this is not the focus of the present study. Sales force contests were also eliminated because these contests seek to maximize the overall performance of all solutions, rather than to identify one or several best solutions (Liu et al. 2007; Terwiesch and Xu 2008). After dropping these observations, our sample includes 1,995 contests with 216,812 submissions.

3.3.2 Sample Data and Measurement

Expertise Variables

Following the perspective that expertise is context dependent, emerging from patterned interactions and practices in specific scenarios (Brown and Duguid 1991; Faraj and Sproull 2000) in the online contest market, the expertise of a solver can be reflected by his emergent and relative contest performance results. We do this by comparing one solver with other solvers in the same community in terms of their participation experience and past performance. In practice, it is hard to define expertise objectively. In our study, a solver's participation experience and past performance can be measured by[13]: the number of contests in which he has participated, *N_Join*, his number of wins, *N_Win*, membership age *MembershipAge* and winning propensity *WinPerJoin*, prior to entering a

[13] In auction studies, a feedback rating is a key indicator of expertise (Snir and Hitt 2003; Banker and Iny 2008). However, in a contest, the feedback rating does not help, since nearly all winners have received positive feedback. This is because winners are chosen based on final performance and there is no information problem. A winner is always relatively better than other competitors. Our data shows that 98% of feedback in this market is positive, with nearly no variance.

new contest. Since *WinPerJoin = N_Win / N_Join*, we can only include two of the three variables in our analysis at the same time. WinPerJoin is an ability measure, which is unique and critical. Although participation experience is important, the winning experience can help a solver understand how to win, so we keep N_Win. Some solvers have a very long membership age, but these extremely high levels of experience are not likely to have a proportional impact on winning probability. Hence, we take the natural log of membership age to reflect the diminishing marginal benefit of increased experience. In summary, in our estimation model, we use *N_Win*, Ln*(MembershipAge)* and *WinPerJoin* to proxy for expertise.

Table 3.2 Expertise Variables Definition

Variable	Definition
N_Win	Total number of wins before current contest.
N_Join	Total number of entered contests before current contest
WinPerJoin	N_Win / N_Join, winning propensity
MembershipAge	Number of days between registration and entering current contest

Table 3.3 Descriptive Statistics of Expertise Variables

Variable	Mean	Std Dev.	Min	Max
N_Join	19.992	66.933	1	1368
N_Win	1.032	5.916	0	163
MembershipAge (Days)	85.4778	142.31	0.0001	813.16
WinPerJoin	0.018	0.070	0	1
No. of Observations		216,812		

Temporal Strategy Variables

The strategies that a solver can fully control are limited. We list all temporal variables in **Table 3.4**. As discussed before, the greatest concern for a solver is when to join the contest *(JoinTime*) and when to submit his solution (*SubmitTime).* In order to perform a comparison across contests with different durations, we standardize these time values with the following equation:

$$G\_SubmitTime = \frac{SubmitTime - StartTime}{Duration} , \tag{3.3}$$

where *G_SubmitTime* is a value between 0 and 1. If *G_SubmitTime=1*, this means that the solver submitted just before the project closed. Using the same method, we calculate *G_JoinTime* and *G_ Δ Time. G_JoinTime* measures the standardized time at which a solver enters a contest. *G_ Δ Time* measures the standardized differences in time lapse between *JoinTime* and *SubmitTime*. *G_ Δ Time* captures not only when to submit, but also the interaction with submission time. So, this value is more informative than *G_JoinTime.*

However, the standardized time value may not reveal a solver's temporal strategy. In practice, if there is only one solver in pool, no matter when he submits, he is always the first and the last to submit. In this circumstance, *JoinTime* and *SubmitTime* are useless in winner determination. To reflect the variation in competition strategy of relatively "early" or "late" submissions, the submission order of a solver (*SubmitOrder*) is more important. We define *G_SubmitOrder as*:

$$G\_SubmitOrder = \frac{SubmitOrder}{N\_Submit} . \tag{3.4}$$

Table 3.4 Strategy Variables Definition

Variable	Definition
JoinTime	The time when a solver declares that he will join the competition
G_JoinTime	(JoinTime-StartTime)/Project Duration
JoinOrder	The number of extant solvers when this solver joins
G_JoinOrder	JoinOrder/N_Submit
SubmitTime	The time when a solver submits his first solution to current contest
G_submitTime	(SubmitTime-StartTime)/Project Duration
SubmitOrder	The number of extant submissions when this solver submits.
G_SubmitOrder	SubmitOrder/N_submit
ΔTime	Submit Time – Join Time
G_ΔTime	ΔTime/Duration

Table 3.5 Descriptive Statistics of Strategy Variables

Variable	Mean	Std Dev.	Min	Max
G_SubmitOrder	0.500	0.284	0	1
G_ΔTime	0.051	0.131	0	0.9995
Number of Observations		216,812		

G_SubmitOrder is also a ratio between 0 and 1. If *G_SubmitOrder=1,* this means that this solver was the last to submit. Using the same method, we have *G_JoinOrder*, which measures the relative time until a solver enters a contest.

To test H2a and H2b, *G_SubmitOrder* and *G_ΔTime* are needed. Descriptive statistics of these two variables are listed in **Table 3.5**.

The correlation coefficients among all variables necessary for the estimation model are listed in **Table 3.6**.

Table 3.6 Correlation Coefficients of Participation Variables

	N_Win	WinPerJoin	Ln(MembershipAge)	G_ΔTime	G_SubmitOrder
N_Win	1.000				
WinPerJoin	0.237	1.000			
Ln(MembershipAge)	0.216	0.165	1.000		
G_ΔTime	0.051	0.056	0.095	1.000	
G_SubmitOrder	- 0.035	- 0.022	- 0.016	0.282	1.000

Project Group Variables

Table 3.7 Descriptive Statistics of Fixed Project Variables

Variable	Mean	Std Dev.	Max	Min
Number of Solver	111.67	319.94	4498	1
Prize Amount (￥)	287.15	361.952	5500	1
Contest Duration (days)	21.28	15.81	104.11	0.0021
Number of Projects			1995	

Table 3.8 Category Means and Features of Projects

Category	Prize(￥)	Duration	No. Solvers	Percentage	Dimensions
Web Building	285.01	20.81	13.53	3.3%	Ideation & Expertise
Translation	67.58	13.37	118.78	1.40%	Expertise
Software Develop	126.84	17.32	8.31	5.45%	Expertise
Q&A	61.70	16.67	22.48	5.15%	Q&A
Media	316.89	16.05	15.49	3.25%	Ideation & Expertise
Web Design	483.57	21.66	18.19	6.65%	Ideation & Expertise
Product Design	545.55	25.32	18.87	5.15%	Ideation & Expertise
Packaging Design	439.73	30.43	31.23	1.9%	Ideation & Expertise
LOGO Design	358.66	22.59	60.56	40.5%	Ideation & Expertise
Interior Design	245.17	17.24	21.46	0.85%	Ideation & Expertise
Graphic Design	288.35	19.56	38.3	13.5%	Ideation & Expertise
Creative Writing	159.21	19.74	71.34	7.5%	Ideation
Naming	136.28	25.32	729.22	9.6%	Ideation

Contest participation history can be grouped by contest projects. There are several project group variables that are fixed for each submission within a project. These variables may impact a solver's perceived probability of winning. These variables include prize amount, contest duration, category, dimensions, etc. Descriptive statistics for these fixed project variables and categorical means are presented in **Table 3.7** and **Table 3.8**.

3.4 Estimation Model

To estimate which kinds of solvers tend to win, and to test H1 and H2, we employ the conditional logit model (Wooldridge 2001, Greene 2002), which is commonly employed in consumer choice research, particularly when the number of alternative choices is large. For an innovation seeker addressing contest project *i*, the value of her utility function from choosing solver *j*'s submission is:

$$U_{ij}^{*} = X_j \boldsymbol{\beta} + \xi_{ij} \quad j = 1,2,\ldots N_i \tag{3.5}$$

where ξ_{ij} accounts for all unobservable attributes that can affect a seeker's preferences. The value provided by solver *j* is determined by X_j, which include solver *j*'s expertise and temporal strategy, with a common parameter vector, $\boldsymbol{\beta}$. Although we cannot observe and measure the utility of each submission, we can conclude that the winning submission offers the highest utility to the seeker. That is, if the seeker ultimately chooses the submission made by solver w_i ($w_i \in \{1,2,\ldots N_i\}$), then we can assume that the utility from that submission $U_{ij} \mid j = w_i$ is the maximum among all N_i utilities. When N_i disturbances, ξ_{ij}, are independently distributed with the Type I extreme value distribution (McFadden 1974), then the probability that w_i is the only winning submission for project *i* is:

$$prob(W_i = w_i) = \frac{\exp X_j \boldsymbol{\beta}}{\sum_{j=1}^{N_i} \exp X_j \boldsymbol{\beta}}, \quad w_i \in \{1,2,...N_i\} \cdot \tag{3.6}$$

To estimate the coefficient vector, $\boldsymbol{\beta}$, we use the maximum likelihood criterion. Variable Win_{ik} is 1 if solver k is the winner, while $Win_{ik} = 0$ for all non-winners. We estimate $\boldsymbol{\beta}$ by maximizing the log-likelihood function:

$$\ln L = \ln \prod_{i=1}^{1995} (W_i = w_i) = \sum_{i=1}^{1995} \sum_{k=1}^{N_i} Win_{ik} \ln \Pr ob(W_i = w_k) \tag{3.7}$$

Estimation model:

$$\begin{aligned} U_{ij}^* &= \beta_{1a}(N\_Wins) + \beta_{1b}(MembershipAge) + \beta_{1c}(WinPerJoin) \\ &\quad + \beta_{1d}(G\_SubmitOrder) + \beta_{1e}(G\_SubmitOrder)^2 + \beta_{1f}(G\_\Delta Time) + \xi \end{aligned}$$

3.5 Results and Analysis

Table 3.9 Conditional Logit Model and Results

Variable	Symbol	Coefficient	Significance	Odds Ratio
N_Win	β_{1a}	1.0442***	<0.001	2.8413
MembershipAge	β_{1b}	- 0.9642	0.219	0.9080
WinPerJoin	β_{1c}	0.1794	0.222	1.1965
G_SubmitOrder	β_{1d}	- 2.0205***	<0.001	0.1325
G_SubmitOrder^2	β_{1e}	2.7364***	<0.001	15.4313
G_ΔTime	β_{1f}	1.9991***	<0.001	7.3824
No. of observations	21,6812 observations grouped in 1995 cases			
Pseudo R^2			8.84%	
LR Chi^2	1671.91*** (Prob>Chi2 is less than 0.0001)			

Note: *~p<0.1, **~p<0.05, ***~p<0.01. Expertise variables have been transformed to percentile scores

Table 3.9 reports the results of which solvers are likely to be the winner of a given contest, based on a conditional logit model. The Pearson correlation coefficients between each term are all lower than 0.3 (**Table 3.4, Table 3.5 and Table 3.6**).

Since we are using a conditional logit model, the influence of each factor needs to be explained in terms of its significance and odds ratio. The above results show that, in terms of our selected expertise variables, winning experience is a good predictor of future wins. The odds ratio of N_Win is larger than 1, so a solver with more winning experience is more likely to win in the future. As such, H1 is supported.

Regarding the temporal strategy, the coefficients on *G_SubmitOrder* and *G_SubmitOrder*^{2} suggest a U-shaped curve (**Figure 3.2**) in winning probability. In other words, very early submissions and very late submissions have a higher probability of winning, compared to those submitted over the intervening period. Therefore, H2a is supported.

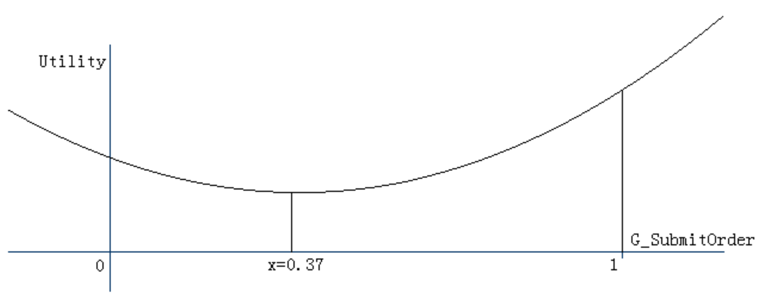


Figure 3.2 Estimated Utility – U shape

Further, a larger *G_ΔTime* is significantly associated with a higher probability of winning. Thus, H2b is also supported.

We should note that one should be cautious in interpreting the results for temporal strategy. Specifically, this result does not imply that submitting earlier or later can definitively increase the probability of winning. The inherent incentive to submit early is that a solver can receive feedback earlier, can discourage other solvers from further pursuing the project, or can learn more from competing submissions to make his own submission more competitive. If a solver simply submits first or last, without concern for the seeker's feedback or without attempting to learn from competing submissions, his probability of winning will not increase at all. However, our finding does suggest that more capable solvers may self-select into submitting earlier or later. For example, given that a late submitter will have observed all prior submissions and that he would likely not bother to submit if he did not believe he could beat others, those who submit late are likely to be of high quality and perceive that they have a high winning probability. On the other hand, a capable solver also has a lot to gain by submitting early because he can discourage other solvers from participating further, while taking advantage of likely feedback to further improve his solution. This has some interesting implications for seekers. For example, when evaluating submissions is costly, a seeker may want to focus more on earlier and later submissions than those submitted over the intervening period.

Given that the project characteristics are fixed within each contest, it is meaningless to add these variables into the conditional logit regression. In order to know how the impact of these features will vary for different projects, we therefore explore sectional regression results for different project categories, dimensions, prize values and durations. The winning ratio and membership age are not reported in this analysis, due to insignificance. Income is also removed from this regression, since it is of little interest here.

Table 3.10 Sectional Results by Category

Variables	LOGO Design (647 Cases)	Graphic Design (235 Cases)	Naming (163 Cases)	Software (82 Cases)
N_Win	0.0332***	0.0168*	0.0075	- 0.0182
G_SubmitOrder	0.7780	- 1.3345	- 5.6585***	1.7127
G_SubmitOrder^2	0.1653	2.2804**	5.5533***	- 0.9925
G_ΔTime	2.5350***	2.2611***	1.8647***	4.2145***
Pseudo R^2	13.33%	9.73%	3.32%	16.26%

Note: *~p<0.1, **~p<0.05, ***~p<0.001.

Table 3.11 Sectional Results by Dimension

Variables	Ideation (342 Cases)	Expertise (137 Cases)	Ideation & Expertise (1413 Cases)	Q&A (93 Cases)
N_Win	0.0030	- 0.0186	0.0207***	- 0.0046
G_SubmitOrder	- 4.4475***	- 1.0155	0.0137	- 3.9954**
G_SubmitOrder^2	4.6086***	0.9489	1.0332**	2.5395*
G_ΔTime	1.8872***	4.0297***	2.5415***	2.9361**
Pseudo R^2	3.38%	11.87%	11.62%	8.07%

Note: *~p<0.1, **~p<0.05, ***~p<0.01.

The results reported in **Table 3.10** show that the submitting strategy has no significant impact on the probability winning in a LOGO design or software development contest, nor does it have a significant impact in any expertise-based projects. However, this factor has a very significant impact in naming or pure ideation-based projects. For ideation-based projects, the probability of winning, based on submitting order, is a symmetric U shape. In other words, the first submission and last submission have an equal chance of winning, and this chance is higher than that of other submissions. However, in **Table 3.11**, *G_SubmitOrder* is not significant, while $G\_SubmitOrder^2$ shows a positive and significant impact. So, for ideation- and expertise-based projects, winners are more likely to emerge from those solvers who submit late. A longer $G\_\Delta Time$ is always helpful, for all kinds of projects. Expertise variables also appear to work differently for different categories. For software projects or purely expertise-based projects, previous performance does not appear to help. This is surprising since the winning result is assumed to be consistent with the expertise-level of the solver. The influence of both expertise and strategic variables is quite different from category to category, so project category, as a project characteristic, can moderate expertise and the impact of strategic on the probability of winning.

To understand how variation in the contest prize and contest duration moderates the above impacts, we split all projects into several sections: higher or lower than average prizes, longer or shorter than average durations (**Table 3.12**). The results show that when the prize is higher, expertise variables play a more important role in predicting winners. Further, winners are more likely to be those who submit late. When the duration is shorter, expertise plays a less important role. However, again, those who submit later are more likely to be winners. In other words, for high prize, short duration projects, learning from other submissions is more beneficial.

Table 3.12 Sectional Results by Prize or Duration

Variables	All 1995 Cases	High Prize Prize>287	Low Prize Prize<287	Long Duration Duration>21	Short Duration Duration<21
N_Win	0.0108***	0.0268***	0.0034	0.0153***	0.0079**
G_SubmitOrder	- 1.3456**	- 0.4769	- 1.7922***	- 2.1892***	- 0.7137
G_SubmitOrder^2	1.9449***	1.5790**	2.1074***	2.7392***	1.3767**
G_ΔTime	2.6157***	2.3481***	2.7777***	2.7633***	2.5140***
Pseudo R^2	8.58%	11.77%	7.08%	8.29%	8.98%

Note: *~p<0.1, **~p<0.05, ***~p<0.01.

To test H3, whether different contest designs can capture different groups of solvers, we calculate the correlation between contest designs (prize and duration) and all available expertise signals. In general, we calculate the correlations between the average expertise of project solvers and contest designs. Besides this, we also check the correlations according to different tiers of solvers. Since project dimensions may impact the solver distribution, we split the correlation results into three sections: those for ideation-based projects, expertise-based project, and ideation- and expertise-based projects, according to the dimensions defined in **Table 3.11**.

There are two additional signals we use in Error! Reference source not found.Error! Reference source not found.: PrizeWon and PrizeWon/PrizeJoined. PrizeWon is a measure of how much prize money a solver has won in the past, which is a measure of experience and which is also the “official” expertise indicator used by TaskCN. PrizeWon/PrizeJoined is another measure of winning propensity that is dependent on prize winnings. PrizeJoined represents the cumulative prizes for all projects that a solver has competed for. In total, we have four experience based signals: N_Join, N_win, PrizeWon, and Membership Age. Further, we have two winning propensity signals: PrizeWon and (PrizeWon/PrizeJoined).

As an example, to interpret the correlation in Error! Reference source not found.Error! Reference source not found., the correlation between solvers' average N_Join and contest duration is negative (ρ = - 0.167) and significant (p<0.01) for ideation-based projects. This means that for longer duration contest, the average N_Join of attracted solvers is significantly lower. A similar result also holds for middle-tier solvers because the correlation is also negative and significant. However, for top- or bottom-tier solvers, variation in duration does not show a significant association with N_Join.

In general, the results show that for ideation-based projects (e.g. naming projects), the distributions of experience-based signals and winning propensity signals are associated with variations in contest prize or duration. Larger prizes and longer duration also usually result in the attraction of solvers with less experience and lower winning propensities. A similar influence also holds for middle-tier solvers and top-tier solvers. Although this might not fit the expectations of innovation seekers, since ideation-based innovation is a highly random process that does not require specific expertise (Terwiesch and Xu 2008), the negative effect on individual performance is not large.

For expertise-based projects (e.g. software development), it is surprising that the prize has nearly no significant correlation with variation in expertise, on average, for any tier. Only middle- and top-tier solvers' winning propensities are negatively and significantly associated with variations in contest duration. So, for expertise-based projects, the assumption of a constant expertise distribution still appears to hold if the duration is fixed.

Table 3.13 Correlations between Contest Designs and Expertise Signals

Expertise Signals		Ideation Based Projects				Expertise Based Projects				Ideation & Expertise Based Projects			
		Average	Bot-tier	Mid-tier	Top-tier	Average	Bot-tier	Mid-tier	Top-tier	Average	Bot-tier	Mid-tier	Top-tier
N_Join	A	-0.045	-0.019	-0.035	-0.035	-0.091	-0.057	-0.062	-0.136	-0.026	-0.032	-0.037	0.040
	D	**-0.167*** **	-0.075	**-0.093***	-0.013	0.108	0.066	0.064	0.067	**-0.054** **	-0.029	**-0.050***	**0.044***
N_Win	A	-0.076	-0.017	-0.021	-0.085	-0.099	-0.042	-0.058	**-0.149***	**-0.077*** **	-0.032	**-0.044***	0.006
	D	**-0.200*** **	-0.071	-0.087	**-0.173** **	0.087	0.058	0.059	0.070	**-0.080*** **	-0.032	-0.041	0.040
Award Won	A	**-0.119** **	-0.017	-0.022	**-0.095***	-0.021	-0.001	-0.035	-0.031	-0.039	-0.021	-0.027	**0.053** **
	D	**-0.178*** **	-0.071	-0.081	-0.081	-0.058	0.036	0.059	-0.052	**-0.052** **	-0.010	-0.031	**0.058** **
Membership	A	0.082	-0.026	-0.001	0.147	-0.072	-0.081	-0.061	-0.044	0.013	-0.055	0.041	0.025
Age	D	**-0.110** **	-0.097	**-0.097***	0.132	-0.145	0.028	0.036	-0.006	-0.040	-0.058	-0.030	-0.031
WinPerJoin	A	**-0.122** **	-0.017	**-0.096***	-0.012	-0.011	0.011	-0.027	-0.037	**-0.110*** **	**-0.061** **	**-0.101*** **	**0.069*** **
	D	**-0.246*** **	-0.071	**-0.180*** **	0.079	**-0.230*** **	-0.092	**-0.165***	**-0.170** **	**-0.187*** **	**-0.081*** **	**-0.142*** **	0.033
AwardWon/	A	**-0.147*** **	-0.016	**-0.077***	0.017	0.003	0.013	-0.003	-0.019	**-0.092*** **	**-0.054** **	**-0.077*** **	**0.053** **
AwardJoined	D	**-0.247*** **	-0.071	**-0.150*** **	0.093	**-0.205** **	-0.066	**-0.140***	**-0.141***	**-0.171*** **	**-0.076*** **	**-0.119*** **	0.024

Note: *~p<0.1, *~p<0.05, ***~p<0.01, A~Award D~ Duration

For ideation- and expertise-based projects (e.g. graphic design), the situation is similar to what we get for ideation-based projects, where both experience-based and winning propensity signals of expertise are negatively and significantly correlated with variations in prize or duration. In addition, the winning propensity of bottom-tier solvers is also influenced. However, in contrast to the other two sections, the correlations between prize/duration and expertise for top-tier solvers are positive and significant. This means that, for larger prizes or longer duration contests, in this section, although the attracted solvers have lower levels of expertise, on average, the top-tier solvers have relatively higher levels of expertise. If the performance of contests is mainly decided by the top-tier distribution (e.g. extreme value model), it is still beneficial for seekers to increase the size of the prize or the contest duration.

Overall, our finding rejects H3 and suggests that the assumption of a fixed expertise distribution is not always valid for online contest markets.

3.6 Discussion and Conclusion

Our study provides empirical evidence that in an online contest market, both a solver's expertise and strategy are associated with his probability of winning. The number of past contest wins for a solver's is a good predictor of his future winning probability. Specifically, a solver with more winning experiences in the past will have a higher probability of winning in the future. In an open environment, solvers who submit earlier have the benefit of receiving earlier feedback from seekers, while later submitters have the benefit of learning more from competing submissions. Our results confirm that those solvers who make early submissions and late submissions are more likely to be winners than those who submit solutions over the intervening period. We also show that those solvers who purposely "wait" to make a submission have higher probability of winning.

We also find that project characteristics can moderate the impact of expertise and strategy on the probability of winning. For naming projects, a higher prize and shorter duration will make the U-shaped probability curve flatter. When the contest prize is smaller or the duration is longer, strategic waiting does not do much to increase the probability of winning.

Finally, we find that the distribution of solvers is not constant across different projects. Solvers with higher levels of prior participation and winning are not necessarily more likely to choose higher-prize projects. In fact, for ideation-based projects, the results show a trend where solvers with more experience choosing projects with relatively smaller prizes. Increasing the prize is a common strategy that seekers currently use in order to attract better solutions. Yet in general, our results show that, by doing so, seekers may actually reach lower quality solvers, on average. However, for the most popular projects; those that require ideation and expertise, larger prizes and longer durations can still help to capture better top-tier solvers. If the overall performance of the contest is decided by the top-tier solver distribution, increasing the size of the prize or extending the contest duration is still beneficial.

In general, the reason that longer durations result in the attraction of lower quality solvers remains unknown. As such, it is unclear whether extending the contest duration is beneficial, though it does seem that an optimal duration exists. How to explain this phenomenon is left for future work.

In this study, while we have identified the importance of relative submission orders for solvers, it is still not clear how the submission order is related to the general submission time. From an innovation seeker's point of view, knowing under what conditions she will receive the most, best solutions is of greatest concern. It would therefore be an interesting future study to explore factors that expedite the arrival of good solutions. Moreover, our results have provided evidence that different projects, even within the same category, attract solvers with different expertise levels. Therefore, an outstanding question remains for future consideration around how to design a project that can attract better quality solvers, given that this is a strategic concern for seekers.

Finally, the validity of our findings is dependent on our specific contest scenario. If the fixed-prize policy or the visibility of submissions were to change, our results would need to be revisited. It might therefore be interesting to employ game theory in tandem with the bidding strategies and mechanisms of auction theory to explain the problems presented in this study.

CHAPTER 4

OPEN EVALUATION FOR OPEN INNOVATION[14]

4.1 Introduction

Innovation is key for a firm to compete and survive in an intensely competitive market. Enhancing innovation performance is at the core of both industrial practice and academic research. Yet, due to resource constraints, it is difficult to greatly improve innovation performance (Chesbrough 2003). Facilitated by the growth of the global Internet and emerging forms of information technology, open innovation (Chesbrough 2003; Hippel 2005; Terwiesch and Ulrich 2009; Terwiesch and Xu 2008) aims to overcome these limitations, thus it has received much attention in recent years. Open innovation posits that firms can actively seek innovative opportunities from the external world, without exhausting their limited internal resources. An example of open innovation strategy is the use of *open innovation contests* or *crowdsourcing contest* (Archak and Sundararajan 2009; Terwiesch and Ulrich 2009; Terwiesch and Xu 2008; Yang et al. 2009), which are launched online by various companies to explore innovative solutions submitted by outsiders.

In an open innovation contest, the innovation seeker (e.g. a firm, an organization or an individual) runs a contest to seek innovative ideas or solutions from external solvers for a specific problem. By employing the collected/collective intelligence of a large pool

---

[14] I co-authored this article with Professor Pei-yu Chen, from the Management Information Systems Department, Fox School of Business, Temple University, and Professor Yuexiao Dong, from the Statistics Department, Fox School of Business, Temple University.

of external solvers, the seeker may achieve faster, more diversified and potentially better ideas or solutions, compared with internal innovation efforts (Bonabeau 2009; Terwiesch and Ulrich 2009). Many firms have successfully adopted open innovation for the purposes of problem solving and new product development. Firms of almost any size can apply open innovation. Large IT firms, such as Google and Netflix, can easily launch self-hosted innovation contest. For firms lacking an IT deployment capacity or channels via which they can reach huge numbers of potential solvers, launching an innovation contests via an established online market is a wise choice. With the extremely low cost of launch, participation and communication via the global Internet, online contests are becoming the most popular choice for open innovation.

A successful innovation project needs to not only generate lots of good ideas or solutions but also needs to identify those solutions that are "exceptional" (Bonabeau 2009; Terwiesch and Ulrich 2009). Experts are typically employed to evaluate candidate innovation solutions, using an internal process. For instance, in order to identify the best solutions amongst portfolio management techniques (Chao and Kavadias 2008; Terwiesch and Ulrich 2009), each solution is scored by several experts according to a set of evaluation criteria that are aligned with the ultimate innovation goal. The solutions with the highest scores are then selected and implemented.

As a resource-intensive process, the evaluation of candidate solutions requires evaluator expertise, labor, time and budget (Rossi et al. 2004). Internal evaluation is efficient only when the number of solutions is small, with a low marginal evaluation cost. Open innovation contests can easily generate a large volume of solutions from the external world, and the cumulative evaluation cost can be tremendous, even with very low marginal evaluation costs. Google’s Project 10^100 is a good case to demonstrate this situation. In September 2008, Google funded the $10M launch of Project 10^100, an open innovation contest that solicited ideas on how to change the world, in the hopes of helping as many people as possible. Before the project was launched, Google planned to finish the evaluation process within 3 month. After two months, Google had received over 154,000 ideas in 25 languages. Due to the excessive number of solutions, Google had to delay the delivery repeatedly (Berndt 2009; Google 2009). In an apology letter sent to all participants, Andy Berndt, the managing director of Google’s Creative Lab, stated: *“We've never managed a project like this and it's taken more time than we ever imagined possible.”*

Finally, Google had to put a team of 3,000 employees and spend eight additional months to evaluate all ideas. While open innovation contests can help a firm to gather more ideas and solutions, faster and at less cost, having too many ideas/solutions is problematic if a seeker has to evaluate all the ideas/solutions internally. Evaluating a single idea may be easy, but evaluating 154,000 ideas is hardly practical, even for Google.

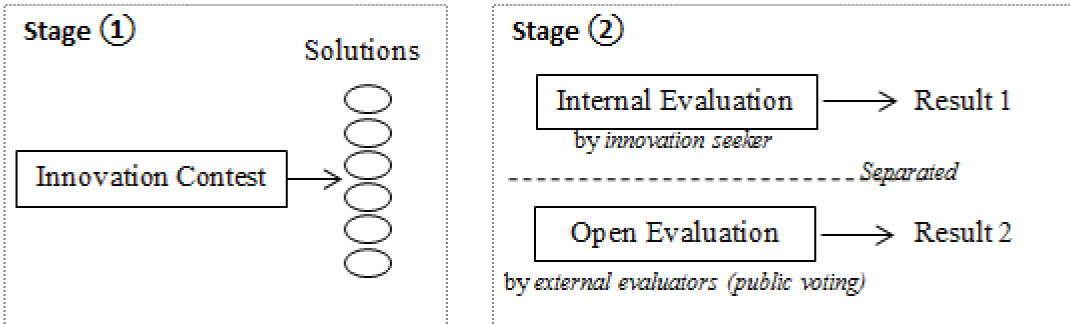
Existing evaluation schemes are limited by the availability of internal resources and dwarfed by the sheer amount of solutions that can be generated in an open innovation contest. This motivates the present study in which we propose a new approach, termed *open evaluation*, with the aim of increasing the evaluation efficiency and lowering the evaluation cost for solutions obtained through open innovation. Our proposal greatly enhances the existing innovation literature, as external intelligence is employed for the purposes of both innovation and evaluation. In this paper, we first introduce the framework of the *open evaluation* system that has been implemented by a large online contest market. We then carefully study three key elements of open evaluation: the collection mechanism by which the evaluators are aggregated, the evaluation method used by the external evaluators, together with the objective criteria provided by the seeker, and the interaction among solutions/evaluations. Performance metrics and performance models are then proposed and examined using a large-scale empirical dataset. Our empirical analysis reveals several interesting findings about different aspects of open evaluation. First, leveraging a prediction market aggregation mechanism, our results suggest that more evaluators can translate to better performance, as their contributions can be channeled into higher collective intelligence. In particular, more evaluators can be successfully aggregated if a seeker provides appropriate incentives, such as a higher evaluation prize. Second, although the objective criteria provided by the seeker are an important factor for meaningful evaluation of result, the usefulness of criteria information depends on the criteria format and the project type. For instance, visual criteria seem to be very helpful, but textual criteria show no significant impact for graphic design projects. Surprisingly, implicit criteria information, such as a seeker's background, is harmful to evaluation, since evaluators seem to read into it too much. Last, but not least, a careful examination of the interaction between solutions and evaluations reveals that a herding effect can emerge, induced by public voting, when later voters follow prior other's opinions, in order to save on decision making costs. This herding results in the trend of a higher evaluation disparity, which is associated with a lower performance in open evaluation.

4.2 ZBJ Open Evaluation System

In this section, we introduce the open evaluation system, *Shi Shi Cai*, used by ZBJ Network (Zhubajie.com), currently the world's largest online contest market. Founded in China in 2006, the market has hosted over 50,000 contests and aggregated over 4 million diversified problem solvers, as of the end of 2010.

We focus our study on the ZBJ online contests market because it has several different features from other online contests markets, and these features provide us with a unique opportunity to examine the performance of open evaluation. **Figure 4.1** describes the workflow of the ZBJ open evaluation system with an open innovation contest. There are three main stages:

- Stage 1. The innovation seeker launches a contest and external problem solvers submit multiple solutions. In our example, 6 solutions were generated.
- Stage 2. Two evaluation procedures are executed separately to evaluate all solutions. The innovation seeker carries out the internal evaluation and the chosen solution wins the contest. In addition, the solutions are made accessible to the external evaluators and public voting for the best solution is solicited. An evaluation prize is awarded to those external evaluators who successfully predict the winner that is chosen by the innovation seeker. If multiple evaluators make the correct prediction, they will share the evaluation prize. This external evaluation process is what we refer to as open evaluation. It is worth noting that later voters in the external evaluation process can see the votes entered by earlier deciders, though no external voter can see the result of the internal evaluation.
- Stage 3. Two types of evaluation results are gathered from stage 2. In our example, the internal evaluation chooses Solution #3 as the winning solution. In the external evaluation, 30 out of 50 external evaluators correctly predict this winning solution by voting for Solution #3, and thus they will share the evaluation prize.



Stage 3 Evaluation Result Analysis (An Example)

Solution #	Result 1 *(Internal Evaluation)*	Result 2 *(Open Evaluation)*
1	lose	11
2	lose	5
3	**WIN**	**30**
4	lose	0
5	lose	3
6	lose	1

30 evaluators made correct predictions and would share the evaluation prize

Figure 4.1 Workflow of ZBJ Open Evaluation System

This system was originally developed by ZBJ to identify suspicious winning solutions, which were actually submitted by seekers. In our example, if the seeker chooses Solution #4 as the winner, which fails to gather many votes from the external evaluators, the system will generate a fraud alert to the market moderators. As a result, this method increases the seeker's cost of fraudulent behavior. We have a different focus here, as we want to know whether open evaluation can aid the internal evaluation process. That is, we wish to determine what the factors are that affect the efficacy of open evaluation. To facilitate evaluating large volume of solutions, with better performance, at lower cost, we hope to understand the mechanism of the above open innovation system, and to derive implications for the design of a better open evaluation system.

4.3 Elements of Open Evaluation

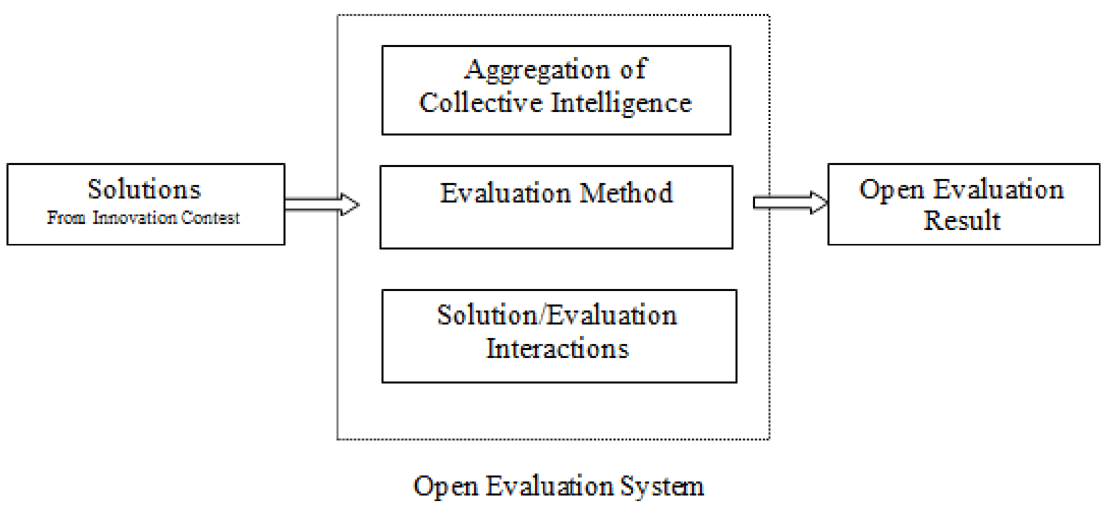


Figure 4.2 Framework of an Open Evaluation System

Figure 4.2 depicts the framework of an open evaluation system. If we treat the open evaluation system as a black box, the input will be the multiple solutions submitted by solvers, while the output is the open evaluation result.

Inside this black box, we identify three key components that may influence the evaluation output. *Aggregation of collective intelligence* is the key idea behind both open innovation and open evaluation. We want to see whether collective intelligence could affect the efficacy of open evaluation, and what factors may facilitate collective intelligence. The *evaluation method* deals with how the external evaluation process is implemented. A very important issue here is how the objective criteria provided by the seeker may affect open evaluation performance. The *solution/evaluation interactions* component focuses on the interplay between the solutions and the evaluations. Here, we want to study the factors that drive this interplay. In the following section, we review the relevant literature and develop hypotheses related to each component.

4.3.1 Aggregation of Collective Intelligence

Collective intelligence is a shared or group intelligence that emerges from aggregated individuals via some mechanism, such as collaboration or competition. (Bothos et al. 2009; Surowiecki 2004; Watkins 2007). Groups of people relax the bounded rationality constraint faced by individuals, and the potential resources of the crowd are collected and interacted. As a result, group-based knowledge can be considered a type of super expert, who is remarkably more intelligent than the smartest people in the crowd (Surowiecki 2004). Many popular applications rely on collective intelligence. For example, Gmail uses collaborative filtering to help identifying spam messages. Wikipedia uses direct democracy to aggregate reviewers, who vote and decide on the modification of content. Sports forecasting uses a prediction market mechanism to aggregate collective intelligence and predict soccer game results. The existing literature on collective intelligence has shown that such mechanisms sometimes perform "*better than theorists can explain*" (Bonabeau 2009), decision-making tasks (Bonabeau 2009; Bothos et al. 2009), recommendation systems (Watkins 2007), stock market forecasting (Fama 1970b) and sports forecasting (Spann and Skiera 2009). Although the power of collective intelligence in areas such as idea management is not yet obvious, it has been argued that the design of an appropriate aggregation mechanism is a key factor in the success of collective intelligence-based applications (Bonabeau 2009).

In the ZBJ open evaluation system, an evaluation prize is awarded to evaluators who successfully predict the winning solution, as chosen by the innovation seeker. With this evaluation prize, evaluators can trade their expectations, based on their knowledge. Once individuals can trade their expectations, a prediction market is effectively formed, which reflects the collective intelligence of participants (Spann and Skiera 2003; Spann and Skiera 2009). The fundamental theory behind prediction markets suggests that markets can solve information problems (Hayek 1945). For instance, Spann and Skiera (2009) find that prediction markets can provide more accurate sports forecasting than betting odds and tipsters. When the crowd is large enough in a prediction market, the market becomes competitive and efficient, and can better reflect all available information (Fama 1970a; Smith 1982; Spann and Skiera 2009). Consequently, more evaluators in a prediction market can aggregate more information, provide better statistical results, and reduce the uncertainty in decision making (Bonabeau 2009; Watkins 2007). This leads us to the following hypothesis:

***H1a**: Having more evaluators will increase the performance of open evaluation.*

In the ZBJ market, each contest, on average, can attract over 200 unique evaluators or votes quickly. By counting the votes, an evaluation score is received for each solution. It is important to investigate factors that may affect the aggregation level of collective intelligence, thus we consider the following hypothesis:

***H1b:** A higher evaluation prize will bring more evaluators.*

The evaluation prize impacts evaluators' perceived compensation and their evaluation effort, thus the prize amount should also have a direct impact on the open evaluation performance. However, a fixed evaluation prize level is set at 2% of the contest prize on all ZBJ online contests, thus we are not able to discriminate the direct effect of higher evaluation prizes from direct or indirect effects of higher contest prizes.

4.3.2 Evaluation Method

An evaluation is a systematic determination of the worth or significance of something or someone, using a set of criteria, against standards. Different approaches to conducting evaluations include testing programs, content analysis, accountability evaluation, decision oriented evaluation and consumer oriented evaluation (House 1978; Stufflebeam and Webster 1980). For instance, decision oriented evaluation is common in new product development, where a team of experts evaluate a candidate product portfolio based on a set of attributes. After each candidate is rated, several heuristic procedures or conjoint analysis methods may be used to design an optimal product line while maximizing an objective function, such as social welfare or firm profit (Chao and Kavadias 2008; Green and Krieger 1985; Kohli and Sukumar 1990; McBride and Zufryden 1988; Terwiesch and Ulrich 2009). There literature on different evaluation methods is extensive, so, for the sake of brevity, we refer readers to Krishnan and Ulrich (2001) for an excellent review.

Our proposal for an open evaluation system contributes to the extant repertoire of evaluation methods. As we have seen in **Figure 4.1**, the ZBJ open evaluation system forces each evaluator to predict the winning solution that the seeker is most likely to choose. To an evaluator, this evaluation method simply depends on the same factors that would be used by the innovation seeker. However, with an increasing count of solutions, the required scope and cost of the evaluation search increases and the average effort spent on each solution falls. Under the assumption that evaluation performance is proportional to the evaluation effort, the result of evaluating a single solution becomes less accurate when there are more solutions to consider. Moreover, having more solutions will increase the complexity of the overall evaluation task. Due to the limited absorptive capacity of each individual evaluator (Cohen and Levinthal 1990), the variance of individual's learning/forgetting rate and their productivity will be significantly affected by the task complexity (Nembhard and Osothsilp 2002). Even if each evaluator maintains the same level of evaluation effort for each solution, as in the simple evaluation case (low solution volumes), the evaluation result is still less accurate due to the lower learning capability and higher forgetting rate during the individual evaluation process. For instance, the evaluator may forget some information about a prior solution and make the wrong judgment in the current solution evaluation. When the individual evaluation results are aggregated via ZBJ open evaluation system, the collective result will become less accurate. Thus, we hypothesize:

H2a: *When there are more solutions to evaluate, the performance of open evaluation decreases.*

In the ZBJ market, the seekers are free to post their project descriptions, which disclose information about how the seekers will choose their favored solution. We refer to this disclosed information as *criteria information*. Three types of criteria information are commonly observed: *textual criteria information* is provided when seekers explicitly state the requirements of a favored solution in their project description, *visual criteria information* is provided when seekers explicitly express their preferences in any visual format and *implicit criteria information* pertains to the seekers' background information. The role of criteria information is summarized in **Figure 4.3**. Because of the monetary incentive of an evaluation prize, external evaluators will attempt to learn from the disclosed criteria information to try to maximize their chance of winning. Without this criteria information, votes from the external evaluators may not align with the seeker's strategic goals.

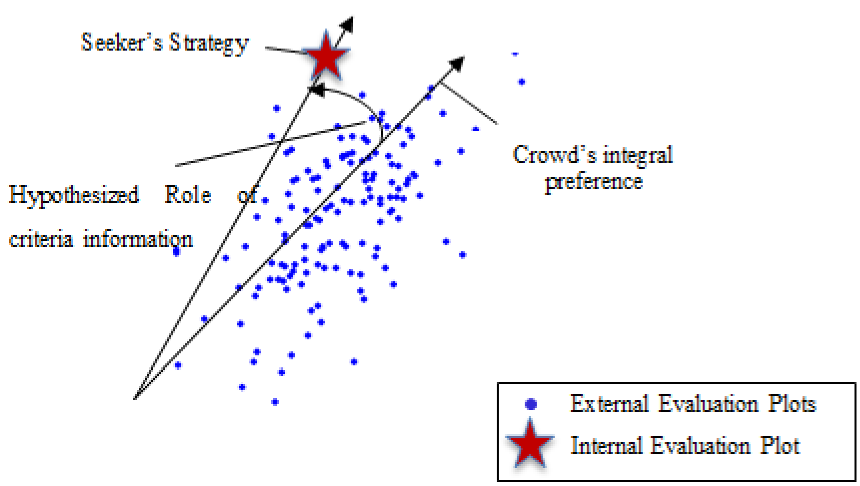


Figure 4.3 Role of Criteria Information

The existing literature about forced ranking and objective criteria information can be extended to the open evaluation context. A forced ranking exercise has been widely used for LOGO design and industrial design, where aesthetics and other holistic product attributes are important (Krishnan and Ulrich 2001). Prior studies have found that, without specifications, individuals tend to evaluate based on personal preferences (Newell et al. 2004). If there are no unified objective criteria, the cumulative evaluation results therefore reflect the evaluators' overall priorities and integral preferences, regardless of the firm's strategic goal. For instance, Dell IdeaStorm asks any Dell user to vote for the ideas generated by the user community, based on personal preferences (Di Gangi and Wasko 2009). No criteria information is needed in this voting as Dell's strategic goal aligns with the preferences of the crowd – Dell's customers. On the other hand, objective criteria are needed when an innovation seeker's goal is not a reflection of evaluators' preferences (Newell et al. 2004). In the ZBJ market, evaluators are aggregated quickly and they have no idea who the seekers are. There is no evidence that these external evaluator's integral preferences are consistent with the innovation seekers', thus objective criteria are needed. More recently, (Terwiesch and Ulrich 2009) have argued that objective criteria are needed when we desire a coherent evaluation result with a large number of evaluators, which is very common in the context of open evaluation. Based on the above arguments, we consider the following hypotheses:

H2b*: Disclosing textual criteria information can increase the performance of open evaluation.*

H2c*: Disclosing visual criteria information can increase the performance of open evaluation.*

***H2d**: Disclosing implicit criteria information can increase the performance of open evaluation.*

4.3.3 Solution/Evaluation Interactions

Each solution is not evaluated alone but is compared with all other candidates in the same pool. It is thus important to understand what factors may influence the solution/evaluation interactions.

We are interested in knowing how the voting distribution may impact the open evaluation performance. To make this idea more concrete, we assume there are two identical contests, A and B. Both contests have an equal number of solutions and are evaluated by an equal number of evaluators. Both the internal evaluation and open evaluation choose Solution #3 as the best submission. As we can see in **Table 4.1**, the only difference between the two scenarios is the distribution of the open evaluation votes. Empirical studies reveal that decision-making is impacted by the information on relationships between alternatives (Davidson et al. 1957; Edwards 1956; Lanzetta and Kanareff 1962).

If we assume that each evaluator makes independent predictions, based on his or her own judgment, the cumulative open evaluation result may be viewed as a proxy for the solution quality. Prior studies have shown that people need to collect information to discriminate options (Harvey and Bolger 2001), and the information cost impacts the uncertainty of choices (Lanzetta and Kanareff 1962). Compared with contest B, the open evaluation results of contest A exhibit a larger disparity. The cost of identifying the best solution in contest A may seem to be lower, and one may expect better open evaluation results in contest A because of its larger disparity.

Table 4.1 Contest A vs. Contest B

Solution #	Contest A		Contest B	
	Result 1 *(Internal Evaluation)*	Result 2 *(Open Evaluation)*	Result 1 *(Internal Evaluation)*	Result 2 *(Open Evaluation)*
1	Lose	11	Lose	11
2	Lose	5	Lose	8
3	**WIN**	**30**	**WIN**	**12**
4	Lose	0	Lose	9
5	Lose	3	Lose	10
6	Lose	1	Lose	0
Total	1 Winner	50	1 Winner	50

The public voting policy of the ZBJ open evaluation system allows each evaluator to see the existing evaluation results of other evaluators before he/she decides which solution to vote for. This public voting policy makes the previous assumption of independent voters unlikely to be true. In fact, "voters are known to be influenced by opinion polls to vote in the direction that the poll predicts will win" (Banerjee 1992). An obvious reason for this is that other voters may have some important information. If enough earlier evaluators have chosen a particular solution, a later evaluator may eventually ignore any private information and just follow everyone else. This outcome is known as an informational cascade, wherein the perception of others' private knowledge cascades across all future evaluators to produce a herding effect (Bikhchandani et al. 1992; Bikhchandani et al. 1998). Similar scenarios have been observed in the recommendation literature, where the usage of online recommendations based on previous customer reviews can generate sales, but diversity of sales decreases significantly (Fleder and Hosanagar 2009; Fleder and Hosanagar 2007). The above arguments suggest that a large disparity among votes could be an indicator of a strong herding effect.

However, as all evaluators that make correct predictions will share the prize together in the ZBJ open evaluation system, a solution with more votes may be a safer bet, but it is also less attractive, given that it will result in a smaller prize. It has been noticed that in sports gambling, such as at horse tracks, that bettors may prefer high-variance, low-probability bets (Golec and Tamarkin 1998). Similarly, Moe and Trusov (2011) have found that people in online social communities tend to provide different product ratings to distinguish themselves from previous reviewers. The above arguments suggest that a strong herding effect may also lead to a lower disparity. Overall, it is unclear whether a herding effect, from a public voting policy, will increase the evaluation disparity.

How will a herding effect impact the performance of open evaluation? In the extreme case, if a voter only follows others' opinions, he/she is simply a chance taker, rather than an evaluator. When a later evaluator ignores his private information, the collective intelligence of the crowd is discounted. Watkins (2007) suggests that it is better to make each decision independently, and hints that the herding effect is always bad for evaluation performance. Although herding behavior gives evaluators the benefit of saving in search costs, there is no benefit to the open evaluation process, since some private information is lost. Given strong empirical evidence about the existence of the herding effect, we surmise that a herding effect results in a higher disparity, thus we have:

H3a: *When the evaluation disparity is higher, the performance of open evaluation decreases.*

People tend to follow others' decision to save on search and decision costs, as well as to reduce the risk of making a wrong decision, particularly when there are many alternatives (Tucker and Zhang 2011). The herding effect should be more significant when there are lots of solutions and the search/decision cost is high. As such, this leads us to hypothesize that:

H3b: *When there are more solutions to evaluate, the evaluation disparity will become higher.*

In the above, we use the evaluation disparity as a proxy for the presence of a herding effect. One interesting direction for future research would be to utilize the sequential decisions of evaluators to test for the presence of a herding effect directly.

4.4 Performance Measurement Theory

In this section, we describe how to construct a performance measurement under the open evaluation framework. This is an important contribution, as no such measurement is currently available in the existing open evaluation literature. The fundamental theory underlying our performance measurement is to compare a seeker's evaluation result with the open evaluation result. We take this approach for two reasons. First, in the ZBJ open evaluation system, crowds are aggregated to make predictions about which solution will be chosen by the seeker. In the prediction market literature, comparing predictions with real outcomes is a dominant method. Second, as discussed in the criteria information section, seekers have no previous relationship with the evaluation crowds, evaluators are not the seeker's customers and most projects are not customer oriented. Further, the goal of our study is to determine the feasibility of replacing internal evaluation effort with open evaluation. Hence, we can use the seeker's evaluation result as a benchmark to measure the open evaluation performances. We propose two performance metrics: the *precision* metric and the *hit-or-miss* metric.

4.4.1 The Precision Metric

In open evaluation, the focus is on the differences between open evaluation results and the internal evaluation results. A similar type of focus is found in the literature dealing with information retrieval systems, where performance is measured by the fit between algorithm-based ranking results and expert-based ranking results. There is a long stream of research (Baeza-Yates and Ribeiro-Neto 1999; Berry et al. 1995; Joachims 2002a; Salton and Buckley 1997) on different performance measurement metrics for information retrieval systems, among which precision is one of the most important. We will therefore extend the idea of precision into the open evaluation framework.

A ranking method and a precision calculation method are needed to derive the precision metric in the information retrieval system (Agichtein et al. 2006). Algorithm-based scores are generated for the purpose of ranking. The well-known Google PageRank algorithm (Page et al. 1999), for instance, could be used to generate such a score. Widely accepted precision calculation methods for information retrieval systems include precision at K, normalized discounted cumulative gain (NDCG) and mean average precision (MAP) (Agichtein et al. 2006). User click-through information has been used to measure performance in many recent studies as well (Agichtein et al. 2006; Joachims 2002a; Joachims 2002b; Joachims et al. 2005).

In open evaluation, each evaluator will vote for the solution that he predicts will win, and the number of votes received by one particular solution is naturally our vote-based score. We will rank the solutions by their vote-based scores, from highest to lowest. In a given contest, we denote the total number of solutions as $N_{solution}$ and the votes received by the *i*th ranked solution as $N_{vote}(i)$. Thus $N_{vote}(1)$ is the number of votes received by the solution with the most votes, $N_{vote}(2)$ is the number of votes received by the solution with the second most votes, etc. For the precision calculation method, we require that it satisfies the following conditions, simultaneously: (1) when the solution with the most votes in open valuation is the winner of the internal evaluation, the precision equals 1; (2) when the winner of the internal evaluation receives zero vote in open valuation, the precision equals 0; (3) compared with lower ranked solutions, higher ranked solutions should have a higher weight in the precision calculation. The last point above is based on the consideration that users give more attention to top ranked results, due to their having limited resources (Agichtein et al. 2006; Baeza-Yates et al. 2005). For instance, if the winner of the internal evaluation is ranked second, with 5 votes, and the top ranked solution has 6 votes, then the precision in this case should be higher than the precision of a similar case where the winner of the internal evaluation is ranked second, with 5 votes, and the top ranked solution has 10 votes. Based on these arguments, we define precision as follows:

$$\text{Precision} = \frac{\sum_{i=k}^{N_{solution}} N_{vote}(i)}{\sum_{i=1}^{N_{solution}} N_{vote}(i)}, \tag{4.1}$$

where k denotes the position of the winning solution among all the ranked solutions.

4.4.2 The Hit-or-Miss Metric

Hit-or-miss for a single prediction and hit rate (average hit-or-miss) for overall predictions are commonly used metrics in prediction markets. Take sports forecasting as an example. A soccer game has two alternative results: home team win or away team win (if there is a tie, a penalty kick is used to decide a winner). The prediction must be one of these two alternatives. The hit-or-miss value is 1 if the prediction is correct and 0 otherwise. In open evaluation, there are many solutions in one contest and it is difficult to predict exactly which solution is going to win. Our strategy here is to divide all alternatives into two groups: the *high group (H)*, which consists of the half of the solutions that have more votes, and the *low group (L)*, which consists of the other half of the solutions, having fewer votes. Hit-or-miss is then decided based on the grouped result, which is similar to the classical setting with two alternatives. The hit-or-miss value then is 1, if the winning solution from the internal evaluation is in the high group, and 0 otherwise. **Table 4.2** demonstrates this procedure using the raw data from **Figure 4.1**.

Table 4.2 Original Result Vs. Grouped Result

Solution #	Result 1 *(Internal Evaluation)*	Result 2 *(Open Evaluation)*	High (H) Group and Low (L) Group	Hit (1) or Miss (0)?
3	**WIN**	**30**	H (46 votes)	Hit (1)
1	Lose	11		
2	Lose	5		
5	Lose	3	L (4 votes)	
6	Lose	1		
4	Lose	0		

Based on the algorithm, the two metrics represent different information. Precision measures the distance from a perfect open evaluation result, while hit-or-miss or hit rate pertains more to the chance of deleting half the solutions without losing the best solution.

4.5 Data Analysis

4.5.1 Measurement

Our data is collected from the ZBJ open evaluation system and we consider all the innovation contests conducted in 2010. Multi-winner contests are removed and we only consider those cases where a single winner was chosen by internal evaluation. Some contests allow the solvers to provide multiple solutions, in which case we count all the solutions from the same solver as one solution, and the votes corresponding to these solutions are accumulated to a single value. We end up with a total of 1,464 contests, 102,807 candidate solutions, and 325,070 external votes. Below is a list of the variables.

- *Number of Solutions* ($N_{solution}$) records how many solutions are generated by the particular contest.
- *Number of Evaluators* ($N_{evaluator}$) counts how many evaluators the particular contest has.
- *Textual Criteria* ($C_{textual}$) is the percentage of textual criteria information provided in the contest description.
- *Visual Criteria* (C_{visual}) is a binary variable that indicates whether the seeker provided explicit visual criteria information in the contest description.
- *Implicit Criteria* ($C_{implicit}$) is the percentage of seeker background information or other implicit criteria information provided in the contest description.

- *Evaluation Prize* ($P_{evaluation}$) is the evaluation prize that the winning evaluators will share.
- *Number of Votes* (N_{vote}) counts how many evaluators predict a particular solution will win.
- *Evaluation Disparity* (*Disparity*) measures the disparity among the number of votes in the particular contest, and is calculated as

$$Disparity = \frac{N_{solution}+1}{N_{solution}-1} - \frac{2}{u \times N_{solution} \times (N_{solution}-1)} \sum_{i=1}^{N_{solution}} i \times N_{vote}(i), \quad (2)$$

  where $N_{vote}(i)$ denotes the votes received by the ith ranked solution, such that $N_{vote}(1)$ is the number of votes received by the solution with the most votes, and u is the average of all $N_{vote}(i)$.

The definition of the evaluation disparity above is adapted from the well-known Gini coefficient (Gini 1912; Yitzhaki 1979), which is used to measure the inequality of the distribution in income, and other economics/sociology measurements. The Gini coefficient provides a standardized value between 0 and 1, with 0 implying total equality and 1 implying maximal inequality. In our case, a larger disparity measure means the votes are distributed less evenly.

Descriptive statistics of the above variables are listed in **Table 4.3**. The Pearson correlations of the above variables are provided in **Table 4.4**.

Table 4.3 Descriptive Statistics

Variable	Definition	Mean	Std. Dev.	Min	Max
$N_{solution}$	The number of solutions	58.54	87.78	10	1449
C_{texual}	Textual criteria (%, between 0 and 1)	19.33%	20.51%	0	1
C_{visual}	If has visual criteria (Yes=1, no=0)	o.41	0.49	0	1
$C_{implicit}$	Implicit criteria (%, between 0 and 1)	11.15%	14.92%	0	1
$P_{evaluation}$	Evaluation prize amount (￥)	11.97	22.66	0.26	600
$N_{evaluator}$	The number of evaluators for a contest	**206.53**	220.21	10	1739
N_{vote}	The number of votes for a solution	3.17	8.80	0	207
$(N_{vote})_{win}$	The number of votes for the winner	12.85	25.01	0	283
Disparity	Evaluation disparity for a contest.	0.65	0.14	0.24	0.94
Control Variables					
Duration	Number of days open for evaluation	14.40	12.33	1	93
Category	Contest project types	Graphic Design (82.1%), Naming (7.1%), Creative writing (7.0%), Website development (2.4%), Software (1.4%)			
Performance Metrics					
Hit-or-Miss	Whether the H group (top 50% of higher voted solutions) hits the winning solution. (0 or 1)	52.3%	0.49	0	1
Precision	See equation 2	0.30	0.36	0	1
#Observations	102,807 solutions with 325,070 votes, grouped by 1,464 contests				

Table 4.4 Pearson Correlation Matrix

	Precision	Hit	$N_{solution}$	C_{texual}	C_{visual}	$C_{implicit}$	$P_{evaluation}$	$N_{evaluator}$
Hit-or-Miss	0.741							
$N_{solution}$	-0.157	-0.089						
C_{texual}	0.003	-0.009	0.064					
C_{visual}	0.119	0.086	-0.156	-0.245				
$C_{implicit}$	-0.116	-0.097	0.122	0.019	-0.234			
$P_{evaluation}$	-0.006	0.029	0.124	-0.074	0.054	-0.050		
$N_{evaluator}$	-0.051	0.005	0.444	-0.052	-0.008	0.049	0.375	
Disparity	-0.220	-0.108	0.491	0.089	-0.217	0.171	-0.017	0.190

4.5.2 Methodology

Models M1 and M2 focus on the open evaluation performance. Linear regression is used in M1 with the precision metric as the response variable, and logistic regression is used in M2 with the hit-or-miss metric as the response variable.

M1:

$$\Pr ecision = \beta_0 + \beta_{2a} Ln(N_{solution}) + \beta_{2b,c,d} C_{textual/visual/implicit} + \beta_{1b} Ln(N_{evaluator}) + \beta_{3a} Disparity + Controls$$

M2:

$$Logit(hit - or - miss) = \beta_0 + \beta_{2a} Ln(N_{solution}) + \beta_{2b,c,d} C_{textual/visual/implicit} + \beta_{1b} Ln(N_{evaluator}) + \beta_{3a} Disparity + Controls$$

In the above models, $N_{solution}$ *and* $N_{evaluator}$ are log-transformed to achieve better model fit. The control variables include contest category and contest description length.

Linear model M3 is used to systematically model the factors that may affect the number of evaluators.

M3: $Ln(N_{solution}) = \beta_0 + \beta_{1a} Ln(P_{evaluation}) + Controls$

In addition to contest category and contest description length, contest duration is also controlled since intelligence aggregation takes time.

To test the relationship between the disparity and $N_{solution}$, we also propose model M4.

M4: $Disparity = \beta_0 + \beta_{3b} Ln(N_{solution}) + Controls$

To account for potentially unobserved factors that may contribute to both open evaluation performance and the predictors in the above models, we use the seemingly unrelated regression (SUR) method.

4.6 Results

The overall correlation between internal evaluation and open evaluation is 0.153, with a p value < 0.0001. The open evaluation is effective and is better than a random guess, as the overall hit rate of the ZBJ open evaluation system is 52.3% (a random hit rate would be 50%). From the results in **Table 4.5**, we see that when $N_{solution}$ is smaller than 20, the cumulative hit rate is 57.6%, which is higher than a sports forecast based prediction market (Spann and Skiera 2009). However, when $N_{solution}$ is larger than 20, although still better than random guess, the cumulative hit rate becomes much lower. This result roughly supports H3a, which suggests that a higher number of solutions will be associated with lower open evaluation performance.

Table 4.5 Hit rate Descriptive Summary

Application	Hit Rate		
ZBJ open evaluation	52.3%,	Overall	
	57.6%,	if $N_{solution}$<20	By $N_{solution}$
	51.7%,	if 20<$N_{solution<}$30	
	51.5%,	if 30< $N_{solution}$	
	36.5%,	Naming	By Category
	53.1%,	graphic designs	
	70.0%,	software development	
Sports forecast (Spann and Skiera 2009)	54.3%,	based on prediction market	
	53.7%,	based on betting odds	
	42.6%,	based on tipster	

Table 4.6 Overall Regression Result

Variable	M1: Hit-or-Miss	M2: Precision	M3: $Ln(N_{evaluator})$	M4: Disparity
Constant	0.364*** (0.130)	0.449***(0.072)	3.768***(0.459)	0.262***(0.013)
$Ln(N_{solution})$	- 0.176* (0.100)	- 0.044***(0.021)	--	0.108***(0.003)
Criteria				
$C_{textual}$	0.274 (0.289)	0.070 (0.049)	--	--
C_{visual}	0.207* (0.127)	0.047** (0.022)		
$C_{implicit}$	- 0.984** (0.405)	- 0.151** (0.069)		
$Ln(N_{evaluator})$	0.304***(0.088)	0.058*** (0.015)	--	--
$Ln(P_{evaluation})$	--	--	0.495***(0.022)	--
Disparity	- 1.144** (0.584)	- 0.328***(0.101)	--	--
Category				
Graphic Design	0.000	0.000	0.000	0.000
Naming	- 0.144** (0.065)	- 0.123***(0.046)	1.231*** (0.082)	0.052***(0.012)
Software Dev.	0.136 (0.113)	- 0.011 (0.080)	- 0.535***(0.169)	-.101***(0.022)
Web App Dev.	0.063 (0.087)	0.089 (0.062)	- 0.903***(0.130)	-0.008 (0.017)
Creative Writing	- 0.001 (0.054)	0.017 (0.038)	- 0.176** (0.039)	-0.006 (0.011)
Duration	--	--	0.007***(0.001)	--
Ln(Description Length)	0.013 (0.016)	- 0.009 (0.012)	--	--
Prob >F	0.0000	0.0000	0.0000	0.0000
R^2	4.15%	8.63%	41.58%	49.67%
No.of observations	1,464			

Note: *~p <0.1; **~p <0.05; ***~p <0.01

4.6.1 Overall Analysis

The results of M1 through M4 are summarized in **Table 4.6**.

In both M1 and M2, the coefficients of $N_{evaluator}$ are positive and significant, which means that more evaluators will improve the open evaluation performance and H1a is supported. In model M3, the coefficient of $P_{evaluation}$ is positive and significant. So H1b is supported. A roughly two-unit increase in the log of the evaluation prize will cause a one-unit increase in the log of the evaluator count. In other words, if the evaluation prize is doubled, the aggregated number of evaluators will increase 50%.

The coefficients on $N_{solution}$ are negative and significant in both models M1 and M2, which means that the overall evaluation performance decreases when there are more solutions to evaluate. Thus, H2a is supported. The three criteria information variables show very different impacts on performance. The use of textual criteria exhibits no significant impact on open evaluation performance, while the use of visual criteria can significantly improve performance. Actually, more than 80% of the contests are graphic designs, where the criteria information expressed in a visual format can be of great help, while textual descriptions may be less useful in conveying aesthetic judgment. Surprisingly, the use of implicit criteria information shows a significant negative impact. One explanation is that the evaluators read too much into the implicit information, such that it becomes misleading. Evaluators have no prior relationship with innovation seekers, so this result suggests that they should "*judge not of men and things at first sight*." Thus H2c is supported, H2d is rejected, and more evidence is needed to evaluate H2b.

The coefficients on the evaluation disparity in M1 and M2 are both negative and significant. This means that a higher evaluation disparity is associated with lower open evaluation performance, so H3a is supported. We suspect that this correlation is caused by a herding effect. When later voters start to follow earlier voters, instead of using their own judgment, this herding effect will produce an increasing evaluation disparity, while simultaneously decreasing the evaluation performance. The coefficient on $N_{solution}$ is negative and significant in M4, which suggests that the evaluation disparity becomes higher when there are more solutions, thus H3b is supported. This implies that the herding effect is stronger when there are more solutions. This is reasonable, as more solutions will increase the evaluation cost for evaluators, making them more likely to follow others' opinions (Tucker and Zhang 2011).

4.6.2 Sectional Analysis

In models M1-M4, category plays a significant role. We begin our sectional analysis by first summarizing the sectional descriptive statistics in **Table 4.7**, where the average values based on all contests within one particular category are reported. For open evaluation performance, projects associated with naming receive the lowest hit rate and precision. A unique feature of the naming category is that naming projects are purely ideation-based, the output of which is highly random (Terwiesch and Xu 2008). For this type of project, although seekers may provide some criteria information, there are nearly no "objective" criteria that most people can agree on. Furthermore, contest categories can be ordered by the evaluator population, as follows:

$$Naming > Graphic\ Desing > Creative\ Writing > Software\ Development > Website\ Development$$

This confirms the earlier finding that naming and graphic design projects usually have more solutions than other type of projects (Yang et al. 2009). This order also suggests that a project with a lower evaluation cost can attract more evaluators. Graphic design and naming projects, both of which are highly ideation based, receive more solutions than other types of projects and they exhibit the highest disparities.

Table 4.7 Sectional Descriptive Analysis

Variable	Graphic Design	Naming	Software Dev.	Web App Dev.
Hit rate	53.1%	36.5%	70.0%	57.2%
Precision	0.318	0.101	0.383	0.425
$N_{solution}$	45.52	212.33	16.35	19.46
$N_{evaluator}$	190.67	385.79	104.55	191.6
$P_{evaluation}$	11.32	5.88	13.82	39.24
Disparity	0.642	0.870	0.457	0.564
Dimension	High Ideation	Pure Ideation	High Expertise	Ideation+Expertise

It is worth noting that the average hit rate of naming projects is much lower than 50%, which is an indicator of a significant herding effect. The average disparity in the naming projects is 0.87, which is much higher than in the other projects. There are usually a large number of solutions for naming projects, and the early evaluation results may have a major impact on later evaluators. The software development and web application development projects have higher hit rates and precisions compared to the other categories. This suggests that expertise-based projects may be easier to evaluate.

Table 4.8 Sectional Regression Results (Metric: Precision)

Variable	Graphic Design		Naming		Software Dev.		Website Dev.	
Constant	0.533***	(0.103)	- 0.138	(0.466)	1.220	(0.795)	- 0.060	(0.785)
$Ln(N_{solution})$	- 0.047**	(0.023)	0.014	(0.055)	- 0.093	(0.208)	- 0.031	(0.274)
Criteria								
$C_{textual}$	- 0.013	(0.057)	0.047	(0.090)	0.639***	(0.179)	0.978***	(0.314)
C_{visual}	0.038*	(0.022)	- 0.091	(0.056)	0.327**	(0.134)	0.155	(0.143)
$C_{implicit}$	- 0.181**	(0.077)	- 0.012	(0.114)	- 1.701	(1.694)	0.296	(0.685)
$Ln(N_{evaluator})$	0.072***	(0.017)	- 0.005	(0.057)	- 0.010	(0.097)	0.050	(0.070)
Disparity	- 0.372***	(0.115)	- 0.166	(0.417)	- 0.860*	(0.493)	- 0.283	(0.556)
Ln(Description Length)	- 0.014	(0.013)	- 0.010	(0.012)	- 0.220***	(0.072)	0.009	(0.088)
Prob>F	0.0000		0.7104		0.0057		0.0968	
R^2	6.10%		5.54%		51.95%		27.21%	

Note: *~p <0.1; **~p <0.05; ***~p <0.01

It is interesting to note that software development has the highest hit rate, while website development has the highest precision, on average. This confirms that the two performance metrics contain different information, although they are highly correlated (Pearson correlation=0.741, $p<0.0001$).

The sectional regression analysis results are summarized in **Table 4.8**. As a purely ideation-based category, none of the variables are significant in the model for naming projects. It is also difficult to give objective criteria with which to evaluate the quality of a name, as a good name to one evaluator may be considered a very bad name by another evaluator. One possibility here may be to model naming projects as a random process (Terwiesch and Xu 2008). For the other three models, the coefficients for $N_{solution}$ and disparity are consistently negative, providing strong empirical evidence for H3a and H3b.

Following Terwiesch and Xu's (2008), we view graphic design as a highly ideation-based project category, software development as a highly expertise-based category and web applicaiton development as both ideation- and expertise-based. The R^2 values suggest that the regression models best explain the performance of the open evaluation process when considering projects that require a higher degree of solver expertise. Another interesting finding here is that different types of criteria information behave in different ways, across different categories. The textual criteria do not have a significant effect when it comes to graphic design projects, but they are helpful in the expertise-based category, which includes software and website development projects. This suggests that H2a is partially supported.

4.6.3 Robustness Check

We performed the following robustness checks to ensure the validity of our results. First, scatter plots of observations and performance metrics did not show any pattern, indicating independence of observations (i.i.d.). In the regressions, we constructed White (robust) standard errors to address potential heteroskedesticity. The effect of multicollinearity was also considered based on the variation inflation factors (VIFs) for all models (Dasgupta and Nti 1998). All VIFs are lower than 3.1, which is below the common threshold of 5. We also checked the correlations between variables, to address the low power of the VIF test in large datasets. To avoid potentially unobserved factors that can contribute to both open evaluation performance and the number of evaluators (i.e., errors are likely to be correlated across the above models), we used the seemingly unrelated regression (SUR) method (Runkel 2006). Our result shows that the coefficients obtained from independent regression are different and, as such, simultaneous regression analysis is necessary.

4.7 Discussion and Conclusion

In recent years, collective intelligence has been applied in many contexts with great successes, and sometimes the results are "better than what theorists can explain" (Bonabeau 2009). Open innovation contests are being adopted by an increasing number of firms, as a new innovation approach based on collective intelligence. However, the scope and performance of open innovation have been greatly limited by the high costs associated with evaluating and reviewing generated ideas. Although collective intelligence has shown great power in terms of idea generation, approaches to also employing it for effective idea evaluation have not been previously considered. We propose the idea of 'open evaluation', where collective intelligence is used for evaluation. Below are our main contributions.

First, we introduce an open evaluation system based on the prediction market aggregation mechanism. We identify three key components for any open evaluation system, including aggregation of individual effort to achieve collective intelligence, the evaluation method and the interactions between solutions and evaluations. Our open evaluation system is based on the online innovation contest market and provides a general framework that may be used in many other contexts. Second, we introduce two metrics for open evaluation: hit-or-miss, which has its roots in prediction markets, and precision, which is an extension of a measure used in the information retrieval system literature. Third, we test several interesting hypotheses using a large-scale empirical dataset, deriving many important managerial implications for the design of better open evaluation systems. For example, the usefulness of criteria information depends on the criteria format and the project type. For graphic design projects, visual criteria are very helpful, while textual criteria show no significant impact. For software development and website development, where specific expertise is required, textual criteria becomes more useful. For naming projects, no criteria information is helpful. Implicit criteria or background information may be misleading, and it seems that there is no need to disclose the background information of innovation seekers. Last but not least, this is one of the first studies to investigate the impacts of voting disparity in an open evaluation context. Our evaluation is based on the Gini coefficient, which we identify as a good measure of evaluation disparity. In private voting, a higher disparity is expected to reduce the cost of discriminating solutions. In public voting, our analysis shows that a high disparity is associated with lower performance of the open evaluation process, which provides strong empirical evidence of a herding effect in public voting. To avoid this herding effect and to achieve better open evaluation performance, we recommend private voting, in order to guarantee the independence of each evaluation.

Our results may help open innovation pioneers, such as Google, to systematically generate and identify exceptionally good ideas at much lower costs. Open evaluation has the potential to greatly enhance the popularity of open innovation, and it may be a good, complementary method to Google's 80/20 innovation time-off model. Open evaluation can also be used in domains outside of innovation. For instance, open evaluation may be beneficial for large firms to systematically identify exceptional candidates for job positions, with lower recruitment costs.

4.8 Limitations and Future Work

In this paper, we take the first step toward understanding the feasibility of employing open evaluation to reduce the organizational burden of evaluating a large amount of solutions, generated through open innovation contests. Our study opens several avenues for future investigation. First, prior studies show that a higher disparity in evaluation can save discrimination costs. However, we find strong evidence of a herding effect in public voting, such that a higher disparity is associated with worse open evaluation performance. It will be interesting to see whether a higher disparity will lead to better performance in private voting with no herding effect. Second, in all the contests that we observed in our study, the evaluation prize is set at 2% of the contest prize. The factors driven by the evaluation prize and the contest prize are thus confounded. For instance, a higher evaluation prize will be likely to attract more evaluators, and a higher contest prize will be likely to attract more solutions. This makes it impossible to study the effect of evaluator count while totally controlling for the solution count. It would be interesting to study the effect of the contest prize on open evaluation performance without this confounding. Third, in the current system, the external evaluator can only cast one vote for the winner, which is equivalent to asking the evaluator to disclose the top selection. It would be interesting to consider a more complex system, where the external evaluator can rank their top two or more selections. Other possible extensions might include testing for the existence of a selection bias amongst external evaluators, giving different weights to the votes of external evaluators based on their expertise levels, and using alternative benchmarks to determine the performance of open evaluation processes, other than the internal evaluation result.

CHAPTER 5

CONCLUSION

The open innovation contest is a promising mechanism for innovation seekers, due to the expectation of higher investment returns, as well as faster and potentially better performance. By taking advantage of the worldwide Internet, launching an innovation contest online can further enhance the performance of open innovation due to the potential for a larger pool of talented solvers from all over the world, as well as the lower cost of attracting them. Indeed, in an online market with millions of individuals, a newly launched online contest can reach vast numbers of potential solvers in a very short time, at manageable cost. Online contests for open innovation are becoming popular and have been adopted by many firms. However, despite increased interest in online open innovation contests, it remains unclear how innovation seekers can take advantage of online contests. Our study makes solid contributions to both the literature on open innovation contests in online markets and the use of online contests in practice.

5.1 Theoretical Contribution

As introduced in Chapter 1, five streams of research provide the main theoretical basis for open innovation contests. These include: new product development (NPD), contests in economics, sales contests, collective intelligence, behavioral economics and psychology. These three studies jointly contribute something to the NPD and collective intelligence literatures.

5.1.1 NPD

Prior studies are have stuck to the traditional contest scenario wherein contestants are organized to compete simultaneously and silently. In contrast, we adopt a very different contest scenario that is consistent with the latest contest timeline and format that we have observed in the most popular online contest markets. Specifically, we show that feedback is critical and that it results in a very different contest performance model than what has been predicted by prior studies. We provide an adapted and simple performance model, which implies that having more problem solvers is always beneficial if the seekers have sufficient capacity to perform the evaluation and to exploit the solution. Feedback is allowed in most markets we have observed and we could not find any reason that feedback should be excluded from online contests in the future. Thus, we kindly suggest that any future study on contest performance should consider the impact of feedback.

Besides the above, we provide a contemporary framework/model for overall performance forecasting. This model can effectively predict future performance before a contest is launched in an online market. Compared to prior studies, this framework considers a much wider range of variables, e.g., contest design parameters, project characteristics and market environment variables.

We also provide profound results for the temporal strategies of winners in a contest. The results help us to understand what winners have in common, as well as solvers' selection biases regarding expertise. This knowledge is necessary to inform an optimal contest design, which needs to align with the goals of principals (innovation seekers) and agents (problem solvers). An obvious advantage of a contest is that it can generate large volumes of ideas at minimal cost. Terwiesch and Ulrich (2009) argue that innovation seekers fundamentally need are more, better and diverse ideas. The enormous diversity of ideas also results in a high variance in the probability of winning for each solver, which makes it difficult to predict winners based on any information other than solution quality.

Evaluation is required to systematically determine the winning solution(s) of a contest. As a core aspect of NPD, evaluation has been studied thoroughly in the past (Krishnan and Ulrich 2001). However evaluation is a resource-intensive process that requires expertise, time and labor (Rossi et al. 2004). Although one might hope to generate large volumes of ideas using contests, this results in a commensurate increase in evaluation costs. To dramatically reduce the evaluation cost, we suggest a new approach, which we refer to as open evaluation. This idea is not novel, however, virtually no empirical work has considered it. Ours is one of the first studies to consider the performance of an open evaluation system for open innovation contests, and we identify three key components for any open evaluation system: the collective intelligence aggregation mechanism, the evaluation method and the interactions between solutions and evaluations. Furthermore, we provide metrics to measure the open evaluation performances and examine our hypotheses leveraging a large-scale empirical dataset. We suggest that open evaluation is more appropriate for expertise-based projects, such as those dealing with software development and so on.

In summary, by studying open innovation contests, we transition the traditional NPD approach to a much more open approach, for both the idea generation and idea evaluation steps, employing systematic methods. We hope our studies can make the NPD literature more accessible to a wider range of topics in the future.

5.1.2 Collective Intelligence

The online innovation contest can also be viewed as an aggregation tool to achieve collective intelligence. Consistent with Terwiesch and Ulrich (2009)'s NPD framework, in Figure 1.3, Bonabeau (2009) also found that the applications of collective intelligence for decision making fall in two categories: idea generation and idea evaluation. For idea generation, a contest provides a proper incentive to aggregate problem solvers. Our studies show that innovation contests are very powerful in the aggregation of collective intelligence. On average, a contest can aggregate the effort of over 700 solvers for a naming project. For idea evaluation, a prediction market is also a very powerful tool to aggregate external evaluators. We find that US$2.00 can facilitate the involvement of over 100 evaluators, on average. In actuality, a prediction market can also be viewed as a special form of contest, since the participants of both are competing at intelligent work for a prize. The only difference here is that problem solvers provide solutions to solve problems, while external evaluators provide predictions to aid in evaluation. From this point of view, our study provides empirical proof of the outstanding aggregation power of contests, both in terms of idea generation and idea evaluation.

Watkins (2007) suggests a prediction market is a proper mechanism to aggregate collective intelligence for the open evaluation of candidates, however, very few applications of this sort, or associated studies, exist (Bonabeau 2009). We uniquely examine such a system in practice using empirical data. After the crowd’s intelligence is collectively aggregated, we take a further look into other complementary components, such as the interactions between individuals and solutions. A comprehensive framework explaining how collective intelligence can be aggregated and used for idea evaluation is a unique contribution of our work.

5.2 Overall Managerial Implications

Our studies provide managerial implications to three groups of stakeholders: innovation seekers, problem solvers and market operators.

5.2.1 Innovation Seekers

Any innovation seekers are allowed to launch innovation contests. If the innovation seeker is a large firm, it may collect diversified ideas from its employees and customers by establishing a self-hosted platform, such as Dell's IdeaStrom. If the firm does not have a large number of employees or customers, and it needs innovative ideas from outside its boundaries, launching a contest in an online market is a good option.

If the firm decides to launch a contest in an online market, our performance-forecasting model can help innovation seekers to anticipate the performance of their contest before it is launched. However, for some types of projects, such as aircraft innovations, open innovation contests may not perform well since the requisite expertise is rare. Besides, due to the current prepaid-prize policy, it is risky to launch high value contests since a successful solution cannot be guaranteed. For this type of project, procurement auctions may be a better option.

Innovation seekers need to acknowledge that the contest performance is not only decided by the contest design parameters (e.g. prize, duration, and project description) and project complexity, but that it is also affected by its relationship with other competing contests. In particular, although extending the duration is costless, most contests will not capture new solutions after 30 days. After a contest is launched, the seeker should start evaluating solutions and issuing informative feedback to solvers, guiding them toward the preferred solution, advising them on how to make improvements. However, many solvers choose to submit their solutions late, even though they have registered for the contest very early. This suggests that the innovation seeker should not choose a winner too early in the process.

If there are lots of solutions and the evaluation cost is high, seekers may consider having customers or other people aid in the evaluation process. If the innovation strategy is customer driven, seekers can have customers evaluate and make a selection decision, without providing any specific criteria. Alternatively, seekers can aggregate evaluators from outside the firm using a prediction market mechanism to evaluate all solutions, based on specific criteria. It would be a particularly good idea to have each individual evaluate a small number of ideas at a time. Although this type of public evaluation may save the decision cost for evaluators, we recommend the use of a private evaluation process, in order to avoid a herding effect, which is harmful to the evaluation performance. This is because herd behavior eliminates the independence of each evaluation result, eliminating the benefits of collective intelligence as an aggregator of market information.

5.2.2 Problem Solvers

To a problem solver, the most important concern is how he can win the competition. A problem solver has to make several decisions throughout a contest. First, he needs to consider which contest to participate in, in order to maximize his potential to win. We find that expertise is an important predictor for future winning probabilities, however, it only contributes slightly to the final winning result, since there are so many competitors and the idea-based part of the project is highly random, which makes it very difficult to predict winning results. For expertise-based projects, the solvers that have high levels of expertise have reason to be confident, however, if the project is highly ideation-based, such as a naming project, then past performance or expertise is not very relevant to future performance.

Once a solver decides to participate in a contest, the next decision he needs to make is when to submit his initial solution. If he submits early, he will have more time to communicate with the seeker and to get feedback. Nevertheless, since most markets make solutions visible to all competitors, a solver's idea may be imitated. Although we find that winners are more likely to be the solvers who submit very early or very late, we could not suggest that a solver should take either of these approaches in particular, since a solver submitting late may have spent more effort on their solution, as opposed to having just waited.

5.2.3 Market Operators

The main concern of a market operator in this setting is to make the market attractive to innovation seekers, as well as to make it sustainable. Hence, the market operators need to take whatever actions they can to satisfy innovation seekers' needs. We suggest following considerations.

Contest Performance Forecasting

By using the number of solvers as a performance proxy, we provide a performance model that can accurately predict how many solvers will be attracted to a contest before it is launched. Ideally, it would be great if the market could predict the number of solutions per day, in advance. This action can largely lower the uncertainty for a new innovation seeker, encouraging them to the market. As a consequence, the market can attract more innovation seekers.

However, it is not better to disclose this information to problem solvers since it may make some early-joining solvers perceive that there are many competitors, and, as a result, drive down their perceived value of the contest prize. In terms of late-joining solvers, the forecast number of solvers is very close to the number competitors they will have observed already, thus the forecasting information would not be especially useful.

Use of Feedback System

We show that the use of a feedback system can improve the overall performance of a contest and that it is needed to facilitate better communication between innovation seekers and problem solvers. Although many markets provide feedback systems, the system often has very limited functionality. Thus, the feedback information could be displayed in a more efficient manner. For example, it is necessary to encourage innovation seekers to send feedback to as many solutions as possible. In a feedback message, it is better to include detailed information on how to meet the seeker’s needs, rather than simply saying "I don't like it."

Adaptable Visibility of Solutions

By default, solutions are set to be publicly visible in most markets. The motivation for this strategy is to make it easier for new innovation seekers to perceive the value of the market. However, this strategy allows later joining solvers to imitate earlier submissions. Imitation lowers the diversity of ideas, which is associated with lower innovativeness (Terwiesch and Ulrich 2009). Thus, we would recommend hiding the solutions from public view while the contest remains open. In order to maintain the attractiveness of the market for new innovation seekers, operators can force all solutions to be made publicly visible after the contest has ended.

Market Sustainability

When a contest reaches the point of saturation in terms of the number of solvers, the probability of winning for each problem solver becomes extremely low, on average. This is also true for solvers with high levels of expertise. We find that some solvers with higher level of expertise tend to choose contests with smaller prizes. This is not likely to be an intuitive finding for innovation seekers. However, if a solver is incapable of earning enough prize money to cover his overall effort, this labor market is not a feasible place for a solver to earn a living. Unfortunately, this is true for most problem solvers in these online markets. To increase sustainability, or to make the problem solvers perceive that they can earn higher value, thereby making them more likely to revisit the market, it is necessary to satisfy solvers' needs, beyond specifying a prize. For example, the ZBJ Network found that the primary motivation of many solvers was to learn and practice, not to make money. To align with this goal, operators can provide a virtual honor system, wherein some virtual honors or points will be awarded to top solutions. Although a prize is only rewarded to very few solvers, a greater number of solvers could earn virtual honors and their efforts could then be acknowledged, making them more willing to come back to the market in the future.

Open Evaluation System

We have shown that ZBJ's open evaluation system works better pure chance. At the least, it can provide a recommended order of evaluation if there are lots of solutions. Our examination shows that the accuracy of the open evaluation result needs to be improved. We suggest designing the open evaluation system with following requirements:

- Focused evaluation. At any given time, an evaluator should only need to compare a small number of solutions. This way, each evaluation result will be more accurate since evaluators can concentrate more and the decision cost will be lower. As a consequence, the evaluation result should be more accurate.
- Private evaluation. We have shown that a herding effect is likely present and that it is harmful to the open evaluation performance. In order to aggregate enough private information from evaluators, an independent evaluation process or private evaluation process is necessary.
- Well-organized criteria. We have shown that explicit criteria, whether in visual format or textual format, are helpful with ZBJ's open evaluation data. However, ZBJ does not list these criteria separately, which makes it difficult for the evaluators to find. We suggest that market operators should list the criteria clearly, so the evaluators can better use the information. The background information helps problem solvers understand what the innovation seeker is doing; however, we find this information impedes the accuracy of open evaluation. We would suggest hiding the seeker’s background information in the project description from external evaluators, in the open evaluation system.

The above implications provide detailed guidelines to innovation seekers, problem solvers and market operators. They make the processes of ideation generation and idea evaluation in online markets more systematic. By aggregating collective intelligence efficiently, the cost of adopting open innovation can be dramatically reduced. Innovation diffusion theory (Attewell 1992; Fichman and Kemerer 1999) suggests that our findings can make open innovation more popular in the future.

5.3 Future Research Suggestions

Since the year of 2006, open innovation contests have received increased academic attention. However, this research area is still new and many topics are worth further investigation.

5.3.1 Reverse Auction vs. Innovation Contest

Nearly all types of projects can be completed using reverse auctions or innovation contests. Both modes require compensations and result in solutions, but they have many differences. Viewing them as business models, solvers of contests need to deliver solutions before receiving any compensation, while solvers of auctions only need to send qualification signals and proposals before receiving any compensation. Viewed from the perspective of information economics, there is no information problem for innovation seekers in contests, since all solutions are submitted with no hidden information, while information is asymmetric in reverse auctions, since seekers do not know what will be received before a provider is chosen. Viewed from the risk management perspective, seekers in auctions do not need to pay before a good provider is found, while seekers in contests have to pay before any solution is submitted. Hence, the contest mode is sometimes much more risky than the auction mode. With the acknowledgement of these differences, it is still difficult to choose which mode is optimal for different types of projects. More studies are needed to minimize the decision cost.

5.3.2 Multiple-stage Contest

Most studies to date have focused on one-stage contests, where solvers only need to compete once. However, one-stage contests are becoming less applicable for high value, complex projects. George Tsipolitis, VP of TopCoder, shared with us that in order to operate high value projects, TopCoder often has to split the whole project into small pieces and run each as a sequential contest. Moldovanu and Sela (2006) provide a theoretical model for multi-stage contests; however, very little is known about these types of contests in practice. More studies are needed in this stream.

5.3.3 Payment Policy

Most contest markets require seekers to deposit the full prize amount in advance and the prize is non-refundable. This policy avoids a moral hazard from the innovation seeker's perspective, wherein they could potentially just take the idea and not reward any solvers. This approach provides solvers with a guarantee that, if they win, they will be paid. However, under this payment policy, seekers will have to bear the full risk of failure, especially for complex projects, which may attract too few solutions or only poor solutions. As a result, seekers may not be willing to launch cutting-edge, complex or high value projects under this 'prize non-refundable policy'. The cutting-edge and high value innovation projects are those that are most important to large firms and the most profitable to market owners. Thus, the formulation of appropriate policies to provide adequate insurance to both seekers and solvers will be an interesting question going forward.

5.3.4 Open Evaluation

We take the first step toward understanding the feasibility of employing an open evaluation system to reduce the organizational burden of evaluating the large amount of solutions generated by an open innovation process. Our work opens several avenues for future investigation. First of all, prior studies show that a higher disparity in the evaluation can deliver savings on the cost of discrimination. However, we find evidence of a strong herding effect in public voting, such that the higher disparity is associated with worse open evaluation performance. It will be interesting to see whether the higher disparity will lead to better performance in private voting, where there is no potential for a herding effect. Second, in all the contests that we observed in our study, the evaluation prize was set to 2% of the contest prize. The factors driven by either the evaluation prize or the contest prize are thus confounded. For instance, a higher evaluation prize will be likely to attract more evaluators, and a higher contest prize will be likely to attract more solvers. This makes it impossible to study the effect of the evaluator count while totally controlling for the solution count. It would be interesting to study the effect of the contest prize on open evaluation performance in the absence of such confounding. Third, in the current system, the external evaluator only casts one vote for the winner, which is equivalent to asking the evaluator to disclose their top selection. It would be interesting to consider a more complex system, where the external evaluator can rank their top two or more selections. Other possible extensions include testing for the existence of selection bias amongst the external evaluators, giving different weights to the votes of external evaluators based on their expertise levels and using an alternative benchmark for open evaluation, instead of the internal evaluation results.

In conclusion, we sincerely hope that the above future research ideas can draw the attention of scholars who are interested in open innovation.

REFERENCES CITED

Agichtein, E., Brill, E., and Dumais, S. 2006. "Improving Web Search Ranking by Incorporating User Behavior Information," ACM, pp. 19-26.

Ainslie, G. 1992. *Picoeconomics: The Strategic Interaction of Successive Motivational States within the Person*. Cambridge University Press Cambridge.

Akerlof, G.A. 1991. "Procrastination and Obedience," *The American Economic Review* (81:2), pp. 1-19.

Allio, R. 2004. "Ceo Interview: The Innocentive Model of Open Innovation," *Strategy & Leadership* (32:4), pp. 4-9.

Archak, N., and Sundararajan, A. 2009. "Optimal Design of Crowdsourcing Contests," *ICIS 2009 Proceedings*), p. 200.

Attewell, P. 1992. "Technology Diffusion and Organizational Learning: The Case of Business Computing," *Organization Science* (3:1), pp. 1-19.

Baeza-Yates, R., Hurtado, C., Mendoza, M., and Dupret, G. 2005. "Modeling User Search Behavior,").

Baeza-Yates, R., and Ribeiro-Neto, B. 1999. *Modern Information Retrieval*. ACM press New York.

Banerjee, A.V. 1992. "A Simple Model of Herd Behavior," *The Quarterly Journal of Economics* (107:3), p. 797.

Banker, R.D., and Iny, H. 2008. "Importance of Measures of Past Performance: Empirical Evidence on Quality of E-Service Providers," *Contemporary Accounting Research* (25:2), Summer2008, pp. 307-337.

Bapna, R., Chang, S., Goes, P., and Gupta, A. 2009. "Overlapping Online Auctions: Empirical Characterization of Bidder Strategies and Auction Prices," *MIS Quarterly* (33:4), pp. 763-783.

Barras, R. 1986. "Towards a Theory of Innovation in Services," *Research Policy* (15:4), pp. 161-173.

Berndt, A. 2009. "An Update from the Project 10^100 Team," *The Official Google Blog* (http://googleblog.blogspot.com/2009/01/update-from-project-10100-team.html).

Berry, M.W., Dumais, S.T., and O'Brien, G.W. 1995. "Using Linear Algebra for Intelligent Information Retrieval," *SIAM review* (37:4), pp. 573-595.

Bikhchandani, S., Hirshleifer, D., and Welch, I. 1992. "A Theory of Fads, Fashion, Custom, and Cultural Change as Informational Cascades," *Journal of political Economy*), pp. 992-1026.

Bikhchandani, S., Hirshleifer, D., and Welch, I. 1998. "Learning from the Behavior of Others: Conformity, Fads, and Informational Cascades," *The Journal of Economic Perspectives* (12:3), pp. 151-170.

Blanchard, O.J., and Diamond, P.A. 1992. "The Flow Approach to Labor Markets." National Bureau of Economic Research Cambridge, Mass., USA.

Bonabeau, E. 2009. "Decisions 2.0: The Power of Collective Intelligence," *MIT Sloan Management Review* (Jan), Jan 9th.

Borle, S., Boatwright, P., and Kadane, J.B. 2006. "The Timing of Bid Placement and Extent of Multiple Bidding: An Empirical Investigation Using Ebay Online Auctions," *Statistical Science*), pp. 194-205.

Bothos, E., Apostolou, D., and Mentzas, G. 2009. "Collective Intelligence for Idea Management with Internet-Based Information Aggregation Markets," *Internet Research* (19:1), pp. 26-41.

Brown, J.S., and Duguid, P. 1991. "Organizational Learning and Communities of Practice: Toward a Unified View of Working, Learning, and Innovation," *Organization Science* (2:1), pp. 40-57.

Bstieler, L. 2005. "The Moderating Effect of Environmental Uncertainty on New Product Development and Time Efficiency*," *Journal of Product Innovation Management* (22:3), pp. 267-284.

Chao, R.O., and Kavadias, S. 2008. "A Theoretical Framework for Managing the New Product Development Portfolio: When and How to Use Strategic Buckets," *Management Science* (54:5), pp. 907-921.

Chapelle, O., and Zhang, Y. 2009. "A Dynamic Bayesian Network Click Model for Web Search Ranking," ACM, pp. 1-10.

Che, Y.K., and Gale, I. 2003. "Optimal Design of Research Contests," *American Economic Review* (93:3), pp. 646-671.

Chesbrough, H. 2003. *Open Innovation: The New Imperative for Creating and Profiting from Technology*. Harvard Business Press.

Cohen, W.M., and Levinthal, D.A. 1990. "Absorptive Capacity: A New Perspective on Learning and Innovation," *Administrative Science Quarterly* (35:1), pp. 128-152.

Craswell, N., Zoeter, O., Taylor, M., and Ramsey, B. 2008. "An Experimental Comparison of Click Position-Bias Models," ACM, pp. 87-94.

Dahan, E., and Mendelson, H. 2001. "An Extreme-Value Model of Concept Testing," *Management Science* (47:1), pp. 102-116.

Dasgupta, A., and Nti, K.O. 1998. "Designing an Optimal Contest," *European Journal of Political Economy* (14:4), pp. 587-603.

Davidson, D., Suppes, P., and Siegel, S. 1957. "Decision Making; an Experimental Approach,").

Di Gangi, P.M., and Wasko, M. 2009. "Steal My Idea! Organizational Adoption of User Innovations from a User Innovation Community: A Case Study of Dell Ideastorm," *Decision Support Systems* (48:1), pp. 303-312.

Dupret, G., and Piwowarski, B. 2008. "A User Browsing Model to Predict Search Engine Click Data from Past Observations," ACM, pp. 331-338.

Edwards, W. 1956. "Reward Probability, Amount, and Information as Determiners of Sequential Two-Alternative Decisions," *Journal of Experimental Psychology* (52:3), p. 177.

Fama, E. 1970a. "Efficient Capital Markets: A Review of Theory and Empirical Work," *Journal of finance* (25:2), pp. 383-417.

Fama, E.F. 1970b. "Efficient Capital Markets: A Review of Theory and Empirical Work," *The journal of Finance* (25:2), pp. 383-417.

Faraj, S., and Sproull, L. 2000. "Coordinating Expertise in Software Development Teams," *Management Science* (46:12), p. 1554.

Fichman, R.G., and Kemerer, C.F. 1999. "The Illusory Diffusion of Innovation: An Examination of Assimilation Gaps," *Information Systems Research* (10:3), p. 255.

Filippini, R., Salmaso, L., and Tessarolo, P. 2004. "Product Development Time Performance: Investigating the Effect of Interactions between Drivers," *Journal of Product Innovation Management* (21:3), pp. 199-214.

Fleder, D., and Hosanagar, K. 2009. "Blockbuster Culture's Next Rise or Fall: The Impact of Recommender Systems on Sales Diversity," *Management Science* (55:5), pp. 697-712.

Fleder, D.M., and Hosanagar, K. 2007. "Recommender Systems and Their Impact on Sales Diversity," ACM, pp. 192-199.

Gassmann, O. 2006. "Opening up the Innovation Process: Towards an Agenda," *R&D Management* (36:3), pp. 223-228.

Gini, C. 1912. "Variability and Mutability," *contribution to the study of statistical distributions and relations*).

Girotra, K., Terwiesch, C., and Ulrich, K. 2010. "Idea Generation and the Quality of the Best Idea," *Management Science*).

Glazer, A., and Hassin, R. 1988. "Optimal Contests," *Economic Inquiry* (26:1), pp. 133-143.

Golec, J., and Tamarkin, M. 1998. "Bettors Love Skewness, Not Risk, at the Horse Track," *Journal of political Economy*), pp. 205-225.

Google. 2009. "Google Project 10^100 Faq," *Google Project 10^100*), p. http://www.project10tothe100.com/faq.html.

Green, L., and Myerson, J. 2004. "A Discounting Framework for Choice with Delayed and Probabilistic Rewards," *Psychological Bulletin* (130:5), pp. 769-792.

Green, P.E., and Krieger, A.M. 1985. "Models and Heuristics for Product Line Selection," *Marketing Science*), pp. 1-19.

Gregg, D.G., and Walczak, S. 2003. "E-Commerce Auction Agents and Online-Auction Dynamics," *Electronic Markets* (13:3), pp. 242-250.

Hart, S.H., Moncrief, W.C., and Parasuraman, A. 1989. "An Empirical Investigation of Salespeople's Performance, Effort and Selling Method During a Sales Contest," *Journal of the Academy of Marketing Science* (17:1), p. 29.

Harvey, N., and Bolger, F. 2001. "Collecting Information: Optimizing Outcomes, Screening Options, or Facilitating Discrimination?," *Quarterly Journal of Experimental Psychology: Section A* (54:1), pp. 269-301.

Hayek, F. 1945. "The Use of Knowledge in Society," *The American Economic Review*), pp. 519-530.

Hippel, E. 2005. "Democratizing Innovation," *Cambridge, Mass. MIT Press. S* (17), pp. 165-168.

House, E.R. 1978. "Assumptions Underlying Evaluation Models," *Educational Researcher* (7:3), pp. 4-12.

Howe, J. 2006. "The Rise of Crowdsourcing," *Wired Magazine* (14:6), pp. 1-4.

Joachims, T. 2002a. "Evaluating Retrieval Performance Using Clickthrough Data," Citeseer, pp. 12–15.

Joachims, T. 2002b. "Optimizing Search Engines Using Clickthrough Data," ACM, pp. 133-142.

Joachims, T., Granka, L., Pan, B., Hembrooke, H., and Gay, G. 2005. "Accurately Interpreting Clickthrough Data as Implicit Feedback," ACM, pp. 154-161.

Kagel, J.H., Battalio, R.C., and Green, L. 1995. *Economic Choice Theory: An Experimental Analysis of Animal Behavior*. Cambridge Univ Pr.

Kalra, A., and Shi, M. 2001. "Designing Optimal Sales Contests: A Theoretical Perspective," *Marketing Science*), pp. 170-193.

Karagozoglu, N., and Brown, W. 1993. "Time-Based Management of the New Product Development Process," *Journal of Product Innovation Management* (10:3), pp. 204-215.

Kohli, R., and Sukumar, R. 1990. "Heuristics for Product-Line Design Using Conjoint Analysis," *Management Science*), pp. 1464-1478.

Krishnan, V., and Ulrich, K.T. 2001. "Product Development Decisions: A Review of the Literature," *Management Science*), pp. 1-21.

Lakhani, K.R., Jeppesen, L.B., Lohse, P.A., and Panetta, J.A. 2007. "The Value of Openness in Scientific Problem Solving," *Boston: Harvard University*).

Lanzetta, J.T., and Kanareff, V.T. 1962. "Information Cost, Amount of Payoff, and Level of Aspiration as Determinants of Information Seeking in Decision Making," *Behavioral Science* (7:4), pp. 459-473.

Laursen, K., and Salter, A. 2006. "Open for Innovation: The Role of Openness in Explaining Innovation Performance among U.K. Manufacturing Firms," *Strategic Management Journal* (27:2), pp. 131-150.

Lazear, E.P., and Rosen, S. 1981. "Rank-Order Tournaments as Optimum Labor Contracts," *The Journal of Political Economy* (89:5), pp. 841-864.

Lévy, P. 1997. *Collective Intelligence: Mankind's Emerging World in Cyberspace*. Perseus Books Cambridge, MA, USA.

Liu, D., Geng, X., and Whinston, A. 2007. "Optimal Design of Consumer Contests," *Journal of Marketing* (71:4), pp. 140-155.

Loch, C.H., De Meyer, A., and Pich, M.T. 2006. *Managing the Unknown: A New Approach to Managing High Uncertainty and Risk in Projects*. John Wiley & Sons.

Mataric, M. 1993. "Designing Emergent Behaviors: From Local Interactions to Collective Intelligence," pp. 423–441.

McAfee, R. 1993. "Mechanism Design by Competing Sellers," *Econometrica: Journal of the econometric society* (61:6), pp. 1281-1312.

McBride, R.D., and Zufryden, F.S. 1988. "An Integer Programming Approach to the Optimal Product Line Selection Problem," *Marketing Science* (7:2), pp. 126-140.

Moldovanu, B., and Sela, A. 2001. "The Optimal Allocation of Prizes in Contests," *American Economic Review*), pp. 542-558.

Moldovanu, B., and Sela, A. 2006. "Contest Architecture," *Journal of Economic Theory* (126:1), pp. 70-96.

Mortensen, D.T. 1986. "Job Search and Labor Market Analysis," *Handbook of labor economics* (2), pp. 849-919.

Murphy, W., Dacin, P., and Ford, N. 2004. "Sales Contest Effectiveness: An Examination of Sales Contest Design Preferences of Field Sales Forces," *Journal of the Academy of Marketing Science* (32:2), pp. 127-143.

Murthy, P., and Mantrala, M.K. 2005. "Allocating a Promotion Budget between Advertising and Sales Contest Prizes: An Integrated Marketing Communications Perspective," *Marketing Letters* (16:1), pp. 19-35.

Nembhard, D.A., and Osothsilp, N. 2002. "Task Complexity Effects on between-Individual Learning/Forgetting Variability," *International journal of industrial ergonomics* (29:5), pp. 297-306.

Newell, B.R., Rakow, T., Weston, N.J., and Shanks, D.R. 2004. "Search Strategies in Decision Making: The Success of "Success"," *Journal of Behavioral Decision Making* (17:2), pp. 117-137.

Ockenfels, A., and Roth, A. 2001. "Strategic Late-Bidding in Continuous-Time Second-Price Internet Auctions,").

Ockenfels, A., and Roth, A.E. 2006. "Late and Multiple Bidding in Second Price Internet Auctions: Theory and Evidence Concerning Different Rules for Ending an Auction," *Games and Economic Behavior* (55:2), pp. 297-320.

Page, L., Brin, S., Motwani, R., and Winograd, T. 1999. "The Pagerank Citation Ranking: Bringing Order to the Web,").

Richardson, M., Dominowska, E., and Ragno, R. 2007. "Predicting Clicks: Estimating the Click-through Rate for New Ads," ACM, p. 530.

Rorty, R. 1979. "Philosophy and the Minor of Nature," *Princeton UP*).

Rossi, P.H., Lipsey, M.W., and Freeman, H.E. 2004. *Evaluation: A Systematic Approach*. Sage Publications, Inc.

Roth, A., and Ockenfels, A. 2002. "Last-Minute Bidding and the Rules for Ending Second-Price Auctions: Evidence from Ebay and Amazon Auctions on the Internet," *American Economic Review* (92:4), pp. 1093-1103.

Runkel, M. 2006. "Optimal Contest Design, Closeness and the Contest Success Function," *Public Choice* (129:1), pp. 217-231.

Runkel, M., Studies, C.f.E., and Wirtschaftsforschung, I.-I.f. 2003. *Optimal Contest Design When the Designer's Payoff Depends on Competitive Balance*. CES.

Salton, G., and Buckley, C. 1997. "Improving Retrieval Performance by Relevance Feedback," *Readings in information retrieval*), pp. 355–364.

Sheremeta, R.M. 2009. "Contest Design: An Experimental Investigation," *Economic Inquiry*).

Smith, V. 1982. "Microeconomic Systems as an Experimental Science," *The American Economic Review* (72:5), pp. 923-955.

Snir, E., and Hitt, L. 2003. "Costly Bidding in Online Markets for It Services," *Management Science* (49:11), pp. 1504-1520.

Sonsino, D., Benzion, U., and Mador, G. 2002. "The Complexity Effects on Choice with Uncertainty 肸 xperimental Evidence*," *The Economic Journal* (112:482), pp. 936-965.

Spann, M., and Skiera, B. 2003. "Internet-Based Virtual Stock Markets for Business Forecasting," *Management Science*), pp. 1310-1326.

Spann, M., and Skiera, B. 2009. "Sports Forecasting: A Comparison of the Forecast Accuracy of Prediction Markets, Betting Odds and Tipsters," *Journal of Forecasting* (28:1), pp. 55-72.

Stufflebeam, D.L., and Webster, W.J. 1980. "An Analysis of Alternative Approaches to Evaluation," *Educational Evaluation and Policy Analysis* (2:3), pp. 5-20.

Surowiecki, J. 2004. *The Wisdom of Crowds: Why the Many Are Smarter Than the Few and How Collective Wisdom Shapes Business, Economies, Societies, and Nations*. Random House, Inc.

Swink, M., Talluri, S., and Pandejpong, T. 2006. "Faster, Better, Cheaper: A Study of Npd Project Efficiency and Performance Tradeoffs," *Journal of Operations Management* (24:5), pp. 542-562.

Terwiesch, C., and Ulrich, K. 2009. *Innovation Tournaments: Creating and Selecting Exceptional Opportunities*. Harvard Business School Pr.

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

Tucker, C., and Zhang, J. 2011. "How Does Popularity Information Affect Choices? A Field Experiment," *Management Science, forthcoming*).

Tversky, A., and Kahneman, D. 1974. "Judgment under Uncertainty: Heuristics and Biases," *Science* (185:4754), pp. 1124-1131.

Vadovic, R. 2009. "Early, Late, and Multiple Bidding in Internet Auctions," *Working Papers*).

Van de Ven, A.H. 1986. "Central Problems in the Management of Innovation," *Management Science* (32:5), pp. 590-607.

VanGundy, A.B. 1988. *Techniques of Structured Problem Solving*. Van Nostrand Reinhold Co.

Walsh, J.P. 1995. "Managerial and Organizational Cognition: Notes from a Trip Down Memory Lane," *Organization Science* (6:3), pp. 280-321.

Watkins, J. 2007. "Prediction Markets as an Aggregation Mechanism for Collective Intelligence,").

Williamson, O. 1998. "Transaction Cost Economics: How It Works; Where It Is Headed," *De Economist* (146:1), pp. 23-58.

Yang, J., Adamic, L., and Ackerman, M. 2008. "Crowdsourcing and Knowledge Sharing: Strategic User Behavior on Taskcn," ACM, pp. 246-255.

Yang, Y., Chen, P., and Pavlou, P. 2009. "Open Innovation: An Empirical Study of Online Contests," *ICIS 2009 Proceedings | International Conference on Information Systems*), p. 13.

Yitzhaki, S. 1979. "Relative Deprivation and the Gini Coefficient," *The Quarterly Journal of Economics*), pp. 321-324.

Yoo, Y., and Hill, T. 2010. "Managing Open Innovation: How and What to Open," *IBIT Report*:Februrary).

Zeithammer, R. 2006. "Forward-Looking Bidding in Online Auctions," *Journal of Marketing Research* (43:3), pp. 462-476.

APPENDIX
FEEDBACK IMPACT

To empirically examine whether feedback encourages solvers to exert more effort, we collected feedback data from Zhubajie.com and did an experiment at TaskCN.com. Both websites are among the largest online contest markets in the world.

Empirical proof from Zhubajie.com

Zhubajie.com provides a well-designed feedback system, which allows seekers to easily leave feedback on each submitted solution. In total, we collected data on 1,461contests between 2010 and 2011, and there were 102,813 solutions submitted by 77,170 solvers, in total. Among all 1,461 contests, the seekers of 1,024 (70.1%) contests used feedback at least once, to respond to 8.9% of solvers (6,854 out of 77,170 solvers). On average, each seeker sent 8.9 pieces of feedback during the contest. A solver that did not receive any feedback contributes 1.27 solutions, on average, while a solver that does receive feedback contributes 1.87 solutions, on average. For a solver, the Pearson correlation coefficient between receiving feedback and contributing additional solutions is 0.1709, with a significance of $p<0.001$. Although the content of feedback information is diversified and not controlled, these results show that sending feedback can obviously encourage solvers to exert more effort than the equilibrium.

Field Experiment Evidence from TaksCN.com

We also did an experiment by launching a contest project at TaskCN.com, which is also one of the largest online contest markets in the world. The contest set a single winner prize of ￥500 (around the average market price, approximately equaling US$78). The project was to suggest a creative LOGO for a website. After a duration of 15 days, the contest attracted 34 solvers. Before we sent any feedback, each solver submitted just one prototype.

This experiment gives us the advantage of fully controlling the feedback quality. To best show the power of feedback, we sent high quality information and always told each solver how to improve the submitted solution to make us more satisfied. This is different from Zhubajie feedback data, which also includes negative feedback information, such as “I don’t like it.” Finally, we sent feedback 38 times, in total, and then received 43 improved solutions. The response rate was 100% and each piece of feedback generated 1.13 improved prototypes, on average.

The above results from two different online contest markets show that sending feedback can significantly encourage solvers to provide higher quality solutions. Further, the feedback quality also has a positive impact on the solution quality.