Does the crowd care?

Monetary Rewards, ESG Goals and Participation in Crowdsourcing Competitions

Radboud Universiteit

Master's thesis

Sjors Vlaanderen

s4622812

Supervisor: Dr. S. Brenk

Second examiner: Prof. dr. B. Hillebrand

Radboud University, Faculty of Management

MSc Innovation & Entrepreneurship

Nijmegen, 2024

Contents

Abstract

Harnessing the power of collective intelligence through crowdsourcing competitions requires attracting a diverse range of participants. This thesis investigates how monetary rewards and Environmental, Social, and Governance (ESG) goals impact participation, using Kaggle as a case study. By analyzing 380 competitions and utilizing machine learning (ML) to evaluate ESG aspects, this research highlights that both financial incentives and ESG goals significantly increase the number of participants. Interestingly, high ESG scores can mitigate the impact of monetary rewards, indicating a complex interplay between extrinsic and intrinsic motivators. These findings provide valuable insights for organizations aiming to design effective crowdsourcing competitions by taking into account moderation effects of ESG goals on monetary rewards. This study demonstrates the potential of applying ML techniques to process textual data on crowdsourcing competitions. The research approach offers new insights in the different contextual factors that influence participation, while providing evidence of the applicability of ML methods to access unexplored data regarding crowdsourcing competitions.

1 Introduction

In recent years, both corporate entities and non-governmental organizations are increasingly resorting to crowdsourcing platforms to explore innovative approaches to data-centric challenges (Aitamurto & Chen, 2017; Yan et al., 2017). Crowdsourcing is a participatory online practice where an individual, institution, not-for-profit organization, or company invites a diverse group of individuals with varying levels of expertise and numbers through an open call to voluntarily engage in a task or project (Estellés-Arolas & Gonzalez-Ladron-de-Guevara, 2012). Crowdsourcing competitions (also known as innovation competitions or tournament-based crowdsourcing) are challenges where participants compete to provide the best solution for a specific innovation problem (Acar, 2018).

A study on US firms demonstrates that crowdsourcing stands as the most favored mechanism for open innovation, with approximately 60 percent of companies adopting it to address problem-solving challenges (Bagherzadeh et al., 2021). Crowdsourcing facilitates the rapid generation of fresh development prospects that may lie beyond an organization's regular operations, procedures and capabilities. In the case of for-profit entities (corporations) for example, crowdsourcing allows them to condense innovation cycles, thereby sharpening their competitive edge through accelerated introduction of new products and services to the market (Seltzer & Mahmoudi, 2013). This can help induce a 'first-mover advantage', i.e. capitalize on market opportunities ahead of competitors, thereby boosting profitability (Brown & Eisenhardt, 1995). Similarly, shorter innovation cycles typically require less capital investment over a shorter period. This efficiency in resource allocation reduces operational costs and increases profitability (Tidd & Bessant, 2020).

In recent years, the debate has shifted to urge managers to think beyond profits, to their social purpose (Zumente & Bistrova, 2021). Social innovators and entrepreneurs are encouraged to adopt unique methods to achieve sustainable social change by exploring new avenues (Oeij et al., 2019). In addressing societal needs and challenges, managers can benefit from crowdsourcing to develop better solutions for social problems (Füller et al., 2012). That is, previous research has highlighted that this cross-sectoral and multi-stakeholder approach is especially critical for effectively addressing issues related to social innovation. This makes it especially important to inspire a large group of solvers who, by using different approaches based on their values, education, experiences, and goals, can come up with a wider range of solutions (Jeppesen & Lakhani, 2010; Piazza et al., 2024).

The positive relation with the level of participation is not just limited to social innovation, but goes for innovation goals in general (Daradkeh, 2022; Meng et al., 2021). That is, Wang and Chen (2023) find that more submissions increase the likelihood of high-quality solutions, as participants strive to stand out, leading to more refined and innovative ideas. This competitive environment fosters creativity and critical thinking (Martinez, 2015). Therefore, to maximize the benefits of crowdsourcing platforms, organizations must successfully host competitions with high participation levels by offering incentives that boost intrinsic or extrinsic motivation.

Competition hosts can boost intrinsic motivation by emphasizing the positive societal impact of the competition. That is, when participants see a direct connection between their efforts and their personal values or the greater good, intrinsic motivation increases (Sheldon & Elliot, 1999). In a recent study, Piazza et al. (2024) show that competitors with prosocial motivations are likely to self-select for competitions hosted to address COVID-19-related problems. This suggests that hosts can leverages prosocial motivations, encouraging individuals who are driven by altruism and a desire to contribute to societal welfare to participate. In a business context, societal benefits have been demonstrated to enhance employees' intrinsic motivation, thereby improving their job performance (Lee & Park, 2023). One of the most widely adopted frameworks in business literature for assessing the sustainability of businesses and their activities is ESG (Halid et al., 2023; Hastalona & Sadalia, 2021). ESG, which stands for Environmental, Social, and Governance, provides a thorough framework for evaluating the sustainability and ethical impacts of corporate actions (Martiny et al., 2024).

The connection between ESG in competitions and intrinsic motivation to participate is rooted in self-determination theory (SDT), which emphasizes the role of autonomy, competence, and relatedness in motivating actions that are internally rewarding (Ryan & Deci, 2000). When participants understand how their contributions can make a difference, it enhances their motivation to participate. For instance, highlighting how solutions might contribute to sustainability or social good can align with the personal values of participants (Afuah & Tucci, 2012). Competition hosts can as well boost extrinsic motivation by offering financial incentives such as prizes, grants, or royalties for winning or highly-ranked submissions. According to the principle of operant conditioning from behavioral psychology, behaviors followed by rewards (or reinforcements) are more likely to recur. Monetary rewards act as positive reinforcements, making participants more likely to engage in the competition and strive for success (Skinner, 1965). For competitions, the size and structure of the reward can significantly influence the quantity and quality of participation (Hofstetter et al., 2018).

So far, literature has shown the positive impact of both monetary rewards and ESG goals on participation levels. However, little is known about the interaction effects between these motivating factors in a crowdsourcing context. That is, previous research on crowdsourcing has typically focused on the effectiveness of a single type of incentive, such as monetary or non-monetary rewards, exploring the best way to structure these incentives to maximize participation (Moghaddam et al., 2023). Also, these studies often overlook the diverse motivations of participants, failing to account for the varied and interconnected reasons individuals might choose to participate in such activities (Acar, 2019). Furthermore, a primary reason for a low quality of solutions is the organization's inadequate understanding of what truly motivates the participants (Majchrzak & Malhotra, 2013; Nevo & Kotlarsky, 2020).

This underscores the practical need for research on factors that drive participation. From a managerial (host) perspective, monetary rewards can increase the benefits of crowdsourcing competition by increasing participation whilst also representing a direct cost. As such, Truong et al. (2023) describe the reward setting in crowdsourcing contests as an optimization problem where the host must find the optimal prize amount that maximizes participation while minimizing costs, in which costs are mainly driven by the total prize amount. To set the optimal prize amount, managers as well need to take into account the potential interplay with other factors that might drive participation, such as ESG-related goals.

Thus, this research specifically focuses on the interaction between monetary rewards and ESG goals in their overall effect on participation in crowdsourcing competitions. As such, the main research question is as follows:

To what extent do monetary rewards and ESG goals impact participation in crowdsourcing competitions?

In subsequent chapters, this thesis will delve into the multifaceted relationship between ESG goals and participation in crowdsourcing contests. Chapter 2 provides a comprehensive review of relevant literature, examining prior research on crowdsourcing, motivation theory and participation in crowdsourcing contests. Chapter 3 outlines the methodology employed in this study, detailing the research design, data collection methods, and analytical approach. In Chapter 4, the empirical findings of the research are presented and analyzed, enumerating the key factors influencing participation in crowdsourcing contests. Finally, Chapter 5 offers a discussion of the implications of these findings, along with recommendations for organizations and future research directions in the field.

2 Theoretical framework

Organizations are progressively turning to crowdsourcing contests as a means to solve problems. However, the effectiveness and longevity of these contests rely heavily on ongoing participation and the submission of high-caliber contributions by individuals (Wang et al., 2020). As such, this theoretical framework describes the mechanisms that drive participation in crowdsourcing competitions, specifically (the interplay of) intrinsic and extrinsic motivational factors.

2.1 Crowdsourcing: definition, types, aims and challenges

Crowdsourcing, a term coined by Jeff Howe in a 2006 Wired magazine article (Howe, 2006), represents a shift in how labor is sourced and tasks are accomplished, moving from traditional models to a more distributed framework leveraging the collective intelligence of a "crowd." In its early stages, crowdsourcing was primarily utilized by companies as a tool for specific problem-solving tasks, such as software development and idea generation. Different studies have demonstrated the capability of crowdsourcing to foster social innovation (e.g. Chalmers, 2013; Chesbrough & Di Minin, 2014; Van der Have & Rubalcaba, 2016) in which social innovation refers to a new solution to a societal problem that primarily benefits the wider community (Phills et al., 2008).

Crowdsourcing also comes with challenges - not just for competitors, as well as competition hosts. That is, to tap into the potential of crowdsourcing, competition hosts have to consider and sometimes overcome certain organizational aspects, including dealing with intermediaries (Sieg et al., 2010), problem formulation (Trantopoulos et al., 2017) and other project-related issues when hosting competitions (Lüttgens et al., 2014). Another pivotal factor is *participation*, generally measured by the number of participants. Crowdsourcing depends on a self-selection mechanism where competitors voluntarily choose to participate in a competition (Lakhani et al., 2007). However, as the online landscape expands and offers more choices, participants' time and attention are becoming increasingly scarce resources, influencing where and how they choose to engage (Yin et al., 2022). From a host perspective, reaching a certain level of participation is critical to ensure a broad spectrum of ideas and solutions (Zheng et al., 2011). As Boudreau and Lakhani (2013) underscore, higher participation leverages the collective intelligence and creativity of a more diverse group of participants. This diversity, in turn, improves the quality and robustness of results, thereby fostering innovation (Afuah & Tucci,

2012). Taken together, competitor participation has been proven to be a crucial factor for the success and diversity of crowdsourcing competitions (Frey et al., 2011; Terwiesch & Xu, 2008; Ye & Kankanhalli, 2017; Zhao & Zhu, 2014).

As such, a considerable segment of the literature focuses on understanding what motivates crowd members to participate (Vu et al., 2022; Zheng et al., 2011). In the context of crowdsourcing, motivations vary widely, ranging from seeking recognition, a personal interest in the topic, to opportunities for skill development (Barnes et al., 2015). Previous research indicates that crowdsourcing efficacy often hinges on a balance between intrinsic and extrinsic motivational factors (Kaufmann et al., 2011; Yang et al., 2009; Zheng et al., 2011). Intrinsic motivation involves engaging in an activity for the sheer enjoyment or fulfillment it brings, while extrinsic motivation is driven by the pursuit of a specific, instrumental objective (Reiss, 2012).

2.2 Motivations and incentive mechanisms for participation

Previous studies on crowdsourcing contests have shown that both intrinsic motivation and extrinsic incentives play a role for participants and their efforts in crowdsourcing competitions (Ke & Zhang, 2009; Sun et al., 2012; Zhao & Zhu, 2014). One notable example of an incentive mechanism in crowdsourcing is the possibility gain monetary rewards (Acar, 2018; Terwiesch & Xu, 2008). In crowdsourcing competitions, monetary rewards often take the form of prizes or financial incentives awarded for the best solutions submitted by participants. Cappa et al. (2019) discuss that financial incentives are vital for both attracting new participants and maintaining the engagement of existing ones, emphasizing the significant correlation between monetary compensation and the quality of contributions in crowdsourcing platforms. Adding to this, Li and Hu (2017) argue that monetary rewards effectively enhance commitment to the tasks, thereby influencing both the quantity and quality of the submissions.

Besides extrinsic motivators, participants can be intrinsically motivated (Cappa et al., 2019; Li & Hu, 2017; Zheng et al., 2011). Self-determination theory posits that intrinsic motivation stems from an individual's basic psychological needs, leading to enjoyment and satisfaction derived from enhancing one's competencies (Ryan & Deci, 2000). In the context of crowdsourcing contests, the main drivers of intrinsic motivation for participants are self-achievement, skill development, and enjoyment (Brabham, 2008; Roberts et al., 2006). This type of motivation serves as a psychological catalyst that can increase an individual's effort toward tasks. Furthermore, previous research suggests that when participants perceive their efforts as

contributing to a larger social good, their intrinsic motivation tends to increase (Wang & Chen, 2023). According to Cappa et al. (2019), contributing to social causes through crowdsourcing can raise participants' self-esteem and satisfaction, thereby increasing intrinsic motivation. As such, competition hosts could enhance participants intrinsic motivations by highlighting its social goals.

In management literature, the ESG framework is widely used to capture the social impact of organizations and their activities (Halid et al., 2023; Hastalona & Sadalia, 2021; Tarmuji et al., 2016). ESG, an abbreviation for Environmental, Social, and Governance, provides a thorough approach for evaluating the sustainability and ethical implications of corporate practices (Martiny et al., 2024). This model evaluates a firm's ecological impact ("E"), its engagement with stakeholders and society as a whole ("S"), and the robustness of its governance structures ("G"). Previous research has highlighted that higher ESG performance offers benefits on financial markets, as companies with high ESG scores are generally seen as more attractive to investors (Giese et al., 2019). Furthermore, research has highlighted that ESG-oriented activities as well pay off within the organization, as it contributes to the intrinsic motivation of employees by promoting a sense of purpose and responsibility towards sustainability goals (Halid et al., 2023), higher levels of employee engagement and satisfaction (Tarmuji et al., (2016), and improves overall job performance (Zhang et al., 2024).

Taken together, higher ESG performance fuels intrinsic motivation of employees and by extension, their performance. This raises the question whether these effects as well hold for participation, engagement and performance of agents outside the organization that perform tasks within a crowdsourcing context. At this date, no empirical studies exist that study the link between ESG goals on the intrinsic motivation of competitors in a crowdsourcing context.

Instead, different empirical studies have shown the positive impact of higher scores on ESG outcomes on different crowd initiatives, namely *crowdfunding*. A recent study of Cumming et al. (2024) shows that crowdfunding platforms with higher ESG criteria are more likely to survive over time. Similarly, startups with salient ESG goals raise financing at higher valuations (Mansouri & Momtaz, 2022). Thus, research on crowdfunding show that entities being associated with ESG goals by the 'crowds', are likely to reap better results than entities that have lower ESG scores. Translating these findings to the context of crowdsourcing, highlighting the positive ESG impact of a particular crowdsourcing context could be a potential way for competition host to boost intrinsic motivation to participate.

That is, signaling ESG in the task description and instructions can be a way for competition hosts to attract more competitors, thereby increasing participation. Previous research (Gillier et al., 2018; Steils & Hanine, 2019) describes task instructions and specifically the *framing* of the problem as important ways to incentivize participation. The reason is that participation in a crowdsourcing contest is primarily influenced by the individual's personal intentions (Alam & Campbell, 2017; Brabham, 2008). According to expectancy theory, when reading task instructions, individuals assess how important the rewards are to them and consider the likelihood of achieving high performance if they invest effort (Vroom, 1964). As such, the way task instructions are written serves as a significant communication cue that impacts an individual's cognitive processes (Yin et al., 2022). In the context of crowdsourcing microtasks, research has shown that framing tasks to highlight their positive social impact positively influences work performance on Amazon's Mechanical Turk (Rogstadius et al., 2011). The study found that tasks framed as helping others significantly improved accuracy compared to those framed as for-profit. The findings imply that emphasizing the social value of tasks can be a powerful strategy to boost quality in crowdsourcing environments.

Taken together, empirical findings on what motivates 'the crowd' can offer practical insights for competition hosts. Indeed, intrinsic motivation and external incentives are commonly employed as motivational tactics across various fields (Liang et al., 2018). Although prior research on crowdsourcing has primarily examined their effects separately, Liang et al. (2018) point out that these factors often exist simultaneously in real-world applications. While both have demonstrated positive outcomes independently, they may not always reinforce each other. In fact, certain research indicates that external incentives might diminish the beneficial impact of intrinsic motivation (Cerasoli et al., 2014; Deci et al., 1999; Murayama et al., 2010). Piazza et al. (2024) find that other intrinsic motivations enhance, while extrinsic motivations undermine, the positive effect of prosocial motivations on participation. The authors explore the effects of intrinsic and extrinsic motivations as moderators on the relationship between prosocial motivations and solvers' participation in crowdsourcing competitions for social innovation. They find that intrinsic motivations enhance the positive impact of prosocial motivations on participation, meaning that competitors who enjoy the activity or seek personal satisfaction are more likely to engage in prosocially motivated crowdsourcing competitions. Conversely, extrinsic motivations, such as monetary rewards, can undermine this relationship by shifting competitors' focus from altruistic goals to the external incentives, thereby reducing their intention to participate (Piazza et al., 2024). Liang et al. (2018) find certain 'crowding out'

effects on task efforts of solvers in a crowdsourcing context. The authors find that intrinsic motivation weakens when the level of extrinsic incentives is high. However, the measurement of intrinsic motivation is constructed by assessing the extent to which solvers are motivated by inner needs such as self-improvement, skill development, and enjoyment, but excludes the importance of social benefits. Furthermore, as these findings focus on individual effort rather than participation, the question remains whether these interaction effects as well exist on the willingness to participate. Other previous studies have also explored their combined effect, but focus on human performance within formal organizational settings (Cerasoli et al., 2014).

2.3 Conceptual framework

This research is centered on explaining the effects of monetary compensation alongside ESG goals on participation levels in crowdsourcing competitions. The goal is to understand the interaction between motivational factors that drive participation levels in crowdsourcing competitions. The focus on participation aligns with extreme value theory, which posits that the maximum value observed from a set of independent draws tends to increase as the number of draws increases (de Haan & Ferreira, 2006). In a crowdsourcing contest, as the number of participants (analogous to draws from the solver pool) increases, the probability of the competition host obtaining high-quality solutions (similar to the maximum values among all draws) also increases (Yin et al., 2022).

One way to increase participation is by offering extrinsic incentives, notably monetary rewards. Several empirical studies have shown a positive direct relation between the level of monetary compensation and participation rates in crowdsourcing (e.g. Cappa et al., 2019; Roberts et al., 2006). This relation can be theoretically grounded in consumption value theory, which posits that individuals' decisions to participate in activities are driven by the perceived utility they expect to derive from their actions (Sheth et al., 1991). Sun et al. (2012) note that for complex tasks, the effect of financial rewards on participation levels is minimal. Thus, it can be concluded that while rewards do drive participation, the relationship is non-linear. This nuanced relationship necessitates controlling for other competition-specific factors that might signal the complexity of crowdsourcing tasks, such as duration and competitive constraints. By accounting for these factors, this study arguments that the positive relationship will be upheld, such that:

Hypothesis 1: The level of monetary rewards in crowdsourcing competitions is positively related with participation levels.

Besides extrinsic motivators, Rogstadius et al. (2011) explore how finding purpose in the context of an activity serves as a strong internal motivation to participate. This is echoed in self-determination theory (Gagné & Deci, 2005), which states that people are motivated by causes that advance societal well-being, even when it does not immediately benefit them personally. This suggests that internal motivations, especially those driven by the desire to contribute to the social good, play a crucial role in participation. Rogstadius et al. (2011) and Chandler and Kapelner (2010) demonstrate that participants are more engaged when activities are meaningful and aligned with social causes, even if they do not stand to gain directly.

Yin et al. (2022) describe that reward-oriented strategies in competition descriptions can leverage motivational cues and emotional appeals to enhance the perceived value of participation, thus increasing engagement. The study highlights that detailed, issue-relevant information combined with emotional appeals is particularly effective in motivating competitors. This approach aligns with the expectancy theory, suggesting that clearly articulated rewards can significantly impact solvers' perceived expectancy and valence, driving higher participation rates. Similarly, the work of Zhao and Zhu (2014) introduces the complexity of tasks and the clarity of their social impact as crucial factors that influence participation levels. The results suggest that if potential participants do not perceive their efforts as effectively contributing to the claimed social benefits, or if the tasks are too complex relative to their skill sets, their motivation to participate might decrease regardless of the altruistic appeal. Thus, complexity of tasks might again make the relationship with participation levels more nuanced. Taken together, previous research underscores the importance of carefully crafted competition descriptions in maximizing participation in crowdsourcing contests through strategic use of linguistic cues and motivational elements. The findings imply that emphasizing the social value of a challenge can be a powerful strategy to boost participation in crowdsourcing competitions.

Mansouri and Montaz (2022) use the term *ESG goals* to capture ESG-driven opportunities in a crowdfunding context. They measure ESG goals by quantifying ESG properties from crowdfunding whitepapers, thereby constructing an ESG score per crowdfunding campaign. Furthermore, *ESG impact* considers the contribution of a business and its activities to ESG goals (Mansouri & Momtaz, 2022). Adapting their approach in a crowdsourcing context, the term ESG goals will as well be adopted in the theoretical framework.

As such, the fundamental premise of Hypothesis 2 is the following:

Hypothesis 2: *ESG goals associated with crowdsourcing competitions are positively related with higher participation levels.*

Taken together, motivation to participate can be seen as a multidimensional construct, reflecting competitors' various needs (Zheng et al., 2011). That is, participants can be driven both intrinsically and/or extrinsically. Furthermore, research suggests a complex relation between intrinsic and extrinsic motivators and their overall impact on participation (Suen et al., 2022). Research by Grant (2008) and Moller and Deci (2014) suggests that when tasks are intrinsically rewarding or aligned with personal values, they tend to engage participants more deeply, making external rewards such as money less relevant. This effect is particularly pronounced in activities that hold significant ethical or moral weight, such as volunteering or social activism, where monetary incentives may even be seen as inappropriate or counterproductive. Zhao and Zhu (2014) further illustrate that intrinsic factors like personal growth, enjoyment, and altruism can drive participation efforts independently of external rewards. Their findings are echoed in the broader application of self-determination theory, which posits that autonomy, competence, and relatedness are crucial for fostering intrinsic motivation (Gagné & Deci, 2005). When crowdsourcing platforms effectively cater to these intrinsic needs, they could enhance participant levels significantly.

Based on these insights, it becomes plausible to suggest that the presence of ESG goals in crowdsourcing competitions could negatively influence the impact of monetary rewards on participation. These effects are known as moderator effects, as the relation between monetary rewards and participation might be moderated by the presence of ESG goals. Specifically for organizational research, Bokbo and Russell (1994) describe that moderator effects are typically examined through the inclusion of interaction terms in regression models. The interaction term, representing the product of two predictor variables, tests whether the effect of one predictor on the dependent variable varies as a function of another predictor. This research will examine how the level or presence of ESG goals alters the strength or direction of the effects that monetary rewards have on participation levels. Modeling the interaction between intrinsic and extrinsic motivators as moderator effects is in accordance with previous studies on the effects on participation in crowdsourcing competitions (Liang et al., 2018; Piazza et al., 2024). Building on previous findings, a negative moderator effect is expected, such that:

Hypothesis 3: *ESG goals have a negative moderating effect on the relation between monetary rewards and participation in crowdsourcing competitions.*

Figure 1. Conceptual framework

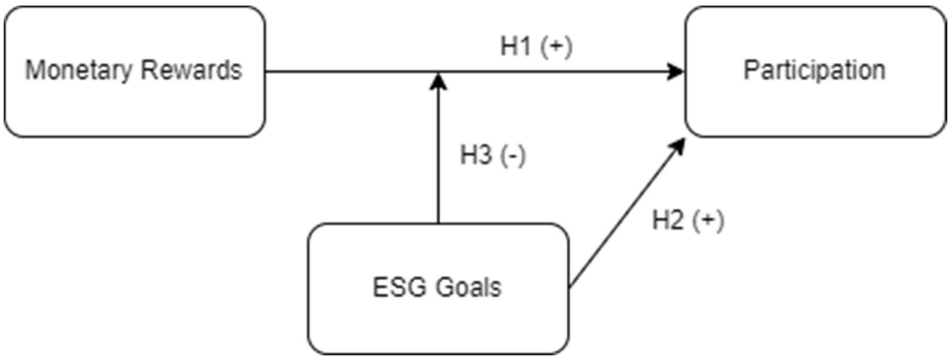


Figure 1 illustrates the hypothesized relationships between monetary rewards, ESG goals and participation in crowdsourcing competitions. Within the framework, both monetary rewards and ESG goals affect the level of participation. Furthermore, it shows that ESG goals might diminish the positive influence of monetary rewards on overall participation to some extent.

3 Methodology

This study delves into the effects of monetary reward and ESG goals on participation in crowdsourcing competition platforms. It aims to quantitatively dissect the impact of monetary incentives and ESG goals on levels of participation. A special focus of this research is on how ESG goals not only directly influence participation levels but also modulate the relationship between monetary rewards and participation levels in crowdsourcing competitions.

3.1 Empirical context: the case of Kaggle

One of these beforementioned crowdsourcing platforms is Kaggle. Kaggle is an online platform that was founded in 2010. As of the end of 2023, Kaggle boasts over 16 million users worldwide, indicating its vast reach and impact across the data science community. Kaggle fosters engagement by offering avenues for individuals to establish prominence within the community (Lampel & Bhalla, 2007; Levina & Arriaga, 2014). Participants gain recognition through indicators of proficiency and expertise, such as badges, points, and rankings. These rankings are based on the number of upvotes, or even better, prizes they win in their competitions. Some competitions offers monetary rewards for winning or achieving a high rank in the competition results. Based on most recent data on competitions between 2010 and 2024, roughly one out of eleven competitions offered monetary rewards, with a median reward of

around 25.000 USD. Around 20 percent offered rewards below 10.000 USD, 60 percent offered rewards between 10.000 USD and 50.000 USD, and around 20 percent offered rewards above 50.000 USD. The top one percent offered rewards lies between 1 and 1,5 million US dollars.

3.2 Data collection

Data on competitions

The data for this study were sourced from the Meta-Kaggle dataset, which contains metadata on Kaggle competitions. The data collection process involved importing and cleaning multiple related datasets to create a comprehensive dataset for analysis. These steps are carried out in STATA and Python.

In STATA, the competitions dataset provided by Kaggle was imported. The dataset provides information on all competitions hosted on the Kaggle platform since 2010, including information on the reward quantities, duration and total submissions. A selection was made of competitions that offered a reward in USD higher than 0 dollars. The data cleaning process as well includes the exclusion of competitions with zero submissions and zero participants, in accordance with Deodhar (2020). This resulted in a dataset with a total of 380 competitions.

Participation

Participation is gauged by the total number of participants per competition. The variable 'TotalCompetitors' in Kaggle's dataset represents the total number of participants who have entered a specific competition. In the context of Kaggle competitions, a participant is counted as a competitor when they make a submission to a competition. This metric does not account for the number of submissions a single participant makes, nor does it include those who might only engage in discussions or view the competition without submitting.[1]

Measuring participation by the number of participants is in line with the methodologies adopted by Chen et al. (2020), Deodhar (2020), Dargahi et al. (2021) and Li and Hu (2017).This variable has been similarly log-transformed to address skewness and improve the linearity of the relationship with monetary rewards. Measuring participation by means of a log-transformation of the number of participants has been established by previous studies, including Chen et al. (2020).

[1] More information about metrics can be found in the official documentation and resources on Kaggle's website: https://www.kaggle.com/docs/competitions

Monetary Compensation

Monetary Compensation is quantified as the total prize money in USD for each competition. This measurement corresponds to previous studies, including Deodhar (2020), Dargahi et al. (2021) and Cappa et al. (2019). This is a continuous variable computed from the 'RewardQuantity' field, which has been logarithmically transformed to normalize its distribution, in line with other research using the Kaggle metadata (Deodhar, 2020; Fang et al., 2021).

ESG scores

On the Kaggle website, additional information on all competitions can be found.[2] This includes a description about the goal of the competition and background information, with an average length of 293 words per description for competitions with a monetary reward. As described by Yin et al. (2022), this information is an important way to attract participation by competition hosts. For example, competition host can use certain linguistic cues to emphasize intrinsic and extrinsic rewards and make emotional appeals (Yin et al., 2022). As such, the competitions' descriptions might hold statements that are in line with ESG goals. As the descriptions of the competition are not included in the Meta Kaggle dataset, a web scraping script is used to extract the descriptions from the competitions' webpage. Web scraping is the process of building and using a software tool that automatically collects information displayed on websites (Boegershausen et al., 2022). It is widely used in scientific research, for example to extract data to gain insight the effects of online marketing (Chong et al., 2017), research current food prices (Hillen, 2019), or to gain insight in public opinion by scraping social media platforms (Batrinca & Treleaven, 2015).

In this research, the web scraping code is created in PyCharm, a coding software that runs on the programming language Python. The software code was written using a combination of 'Beautiful Soup' and 'Selenium', two popular web scraping packages (Taha et al., 2023; Zheng et al., 2015). Running the code results in an automated process of searching the competitions based on their title - taken from the Meta-Kaggle database - extracting the descriptions based on the HTML code that indicates the header 'Descriptions', and copying the descriptions in a new Excel file. A detailed description of the carried out steps can be found in Appendix A.1, the full code will be attached to this thesis as a separate file.

A next step in the data collection is to calculate the environmental-, social- and governance-score per competition, as well as their aggregate ESG score based on the description text. This

[2] https://www.kaggle.com/competitions?listOption=completed

was done by using the Machine Leaning (ML) algorithm developed by Mansouri and Momtaz (2022). The authors constructed an ESG scoring algorithm to analyze the effect of ESG-properties of startups' funding amount in a crowdfunding context. Their Python source code, along with detailed and technical documentation of the machine-learning approach, is provided in Internet Appendix A of Mansouri and Momtaz (2022) and on the *Sustainable Entrepreneurship GitHub* project page.[3]

In general terms, the approach of Mansouri and Momtaz (2022) consists of two steps. As a first step, the authors create an ESG dictionary using a machine-learning method that identifies key phrases in sentences and predicts related phrases to form a list of important terms. Not every term is given the same importance; the importance is determined based on how closely related the terms are to each other. They start with a set of initial words taken from Financial Times articles about ESG investing, which they manually categorize into environmental, social, or governance categories. The approach is based on techniques from previous research by Li et al. (2021) and Mikolov et al. (2013).

As a second step, the authors use the ESG dictionary to count how many times specific ESG-related words appear in startup whitepapers, adjusting for the total number of words in each list. In accordance with Mansouri and Momtaz (2022), the number of E-,S-,G-related terms are counted per description and normalized to the size of the word list. For every competition description i, every E-, S- and G-score is measured as:

$$\zeta_i = \frac{\sum_t 1_{c(t)_i > 0}}{c(n)} \; for \; \zeta = \in (\mathrm{E}, \mathrm{S}, \mathrm{G})$$

Where $c(t)_i$ denotes the count of term t in the description of competition i and $c(n)$ the size of the description text, measured in the total number of words. This approach is based on the approach of Loughran and McDonald (2020) to take into consideration the non-standardized nature of the text. The total ESG score for a competition is the sum of its scores in the E, S, and G categories:

$$ESG_i = \sum_{(E_i, S_i, G_i,)} \zeta_i$$

The method above is applied to the extracted competition descriptions by adapting the Python code provided in their GitHub page. This way, all descriptions are analyzed in one 'run',

[3] https://github.com/sasi2400/sustainableentrepreneurship.org

ensuring the same method is applied. The code provided is further adapted to automatically create an excel-output file with the descriptions and their corresponding E-, S-, and G-scores. The adapted Python Code is attached to this thesis as a separate document, the explanation of the script can be found in Appendix A.2. Next, standardization of the ESG Score is applied in accordance with Mansouri and Momtaz (2022). These transformations help in accurately assessing the impact of ESG factors and are consistent with the log-transformation applied to the dependent variables, which addresses skewness and improves the robustness of the regression models.

The ESG-scoring algorithm developed by Mansouri and Momtaz (2022) is particularly applicable for this research. That is, the authors developed the algorithm to extract ESG-information on startups from whitepapers provided for the Initial Coin Offering (ICO) in crowdfunding campaigns. Given that these whitepapers are designed to spur the interest of the crowd, the linguistic approach and motivation-spurring strategies are likely to be similar to the one applied in a crowdsourcing context described by Yin et al. (2022). Also, the writers provide a webpage link that enables online testing of their ML approach; after inserting your own text the website immediately outputs an ESG score, facilitating transparency and the use of their method, and thereby the method of this thesis as well.[4]

Interaction variables

Interaction variables are constructed to capture the interaction effects between monetary rewards and ESG scores. This involves calculating the product of 'RewardQuantity' and the ESG Score. Each variable is then standardized, ensuring comparability across different dimensions and controlling for variations in text length and other factors.

Alternative ESG scores

As a sanity check, an alternative methodology is adapted to the ESG score per competition to ensure that the approach identifies competitions' ESG properties reliably. Specifically, the ESG BERT model developed by Schimanski et al. (2024) is used for this purpose. ESG BERT stands for "*Environmental, Social, and Governance Bidirectional Encoder Representations from Transformers*."(Mehra et al., 2022, p. 185). It is a specialized version of the BERT model, tailored specifically to analyze and classify text related to ESG domains. The goal is to provide a robust tool for evaluating corporate sustainability disclosures and enhancing transparency and accuracy in ESG measurement. The approach builds on previous methodologies, such as ClimateBERT by Leippold et al. (2022) which developed a model for climate-related text, and

[4] https://www.sustainableentrepreneurship.org/

FinBERT by Huang et al. (2023), which created a financial text model for extracting information from financial documents. In the approach of Schimanski et al. (2024), if a sentence qualifies for either subcategory of ESG, it becomes an ESG sentence. This approach uses Natural Language Processing (NLP), a form of ML, to analyze sentences and compare them to training data. Similar to the approach of Mansouri and Momtaz (2022), the data is standardized, in this case by dividing the sum of ESG sentences by the total of all sentences. One key difference in the methodologies is counting (and standardizing) based on ESG terms (Mansouri & Momtaz, 2022), versus ESG sentences (Schimanski et al., 2024).

On the one hand, the sentence-based approach of Schimanski et al. (2024) can improve accuracy by accounting for the context in which the terms are used. That is, by analyzing entire sentences rather than isolated terms, the model could gain a deeper contextual understanding of how ESG topics are discussed. This allows it to capture nuances and subtleties in language, ensuring that each piece of ESG information is classified accurately. For instance, a sentence provides context that can clarify the meaning of terms, reducing ambiguity from homonyms or words with multiple meanings.

On the other hand, the term-based approach of Mansouri and Momtaz (2022) could be more applicable in a crowdsourcing context as the number of sentences per description is rather limited, especially compared to corporate annual reports. Furthermore, the linguistic terms and information structure of corporate reporting is likely to be less comparable to the language and structure used in a crowdsourcing competition context. Given both that the length and nature of information is likely to be different, the algorithm developed by Schimanski et al. (2024) serves as a second best, and will as such be used as a robustness check on the ESG scores calculated based on the first-best (i.e. that of Mansouri and Momtaz) approach.

To ensure comparability, after applying normalization in accordance with Schimanski et al. (2024), the ESG scores and their interaction terms are normalized in accordance with Mansouri and Momtaz (2022).

Control variables

Three types of control variables are included to ensure the robustness and validity of the regression results. Control variables are selected on their relevance both with respect to the dependent and independent variables, based on prior research by Deodhar (2020). The first type of control variables consist of measures of competitive constraints. Competitive constraints are determined by the competition host with the primary aim of making the competitive landscape

less asymmetric (Deodhar, 2020). Kaggle offers the opportunity to include different competitive constraints in the competition design, such as restrictions on team size and whether a competition allows the merging of teams. Deodhar (2020) finds that in the case of Kaggle, both restrictions are correlated with both monetary rewards and the number of participants. In accordance with his research, TeamConstraint is coded as 1 if the competition limits team size and 0 otherwise. Furthermore, the indicator BanMerger is constructed to account for the possibility of teams combining their efforts. BanMerger takes the value of 1 if the competition host does not allow for team to be merged during the competition and zero otherwise.

Second, this research also incorporates a control for the duration of the tournament, measured by the number of days from the start to the end of the contest (Duration), in line with Deodhar (2020). Based on prior research, the hypothesis is that tournaments that last longer could influence the level of participation (Yang et al., 2009).

Third, the distribution of rewards is considered, which could have important effects on participation (Ales et al., 2017; Terwiesch & Xu, 2008). In the context of Kaggle, the competition host can select the number of winners that will share the reward. In accordance with Deodhar (2020), this variable is log transformed (NumPrize).

3.3 Statistical Analysis and Robustness checks

The main OLS-regression model is specified as:

$$log(TotalCompetitors_i) = \beta_0 + \beta_1 \log(RewardQuantity_i) + \beta_2 \text{ESG Score}_i + \beta_3 (RQ_i \times \text{ESG}_i) + \beta_4 X_i + \epsilon_i$$

Where both the ESG scores ESG Score_i and the interaction variable $(RQ_i \times \text{ESG}_i)$ are normalized. X_i represents a vector of control variables that are included to account for potential confounding factors and isolate the effect of the primary independent variables on the number of competitors. Table 1 defines all variables included in the model.

Table 1 Definitions of variables

Dependent variable	
TotalCompetitors (log)	Logarithm of the number of participants in the competition.
Independent variables	
Reward quantity (log)	Logarithm of the total monetary reward for the competition.
ESG Score (norm.)	Normalized and standardized ESG score.

RQ × ESG (norm.)	Normalized and standardized interaction effect between ESG and reward quantity
Control variables	
Days_Between	The number of days the competition runs.
TeamConstraint	A dummy variable indicating whether there is a team size constraint (1 if constrained, 0 otherwise).
BanTeamMergers	A dummy variable indicating whether team merging is banned (1 if banned, 0 otherwise).
NumPrize	The logarithm of the number of prizes offered in the competition.

Robustness check

To check for robustness of results, two types of additional analyses are carried out. The first test checks for the validity of the ESG scores (including scores on each dimension) per competition. For this, an alternative ML approach is used, based on the BERT ESG algorithm. Appendix A.3 contains a detailed description of the steps carried out in Python to apply the BERT ESG algorithm to the dataset.

Second, an alternative dependent variable is used by replacing the log transformed number of participants by the log transformed number of submissions. In the context of Kaggle, often competitors are allowed to submit multiple solutions. As a result, the number of submissions is on average higher than the number of participants. In the end, both the number of submissions and participants are relevant for competition hosts. Furthermore, the relationship between the number of submissions and the number of participants in a crowdsourcing competition is generally positively correlated, as highlighted by various studies (Boudreau et al., 2011; Zheng et al., 2011). Thus, the additional analysis sheds additional light on the potential effects on submissions and serves as a validity check of the main analysis.

3.4 Ethical Considerations

Throughout the data preparation and analysis phases of this study, ethical considerations were observed to uphold the highest standards of research integrity. The used datasets are structured to contain no information that can be traced back to individuals, ruling out privacy risks. Second, transparency and reproducibility were prioritized. Detailed information on the steps carried out is provided in Appendix A and the full code will be attached to this thesis as a separate file, ensuring that the research can be replicated and verified.

4 Results

In this chapter, the results of all analyses will be outlined. First, the descriptive analyses will be presented. Next, the tests for the assumptions of regression analysis are enumerated, after which the main results will be presented. Lastly, the robustness checks for the results are discussed.

4.1 Descriptive analysis

Table 2 provides an overview of the descriptive statistics of the variables included in the main econometric models.

Table 2 Descriptive statistics of variables

	Mean	Standard Deviation	Min	Max
Dependent variable				
TotalCompetitors (log)	6.672	1.231	3.219	9.189
Independent variables				
Reward quantity (log)	9.886	1.533	4.605	14.221
ESG Score (normalized)	0.000	1.000	-0.857	6.230
RQ × ESG (normalized)	0.000	1.000	-0.373	12.082
Control variables				
BanTeamMergers	0.039	0.195	0.000	1.000
TeamConstraint	0.529	0.500	0.000	1.000
NumPrize (log)	1.180	0.554	0.000	2.565
Days_Between	84.465	49.588	0.417	730.581
Observations	380			

4.2 Test assumptions

In order to confidently do multiple regressions analysis, various assumptions about the data must be met. Four tests have been run, according to Hair (2009), and all assumptions have been proven. The results of the tests and their explanation can be found in Appendix B.

4.3 Main results

The regression results are shown in Table 3. First, the fit of the model with only the control variables is tested, model (1). This results in an R-squared of 0.322. Next, the main direct effects of reward quantity and ESG score are introduced (2), heightening the explanatory power of the model to 0.405. Last, the interaction effect of reward quantity and ESG goals are added in model (3). This complete model has an R-squared of 0.415, indicating both an increase in explanatory

power by adding the interaction effect of ESG Goals with monetary rewards, as well as having a strong model fit.

Table 3 Monetary rewards, ESG scores, interaction effects and total competitors

	DV = *TotalCompetitors (log)*		
	(1)	(2)	(3)
RQ × ESG (normalized)			-0.152**
			(0.059)
ESG Score (normalized)		0.100*	0.147***
		(0.051)	(0.054)
Reward quantity (log)		0.307***	0.349***
		(0.044)	(0.047)
BanTeamMergers	-0.648**	-0.629**	-0.600**
	(0.280)	(0.264)	(0.262)
TeamConstraint	0.723***	0.558***	0.519***
	(0.121)	(0.117)	(0.117)
NumPrize (log)	0.700***	0.224*	0.217*
	(0.109)	(0.122)	(0.121)
Days_Between	0.003***	0.001	0.002
	(0.001)	(0.001)	(0.001)
Constant	5.199***	2.997***	2.578***
	(0.142)	(0.351)	(0.384)
Observations	380	380	380
R-squared	0.322	0.405	0.415

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The results presented in Table 3, will be used to analyze the three hypotheses described in Chapter 2. The results from model (2) will be used for the first two hypotheses. With respect to hypothesis 1, the results confirm a positive and significant overall effect of the reward quantity amount with a coefficient of approximately 0.307. This indicates that increasing the reward quantity positively influences the number of participants. Specifically, a one-unit increase in the log of reward quantity results in an increase of 0.307 units in the log of total competitors. Translating this into practical terms, if the reward quantity doubles (a 100% increase), the number of competitors will increase by approximately 30.7%.

Hypothesis 2 suggested that competitions with higher ESG goals would attract more participants. The coefficient for the ESG score is indeed positive and significant ($p < 0.10$), with a value of 0.100. The results in Appendix C show that the effect of ESG is largely driven by its social component. That is, Table C. 2 shows that the S-score on itself is significant, albeit the coefficient is smaller than the coefficient of the overall ESG score in Table 3. This indicates that higher ESG scores are associated with higher participation levels. To translate the coefficient into practical terms, a one standard deviation increase in the ESG score results in a roughly 10.5% increase in the actual number of competitors. This finding underscores the importance of ESG factors in motivating participants, highlighting that competitors are significantly more likely to engage in competitions that emphasize environmental, social, and governance aspects.

Lastly, hypothesis 3 posited that ESG goals have a negative moderating effect on the relationship between monetary rewards and participation in crowdsourcing competitions. The interaction term between reward quantity and ESG score in model (3) is indeed negative and significant with a value of -0.152. This indicates that the positive impact of increasing rewards is diminished when ESG scores are higher. Thus, competition organizers should balance monetary rewards and ESG goals to optimize participation, as high levels of both may not yield additive benefits due to the interaction effect.

4.4 Robustness checks

BERT ESG

This section presents a robustness check using an alternative algorithm, BERT ESG, to calculate the ESG scores of competitions. The purpose of this analysis is to validate the robustness of the regression results discussed in section 4.3 by comparing them with those obtained using BERT ESG scores, displayed in Table 4.

Table 4 Monetary rewards, BERT ESG Scores, interaction and Total Competitors

	DV = *TotalCompetitors (log)*		
	(1)	(2)	(3)
RQ × BERT ESG (normalized)			-0.208***
			(0.063)
BERT ESG Score (normalized)		0.0402	0.135**
		(0.050)	(0.057)
Reward quantity (log)		0.305***	0.368***

		(0.044)	(0.047)
BanTeamMergers	-0.648**	-0.650**	-0.612**
	(0.280)	(0.265)	(0.261)
TeamConstraint	0.723***	0.581***	0.502***
	(0.121)	(0.117)	(0.118)
NumPrize (log)	0.700***	0.258**	0.260**
	(0.109)	(0.122)	(0.120)
Days_Between	0.003***	0.002	0.001
	(0.001)	(0.001)	(0.001)
Constant	5.199***	2.953***	2.349***
	(0.142)	(0.351)	(0.392)
Observations	380	380	380
R-squared	0.322	0.400	0.417

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

The robustness of the results described in Table 3 can be assessed by comparing the coefficients and R-squared values obtained from the main approach (using the Mansouri and Momtaz algorithm) with those from the BERT ESG algorithm. The comparison indicates that the results are generally consistent, suggesting that the findings are robust across different methods of ESG score calculation. The primary methodology and the BERT ESG approach yield similar coefficients for most variables, indicating that the impact of monetary rewards and ESG scores on participation levels is reliable. The R-squared values are also comparable across models, with the main approach yielding an R-squared of 0.415 and the ESG-Bert model showing 0.417. This indicates that the explanatory power of the models is almost identical, further supporting the robustness of the findings.

The regression results using the ESG-Bert scores, as displayed in Table 4, show some changes in coefficients compared to the main approach (Table 3). Notably, the coefficient for the log of reward quantity remains positive and significant in both models. In the ESG-Bert model, the coefficient is slightly higher (0.368 compared to 0.349), suggesting a marginally stronger effect of monetary rewards on participation when using the alternative ESG calculation method. Similarly, the coefficient for the normalized ESG score is positive and significant in both models, though the magnitude is slightly lower in the ESG-Bert model (0.134 compared to 0.147), indicating a consistent but slightly reduced effect of ESG Goals on total competitors.

Total Submissions

Largely the same effects are found with respect to the number of submissions, as can be seen in Table 5. In accordance with previous findings, the number of submissions increases with the level of monetary rewards. Likewise, a higher ESG score increases the number of submissions.

Table 5 Monetary rewards, ESG Scores, interaction and total submissions

	DV = *TotalSubmissions (log)*		
	(1)	(2)	(3)
RQ × ESG (normalized)			-0.203**
			(0.088)
ESG Score (normalized)		0.258***	0.323***
		(0.076)	(0.081)
Reward Quantity (log)		0.374***	0.430***
		(0.0653)	(0.069)
BanTeamMergers	-0.947**	-0.883**	-0.846**
	(0.413)	(0.393)	(0.391)
TeamConstraint	0.876***	0.631***	0.578***
	(0.179)	(0.174)	(0.175)
NumPrize (log)	0.542***	-0.071	-0.081
	(0.160)	(0.181)	(0.180)
Days_Between	0.002	-0.001	-0.000
	(0.002)	(0.002)	(0.002)
Constant	7.654***	5.058***	4.495***
	(0.210)	(0.522)	(0.573)
Observations	380	380	380
R-squared	0.168	0.254	0.264

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Compared to the main analysis, the R-squared value is lower (0.264 to 0.415), indicating that the model is better suited to predict participation than submissions. One notable difference in this model is the higher constant term. This is expected because the number of total submissions is generally higher than the number of participants, as some participants submit multiple entries. Despite this difference, the effects of the variables are generally similar to those observed in the main analysis.

On the other hand, the number of prizes (NumPrize), which was significant in predicting participation, is no longer significant in the model predicting total submissions. This change

suggests that while the number of prizes can attract participants, it does not significantly influence the decision to submit entries. Overall, the robustness check using total submissions confirms the robustness of the primary results while providing additional insights into the factors that specifically drive submissions. The consistent findings across different models reinforce confidence in the conclusions about the effects of monetary rewards and ESG goals on participation and – to a somewhat lesser extent - submissions in crowdsourcing competitions.

5 Discussion and limitations

5.1 Summary of key findings

This thesis centers around the (moderating) effects of monetary rewards and ESG goals on participation in crowdsourcing competitions, specifically within the context of Kaggle competitions. The sample consists of 380 Kaggle competitions hosted in the period 2010-2024 for which a monetary reward was offered. Information on ESG goals in the competition description was quantified by adapting a machine-learning approach that was developed for assessing ESG properties in a crowdfunding context (see(Mansouri & Momtaz, 2022). The primary findings indicate that both monetary rewards and ESG goals independently positively influence participation levels. However, the interaction between these two factors reveals a more complex relationship, where high ESG scores can partly diminish the impact of monetary rewards on participation.

The results support the notion that monetary rewards are a significant driver of participation, aligning with existing empirical evidence that suggests financial incentives are motivating individuals to engage in crowdsourcing activities (Cappa et al., 2019; Deodhar, 2020). Additionally, the findings extend the understanding of the effect ESG-related goals have in a crowdsourcing context, showing that competitions with higher ESG scores also attract more participants, consistent with the principles of self-determination theory (Ryan & Deci, 2000) and previous research on the intrinsic motivations tied to social and environmental causes (Cappa et al., 2019; Wang & Chen, 2023). However, the negative moderation effect between ESG goals and monetary rewards suggests that when both are present, the motivational impact of monetary rewards is weakened. This supports the "crowding out" theory, where intrinsic motivations (such as those driven by ESG goals) can overshadow the effects of extrinsic motivations (monetary rewards) (Deci et al., 1999; Piazza et al., 2024).

5.2 Theoretical contributions and managerial implications

Theoretical contributions

This thesis offers several contributions to the existing crowdsourcing literature on what drives motivation, and by extension, participation. The findings of this thesis can be related to several motivation theories, that provide the theoretical groundwork for the analysis. At the core is expectancy theory, developed by Vroom (1964), which suggests that individuals are motivated to act based on the expectation that their actions will lead to desired outcomes. Building on this, expectancy-value theory (Eccles & Wigfield, 2023) posits that motivation is determined by both expectancy and value. Here, value encompasses multiple dimensions, including attainment value, intrinsic value, utility value and cost.

Firstly, this thesis adds to the empirical evidence on the relation between monetary rewards and participation in a crowdsourcing context that is theoretically grounded in expectancy-value theory. Previous studies (such as Li & Hu, 2017) apply expectancy-value theory to crowdsourcing contests to explain why monetary rewards increase participation levels. According to Li and Hu (2017), offering higher monetary rewards would be a way to enhance the perceived utility of participating. As such, offering monetary rewards serves as an incentive mechanism that provides external motivation for participation. The findings of this thesis provide further empirical evidence in support of such mechanisms, in accordance with previous findings that focus on participation on other crowdsourcing platforms (Chen et al., 2020; Li & Hu, 2017). This thesis applies the framework in the context of the Kaggle platform for competitions hosted between 2014 and 2024. Similar to empirical studies carried out for other platforms, the findings show that the number of participants is positively related with the level of monetary rewards. The regression results are significant on a 0.01 level and include a vector of control variables to account for potential confounding factors.

Second, the findings contribute to the limited knowledge on how ESG goals of crowdsourcing competitions might increase intrinsic motivation, and by extension, participation. While extrinsic motivation is driven by the pursuit of a specific, instrumental objective, intrinsic motivation involves engaging in an activity purely for the enjoyment or fulfillment it provides (Reiss, 2012). This is grounded in self-determination theory (SDT), that posits that intrinsic motivation stems from an individual's basic psychological needs, leading to enjoyment and satisfaction derived from enhancing one's competencies (Ryan & Deci, 2000). Furthermore, previous research suggests that when participants perceive their efforts as contributing to a larger social good, their intrinsic motivation tends to increase (Wang & Chen, 2023). This

suggests that hosts can leverages prosocial motivations, encouraging individuals who are driven by altruism and a desire to contribute to societal welfare to participate. In a business context, ESG has been demonstrated to enhance employees' intrinsic motivation, thereby improving their job performance (Lee & Park, 2023).

To my knowledge, this study is the first that analyzes ESG properties of crowdsourcing competitions and its effects of participation. Previous studies such as by Piazza et al. (2024) have highlighted that prosocial motivations increases willingness to participate in certain social innovation contests. However, the findings of this particular study is limited to a specific context (i.e. COVID-19) and does not allow for comparison amongst competitions with different subject-related properties. This study makes use of recent advancements in ML techniques that allow for quantifying ESG properties based on textual data. Hereby, a concrete understanding is provided on how ESG goals influence participation in crowdsourcing contests.

Thirdly, this thesis is one of the first studies that sheds light on moderation effects of intrinsic motivators, quantified as ESG goals, on the relation between extrinsic motivations (provided by monetary rewards) and participation levels. Based on consumer value theory (Holbrook, 1999), motivation to participate is a multidimensional construct that reflects the diverse needs of competitors. Participants can be driven by both intrinsic and extrinsic factors. Additionally, research indicates a complex relationship between these motivators and their overall impact on participation. Studies by Grant (2008) and Moller and Deci (2014) suggest that tasks which are intrinsically rewarding or aligned with personal values tend to engage participants more deeply, thereby reducing the relevance of external rewards such as money. This thesis uses ESG as a continuous measure that aims to capture the many nuances in the relation between monetary compensation, motivation and participation. Given the methodological differences with previous research, this thesis offers a more concrete and reliable understanding in the moderation effects between intrinsic and extrinsic motivational factors that drive participation, building further upon self-determination theory.

The last contribution is methodological. A growing body of literature employs state-of-the-art ML techniques to bridge the gap between human communication and computer understanding (Schimanski et al., 2024). This includes ML and NLP techniques to quantify ESG by analyzing textual data. This study is most likely the first to make use of these ML algorithms to quantify ESG scores in descriptions of crowdsourcing competitions. The successful implementation shows the applicability of these algorithms in crowdsourcing literature, especially for

algorithms that were pre-trained on textual data of crowdfunding campaigns (Mansouri & Momtaz, 2022). As the field of machine learning is rapidly developing, this suggests that the crowdsourcing literature could increasingly profit from these ML-techniques to access data regarding factors that influence the success of crowdsourcing competitions.

Managerial implications

The findings are relevant for managers and organizations that seek to benefit from crowdsourcing competitions for innovation purposes. Previous studies have demonstrated that attracting a substantial number of competitors allows competition hosts to obtain a wider array of solutions, thereby enhancing the chances of obtaining suitable ones (Jeppesen & Lakhani, 2010; Terwiesch & Xu, 2008). As such, it is essential to inspire a diverse group of problem solvers who, drawing on their unique values, education, experiences, and goals, can generate a broader spectrum of solutions (Piazza et al., 2024).

This has certain practical implications for competition hosts in solving their optimization problem when designing the competition. Firstly, when designing crowdsourcing competitions, hosts should not solely rely on monetary incentives if they are able to incorporate ESG-related goals to attract a broader range of participants. Highlighting the ESG impact of the competition can tap into the intrinsic motivations of participants, encouraging those who are driven by a desire to contribute to societal good to take part. This dual approach can lead to a more engaged and diverse pool of solvers, enhancing both the quality and quantity of participation.

Secondly, competition hosts need to carefully balance the amount of monetary rewards and the emphasis on ESG goals. While both are important, their combined effect can be less straightforward. Hosts should experiment with different levels of financial incentives and ESG messaging to find the optimal combination that maximizes participation without incurring excessive costs. This balance depends on the budgetary constraints of the host organization. It also depends on the genuine applicability of ESG goals to the competition and the host organization, as noted above.

For instance, a host organization with a low ESG score might profit more from offering high monetary rewards to attract participants, as it cannot rely on strong intrinsic motivators related to ESG. These organizations need to compensate for their lack of ESG credibility by providing substantial extrinsic incentives. Conversely, a host organization with high ESG scores might be less dependent on higher monetary rewards. Such organizations can substitute extrinsic motivations by effectively highlighting the ESG goals served by the competition, thereby

attracting participants who are intrinsically motivated by the desire to contribute to social and environmental causes. By using data from past competitions, hosts can analyze patterns and adjust accordingly, ensuring that their strategy aligns with both their financial constraints and their ESG commitments.

5.3 Limitations and future research directions

The observational nature of the data introduces biases inherent in such designs and cannot fully account for all possible influences on participation. Without experimental manipulation or randomized control trials, it is not possible to entirely rule out the possibility that the relationships observed are driven by underlying factors not captured in the dataset. Consequently, while the findings provide valuable insights into the associations between ESG goals, monetary rewards, and participation, they should be interpreted with caution. The inclusion of control variables strengthens the analysis, yet future research employing experimental designs or longitudinal data would be essential to establish causal links and provide a more comprehensive understanding of the dynamics at play in crowdsourcing competition. Another limitation of this study is that the data is taken from one crowdsourcing platform, namely Kaggle. To further generalize the findings of this study, the same method should be applied to other crowdsourcing platforms to rule out platform-specific factors such as the maximum amount of team members per team, or the overall visibility of a competitions' description. Another limitation of this study concerns the suitability of the algorithm used for weighing ESG terms, which was originally designed for a crowdfunding context. The methodology employed in this research is based on algorithms and term weightings tailored to the specific dynamics of crowdfunding platforms, where the motivations and behaviors of participants might differ from those in crowdsourcing competitions. Although both contexts involve engaging a 'crowd,' the factors that drive participation could vary. As such, the current algorithm's term weights may not fully capture the nuances and priorities of crowdsourcing participants. This potential misalignment highlights the need for future research to tweak and refine the methods used, such as adjusting term weights and counting methodologies, to better align with what matters most in crowdsourcing contexts.

References

Acar, O. A. (2018). Harnessing the creative potential of consumers: money, participation, and creativity in idea crowdsourcing. *Marketing Letters*, *29*(2), 177-188. https://doi.org/10.1007/s11002-018-9454-9

Acar, O. A. (2019). Motivations and solution appropriateness in crowdsourcing challenges for innovation. *Research Policy*, *48*(8), Article 103716. https://doi.org/10.1016/j.respol.2018.11.010

Afuah, A., & Tucci, C. L. (2012). CROWDSOURCING AS A SOLUTION TO DISTANT SEARCH. *Academy of Management Review*, *37*(3), 355-375. https://doi.org/10.5465/amr.2010.0146

Aitamurto, T., & Chen, K. P. (2017). The value of crowdsourcing in public policymaking: epistemic, democratic and economic value. *Theory and Practice of Legislation*, *5*(1), 55-72. https://doi.org/10.1080/20508840.2017.1282665

Alam, S. L., & Campbell, J. (2017). Temporal Motivations of Volunteers to Participate in Cultural Crowdsourcing Work. *Information Systems Research*, *28*(4), 744-759. https://doi.org/10.1287/isre.2017.0719

Ales, L., Cho, S. H., & Körpeoglu, E. (2017). Optimal Award Scheme in Innovation Tournaments. *Operations Research*, *65*(3), 693-702. https://doi.org/10.1287/opre.2016.1575

Bagherzadeh, M., Markovic, S., & Bogers, M. (2021). Managing Open Innovation: A Project-Level Perspective. *Ieee Transactions on Engineering Management*, *68*(1), 301-316. https://doi.org/10.1109/tem.2019.2949714

Barnes, S. A., Green, A., & de Hoyos, M. (2015). Crowdsourcing and work: individual factors and circumstances influencing employability. *New Technology Work and Employment*, *30*(1), 16-31. https://doi.org/10.1111/ntwe.12043

Batrinca, B., & Treleaven, P. C. (2015). Social media analytics: a survey of techniques, tools and platforms. *Ai & Society*, *30*(1), 89-116. https://doi.org/10.1007/s00146-014-0549-4

Bobko, P., & Russell, C. J. (1994). ON THEORY, STATISTICS, AND THE SEARCH FOR INTERACTIONS IN THE ORGANIZATIONAL SCIENCES. *Journal of Management*, *20*(1), 193-200. https://doi.org/10.1177/014920639402000111

Boegershausen, J., Datta, H., Borah, A., & Stephen, A. T. (2022). Fields of Gold: Scraping Web Data for Marketing Insights. *Journal of Marketing*, *86*(5), 1-20. https://doi.org/10.1177/00222429221100750

Boudreau, K. J., Lacetera, N., & Lakhani, K. R. (2011). Incentives and Problem Uncertainty in Innovation Contests: An Empirical Analysis. *Management Science*, *57*(5), 843-863. https://doi.org/10.1287/mnsc.1110.1322

Boudreau, K. J., & Lakhani, K. R. (2013). Using the Crowd as an Innovation Partner. *Harvard Business Review*, *91*(4), 60-+.

Brabham, D. C. (2008). Crowdsourcing as a model for problem solving: An introduction and cases. *Convergence*, *14*(1), 75-90.

Brown, S. L., & Eisenhardt, K. M. (1995). PRODUCT DEVELOPMENT - PAST RESEARCH, PRESENT FINDINGS, AND FUTURE-DIRECTIONS. *Academy of Management Review*, *20*(2), 343-378. https://doi.org/10.2307/258850

Cappa, F., Rosso, F., & Hayes, D. (2019). Monetary and Social Rewards for Crowdsourcing. *Sustainability*, *11*(10), Article 2834. https://doi.org/10.3390/su11102834

Cerasoli, C. P., Nicklin, J. M., & Ford, M. T. (2014). Intrinsic Motivation and Extrinsic Incentives Jointly Predict Performance: A 40-Year Meta-Analysis. *Psychological Bulletin*, *140*(4), 980-1008. https://doi.org/10.1037/a0035661

Chalmers, D. (2013). Social innovation: An exploration of the barriers faced by innovating organizations in the social economy. *Local Economy*, *28*(1), 17-34. https://doi.org/10.1177/0269094212463677

Chen, L., Xu, P., & Liu, D. (2020). Effect of Crowd Voting on Participation in Crowdsourcing Contests. *Journal of Management Information Systems*, *37*(2), 510-535. https://doi.org/10.1080/07421222.2020.1759342

Chesbrough, H., & Di Minin, A. (2014). Open social innovation. *New frontiers in open innovation*, *16*, 301-315.

Chong, A. Y. L., Ch'ng, E. B., Liu, M. J., & Li, B. Y. (2017). Predicting consumer product demands via Big Data: the roles of online promotional marketing and online reviews. *International Journal of Production Research*, *55*(17), 5142-5156. https://doi.org/10.1080/00207543.2015.1066519

Cumming, D., Meoli, M., Rossi, A., & Vismara, S. (2024). ESG and crowdfunding platforms. *Journal of Business Venturing*, *39*(1), Article 106362. https://doi.org/10.1016/j.jbusvent.2023.106362

Daradkeh, M. (2022). Innovation in Business Intelligence Systems: The Relationship Between Innovation Crowdsourcing Mechanisms and Innovation Performance. *International Journal of Information Systems in the Service Sector*, *14*(1). https://doi.org/10.4018/ijisss.302885

Dargahi, R., Namin, A., Ketron, S. C., & Saint Clair, J. K. (2021). Is self-knowledge the ultimate prize? A quantitative analysis of participation choice in online ideation crowdsourcing contests. *Journal of Retailing and Consumer Services*, *62*, 102644.

de Haan, L., & Ferreira, A. (2006). Extreme and Intermediate Order Statistics. *Extreme Value Theory: An Introduction*, 37-63.

Deci, E. L. (1971). EFFECTS OF EXTERNALLY MEDIATED REWARDS ON INTRINSIC MOTIVATION. *Journal of Personality and Social Psychology*, *18*(1), 105-&. https://doi.org/10.1037/h0030644

Deci, E. L., Koestner, R., & Ryan, R. M. (1999). A meta-analytic review of experiments examining the effects of extrinsic rewards on intrinsic motivation. *Psychological Bulletin*, *125*(6), 627-668. https://doi.org/10.1037/0033-2909.125.6.627

Deodhar, S. J. (2020). Interplay between Constraints and Rewards in Innovation Tournaments: Implications for Participation. *International Journal of Cooperative Information Systems*, *29*(1-2), Article 2040004. https://doi.org/10.1142/s0218843020400043

Eccles, J. S., & Wigfield, A. (2023). Expectancy-Value Theory to Situated Expectancy-Value Theory: Reflections on the Legacy of 40+Years of Working Together. *Motivation Science*, *9*(1), 1-12. https://doi.org/10.1037/mot0000275

Estellés-Arolas, E., & Gonzalez-Ladron-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, *38*(2), 189-200. https://doi.org/10.1177/0165551512437638

Fang, T. P., Wu, A., & Clough, D. R. (2021). Platform diffusion at temporary gatherings: Social coordination and ecosystem emergence. *Strategic Management Journal*, *42*(2), 233-272. https://doi.org/10.1002/smj.3230

Frey, K., Lüthje, C., & Haag, S. (2011). Whom Should Firms Attract to Open Innovation Platforms? The Role of Knowledge Diversity and Motivation. *Long Range Planning*, *44*(5-6), 397-420. https://doi.org/10.1016/j.lrp.2011.09.006

Füller, J., Hutter, K., & Fries, M. (2012). Crowdsourcing for goodness sake: impact of incentive preference on contribution behavior for social innovation. In *Interdisciplinary*

approaches to product design, innovation, & branding in international marketing (pp. 137-159). Emerald Group Publishing Limited.

Gagné, M., & Deci, E. L. (2005). Self-determination theory and work motivation. *Journal of Organizational Behavior*, *26*(4), 331-362. https://doi.org/10.1002/job.322

Giese, G., Lee, L.-E., Melas, D., Nagy, Z., & Nishikawa, L. (2019). Foundations of ESG investing: How ESG affects equity valuation, risk, and performance. *The Journal of Portfolio Management*, *45*(5), 69-83.

Gillier, T., Chaffois, C., Belkhouja, M., Roth, Y., & Bayus, B. L. (2018). The effects of task instructions in crowdsourcing innovative ideas. *Technological Forecasting and Social Change*, *134*, 35-44. https://doi.org/10.1016/j.techfore.2018.05.005

Grant, A. M. (2008). The significance of task significance: Job performance effects, relational mechanisms, and boundary conditions. *Journal of Applied Psychology*, *93*(1), 108-124. https://doi.org/10.1037/0021-9010.93.1.108

Hair, J. F. (2009). Multivariate data analysis.

Halid, S., Rahman, R. A., Mahmud, R., Mansor, N., & Wahab, R. A. (2023). A literature review on ESG score and its impact on firm performance. *International Journal of Academic Research in Accounting Finance and Management Sciences*, *13*(1), 272-282.

Han, S., & Anderson, C. K. (2021). Web Scraping for Hospitality Research: Overview, Opportunities, and Implications. *Cornell Hospitality Quarterly*, *62*(1), 89-104. https://doi.org/10.1177/1938965520973587

Hastalona, D., & Sadalia, I. (2021). Literature review ESG and sustainability finance. *Budapest International Research and Critics Institute (BIRCI-Journal): Humanities and Social Sciences*, *4*(3), 3548-3557.

Hillen, J. (2019). Web scraping for food price research. *British Food Journal*, *121*(12), 3350-3361. https://doi.org/10.1108/bfj-02-2019-0081

Hofstetter, R., Zhang, J. Z., & Herrmann, A. (2018). Successive Open Innovation Contests and Incentives: Winner-Take-All or Multiple Prizes? *Journal of Product Innovation Management*, *35*(4), 492-517. https://doi.org/10.1111/jpim.12424

Holbrook, M. B. (1999). Consumer value. *A framework for analysis and research*.

Howe, J. (2006). The rise of crowdsourcing. *Wired magazine*, *14*(6), 176-183.

Huang, A. H., Wang, H., & Yang, Y. (2023). FinBERT: A Large Language Model for Extracting Information from Financial Text. *Contemporary Accounting Research*, *40*(2), 806-841. https://doi.org/10.1111/1911-3846.12832

Jeppesen, L. B., & Lakhani, K. R. (2010). Marginality and Problem-Solving Effectiveness in Broadcast Search. *Organization Science*, *21*(5), 1016-1033. https://doi.org/10.1287/orsc.1090.0491

Kapelner, A., & Chandler, D. (2010). Preventing satisficing in online surveys. *Proceedings of CrowdConf*.

Kaufmann, N., Schulze, T., Veit, D., & Assoc Informat, S. (2011). More than fun and money. Worker Motivation in Crowdsourcing - A Study on Mechanical Turk. *Amcis 2011 Proceedings*.

Ke, W. L., & Zhang, P. (2009). Motivations in Open Source Software Communities: The Mediating Role of Effort Intensity and Goal Commitment. *International Journal of Electronic Commerce*, *13*(4), 39-66. https://doi.org/10.2753/jec1086-4415130403

Lakhani, K. R., Jeppesen, L. B., Lohse, P. A., & Panetta, J. A. (2007). The value of openess in scientific problem solving.

Lampel, J., & Bhalla, A. (2007). The Role of Status Seeking in Online Communities: Giving the Gift of Experience. *Journal of Computer-Mediated Communication*, *12*(2), 434-455. https://doi.org/10.1111/j.1083-6101.2007.00332.x

Lee, C., & Park, S. (2023). Impact of ESG on Employees' Intrinsic Motivation across generations. In (Vol. 16, pp. 97-119): K*orea Review of International Studies*.

Leippold, M., Bingler, J. A., Kraus, M., & Webersinke, N. (2022). ClimateBert: A pretrained language model for climate-related text.

Levina, N., & Arriaga, M. (2014). Distinction and Status Production on User-Generated Content Platforms: Using Bourdieu's Theory of Cultural Production to Understand Social Dynamics in Online Fields. *Information Systems Research*, *25*(3), 468-488. https://doi.org/10.1287/isre.2014.0535

Li, D., & Hu, L. Y. (2017). Exploring the effects of reward and competition intensity on participation in crowdsourcing contests. *Electronic Markets*, *27*(3), 199-210. https://doi.org/10.1007/s12525-017-0252-7

Li, K., Mai, F., Shen, R., & Yan, X. Y. (2021). Measuring Corporate Culture Using Machine Learning. *Review of Financial Studies*, *34*(7), 3265-3315. https://doi.org/10.1093/rfs/hhaa079

Liang, H. G., Wang, M. M., Wang, J. J., & Xue, Y. J. (2018). How intrinsic motivation and extrinsic incentives affect task effort in crowdsourcing contests: A mediated moderation model. *Computers in Human Behavior*, *81*, 168-176. https://doi.org/10.1016/j.chb.2017.11.040

Loughran, T., & McDonald, B. (2020). Textual Analysis in Finance. *Annual Review of Financial Economics, Vol 12, 2020*, *12*, 357-375. https://doi.org/10.1146/annurev-financial-012820-032249

Lüttgens, D., Pollok, P., Antons, D., & Piller, F. (2014). Wisdom of the crowd and capabilities of a few: internal success factors of crowdsourcing for innovation. *Journal of Business Economics*, *84*, 339-374.

Majchrzak, A., & Malhotra, A. (2013). Towards an information systems perspective and research agenda on crowdsourcing for innovation. *Journal of Strategic Information Systems*, *22*(4), 257-268. https://doi.org/10.1016/j.jsis.2013.07.004

Mansouri, S., & Momtaz, P. P. (2022). Financing sustainable entrepreneurship: ESG measurement, valuation, and performance. *Journal of Business Venturing*, *37*(6), Article 106258. https://doi.org/10.1016/j.jbusvent.2022.106258

Martinez, M. G. (2015). Solver engagement in knowledge sharing in crowdsourcing communities: Exploring the link to creativity. *Research Policy*, *44*(8), 1419-1430. https://doi.org/10.1016/j.respol.2015.05.010

Martiny, A., Testa, F., Taglialatela, J., & Iraldo, F. (2024). Determinants of Environmental Social and Governance (ESG) Performance: A Systematic Literature Review. *Journal of Cleaner Production*, 142213.

Mehra, S., Louka, R., & Zhang, Y. (2022). Esgbert: Language model to help with classification tasks related to companies environmental, social, and governance practices. *arXiv preprint arXiv:2203.16788*.

Meng, Q. L., Hang, Y., Chen, X. J., & Cao, Q. Y. (2021). Value co-creation mechanisms of multi-agent participation in crowdsourcing innovation: A grounded theory study. *Journal of Intelligent & Fuzzy Systems*, *41*(4), 4995-5006. https://doi.org/10.3233/jifs-189986

Mikolov, T., Sutskever, I., Chen, K., Corrado, G. S., & Dean, J. (2013). Distributed representations of words and phrases and their compositionality. *Advances in neural information processing systems*, *26*.

Moghaddam, E. N., Aliahmadi, A., Bagherzadeh, M., Markovic, S., Micevski, M., & Saghafi, F. (2023). Let me choose what I want: The influence of incentive choice flexibility on the quality of crowdsourcing solutions to innovation problems. *Technovation*, *120*, Article 102679. https://doi.org/10.1016/j.technovation.2022.102679

Moller, A. C., & Deci, E. L. (2014). The psychology of getting paid: An integrated perspective. In *The psychological science of money* (pp. 189-211). Springer.

Murayama, K., Matsumoto, M., Izuma, K., & Matsumoto, K. (2010). Neural basis of the undermining effect of monetary reward on intrinsic motivation. *Proceedings of the National Academy of Sciences of the United States of America*, *107*(49), 20911-20916. https://doi.org/10.1073/pnas.1013305107

Nevo, D., & Kotlarsky, J. (2020). Crowdsourcing as a strategic IS sourcing phenomenon: Critical review and insights for future research. *Journal of Strategic Information Systems*, *29*(4), Article 101593. https://doi.org/10.1016/j.jsis.2020.101593

Oeij, P. R. A., van der Torre, W., Vaas, F., & Dhondt, S. (2019). Understanding social innovation as an innovation process: Applying the innovation journey model. *Journal of Business Research*, *101*, 243-254. https://doi.org/10.1016/j.jbusres.2019.04.028

Phills, J. A., Deiglmeier, K., & Miller, D. T. (2008). Rediscovering social innovation. *Stanford Social Innovation Review*, *6*(4), 34-43.

Piazza, M., Mazzola, E., & Perrone, G. (2024). Solvers' participation in crowdsourcing initiatives for social innovation: Exploring interactions among motivational forces. *Creativity and Innovation Management*. https://doi.org/10.1111/caim.12610

Reiss, S. (2012). Intrinsic and Extrinsic Motivation. *Teaching of Psychology*, *39*(2), 152-156. https://doi.org/10.1177/0098628312437704

Roberts, J. A., Hann, I. H., & Slaughter, S. A. (2006). Understanding the motivations, participation, and performance of open source software developers: A longitudinal study of the Apache projects. *Management Science*, *52*(7), 984-999. https://doi.org/10.1287/mnsc.1060.0554

Rogstadius, J., Kostakos, V., Kittur, A., Smus, B., Laredo, J., & Vukovic, M. (2011). An assessment of intrinsic and extrinsic motivation on task performance in crowdsourcing markets. Proceedings of the international AAAI conference on web and social media,

Ryan, R. M., & Deci, E. L. (2000). Intrinsic and extrinsic motivations: Classic definitions and new directions. *Contemporary Educational Psychology*, *25*(1), 54-67. https://doi.org/10.1006/ceps.1999.1020

Schimanski, T., Reding, A., Reding, N., Bingler, J., Kraus, M., & Leippold, M. (2024). Bridging the gap in ESG measurement: Using NLP to quantify environmental, social, and governance communication. *Finance Research Letters*, *61*, Article 104979. https://doi.org/10.1016/j.frl.2024.104979

Seltzer, E., & Mahmoudi, D. (2013). Citizen Participation, Open Innovation, and Crowdsourcing: Challenges and Opportunities for Planning. *Journal of Planning Literature*, *28*(1), 3-18. https://doi.org/10.1177/0885412212469112

Sheldon, K. M., & Elliot, A. J. (1999). Goal striving, need satisfaction, and longitudinal well-being: The self-concordance model. *Journal of Personality and Social Psychology*, *76*(3), 482-497. https://doi.org/10.1037/0022-3514.76.3.482

Sheth, J. N., Newman, B. I., & Gross, B. L. (1991). WHY WE BUY WHAT WE BUY - A THEORY OF CONSUMPTION VALUES. *Journal of Business Research*, *22*(2), 159-170. https://doi.org/10.1016/0148-2963(91)90050-8

Sieg, J. H., Wallin, M. W., & von Krogh, G. (2010). Managerial challenges in open innovation: a study of innovation intermediation in the chemical industry. *R & D Management*, *40*(3), 281-291. https://doi.org/10.1111/j.1467-9310.2010.00596.x

Skinner, B. F. (1965). *Science and human behavior*. Simon and Schuster.

Steils, N., & Hanine, S. (2019). Recruiting valuable participants in online IDEA generation: The role of brief instructions. *Journal of Business Research*, *96*, 14-25. https://doi.org/10.1016/j.jbusres.2018.10.038

Suen, T. Y. T., Cheung, S. K. S., Wang, F. L., & Hui, J. Y. K. (2022). Effects of Intrinsic and Extrinsic Motivational Factors on Employee Participation in Internal Crowdsourcing Initiatives in China. *Sustainability*, *14*(14), Article 8878. https://doi.org/10.3390/su14148878

Sun, Y. Q., Fang, Y. L., & Lim, K. H. (2012). Understanding sustained participation in transactional virtual communities. *Decision Support Systems*, *53*(1), 12-22. https://doi.org/10.1016/j.dss.2011.10.006

Taha, A., Taha-Mehlitz, S., Bach, L., Ochs, V., Bardakcioglu, O., Honaker, M. D., & Cattin, P. C. (2023). Robotic colorectal surgery: quality assessment of patient information available on the internet using webscraping. *Computer Assisted Surgery*, *28*(1), Article 2187275. https://doi.org/10.1080/24699322.2023.2187275

Tarmuji, I., Maelah, R., & Tarmuji, N. H. (2016). The impact of environmental, social and governance practices (ESG) on economic performance: Evidence from ESG score. *International Journal of Trade, Economics and Finance*, *7*(3), 67.

Terwiesch, C., & Xu, Y. (2008). Innovation contests, open innovation, and multiagent problem solving. *Management Science*, *54*(9), 1529-1543. https://doi.org/10.1287/mnsc.1080.0884

Tidd, J., & Bessant, J. R. (2020). *Managing innovation: integrating technological, market and organizational change*. John Wiley & Sons.

Trantopoulos, K., von Krogh, G., Wallin, M. W., & Woerter, M. (2017). EXTERNAL KNOWLEDGE AND INFORMATION TECHNOLOGY: IMPLICATIONS FOR PROCESS INNOVATION PERFORMANCE. *Mis Quarterly*, *41*(1), 287-+. https://doi.org/10.25300/misq/2017/41.1.15

Truong, N. V. Q., Dinh, L. C., Stein, S., Tran-Thanh, L., & Jennings, N. R. (2023). Efficient and adaptive incentive selection for crowdsourcing contests. *Applied Intelligence*, *53*(8), 9204-9234. https://doi.org/10.1007/s10489-022-03593-2

Van der Have, R. P., & Rubalcaba, L. (2016). Social innovation research: An emerging area of innovation studies? *Research Policy*, *45*(9), 1923-1935.

Vroom, V. H. (1964). Work and motivation.

Vu, T., Magis-Weinberg, L., Jansen, B. R. J., van Atteveldt, N., Janssen, T. W. P., Lee, N. C.,…Meeter, M. (2022). Motivation-Achievement Cycles in Learning: a Literature Review and Research Agenda. *Educational Psychology Review*, *34*(1), 39-71. https://doi.org/10.1007/s10648-021-09616-7

Wang, R., & Chen, B. (2023). A Configurational Approach to Attracting Participation in Crowdsourcing Social Innovation: The Case of Openideo. *Management Communication Quarterly*, *37*(2), 340-367, Article 08933189221108360. https://doi.org/10.1177/08933189221108360

Wang, X., Khasraghi, H. J., & Schneider, H. (2020). Towards an Understanding of Participants' Sustained Participation in Crowdsourcing Contests. *Information Systems Management*, *37*(3), 213-226. https://doi.org/10.1080/10580530.2020.1696586

Yan, W. C., Zhou, W. J., Letizia, P., & Bichescu, B. (2017, Nov 18-21). Crowdsourcing Data Science for Innovation.*International Conference on Data Mining Workshops* [2017 17th ieee international conference on data mining workshops (icdmw 2017)]. 17th IEEE International Conference on Data Mining (ICDMW), New Orleans, LA.

Yang, Y., Chen, P.-Y., & Pavlou, P. (2009). Open innovation: An empirical study of online contests. *ICIS 2009 proceedings*, 13.

Ye, H., & Kankanhalli, A. (2017). Solvers' participation in crowdsourcing platforms: Examining the impacts of trust, and benefit and cost factors. *Journal of Strategic Information Systems*, *26*(2), 101-117. https://doi.org/10.1016/j.jsis.2017.02.001

Yin, X. C., Zhu, K., Wang, H. W., Zhang, J. P., Wang, W., & Zhang, H. (2022). Motivating participation in crowdsourcing contests: The role of instruction-writing strategy. *Information & Management*, *59*(3), Article 103616. https://doi.org/10.1016/j.im.2022.103616

Zhang, T. X., Zhang, J., & Tu, S. Y. (2024). An Empirical Study on Corporate ESG Behavior and Employee Satisfaction: A Moderating Mediation Model. *Behavioral Sciences*, *14*(4), Article 274. https://doi.org/10.3390/bs14040274

Zhao, Y. C., & Zhu, Q. H. (2014). Effects of extrinsic and intrinsic motivation on participation in crowdsourcing contest A perspective of self-determination theory. *Online Information Review*, *38*(7), 896-917. https://doi.org/10.1108/oir-08-2014-0188

Zheng, C. M., He, G. M., & Peng, Z. J. (2015). A Study of Web Information Extraction Technology Based on Beautiful Soup. *Journal of Computers*, *10*(6), 381-387. https://doi.org/10.17706/jcp.10.6.381-387

Zheng, H. C., Li, D. H., & Hou, W. H. (2011). Task Design, Motivation, and Participation in Crowdsourcing Contests. *International Journal of Electronic Commerce*, *15*(4), 57-88. https://doi.org/10.2753/jec1086-4415150402

Zumente, I., & Bistrova, J. (2021). ESG importance for long-term shareholder value creation: Literature vs. practice. *Journal of Open Innovation: Technology, Market, and Complexity*, *7*(2), 127.

Appendix A Python code

Appendix A.1 Web scraping

Below the preparation process and the python code for the web scraping of the Kaggle competition pages will be explained. This process is in line with the steps explained by Han & Anderson (2021), who use web scraping on URL's that have interactive webpages, which is also the case for the Kaggle competition pages.

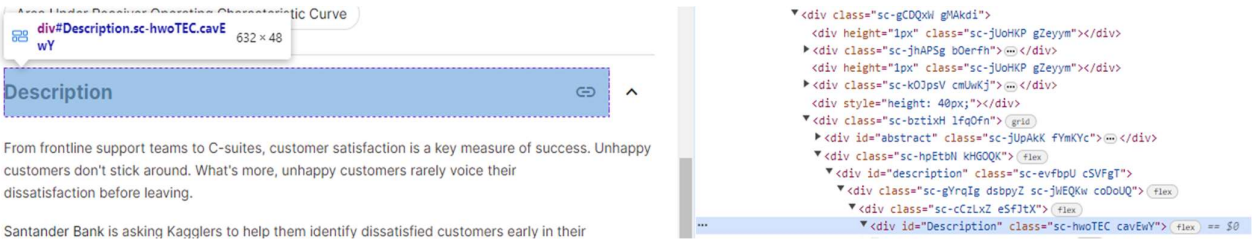
1. Downloading/installing Python and an Python IDE (in this case PyCharm) to create and run the code. IDE's are programs that allow for interpretation of the Python programming language – the program Python on itself cannot be used to execute the code below, its installation only facilitates the usage of said language by PyCharm.
   - https://www.python.org/downloads/
   - https://www.jetbrains.com/pycharm/download/
2. Downloading/installing a WebDriver. In this study a driver for Google Chrome is used. WebDrivers are software programs that take commands from e.g. a Python code and execute these commands directly in the web browser. Here it is important that the driver version corresponds with the current version of the installed browser. Also it is important to install the WebDriver in the same file directory as the Python Code, so it can be accessed by PyCharm.
   - https://googlechromelabs.github.io/chrome-for-testing/

Python code

3. *Line 1-12:* The first part of the code imports Python packages. Packages are existing lines of code with specific functionalities, ranging from simple (e.g. importing the current time from the operating system) to complex (e.g. a NLP algorithms such as BERT). They are installed in the *Terminal* of Pycharm, whereafter they can be used ("called upon") by importing them in the specific project. Below is a short description of the packages used.
   a. *OS* is used to access files within the operating system.
   b. *pandas* is used to import and create Excel and CSV files.
   c. *selenium* is the package created specific for dynamic and interactive webpages. Multiple subfunctions of selenium are used.
   d. *logging* is used to save data in between commands.
   e. *BeautifulSoup* is used to read the HTML code of the website.

4. *Line 13-26:* Next, the titles of the competitions with a monetary reward are loaded in, using the file path specified to the CSV document.

5. The parameters of the function that extracts description text are determined by inspecting the competitions' webpage HTML code (by right clicking and selecting Inspect or pressing F12 on a keyboard). The screenshot below shows that the header of the descriptions is also called "Description" in the HTML code.



6. *Line 27-40:* After that, the function that searches for the text of the Description within the webpage is defined. This is combination of the parameter identified in step 5 and the BeautifulSoup package. An output is added in case a competition page does not have a description. However, the output file shows that every competition has a description text.

7. *Line 41-44:* Next, the path of the WebDriver is setup, ensuring that the code can access and use the WebDriver.

8. *Line 45-70:* Next, the main function of the code is defined. This is a combination of the package Selenium using the installed WebDriver to navigate, click and type on the Kaggle website. It loops through the titles loaded in at step 4, and types these into the search box of Kaggle. After clicking and loading the corresponding competitions' webpage the

9. *Line 71-81:* Then, the defined main function is run. First an empty data frame is created to store the descriptions, whereafter the main function is executed for every title. The title and the description of the competition that are extracted are then stored in the created data frame. When the function has looped through every title, the Chrome Driver is closed down.

10. *Line 82- 87:* Finally, the stored data from the previous step is stored in an data frame suitable for Excel output. Then the folder to store the excel file is defined, whereafter the Excel file is created using the generated data frame. To indicate the process has finished, a message is printed.

Appendix A.2 ESG Score – Mansouri & Momtaz (2022)

Below, the steps in the Python code for the ESG Scoring method, adapted from Mansouri and Momtaz (2022), will be listed and explained. The complete Python code will be attached as a separate document.

1. *Line 1-4:* First the required Python packages are imported. These are mainly standard packages, such as *pandas* and *numpy*, to handle excel and CSV files.

2. *Line 5-10:* The next part imports the ESG word-list, created by Mansouri and Momtaz (2022). As mentioned in the chapter 3.2, this word-list is based on relevant documents, including news-papers, journal articles, annual reports, using a ML approach. The word-list can be downloaded via the following link:

   - https://www.dropbox.com/s/e28dihonntg8o82/expanded_dict.csv?dl=1

3. *Line 11-41:* Thirdly, the function that divides the competition descriptions into words is defined. This uses the 'ngrams' package. After this, the list of words is controlled for irrelevant data, such as telephone numbers and email addresses, using the 'cleaning' package.

4. *Line 42-66:* Then, the main function is defined. This involves the counting of words and dividing it by the relevant words taken from the word list created by Mansouri and Momtaz (2022), where the weight of the word is adjusted by the size of the word list.

5. *Line 67-85:* After, the main function is integrated in a loop that runs through all descriptions, and creates a data frame with every E-, S-, and G-Score per description.

6. *Line 86-90:* Finally, the results are loaded into a new excel file and a message is printed to indicate the analysis has finished.

Appendix A.3 BERT ESG Score

The Python code for the ESG analysis, using the BERT NLP model, will be explained below. Here it is important to note that this model has three separate version, each for one of the components of ESG; *Environmental, Social* and *Governance.* This is because the BERT models are each trained individually with their own training data and can therefore not be used together.

1. *Line 1-6:* The required packages are imported. The main difference with the method explained in Appendix Appendix A.2 is that this uses the 'transformers' package, which includes very sophisticated NLP models. These lines of code serve as a base model that can be trained depending on the context. Also, 'parsing' packages are imported that are used to divide large pieces of text into sentences that can be analyzed by algorithms.

2. *Line 7- 16:* Next, the subpackages from 'transformers' are loaded and defined, after which the file path to the input file is integrated.

3. *Line 17-43:* The main function is then defined, utilizing the 'pipeline' subpackage to classify sentences in the descriptions as *Environmental*, *Social* or *Governance.* Essentially, the pipeline function is an algorithm that compares sentences in the descriptions to sentences that are labeled as *Environmental*, *Social* or *Governance* in the 'training' data of the BERT model and 'marks' sentences it deems corresponding. The function also includes the option to input a specific range of competitions for testing purposes.

4. *Line* 55-70: Then, the main function is integrated in a loop that cycles through every competition description and stores both the number of marked sentences in a data frame, as well as the marked sentences in a separate frame.

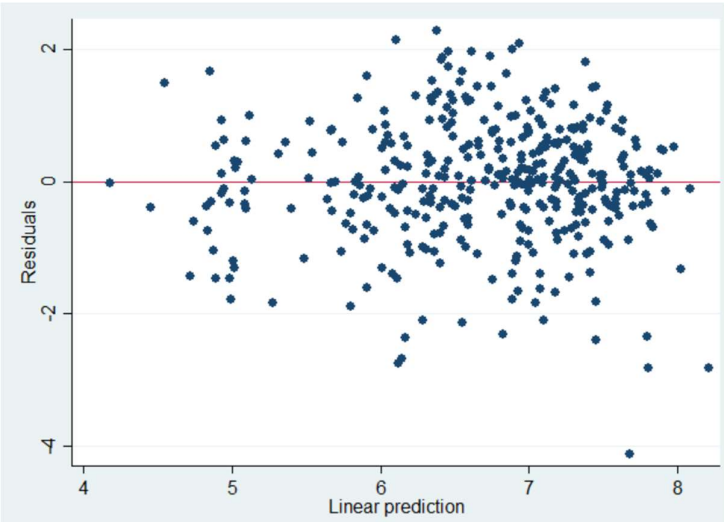
5. *Line 44-54:* The data frames are used to create two new files. The number of marked sentences are used for the main analysis of the thesis, the document with the marked sentences is to check for correctness of the algorithm.

6. *Line 71:* Finally, a message is printed to indicate the process has finished.

Appendix B Assumption analysis

In order to confidently perform multiple regression analysis, various assumptions about the data must be met. Firstly, the linear relationship between the dependent variable and the independent variables is checked. To determine a linear relationship, the pattern of residuals should be randomly scattered around zero. Looking at Figure B. 1, this statement holds.

Figure B. 1 Residuals vs Fitted plot



For homoscedasticity, the Breusch-Pagan / Cook-Weisberg test indicated that the null hypothesis of constant variance could not be rejected (chi2(1) = 0.08, p = 0.7822), suggesting no evidence of heteroskedasticity and confirming that the variance of the residuals is constant (Figure B. 2).

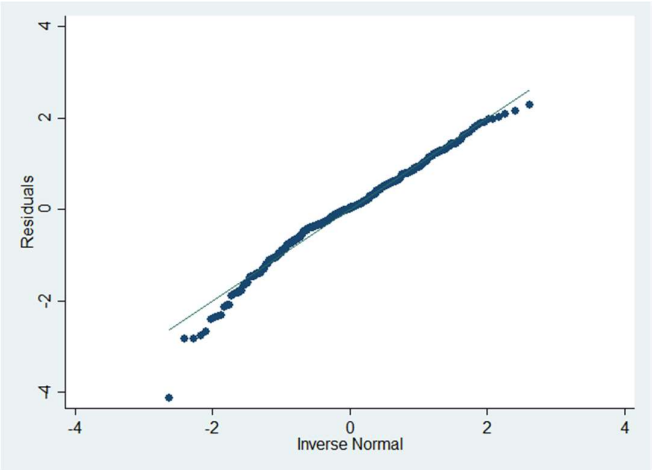
Figure B. 2 Breusch-Pagan / Cook-Weisberg test

```
. estat hettest

Breusch-Pagan / Cook-Weisberg test for heteroskedasticity
         Ho: Constant variance
         Variables: fitted values of logtotalcompetitors

         chi2(1)      =     0.08
         Prob > chi2  =   0.7822
```

To test for normality, a Q-Q plot was created from the residuals (Figure B. 3), and as per Hair (2009) the plot showed observations at approximately a 45-degree angle, thus meeting the normality assumption. Lastly, to verify the absence of multicollinearity, the Variance Inflation Factor (VIF) was tested. As can be seen in Table B 1, the highest VIF value was 2.14, indicating no multicollinearity issues among the independent variables.

To test for normality, a Q-Q plot was created from the residuals (Figure B. 3), and as per Hair et al. (Hair, 2009) the plot showed observations at approximately a 45-degree angle, thus meeting the normality assumption.

Figure B. 3 Q-Q plot of Residuals



Lastly, to verify the absence of multicollinearity, the Variance Inflation Factor (VIF) was tested. As can be seen in Table B 1, the highest VIF value was 2.14, indicating no multicollinearity issues among the independent variables.

Table B 1 VIF values

Variable	VIF	1/VIF
Independent variables		
TotalCompetitors (log)	2.14	0.468
ESG Score (normalized)	1.46	0.684
RQ × ESG (normalized)	1.24	0.806
Control variables		
BanTeamMergers	1.1	0.911
TeamConstraint	1.44	0.696
NumPrize (log)	1.88	0.531
Days_Between	1.17	0.857
Mean VIF	**1.49**	

Appendix C Results of E-, S- and G-scores separately

Table C. 1 Monetary rewards, Environmental scores, interaction effects and total competitors

	DV = *TotalCompetitors (log)*		
	(1)	(2)	(3)
RQ × E-score (normalized)			-0.118**
			(0.057)
E-Score (normalized)		0.051	0.102*
		(0.050)	(0.056)
Reward Quantity (log)		0.307***	0.328***
		(0.044)	(0.045)
BanTeamMergers	-0.648**	-0.631**	-0.631**
	-0.28	(0.266)	(0.264)
TeamConstraint	0.723***	0.577***	0.568***
	-0.121	(0.117)	(0.117)
NumPrize (log)	0.700***	0.249**	0.252**
	-0.109	(0.121)	(0.121)
Days_Between	0.003***	0.001	0.001
	-0.001	(0.001)	(0.001)
Constant	5.199***	2.941***	2.746***
	-0.142	(0.351)	(0.362)
Observations	380	380	380
R-squared	0.322	0.401	0.408

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table C. 2 Monetary rewards, Social scores, interaction effects and total competitors

	DV = *TotalCompetitors (log)*		
	(1)	(2)	(3)
RQ × S-score (normalized)			-0.123**
			(0.059)
S-Score (normalized)		0.090*	0.135**
		(0.0517)	(0.056)
Reward Quantity (log)		0.301***	0.330***
		(0.044)	(0.046)
BanTeamMergers	-0.648**	-0.627**	-0.589**
	-0.280	(0.264)	(0.264)
TeamConstraint	0.723***	0.568***	0.529***
	-0.121	(0.117)	(0.117)
NumPrize (log)	0.700***	0.230*	0.216*
	-0.109	(0.122)	(0.121)
Days_Between	0.003***	0.001	0.002
	-0.001	(0.001)	(0.001)
Constant	5.199***	3.035***	2.757***
	-0.142	(0.353)	(0.376)
Observations	380	380	380
R-squared	0.322	0.404	0.411

Standard errors in parentheses
*** p<0.01, ** p<0.05, * p<0.1

Table C. 3 Monetary rewards, Governance scores, interaction effects and total competitors

	DV = *TotalCompetitors (log)*		
	(1)	(2)	(3)
RQ × G-score (normalized)			-0.0703
			(0.054)
G-Score (normalized)		0.056	0.082
		(0.051)	(0.054)
Reward Quantity (log)		0.308***	0.317***
		(0.041)	(0.045)
BanTeamMergers	-0.648**	-0.694***	-0.708***
	-0.28	(0.266)	(0.266)
TeamConstraint	0.723***	0.587***	0.587***
	-0.121	(0.116)	(0.116)
NumPrize (log)	0.700***	0.233*	0.236*
	-0.109	(0.123)	(0.122)
Days_Between	0.003***	0.001	0.001
	-0.001	(0.001)	(0.001)
Constant	5.199***	2.948***	2.858***
	-0.142	(0.351)	(0.358)
Observations	380	380	380
R-squared	0.322	0.401	0.404

Standard errors in parentheses *** p<0.01, ** p<0.05, * p<0.1