Crowdsourcing for Innovation

Investigating the interaction- and contributor-related characteristics that lead to team success for offline crowdsourcing for innovation

L.A. Bijker
4176723

Master's thesis
Sustainable Business & Innovation
Faculty of Geosciences, Utrecht University

Supervisor: Dr. M.M.H. Chappin

2nd reader: Dr. L. Piscicelli

Date: 27-01-2023

Word count: 22325

Abstract

Companies increasingly turn to crowdsourcing for innovation, as it allows them to tap into the collective intelligence of the crowd to generate new ideas and solutions more effectively. Not only can this be relevant for the organisation to stay competitive, but also it has the potential to play a significant role in addressing global challenges such as climate change, as the complexity and wickedness of such challenge requires a wide range of perspectives that are not limited to one sector or discipline. It comes then without a surprise that many scholars are interested in what makes these types of events successful in relation to the ideas generated or the real-life impact made. Although the factors that contribute to success are widely researched in the context of online crowdsourcing, research lacks on offline forms of crowdsourcing**.** Also, whereas most crowdsourcing initiatives revolve around individual contributors solving problems, research about crowdsourcing teams is scarce. In addition, most of those factors in relation to team success are studied in isolation while there are many reasons to believe that these success factors show interdependencies. Therefore, this study aims to fill this gap by focusing on the interplay between contributor- and interaction-related conditions that influence team success. This study applies a deductive configurational design using Qualitative Comparative Analysis, which is inherently suitable for studying interrelationships. The sample consists of 19 crowdsourcing teams that participated in the Circular Challenge, an offline team-based crowdsourcing for innovation event that revolves around commercialising excess waste streams in the context of the circular economy. Structured interviews and self-completion surveys were conducted among representatives of these teams. This was analysed with csQCA. The findings show that contributor-related and interaction-related conditions are important for team success and that crowdsourcing teams can become successful through an internal pathway and through an external pathway. Moreover, it is theorised that interaction-related and contributor-related conditions show a substitutive relationship. This study provided a methodological contribution by confirming the relevance of studying interrelationships within this field. The results provide an avenue for future research through the suggestion that there is an interplay between contributor-related and interaction-related factors. The findings have several practical implications for crowdsourcing organisers and teams participating.

Keywords: crowdsourcing, innovation, crowd contest, innovation challenge, interaction, contributor

Content

List of tables

List of figures

1. Introduction

Radical innovations and systemic changes are needed to counteract the effect of climate change and to maintain Earth's habitability (Steffen et al., 2018). Governmental organisations, firms and NGOs have turned to crowdsourcing as a means of solving such problems (e.g. Boudreau Lakhani, 2013; Armisen & Majchrzak, 2015), as the ”wisdom of the crowd” has the potential to quickly provide novel input of high quality to sought solutions (e.g. Chesbrough, 2003; Jain, 2010). Although crowdsourcing comes in many scents and flavours (e.g crowdfunding, crowd voting, open innovation, open source software) (Ali-Hassam & Alam, 2016), all forms share the same recipe: crowdsourcing revolves around the idea of outsourcing a task to a ‘’crowd’’ instead of to a designated ‘’agent’’, like an organisation, team, or individual as a contractor. As the call for the outsourced task is open, in principle everyone can contribute (Howe, 2006; Jeppesen & Lakhani, 2010).Crowdsourcing has great potential for innovation, as organisations can provide themselves access to much knowledge previously unavailable, which is distributed among many diverse stakeholders (Chanal & Caron-Fasan, 2010; Schenk & Guittard, 2011). It can provide them with richer content and better solutions in a creative and cost-effective way than they could make themselves (Chesbrough, 2003; Jain, 2010; Parameswaran & Whinston, 2007). Many firms leverage the crowd’s potential by calling for ideas on platforms that specialise in crowdsourcing. Platforms such as Innocentive or Ninesigma have set out thousands of problems to be solved, and as these platforms report that the majority of their challenges are solved successfully (Natalicchio, Petruzzelli, Garavelli, 2014), the crowd seems to be a driver of innovation. There are several well-known types of crowdsourcing, innovation contests and hackathons among them. Innovation contests are typically used to solve innovative or challenging problems in the form of an open call to a crowd (Afuah & Tucci, 2012; Blohm, Zogaj, Bretschneider & Leimeister, 2018; Boudreau & Lakhani, 2013; Jeppesen & Lakhani, 2010). Hackathons are time-bounded multi-day events, during which people work in teams on a provided challenge (Trainer, Kalyanasundaram, Chaihirunkarn, Herbsleb, 2016). Initially, hackathons were meant to solve computer problems; the word itself, etymologically compounded from “Hack" and “Marathon” gives away that it originates from the field of information technology (Soltani, Pessi, Ahlin, Wernered, 2014). However, these now exist in many other fields (e.g. Climathon, Hackhealth, Hacksforhumanity). A commonly studied topic is the conditions that lead to contributor success in crowdsourcing initiatives (e.g. Javadi Khasraghi & Hirschheim, 2021; Bullinger, Neyer, Rass & Moeslein, 2010). The studied factors for team success can be divided into three categories following the framework of Schemmann (2018): *contributor*-related characteristics (motivations, expertise, effort and engagement), *interaction*-related characteristics (between the contributors as well as between the knowledge-seeker and the contributor), and *task*-related characteristics (the formulation, framing, transparency and accessibility) (Schemmann, 2018; Schemmann, Herrmann, Chappin, Heimeriks, 2016).

There are predominantly studies about the success of contributors within *online* forms of crowdsourcing for innovation. This is not surprising, as the Internet and Web 2.0 tools have enabled organisations to reach more people than ever before, from all around the world, with a minimum amount of resources. However, there are also offline variants of crowdsourcing. Although the literature on Hackathons could provide insights into success factors for contributors, there are often two limitations to the generalisability of these studies. Firstly, these events are time bound and relatively short by nature (typically 1-3 days), which a lot of innovation contests are not. Hackathons are still predominantly in the field of IT, although a variety of applications exist nowadays. This means that there is a gap within the literature on the success factors for forms of crowdsourcing for innovation that are *offline* and *team-based.* The studies that *have* been done on these factors, focus on the sole influence of those factors on the dependent variables, and not on the interplay of those factors. Therefore, it is worth investigating under what (combinations of) conditions these offline teams that participate in crowdsourcing for innovation are successful. As interaction-related and contribution-related factors have the strongest link with contributing *teams,* this study aims to shed light on those combinations of contributor-related and interaction-related factors that contribute to team success within crowdsourcing for innovation. The research question becomes:

What combinations of contributor-based and interaction-based characteristics lead to team success within offline team-based crowdsourcing for innovation?

This study applies a configurational approach, using Qualitative Comparative Analysis (Ragin, 1987). The sample of investigation consists of teams participating in an innovation contest “Circular Challenge” in the Netherlands. Data about the teams will be collected through structured interviews with individual team members. Crisp set QCA (csQCA) is used to analyse the data.

This study contributes to the existing literature in several ways. First of all, within the domain of crowdsourcing for innovation, it will provide novel information about interaction- and contributor-related characteristics that are essential for team success, specifically in the context of *offline team-based crowdsourcing*. To date, limited studies have been done on this topic. Also, the application of QCA within this field is novel, as QCA has only been applied a handful of times within the crowdsourcing for innovation domain (Smeets, Chappin & Kaashoek, 2012; Xu, Wu & Hamari, 2022). Managerial implications can culminate out of this study: it has the potential to help organisations that host crowdsourcing initiatives to shape their governance and team structures so as to set up participating teams for success. This study is relevant on a macro scale as it could contribute to the largest problems of our time needing to be solved quickly; offline team-based crowdsourcing has the potential to contribute to creative and high-quality solutions fast and effectively. The structure of this report is as follows. First, the theoretical background is explained. This is followed by a section about the methodology, a section about the results, the discussion, theoretical implications, practical recommendations, limitations & future research and the appendices.

2. Theoretical Framework

This study's theory is based on research on innovation contests, hackathons, and the wider field of innovation management. The following section aims to provide an understanding of the relationship between contributor-related and interaction-related characteristics of crowdsourcing teams and team success. First, it gives a general overview of the field of crowdsourcing. Then, it elaborates upon the subfield of crowdsourcing for innovation, after which specific forms within the research context are given: offline team-based crowdsourcing for innovation. Finally, the current research on team success and its influencing factors are explained, as it is the interdependencies between these factors that are the main focus of this study.

2.1 Crowdsourcing

The wide field of crowdsourcing

Although crowdsourcing has been a known concept for roughly two decades, the scientific world still lacks a common definition of the concept (Estellés-Arolas & González-Ladrón-de-Guevara, 2012; Zhao & Zhu, 2014). Crowdsourcing research is a dynamic and vibrant research field that has been gradually growing over the years. Several scholars have tried to grasp the concept and its evolution by performing literature reviews (Zhao & Zhu, 2014; Hossain & Kauranen, 2014; Ghezzi, Gabelloni, Martini & Natalicchio, 2018; Karachiwalla & Pinkow, 2021). Many researchers attempted to create their definitions of crowdsourcing (Hopkins, 2011), each departing from different theoretical bases and practices. This is also the case for many typologies and categorisations (Ali-Hassan & Hallam, 2016; Garavelli, Peruzzelli, Natallichio, 2014; Karachiwalla & Pinkow, 2021). This has resulted in an unstructured evolution of the field, with several strands of research authors are working on (Ghezzi et al., 2018), crossing the boundaries of innovation and technology theory (Geiger, Rosemann, Fielt & Schader, 2012; Whitla, 2012). To date, the term still entails many different elements and practises, and thus the definitions of the concept remain fuzzy and unclear (Estellés-Arolas & González-Ladrón-de-Guevara, 2012; Zhao & Zhu, 2014; Whitla, 2012). This study takes the earliest definition created by (Howe, 2006).They regard crowdsourcing as the process of outsourcing a task traditionally performed by a designated agent (e.g. an employee, or contractor) to a crowd consisting of an undefined but large group of people through an open call.

Crowdsourcing for innovation

As previously mentioned, there are countless different subgroups of crowdsourcing, like crowdfunding, crowd voting, open innovation, citizen science, and open source software (Ali-Hassam & Alam, 2016). However, this study focuses on crowdsourcing for innovation purposes. Crowdsourcing can be regarded as a form of open innovation (Chesbrough, 2003). Central to the idea of open innovation is the use of ideas

external to the organisation for innovation purposes (Chesbrough, 2006), also called 'inbound innovation' (Dahlander & Gann, 2010). However, within crowdsourcing for innovation, it is not other firms or universities the external ideas come from but rather from a crowd (West & Bogers, 2014). Crowdsourcing has great innovation potential, as organisations can provide themselves with access to much knowledge previously unavailable, distributed among many diverse stakeholders (Chanal & Caron-Fasan, 2010; Schenk & Guittard, 2011). It can provide them with richer content and better solutions than they could make themselves in a creative and cost-effective way (Chesbrough, 2003; Jain, 2010; Parameswaran & Whinston, 2007). Examples of crowdsourcing for innovation are innovation contests, tournament-based crowdsourcing or broadcast search (Karachiwalla & Pinkow, 2021; Afuah & Tucci, 2012; Boudreau & Lakhani, 2013; Terwiesch & Xu, 2008). Regardless of the terminology used, such types of events are usually structured in the following way. First, a call for solving a specific problem is set out by a knowledge seeker or intermediary organisation. Then, the crowd can contribute its input. Thirdly, the knowledge seeker evaluates the ideas and chooses the winning idea. Often, the winning contributor receives a reward, which can be in a monetary or non-monetary form (Afuah & Tucci, 2012; Blohm et al., 2018; Schemmann, 2018; Karachiwalla & Pinkow, 2021). Well-known online platforms such as Innocentive or Ninesigma have set out thousands of problems to be solved, and these platforms report that the majority of their challenges are solved successfully (Natalicchio et al., 2014).

Offline team-based crowdsourcing

The origins of offline team-based crowdsourcing for innovation can be traced back to the shift from the manufacturer-active to the customer-active paradigm in the late 1970s (Bilgram, Brem, Voigt, 2008; von Hippel, 1978; Foxall and Tierney 1984). This has led to the development of user-centric innovation communities, which revolve around lead-user testing and new product development for a firm (Bilgram, Brem, Voigt, 2008; Franke and Shah, 2003; Hienerth, 2006). Despite the fact that these users were mostly involved individually, most of these users preferred innovation in groups (Franke and Shah, 2003; Lüthje, Herstatt, Von Hippel, 2005; Füller & Matzler, 2007). The nature of these communities changed, as with the arrival of Web 2.0 applications, many communities changed with regards to their locus of operation from offline to online (Bilgram, Brem, Voigt, 2008; O'Reilly, 2005). Although the centre of gravity of crowdsourcing initiatives as well as the research focusing on these initiatives lay in the online sphere, some forms of crowdsourcing stayed offline. One of these forms is Hackathons. Hackathons are time-bound events to solve a problem (Jaribion, Khajavi, Järvihaavisto, Nurmi, Gustafsson, & Holmström, 2021). The term hackathon comes from "Hack" and "Marathon", originating from the programmer world. Hackathons were initially understood as ''coding events''. However, the concept is now also applied for innovation purposes in other fields, whether it is environmental related (e.g. Climathon), health-related (e.g. Hackhealth), or social-related (e.g. Hacksforhumanity) (Nolte, Chounta & Herbsleb, 2020).

2.2 Team success

As with any innovation process, understanding what leads to a successful output is of high interest. Within offline team-based crowdsourcing for innovation such successful output revolves around team performance. However, this has been studied in many different ways. As crowdsourcing for innovation is often used by organisations looking for new or innovative solutions to their problems, success with respect to the ideas generated is often central to these studies. Some studies focus on the number of ideas. For example, on a contest level, Walter & Black (2011) find that the duration of the contest and strong brands lead to more contributed ideas. Many studies revolve around the quality of the idea (e.g. Javadi Khasraghi & Hirschheim, 2021; Walter & Black, 2011; Lykourentzou, Antoniou, Naudet & Dow, 2016; Bullinger, Neyer, Rass & Moeslein, 2010, Wang, 2021). Although idea quality is often central to studies analysing contributor success, many studies also focus on the realisation of the idea. It is not surprising that such success has been investigated, as ideas are only worth something when executed (Levit, 1963). Types of idea realisation include idea implementation (Schemmann, Hermann, Chappin, Heimeriks, 2016), partial idea implementation (Schemmann, Chappin & Herrmann, 2017), and establishment of a startup (Kitsios & Kamariotou, 2018). Moreover, it can also concern the continuation of idea development after the innovation contest has finished, such as sustained participation (Langner & Seidel, 2015; Seidel & Lagner, 2015; Shah, 2006; Javadi Khasraghi, Wang, Li, 2020), or project continuation (Pe-Than, Nolte, Filippova, Bird, Scallen, & Herbsleb, 2018). As Mathieu, Maynard, Rapp and Gilson (2008) noted, team performance heavily relies on the specific context, specifically relating to the tasks executed and the goals that need to be achieved. As described in the section above, both the quality of ideas and their potential implementation are important factors in quantifying the success of an innovation event. For this reason, both idea quality and project continuation (as a measure of the implementation of ideas) are defined as outcome conditions in this study.

As in this study's research context, idea quality is of paramount importance, this is taken as an outcome condition. Another outcome condition that was taken is project continuation, as it is deemed important to know what conditions lead teams to continue their project after the crowdsourcing initiatives, potentially having a real-life impact.

2.3 Factors for contributor success

The wider field of crowdsourcing for innovation has been developing gradually since the earliest mention of the concept in the mid-00s (Howe, 2006a). However, there currently is a notable but small stream of literature about the factors that are important for contributor success within the context of offline team-based crowdsourcing for innovation specifically. These factors can differ with respect to the crowdsourcing design elements they are related to. The factors can be related to the *contributors* involved, the *interaction* of the contributors, and the *task* outsourced to the crowd (Schemmann, 2018; Schemmann, Herrmann, Chappin, Heimeriks, 2016). Contributor-related factors could be, for example, previous bonds and

familiarity with the knowledge-seeking organisation (Chan, Li, Zhu, 2015), personality compatibility within a team (Lykourentzou, Antoniou, Naudet, Dow, 2016), or various types of diversity (Wang, 2021; Nolte, Chounta, Herbsleb, 2020; Riedl & Woolley, 2017). Interaction-related factors could be, for example, attention paid to other contributors' ideas (Schemmann et al., 2016), and interaction between contributors and between contributors and knowledge seekers (Chan, Li, Zhu, 2015). Task-related factors could be, for example, the nature of the tasks (Seidel & Langner, 2015), task formulation and task framing (Schemmann, 2018). The emphasis on the identification of important factors laid mainly on finding those closest to the studies context of offline team-based crowdsourcing for innovation. For example, those found in studies about offline user-centric innovation, offline hackathons, or other offline team-based forms of crowdsourcing for innovation initiatives. As it turned out there was a limited number of studies done in this specific context, thus an investigation was done into adjacent and overlapping fields of crowdsourcing for innovation. For example, forms in which the contributors are individuals instead of teams, and forms that are online instead of offline. Whenever research on factors here was also lacking, the factors were found in overarching fields. For example, crowdsourcing, innovation management, or organisational psychology. An overview of the factors studied in the literature can be found in table 1. This list is non-exhaustive, as no systematic literature was done. This study focuses on both contributor-related factors and interaction-related factors, as multiple scholars highlighted the need for further research that includes interaction-related characteristics of ideators, besides contributor-related characteristics (Schemmann, Herrmann, Chappin, Heimeriks, 2016).

These factors included were chosen because of their presumed relationship either with one of the outcome dimensions, idea quality or project continuation or with other factors (e.g. through moderating or mediating relationships), based on literature.

The factors chosen with regard to contributors are *psychological safety, teamwork*, *cognitive diversity* and *gender diversity.* The factors chosen with regard to interaction-related characteristics are *team interaction* and *client interaction.* The further paragraphs will elaborate on the chosen factors and will justify the selection on the basis of the interrelationships.

Table 1: An overview of factors that influence team success

Reference	Channel	Type of contributor	Level of analysis	Independent Variable	Dependent Variable
Boons & Stam (2019)	Online	Individual	Individual	Related perspectives, Unrelated perspectives	Idea quality
Schemmann, Herrmann, Chappin, Heimeriks (2016)	Online	Individual	Individual	Ideator motivation, Attention paid to other ideas, Idea popularity	NPD: Idea implementation or rejection
Schemmann, Chappin, Herrmann (2017)	Online	Individual	Individual	Idea development Input to other ideas Positive attention Pre-ideation attention	Idea Implementation

Javadi Khasraghi & Hirschheim (2021)	Online	Team	Team	Discussion-forum performance, Solution sharing performance	Competition Performance
Chan, Li, Zhu (2015)	Online	Individual	Individual	Past ideation participation, Peer to Peer interaction, Peer to Firm interaction	Idea Generation, Subsequent Idea Generation
Khasraghi, Wang, Li (2020)	Online	Individual	Individual	Structural capital, Familiarity with organisation, Experience with organisation	Sustained participation
Armisen & Majchrzak (2015)	Online	Individual	Individual	Discussion thread variety, # Collaborative versus argumentative posts, # Prior participant posts	Innovative idea generation
Zheng, Hou, Li (2014)	Online	Individual	Individual	Intrinsic Motivation, Extrinsic Motivation	Participation
Lykourentzou, Antoniou, Naudet, Dow (2016)	Online	Team	Team	Personality compatibility	Team performance, Individual perceptions
Riedl & Woolley (2017)	Online	Team	Team	Burstiness Diversity of information	Team performance
Dissanayake, Zhang, Gu (2015a)	Online	Team	Team	Alignment social and intellectual capital	Team performance
Dissanayake, Zhang, Gu (2015b)	Online	Team	Team	Member centrality, Skill level, Skill alignment	Team performance
Wang (2021)	Online	Team	Team	Expertise diversity, Winning ideas diversity, Winning challenges diversity, Geographic diversity, Tenure diversity,	Team performance
Fuger, Schimpf, Füller, Hutter (2017)	Online	Team	Team	Team structure: user role heterogeneity	Team performance
Nolte, Chounta, Herbsleb (2020)	Offline	Teams	Team	Technical preparation, Hackathon win, Skill diversity, Continuation intentions, Fit of technical capabilities, # technologies used in the project	Continuation: short-term and long-term
Langner & Seidel (2015)	Online	Individual	Individual	Member identification with firm	Sustained participation
Seidel & Langner (2015)	Online	Individual	Individual	Nature of design tasks (within community)	Sustained participation
Shah (2006)	Online	Individual	Individual	Fun, Challenge, Reciprocity, Fairness	Participation

Javadi & Gebauer (2019)	Offline	Individual	Individual	Information Diversity	Idea Integration
Soltani, Pessi, Ahlin, Wernered (2016)	Offline	Team	Contest	Problem area definition Reward Skill diversity Competence diversity Mentor availability Expertise communication Jury expertise Hackathon entry requirements	Hackathon Success
Walter & Black (2011)	Online	Individual	Contest	Rewards Duration Description Length Specificity Answer type Brand strength Market Maturity	Idea quantity Idea quality
Majchrzak & Malhotra (2016)	Online	Individual	Contest	Order of knowledge sharing	Innovative outcomes
Zaamout & Barker (2018)	Online	Team	Team	Member behaviour: Impact Activity Policing/Rowdiness	Team Contribution Quality

Gender diversity

Although the link between diversity and innovation is not extensively studied in the context of innovation crowdsourcing specifically (Jones, Chace, & Wright, 2020; Joecks, Pull & Vetter, 2013), many studies in the innovation field show that gender diversity is important for team success. A study by the Boston Consulting Group found that gender diversity positively correlated with the innovation performance of management teams (Lorenzo, Voigt, Tsusaka, Krentz, Abouzahr, 2018). Moreover, gender diversity in R&D teams positively correlates to radical innovation output (Díaz-García, González-Moreno & Jose Sáez-Martínez, 2013), or, on a firm level, leads to higher innovation potential (Østergaard, Timmermans & Kristinsson, 2011). Similarly, gender diversity within new venture teams was found to positively affect the venture's innovation performance (Dai, Byun & Ding, 2019). The rationale underlying the positive effects of gender diversity on innovation performance is that teams with a greater presence of women will increase the differentiation of the knowledge base (Dai, et al., 2019), for two reasons. The first is that women exhibit, on average, a higher sensitivity to nuances and cues (Darley & Smith, 1995). They are also more accepting towards information that is new and differs from the existing mental models (Chung & Monroe, 1998), which enables them to be more open to new information and thereby knowledge differentiation. The second reason is that team resources are likely to be diversified concerning human and social resources, as women often differ from men in their experiences regarding socialisation, career trajectories and social

networks (Mateos de Cabo, Gimeno, & Nieto, 2012; Singh, Terjesen, & Vinnicomb, 2008). To see whether gender diversity is also important in this research context it is included as a condition.

Cognitive diversity

The relationship between cognitive diversity and team functioning has been widely studied (Mello & Rentsch, 2015) and many scholars found that cognitive diversity has a positive effect on team performance (e.g. Horwitz & Horwitz, 2007; van Dijk, van Engen, & van Knippenberg, 2012). Within the field of crowdsourcing for innovation, similar relationships have been found. For example, it has been found that when expertise diversity is high that contributors have a higher chance of success within online crowdsourcing contests (Wang, 2021; Riedl & Wooley, 2017), or that continuing their project in offline innovations contest is more likely when knowledge and skill diversity is high (Nolte, Chounta, Herbsleb, 2020). Javadi & Gebauer (2019) found that information diversity among group members was important for idea integration, which is important for innovation. Boons & Stam (2019) found that contributors with diversity regarding their related and unrelated perspectives are more likely to integrate ideas from those fields, thereby creating valuable ideas for crowdsourcing initiatives.

Although cognitive diversity has been widely studied, there is a multitude of conceptualisations of the term cognitive diversity and operational definitions, therefore, vary strongly (Mello & Rentsch, 2015). For example, the term can refer to differences in personal and professional backgrounds (Colón-Emeric et al., 2006), personalities, values and attitudes (Harrison, Price, Gavin & Florey, 2002; Tegarden, Tegarden & Sheetz, 2009), differences in cognitive processes (Kurtzberg, 2005) and diversity in knowledge, skills and capabilities (Martins, Schilpzan, Kirkman, Ivanaj & Ivanaj, 2013). This study aims to explain contributor success based on multiple dimensions of cognitive diversity, as Martins et al. (2013) suggested. The dimensions taken are the following: *educational diversity, nationality diversity, and job diversity.*

Educational diversity has often been used for examining cognitive diversity (e.g. Dahlin, Weingart & Hinds, 2005; Simons, 1995). Within a wider context, educational diversity was found to positively influence a firm's innovation performance. For example, Mohammadi, Broström, and Franzoni (2017) found that workforce diversity in terms of educational background positively correlates to a firm's innovation performance, and Østergaard et al. (2011) find that education diversity within firms leads to higher innovation potential. Within crowdsourcing for innovation, this is also the case. For example, Boons & Stam (2019) found that individuals within crowdsourced idea challenges contribute higher quality ideas when they possess both related and unrelated perspectives regarding their educational background. The underlying mechanism is that having knowledge in diverse educational backgrounds enables the contributor to better identify, assimilate and apply external knowledge for innovation purposes (Bogers et al., 2018; Østergaard et al., 2011), which is known as absorptive capacity (Cohen and Levinthal, 1990).

To see whether educational diversity is also important in this research context it is aggregated with the other two forms of diversity to form the construct of cognitive diversity.

Nationality diversity refers to diversity concerning growing up in different countries, as cultural differences have been shown to have an effect on norms, values, and even communication styles and cognitive processes (Gibson & Gibbs, 2006). Literature on the effect of demographic diversity on team performance has been two-sided, with one side claiming that it hinders performance as the differences with regard to communication, values and norms induce social categorisation, thereby risking a lower level of knowledge exchange (Bell, Villado, Lukasik, Belau, Briggs, 2011). However, other scholars have found promising results with regard to the effect on innovation performance. For example, geographic diversity was positively related to team success in the context of online crowdsourcing for innovation, as team members were encouraged to collaborate by learning about each other's diverse perspectives (Wang, 2021). In a wider context, current insights about the effect of cultural differences on team performance are ambiguous and, therefore, multiple scholars have called for a further investigation of the role of culture in team performance (Klein, McHugh, 2005; Salas, Cook, Rosen, 2008). To see whether nationality diversity is also important in this research context it is aggregated with the other two forms of diversity to form the construct of cognitive diversity.

Job diversity refers to the diversity concerning the work experience that members of a group have, whether it regards functional background or tenure (van Dijk, Engen, van Knippenberg, 2012). It was found to have a positive effect on team success. Wang (2021) finds that teams that are diverse in terms of their expertise are more likely to win online crowdsourcing contests. In a broader context, several meta-analyses show that job-related diversity was positively related to innovative performance (Hülsheger, Anderson, Salgado, 2009; van Dijk et al., 2012). Job-related diversity has not been studied specifically in crowdsourcing for innovation. However, some scholars suggested that job-related diversity is particularly important in team innovation when creative thinking is required (van Dijk et al., 2012), and multiple scholars have called for the inclusion of work experience in diversity concepts (Boons & Stam, 2019; Martins et al., 2013). Thus, to see whether job diversity is also important in this research context it is aggregated with the other two forms of diversity to form the construct of cognitive diversity.

Psychological safety

Psychological safety revolves around mutual respect and trust among team members so that they have a sense of confidence towards the idea that no one will be embarrassed, rejected or punished for speaking up or making mistakes (Edmondson, 1999). Through the alleviation of concerns about possible negative reactions to member's actions related to learning behaviour (e.g. taking risks, making mistakes), it has often been found that psychological safety facilitates learning behaviour in teams and thereby indirectly influences team performance (e.g. Edmondson, 1999; Kostopoulos & Bozionelos, 2011; Newman, Donohue & Eva, 2017). Within the field of innovation, there has been a growing amount of evidence that

psychological safety also leads to more creativity and innovation within organisations (Carmeli, Reiter-Palmon, Ziv, 2010; Kark & Carmeli, 2009). On the team level, psychological safety has been found to influence research and development outcomes (Gu et al., 2013; Post, 2012), knowledge creation (Choo, Linderman, & Schroeder, 2007), and creative team performance strongly (Kessel, Kratzer, and Schultz, 2012).

Besides having an effect on the aforementioned factors, psychological safety has also been shown to moderate the relationship between certain types of diversity and performance. For example, at the team level it was found that psychological safety moderates the relationship between expertise diversity and team performance (Martins, Schilpzand, Kirkman, Ivanaj, and Ivanaj, 2013; Reynolds & Lewis, 2018; Cho, 2022), so that the relationship between expertise diversity and team performance is more positive when psychological safety is high, and more negative when psychological safety is low. Similarly, it was found that psychological safety moderates the relationship between nationality diversity and team performance. The relationship between nationality diversity and team performance is more positive when psychological safety is high and more negative when psychological safety is low (Kirkman, Cordery, Mathieu, Kukenberger, 2013). The mechanism underlying the moderating effect of psychological safety on the relationship between cognitive diversity and team performance is that through the alleviation of negative concerns about group conflict related to different viewpoints, psychological safety contributes to an environment in which challenging perspectives and seeking elaboration of ideas is normalised. This in turn leads to a higher likelihood of integrating different perspectives into team solutions (Edmondson, 1999; Gibson & Gibbs, 2006). As this condition affects not only the outcome condition of team success but also other relationships, it is well suited for analysing it in relation to other conditions, which is the methodological approach taken by this study.

Teamwork

Teamwork is a multifaceted construct, revolving around the in-group communication, coordination of efforts, the balance of member contributions, the degree of mutual support, the individual team members' efforts, and group cohesion within a group that works together on a set of tasks (Hoegl & Gemuenden, 2001). For several decades the concept has been studied, and in relation to team performance, many theoretical frameworks have been developed (Salas, Stagl, Burke, & Goodwin, 2008; McEwan, Ruissen, Eys, Zumbo, Beauchamp, 2017). Teamwork has been shown to positively affect team performance, as was found by many scholars (Lohmann, Pratt, Benckendorff, Strickland, Reynolds & Whitelaw, 2019). Higher levels of teamwork are also found to influence the innovation output of teams. The underlying mechanism for this is that through team processes (e.g. better communication and coordination of tasks) groups become more effective and efficient in the overall process of reaching their goals (Hoegl & Gemuenden, 2001). To see whether this condition is also important in this research context, it is included in this study.

Team Interaction

Team interaction revolves around the interaction between contributing teams in crowdsourcing events. Several types of interaction are described as having a positive effect on team performance. One form is where the focus is on other ideas. For example, Schemmann, Chappin and Herrmann (2017) found that online ideators are likely to be more successful when paying more attention to other contributors' ideas. Likewise, Javadi, Khasraghi & Hirschheim (2021) found that contributors that often shared solutions with other teams or participated in discussions scored higher on team performance in the innovation competition. Chan, Li and Zhu (2015) found that contributors that commented intensively on other contributors' inputs within the Dell Ideastorm community, led to contribute more subsequent idea suggestions. Interestingly, Bullinger, Neyer, Rass & Moeslein (2010) found that both very high and very low cooperative behaviour leads to higher idea quality. Fuger, Schimpf, Füller & Hutter (2017) found that teams that have higher-quality ideas more often consist of a higher proportion of collaborators than teams with low-quality ideas. Another form concerns interaction where the focus lies on one's own idea. For example, Boss, Kleer, & Vossen (2019) found that contributors that gained constructive feedback led them to produce higher quality ideas within innovation contests. An often-given explanation for the importance of team interaction on team success is that teams that interact with other teams through collaboration, communication and feedback gain access to diverse information and resources that lead to a performance advantage (Benefield et al. 2016; Baker & Salas, 1992). To see whether this condition is also important in this research context, it is included in this study.

Client Interaction

Client interaction revolves around the interaction between the contributor and the knowledge-seeker within an innovation challenge. Client interaction was found to be important for team success in a limited amount of studies within the context of crowdsourcing for innovation (Langner & Seidel, 2015; Boons et al., 2015, Dahlander & Piezunka, 2014). A mechanism underlying this effect is that of a ''porous boundary'' between the client and the contributor, which can be seen as a type of interaction in which information flows naturally and organically. Such interactions provide motivation for the contributor, which can lead to sustained participation after the crowd contest. The latter is also influenced by the contributor receiving feedback from the client (Boons et al., 2015, Dahlander & Piezunka, 2014).

Accordingly, within firm-internal crowdsourcing, Zhu, Kock, Wentker, and Leker (2019) found that expert feedback was especially important for the quality of ideas created. Chan, Li and Zhu (2015) found that a high level of peer-to-firm interaction in an online crowdsourcing community led contributors to generate more ideas. The mechanisms are multifold; a high level of client interaction can lead to high-quality ideas through the mechanism of knowledge integration, but it can also sustain participation through the motivation of contributors. To see whether this condition is also important in this research context, it is included in this study.

3. Methodology

This section explains the research design, the sampling strategy, the data collection, the data analysis, and the operationalisation of all constructs.

3.1 Research design

This study applies a configurational approach by using Qualitative Comparative Analysis (QCA), a research approach and a data analysis technique (Schneider & Wagemann, 2010a; 2010b; Ragin, 1987; Berg-Schlosser, De Meur, Rihoux, Ragin, 2009). This approach was chosen as it very well fits the main objective of this study due to its focus on the identification of causal patterns within the data of the cases under examination and finding a meaningful explanation for those patterns (Ragin, 1987).

QCA makes it possible to express outcome conditions of interest in terms of causal relationships of necessary and sufficient conditions (Rihoux & Ragin, 2008). A necessary condition has to be present for the outcome to occur (in soccer terms this could be, for example, the football has to cross the line to score a point); a condition is sufficient if the outcome occurs as soon as the condition is present (e.g. kicking the ball with the foot on the goal), but other conditions can also be sufficient (e.g. heading the ball towards the goal) (Rihoux & Ragin, 2008). Key characteristics of QCA are equifinality, conjunctural causation and causal asymmetry (Berg-Schlosser et al., 2009; Ragin, 1987). Equifinality means that there can be multiple combinations of conditions, called pathways, to a particular outcome. Conjunctural causation means that certain combinations can be necessary for the outcome to occur so that separate conditions' effects would not result in the same outcome. Causal asymmetry means that causal relationships do not necessarily work equally strongly towards a present or an absent outcome condition (Berg-Schlosser et al., 2009; Ragin, 1987). QCA comes with its own distinct terminology that shall be used, meaning that conditions and the outcome are the names of the variables of interest instead of independent and dependent variables.

Although QCA contains elements from both an inductive and deductive approach (Thomann & Magetti, 2020), this study will predominantly have a deductive nature. Among the aims of QCA are testing existing hypotheses and theories and developing new arguments (Schneider & Wagemann, 2010b; Berg-Schlosser et al., 2008; Ragin & Rihoux, 2004). This study is deductive in the sense that it bases the investigated conditions on the literature in the field. However, QCA also contains an element of inductive nature, as inherent to the approach is the idea of iterative data collection and data analysis, or moving "between ideas and evidence" (Ragin, 2004, p.126). Preliminary empirical findings can inform both the case selection and the data analysis also called the "analytical moment" (Ragin, 2000; Schneider & Wagemann, 2010a). For example, concerning the latter, the outcome and conditions can be re-conceptualised and as well as membership requirements for both (Schneider & Wagemann, 2010a; Thomann & Magetti, 2020). Another inductive element of QCA is its exploratory feature. Due to its

configurational approach, it can shed light on relationships previously unstudied and therefore the results have the potential to act as a starting point for the theoretical development or reexamination of existing theories (Thomann & Magetti, 2020; Berg-Schlosser et al., 2009).

There are several other beneficial features of QCA. It is particularly useful to study medium-N-sized datasets through its strong case orientation (Greckhamer, Misangyi & Fiss, 2013; Herrmann & Cronqvist, 2009). Also, it is particularly useful for establishing propositions around complex interdependencies formulated in set-theoretical terms (Emmenegger, Kvist, Skaaning, 2013). This is the main advantage of regression methods, as these mostly focus on relationships in isolation (Schneider & Wagemann, 2010a). Due to these reasons, the configurational approach of QCA was chosen as the most fitting research design, promising a detailed insight into the causal relationships between outcome and the different variables.

3.2 Sampling strategy

As the research goal is to investigate what makes innovation teams achieve success within the context of crowdsourcing for innovation, the participating teams in such an innovation challenge are the unit of analysis of this study (Bryman, 2013). The research context within this study is called the “Circular Challenge”, an innovation challenge organised by the so-called Circular Economy hotspot BlueCity in Rotterdam, the Netherlands (https://www.circularchallenge.nl/). This six-week event aims for teams to create novel and viable business models for excess waste streams provided by participating knowledge-seeking organisations. During the process, they are coached by the organisers, their clients, and other teams. At the end of the six-week event, they present their idea to the clients, other teams, and friends and family during the final pitch night. A winner is elected by an expert jury and this team wins a temporary free working space at the location of BlueCity, including coaching on their idea.

Population

The total population of this study consists of 55 innovation teams, as the event has been organised multiple times per year since 2015 and per edition generally four to six teams participate. The teams mostly consist of people in the late phase of their studies or who are already working full-time jobs. The average age lies between 25-30 years old. Generally, the participants do not have any prior connection to one another; if they do, the organisation ensures they are placed in separate teams. The event organisers try to create multidisciplinary teams, as the organisers reported to aim for placing people with a background in business design with someone with a design background and someone with a more hands-on practical background. They also aim to create diverse teams regarding gender, nationality, and personality, although there were no clear criteria for this. All teams get assigned a client and a corresponding waste stream and these are unique for every team within their edition of the Circular Challenge. The type of clients and waste streams

vary, but in the past years many clients originated from the water management of waste processing industries with corresponding waste streams Some editions had overarching themes for the type of waste streams, while others had not. An overview of all editions and teams that participated in the circular challenge can be seen in table 2. The teams marked bold are the ones from the sample in this study.

Table 2. Overview of editions of the Circular Challenge, themes and participating teams

Edition	**Theme**	**Teams**
2022 December	Water	Edu Box; Emerge; Brijnwater; Future Abiding Technology; Bodemverbeteraar voor de landbouw
2022 July	Waste Processing	**Gypxel; P; Loopy Foam;** Wormpost
2022 April	Water	**Why Knot; Compeat;** Bubbles; **Bagger It Up;** Soak It Up; **Circularden**
2021 December	Water	Patchup; Waterloop; Connecting Nature; The Grassroute; **Hedgehok**
2021 July	Fibre	**Kei-Pot; Grassrope (Grasslagers); The Green Side;** Alclean; **Cirqua; Circulizer; Stain Alive**
2020 September	No theme	Fruiting Produce Life Extender; **Flip The City**
2020 July	High Rise	**Muuras;** The FuturePlastic; Veggie Water;
2018	Plastic	GymPlastic; Plastic it Playgrounds; Rebound; Port-tray
2018 May	No theme	**Total Loss, Total Gain (Cliq);** Blue Booster; Bluecarpet; Renoleum; Ongedeerd Verweerd; **Waterweg**
2017	No theme	Bakesight; Team Repurposing; Disposables; Team F.air; Blue Roof
2016 September	No theme	Flake It Till You Make It; Wandmore; **Pulp Fiction;** Team Eigen Haard
2015	No theme	Foodminds; Chaingers; Hapje Amsterdam; Eco Flow; Circular Square; Larven Lovers; Circular Flow; Planq

Sampling

The type of sampling used during this study was non-probability sampling, as not all population members had an equal chance to participate in this study (Bryman, 2013). More specifically, a combination of *purposive sampling, convenience sampling,* and *snowball sampling* was used.

Purposive sampling was used so that both winning and non-winning teams were sampled for this study, as investigating cases with both a present and an absent outcome is important for QCA (Rihoux & Ragin, 2008). Both winning teams and non-winning teams were approached by the researcher during the pitch nights of the Circular Challenge editions from December 2021, March 2022, and July 2022. This resulted in 11 teams that were contacted via the circular challenge, 8 of which wanted to participate.

Convenience sampling was used as it turned out that the accessibility to the teams was very low. The teams were reached through communication with the organisers of the Circular Challenge, and the combination of publicly available data about the teams and the participants.

Regarding the former, the organisers sent two e-mail requests to previous teams. One was embedded in the regular newsletter and the other was sent specifically to all participants that participated in the 2020 and 2021 editions. This resulted in 3 teams signing up to participate in the study.

Regarding the latter, publicly available data were combined to reach participants. Data from the channels of the Circular Challenge and BlueCity (e.g. websites, LinkedIn pages, Youtube channels) were combined to approach 10 participants, 6 of whom wanted to participate. Ultimately, this type of sampling resulted in 13 teams and 9 of which wanted to participate.

Snowball sampling resulted in two teams that were found by asking interviewees that already participated. These both wanted to participate. See table 3 for an overview of the participating teams ordered by type of sampling.

The researcher aimed at sampling ≥ 34 cases, to ensure a sufficient level of reliability in this study. Investigating six conditions in the analysis requires a minimum of 34 cases to be studied to satisfy the ideal requirement of a 1% reliability threshold (Axel, Cambré and Rihoux, 2013, pp.38-40). However, due to the lack of data access, it turned out to be impossible to approach this many teams in the first place. 26 teams were approached, and 19 teams participated, meaning that 73% of the approached teams were willing to participate in this study.

Table 3 Overview of sampled teams per sampling type

Sampling method	**Contacted (# teams)**	**Participation (# teams)**	**Willingness to participate (per cent)**
Purposive Sampling	11	8	
Convenience Sampling	13	9	
Snowball Sampling	2	2	
	26	19	73%

Sample

The sampling ultimately resulted in a sample that consisted of 19 out of 55 potential teams. According to Axel et al. (2013, pp.38-40), using five conditions in the analysis requires a minimum of 17 cases to be studied to satisfy a reliability of 90%, meaning there is a maximum 10% chance that results could be found with randomised datasets. With five out of the six initial conditions being used for the data analysis and 19 teams studied, this requirement was met. The sample consisted of 7 winners, five runners-up and seven teams without a rank. The teams stemmed from 7 different editions of the Circular Challenge. The editions differ slightly from one another concerning the format. In three editions there were less structured feedback

sessions in which teams and clients interact. However, personal communication with the organisers and some teams showed that it was still normal to have regular contact with the client, and interaction between teams was also possible as these often worked at the same physical location. Therefore, the small differences concerning the format are thought to have a minimal effect on the comparability of the sample. An overview of the teams, clients and waste streams can be seen in table 4.

Table 4. Overview of the sampled teams, their clients and waste streams

Team name	**Client**	**Waste stream**	**Edition**
Bagger It Up	Province of Gelderland	Dredged material	April 2022 (water edition)
Circularden	Water board Rijnland	Reclaimed wooden river revetment	April 2022 (water edition)
Circulizer	Water board Rijnland	Ammonia from wastewater	July 2021 (Fibre edition)
Cirqua	Water board Vallei & Veluwe, Pharmafilter	Hospital wastewater	July 2021 (Fiber edition)
Compeat	Water board Noorderzijlvest	Floating pennywort	April 2022 (water edition)
Flip the City	Province Zuid-Holland, Municipality of Rotterdam, Water board Rijnland	Duckweed	September 2020 (no theme)
Grassrope	Water board Rivierenland	Mowing waste	July 2021 (Fibre edition)
Gypxel	Reinis, Municipality of Nissewaard	Gypsum	July 2022 (Waste processing edition)
Hedgehok	Water board De Stichtse Rijnlanden	Digestate	December 2021 (water)
Keipot	Water board Rijn en IJssel	KEI sludge	July 2021 (Fibre)
Loopyfoam	Cyclus, Municipality of Krimpenerwaard	Used furniture foam	July 2022 (Waste processing)
Muras	Water board Schieland & Krimpenerwaard	Wastewater of high rise buildings	July 2020 (High rise)
Pulp Fiction	Waternet	Cellulose from sewage water	September 2016 (no theme)
Stain Alive	AquaMinerals	Iron-lime sludge	July 2021 (Fibre)
Termes	Avalex, HVC	Discarded chipboard	July 2022 (Waste processing)
The Green Side	Water board Hollandse Delta	PFAS contaminated dredge	July 2021 (Fibre)

Total Loss, Total Gain (Cliq)	Consortium of insurance companies	Total loss cars	May 2018 (no theme)
Waterweg	Water board Delfland	Dredged material	May 2018 (no theme)
WhyKnot	Water board Schieland & Krimpenerwaard	Japanese Knotweed	April 2022 (water)

Sampling bias

Generalisability, or external validity, is high when inferences about the sample under investigation can be extended to a much wider population (Bryman, 2012). Some requirements for this might not have been met due to several types of sampling bias.

Survivorship bias is the type of bias that occurs when members of the population are investigated that passed through some form of selection process (Bryman, 2012). In this study, it seems that successful teams were more likely to participate, as most teams sampled either were winners (7 teams) or were runner-up (5 teams). The underlying mechanism might have been voluntary response bias due to convenience sampling, meaning that members of a population can self-select (Bryman, 2012). Although efforts have been made to select teams on the basis of their level of success, as well as on their combination of conditions, this might have a negative effect on the external validity, as the sample might not be fully representative of the full population (Bryman, 2012).

Within QCA, *limited diversity* in the sample poses a risk for generalising findings. Limited diversity means the observed data is much less rich than the ‘’logical space’’. Concretely, this means that the amount of possible configurations of the conditions greatly exceeds the number of observed configurations of conditions, as presented within the sample’s cases (Rihoux & Ragin, 2008). This results in logical remainders, combinations of conditions that have not been presented in the sample by empirical data (Rihoux & Ragin, 2008). To clarify the external validity of the results, justifications about decisions made with such logical remainders, simplifying assumptions, as well as transparency regarding the implications of these decisions were included in this study (e.g Thomann & Maggetti, 2020; Schneider & Wagemann, 2010a). These can be found in the subchapter on data analysis.

3.3 Data Collection

The data collection strategy was chosen on the basis of the research aim and the methodological approach used. These both required gathering data on theorised conditions and the interrelationships between these conditions. Gathering data through a self-completion survey that was embedded in a structured interview proved very useful for this aim. Gathering data through a self-completion questionnaire was favourable, as the quantitative data obtained enabled dichotomisation of the conditions. This is an important criterium for using QCA, as problems regarding dichotomisation can lead to a loss of cluster information, potentially leading to contradictory observations (Herrmann & Cronqvist, 2009). The qualitative data obtained through the structured interviews was useful because it provided the researcher with in-depth knowledge about the cases (e.g. conditions, relationships) (Bryman, 2012). This is highly relevant for QCA, as it not only helps to deeply understand the case but because it serves several aspects of the data analysis process. It informs the (re)operationalisation of constructs, memberships scoring and the calibration decisions and the treatment of logical remainders and contradictory observations (Ragin, 1987; Rihoux & Ragin, 2008). Additionally collecting data via structured interviews also increases the measurement validity of the survey, as in-depth case knowledge helps the researcher to minimise measurement error (Schneider and Wagemann, 2012). Also, structured interviews were useful because they allowed for potentially higher response rates than other data collection methods (e.g. surveys). The data collection methods were therefore very fitting for the configurational approach taken and the research aim to study conditions and their interrelationships.

The development of the structured interviews and the self-completion surveys were done in parallel, following a largely similar process. Although the steps of the data collection and the data analysis have been carefully presented in figure 1, an explanation of the several steps will be given. The interview guide has been attached as a separate document, Appendix G.

First, all theoretical constructs were translated into operational definitions so that these could be measured (Forza, 2002).

Second, the survey and the interview guide were made. Regarding the survey, scales with statements for the conditions of psychological safety, teamwork, team interaction, and client interaction were developed. As some scales were adapted from other scholars, attention was paid to re-formulating the statements to fit this research context (Bryman, 2012). Also, reversely formulated questions were integrated, to counteract a potential acquiescence bias (Bryman, 2012). The software Qualtrics, was used for the administration of the surveys. To provide deep case insights that can be aggregated with other data, a structured interview was chosen over other types (Bryman, 2012). The interview guide was made taking into account providing a thorough introductory statement about the research, and information about informed consent (Bryman, 2012). Also, attention was paid to guiding the interviewees through the structure of the interview (Bryman, 2012).

Third, the interview and survey were pilot-tested to refine its administrative procedures and its formulation and structure (Bryman, 2012). This was done with a member of the population that was not part of this study, and with two peers.

Fourth, data was collected. Prior to data collection, the participants were informed of their rights with respect to their participation in the research and asked whether the interview sound could be recorded (Bryman, 2012). Informed consent forms were sent via email before the interviews took place and these were explained before the interviews started. The consent forms were read and returned either by signing the form itself or by replying with an email confirming one's consent. The standard Utrecht University consent form was used. The template consent form can be found in appendix A. Otter was chosen as sound recording and transcription software for the interview, and the interviews were held through a video call via Google Meet. The surveys were self-completed during the interviews, as these were quicker to administer (Bryman, 2012). The survey statements were presented by the researcher on-screen through screen sharing, the statements and the answers were read out loud, and the answers were filled out by the researcher. The interviewees could clearly see the answers filled out. Data collection lasted from the 12th of July until the 27th of September 2022. Although most interviews were conducted in English, one interview was conducted in Dutch. The existing interview guide was used and translations were done at the moment by the researcher. This Dutch interview was transcribed using Descript instead of Otter. It is not expected that this has had any influence on the validity of the findings, as the researcher is a native Dutch speaker.

Fifth, the data was prepared for analysis. The interview data was coded with Nvivo. The codes were based on the conditions in this study. The survey data was imported into SPSS, where the overarching variables and labels were created. A check was done for reverse-scored answers. Assessments of reliability and validity were done by tests for factor analysis and Cronbach's Alpha. These were interpreted with full caution, as these are measures normally only applicable on the conditions that the sample size produces significant results.Descriptive statistics were done to provide insights into the characteristics of the survey data and to inform the calibration of the memberships.

Sixth, the interview data and the survey data were merged. A data matrix was made (Excel) to systematically order the qualitative and quantitative data per case and per condition.

Seventh, the data was analysed. This process is explained in the following chapter.

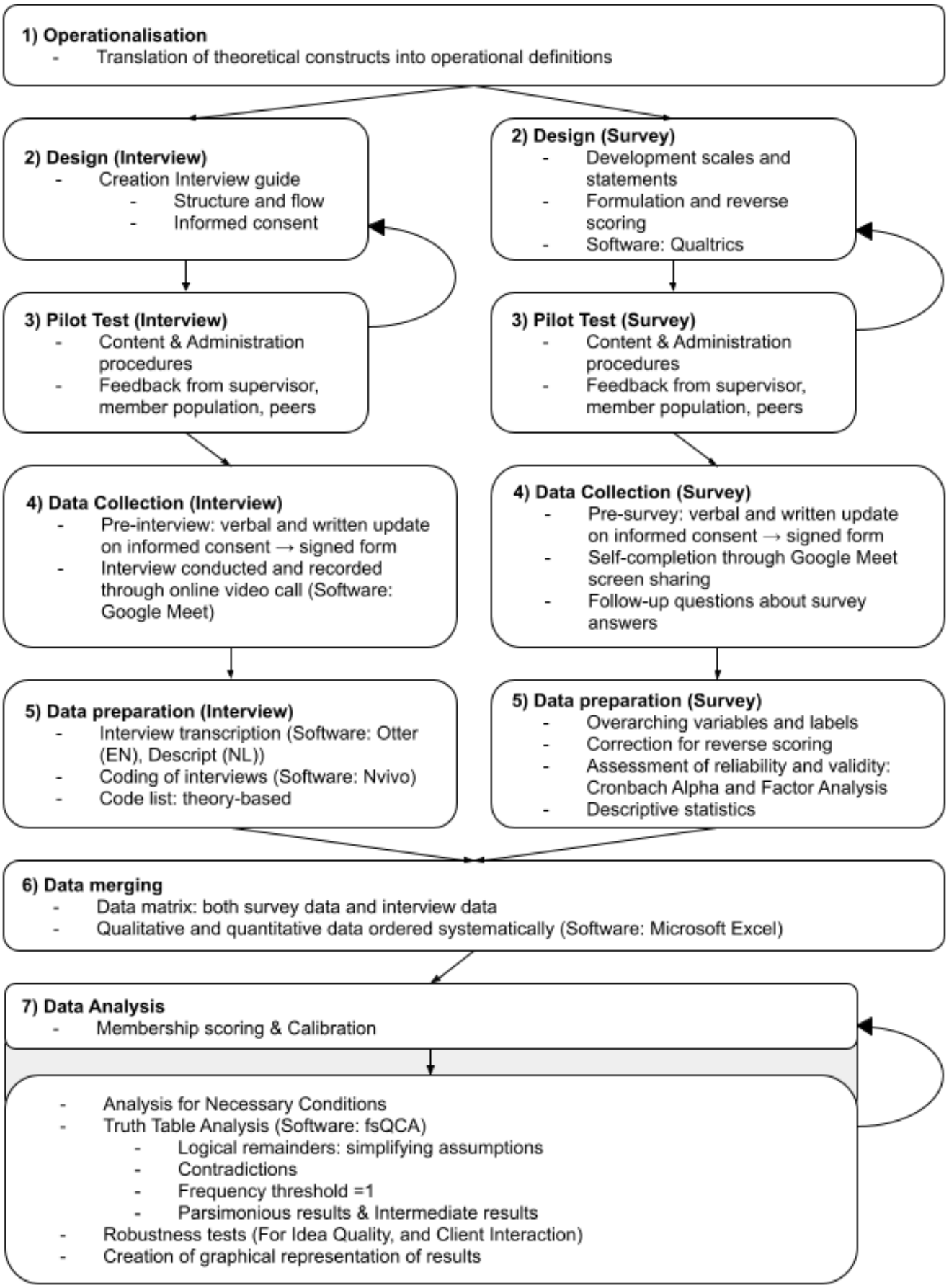

Figure 1. Data collection and data analysis process, adapted from Forza (2002).

3.4 Data analysis

The process of data analysis might be slightly unusual for the newcomer in QCA. As was mentioned in the section about research design, inherent to the methodological approach of QCA is the iterative process between findings and the data (Rihoux & Ragin, 2004). Preliminary empirical findings can inform steps in the further analyses of the data, like the reconceptualisation of the conditions, calibration thresholds and thereby membership scores for the cases (Schneider & Wagemann, 2010a; Thomann & Maggetti, 2020). In addition, this study adds to this uniqueness through the combination of quantitative and qualitative data. To guide the reader, the separate steps and the methodological decisions made have clearly been elaborated upon (Thomann & Magetti, 2020). The software program fsQCA 3.0 was used as well as the corresponding user manual explaining all the operations related to the program (Ragin, 2018). The iterative process consisted of the following steps:

First, membership scoring was done. The merged qualitative and quantitative data allowed for a thorough interpretation of the cases with respect to all the conditions, so that membership scores could be determined. A standard practice of csQCA is the setting of calibration thresholds, i.e. the determination of a value above and below which the cases have full or zero membership in the conditions (Rihoux & Ragin, 2008). Although these calibration thresholds ideally should be informed by theory, scholars have been confronted by a lack of prior knowledge available as was noted by Chappin, Cambré, Vermeulen, Lozano (2015). This means that the thresholds are set on empirical data within the study. This is acceptable, as long as these thresholds are clearly substantiated (Ragin 2008, Schneider & Wagemann, 2007). The section on the operationalisation of the construct provides the corresponding rationale for the calibration decisions. With all membership scores determined a data table was constructed in fsQCA. This table can be found in appendix B.

Second, an analysis of the necessary conditions was done. Data analysis should always start with such an analysis, as its output might result in the decision of eliminating conditions in the subsequent truth table analysis (Schneider and Wagemann, 2007). Conditions that exceed the value of 0,90 can be labelled a necessary condition (Ragin, 2006), meaning it has to be present for the outcome to occur, but the presence of a single presence condition does not guarantee the outcome (Rihoux & Ragin, 2008). The analysis for necessary conditions was done for every form of team success used in this study as well as idea quality and project continuation.

Third, a truth table analysis was performed. Both an analysis for parsimonious results and intermediate results was done, as these give insights into the relative importance of the conditions.
An analysis for parsimonious results aims to identify the simplest and most efficient explanation of the outcome, also known as the "minimal solution". The analysis for intermediate conditions aims at finding a more comprehensive set of solutions that can explain the outcome. The types of analyses differ with respect to the decisions surrounding logical remainders. For the former, the software program fsQCA

makes simplifying assumptions, whereas for the latter the program makes decisions on the basis of the easiest counterfactuals (Schneider & Wagemann, 2010a). In the first step, it turned out that logical remainders were present (17 out of 32 rows). These rows were deleted, the frequency threshold being one. To determine whether outcomes were sufficient, consistency thresholds of 0,80 were used, which is regarded as acceptable following prior research (Fiss, 2011; Chappin et al., 2015). Decisions around prime implicants were theory-based so that the presence of combinations of conditions was prioritised over the combinations including the absence of conditions. For the intermediate analysis, all conditions were expected to be present for the outcome to occur. The output of both analyses was combined in a representation devised by Fiss (2011) to serve the interpretation of the results. Contradictory observations were also dealt with. These occur when cases that are identical in terms of the scoring on their causal conditions differ in the outcome they present (Marx, Cambré, Rihoux, 2013). The configuration of conditions in such a situation then does not explain the variation in the outcome and therefore serves as a starting point for further model investigation (Ragin, 1987, pp. 113-118; Rihoux & De Meur, 2009, pp.48-50). Following this reasoning, whenever contradictions occurred, the data was revisited with respect to the membership scoring. These steps formed the main part of the iterative process of QCA. The results of this study were examined with several measures that are specific to QCA; tests for *consistency* and *coverage*. *Consistency* is "the degree to which instances of an outcome agree in displaying the causal condition" (Ragin, 2008: p.44). High consistency scores are better, as those indicate a large proportion of cases with a certain configuration of factors display the same outcome (Ragin, 2008). Consistency scores of 0,80 are regarded as acceptable following prior research (Fiss, 2011; Chappin, Cambré, Vermeulen, Lozano, 2015). Moreover, variables that exceed the value of 0,90 can be labelled a necessary condition (Ragin, 2006). A necessary condition is one that has to be present for the outcome to occur, but the presence of a single presence condition does not guarantee the outcome (Rihoux & Ragin, 2008). *Coverage* is "the way the respective terms of the minimal formulas cover the observed cases" (Rihoux & Ragin, 2009: p.64). Specifically, three measures of coverage shall be applied: raw coverage, unique coverage, and solution coverage. Raw coverage assesses to what degree a solution term covers the cases with a certain outcome; unique coverage assesses to what degree a solution term uniquely covers the cases with a certain outcome, meaning that no other terms cover those cases; solution coverage assesses to what degree the cases are covered by all terms (Ragin, 2006; 2008; Rihoux & Ragin, 2008). The solution consistency measures the degree to which membership in the solution is a subset of membership in the outcome (Rihoux and Ragin, 2008). In other words, it shows what fraction of the teams with this particular configuration of conditions show the same outcome. The solution coverage measures the proportion of membership in the outcome that is explained by the complete solution (Rihoux and Ragin, 2008). Robustness tests were done to check for the sensitivity of the results. These were done for the outcome of idea quality and for client interaction. These can be found in appendices D and E.

3.5 Operationalisation

This part revolves around the operationalisation of all conditions. First, the outcome condition is operationalised, and then the causal conditions are operationalised. This chapter also contains an overview of the conditions, measures, and calibration (see table 8).

Outcome condition

Team Success

The outcome condition in this study is Team Success. Two separate definitions were chosen for this study, as described in the theoretical framework: *idea quality* and *project continuation*.

Idea Quality

Idea quality is measured by the *position in the challenge*. To measure the position within the challenge, jury evaluations are used. Taking the scores or the ranking of ideas by the contributors is a standard practice within the research (Wang, 2021; Javadi Khasraghi & Hirschheim, 2021). Within this innovation contest, not all contributions are ranked in order, but an expert jury does determine a winner of the contest and a runner-up. Winners were considered to have a high position in the challenge and thus a high level of idea quality, whereas the other teams were considered to have a low position in the challenge and thus a low level of idea quality. Table 5 provides an overview of the calibration below.

Table 5. Calibration of idea quality

Position in the challenge		**Idea Quality**	**Team Success**
Position	**Membership**	**Membership**	**Membership**
Winner	1	1	1
Winner	1	1	1
Runner-up	0	0	0
Runner-up	0	0	0
Other	0	0	0
Other	0	0	0

Project Continuation

The second operationalisation of team success is *project continuation*. Project continuation is also often taken by scholars as a dependent variable for studying contributor success in innovation challenges (e.g. Pe-Than et al., 2018; Langner & Seidel, 2015; Seidel & Lagner, 2015; Shah, 2006; Javadi Khasraghi, Wang, Li, 2020). Project continuation was considered high when a team continues working on the project three months after finishing the challenge. Such a measure was also used by Nolte et al. (2020). Teams that continued in any form three months after the challenge were considered to have a high level of project continuation, whereas teams that did not were considered to have a low level of project continuation. See table 6 below for the calibration method.

Table 6. Calibration of project continuation

Project Continuation 3 months after the Circular Challenge	**Project Continuation**	**Team Success**
Yes	High	1
No	Low	0

Causal conditions

This subchapter starts with presenting the descriptive statistics of the survey conditions, and continues by giving an overview of the conditions, measures, and calibration. The data in the tables will be elaborated upon in the paragraphs to come.

Table 7. Descriptive statistics survey scales

Condition	**Mean**	**N**	**SD**	**Minimum**	**Maximum**	**Median**
Psychological Safety	5,96	19	0,71	4,00	6,80	6,00
Teamwork	5,32	19	0,83	3,60	7,00	5,60
Team Interaction	4,95	19	0,86	3,67	6,67	5,25
Client Interaction	5,59	19	0,75	4,13	7,00	5,50

Table 8. Overview of conditions, measures and calibration

Conditions	**Measure**	**Calibration**
Outcomes		
Idea Quality	Position in the competition	Present: winning Absent: runner-up or other
	Prototype quality *(Only applicable to robustness test)*	*Present: high fidelity prototype* *Absent: low-fidelity prototype*
Project Continuation	Long-term continuation	Present: any form of project continuation three months after the end of the Circular Challenge Absent: discontinuation before three months
Causal Conditions		
Gender diversity	Percentage of women versus men	Present: percentages women or men between 40-60% Absent: percentage women or men smaller than 40%, or larger than 60%
Cognitive diversity	Educational Diversity: Shannon-Wiener Index of Heterogeneity	Present: scores ≥ 1 Absent: scores <1
	Nationality Diversity: Shannon-Wiener Index of Heterogeneity	Present: scores ≥ 1 Absent: scores < 1
	job diversity: percentage of team members with ≥ 1 year work experience	Present: scores ≥ 50% Absent: scores < 50%
Psychological Safety	Survey score Interview data	Present: scores ≥ 5,00 Absent: scores < 5,00
Teamwork	Survey score Interview data	Present: scores ≥ 5,00 Absent: scores < 5,00 (two interventions)
Team Interaction	Survey score Interview data	Present: scores ≥ 5,00 Absent: scores < 5,00
Client Interaction	Survey score Interview data	Present: scores ≥ 5,30 Absent: scores < 5,30 (two interventions)

Cognitive Diversity

Cognitive diversity consists of the dimensions of *educational diversity*, *cultural diversity,* and *job diversity. Educational diversity* refers to the heterogeneity in terms of educational backgrounds in the group. To determine the educational background team members had, educational categories were used that were based on the broad fields of education, as described in the International Standard Classification of Education (ISCED), a reference classification framework for organising educational programmes and related qualifications that was developed by the UNESCO Institute for Statistics (2015). Eleven educational categories were distinguished, which are shown in table 9 below.

The assignment of team members' educational backgrounds to educational categories was done by matching the interview data with the corresponding criteria for the broad fields of education (UNESCO Institute for Statistics, 2015). More specifically, interviewees were asked to mention each of their team members' highest-finished educational programme or to describe their study background when they did not know the name of the study and on the basis of this they were categorised. Generally, this was straightforward, but in some cases, study programmes were dual interpretable. Sometimes it occurred that team members had done two study programmes as their highest finished education. As the knowledge diversity in the group was likely to be higher when including a person that had done e.g. two master programmes, this educational diversity was accounted for. Specifically, this meant that such members were assigned another educational category. Moreover, sometimes members were pursuing their bachelor's degrees at the time. In those cases, this programme was used for the assignment to an educational category, instead of using their finished high school education, as those backgrounds then functioned as their largest knowledge base (Bogers et al., 2018; Østergaard et al., 2011). Educational diversity was calculated with the Shannon-Wiener Index of Heterogeneity, which is a commonly used index originating from the literature on biological diversity, and ecological monitoring, and is now also widely used in other fields (Spellerberg & Fedor, 2003). The determination of a high level of educational diversity for the Shannon-Wiener index largely depends on the size of the population and the number of species. Larger populations with more species result in higher values for the index. For smaller populations, the value tends to be lower, due to the limited size and the limited amount of species that could theoretically occur. For example, the maximum value for a group of four entities with four species is 1,39. In this research, the values for educational diversity ranged from 0,56 to 1,39 (mean: 1,04). As theoretical data concerning fitting calibration levels were not available, a relative threshold was used, meaning that teams with an educational diversity lower than 1 were scored as having a low level of educational diversity.

Table 9. Educational categories according to the International Standard Classification of Education (ISCED)

Educational category
Generic programmes and qualifications
Education
Arts and Humanities
Social sciences, journalism and information
Business, administration and law
Natural sciences, mathematics, and statistics
Information and Communication
Engineering, manufacturing and construction
Agriculture, forestry, fisheries and veterinary
Health and welfare
Services

Nationality diversity refers to the heterogeneity in terms of nationality in the group. The majority of the teams had at least two different nationalities in the group and therefore the spoken language was generally English. The most frequent nationality was still Dutch, with 48 out of 80 participants coming from the Netherlands. There were 20 other nationalities among the teams, representing countries across all continents except for Oceania. The membership for nationality diversity was also measured with the Shannon-Wiener index of Heterogeneity. The values for this measure ranged from 0 for all-Dutch groups, to 1,39 (mean 0,76; median 0,56) for groups with only different nationalities. Just like for educational diversity, the value of 1 was set as a threshold, meaning that groups that scored higher than this value were regarded as having a high level of nationality diversity. In total, eight groups were scored as having a high membership and eleven groups were scored as having a low membership.

Job diversity refers to the diversity in the group concerning work experience. Some teams consisted of both students and professionals that already got a full-time jobs prior to the challenge, and others only consisted of students without prior work experience. Only full-time jobs prior to the challenge were considered. Internships were not regarded as having work experience, just like side jobs (e.g. working in a bar or supermarket). Some designers reported being self-deployed as freelance designers and these

were regarded as having work experience. Teams in which at least 50% of the team members had a full-time job before were regarded as having a high level of job diversity., This resulted in eight teams obtaining a high membership on job diversity. All dimensions had equal importance in the calibration of cognitive diversity. When at least two out of three conditions were high, cognitive diversity was also scored high. When there was only one condition that scored high, or when there were none that scored high, then cognitive diversity was scored low. See table 10 below.

Table 10. Calibration of cognitive diversity

Educational diversity		**Cultural diversity**		**job diversity**		**Cognitive diversity**	
High	1	High	1	High	1	High	1
High	1	High	1	Low	0	High	1
High	1	Low	0	High	1	High	1
High	1	Low	0	Low	0	Low	0
Low	0	High	1	High	1	High	1
Low	0	High	1	Low	0	Low	0
Low	0	Low	0	High	1	Low	0
Low	0	Low	0	Low	0	Low	0

Gender Diversity

Gender diversity revolves around the percentage of women and men in the group. All teams except for one had both male and female group members. Out of 80 participants in the sampled teams, 47 were female and 33 were male. On average 60% of the team members across the teams was female. Taking into account non-binary gender identifications did not prove necessary, as during the interviews this was not mentioned by either of the interviewees. The calibration was done per team size on basis of the percentage of women and men in the group. Teams that consisted of a number of women between 40-60% were regarded as gender diverse. Likewise, the same holds for the percentage of men.

Table 11. Calibration of gender diversity

Male [%]	Female [%]	Gender Diversity
40-60%	60-40%	1
>60%, or <40%	<40%, or >60%	0

Psychological Safety

To measure the condition of psychological safety, the scales of Edmondson (1999) and Xu & Yang (2010) for team psychological safety were combined and used. Some items underwent minor adaptations in the formulation to fit this research context, and statements that were non-relevant or double were removed before merging the scales. The previous actions resulted in an initial scale made out of 7 items. However, two statements were removed after performing reliability tests with the data using Cronbach's Alpha. This is a measure of the internal consistency of a survey scale, with a value of 0.5 or higher indicating acceptability in preliminary research and a value of 0.7 or higher indicating a threshold of good reliability (Forza, 2002; Nunnally & Bernstein, 1979). In general, a Cronbach's Alpha of 0.8 or higher is considered to indicate a high level of scale reliability (Nunnally & Bernstein, 1979). Removing two items from the initial scale changed Cronbach's Alpha value from 0,364 to 0,615, indicating the internal consistency of this scale is acceptable. As deleting a third item *(Q2.5: It was safe to take a risk on this team)* would not result in a much larger Cronbach's Alpha (0,678), and the topic of this statement revolved around risk-taking which is central to the idea of psychological safety according to Edmondson (1999), this item was left in the scale. A factor analysis and a scree plot were also done to assess the covariance across the items in the survey scale (Bryman, 2012). This plot confirmed the mentioned two items could best be removed. Although the mentioned tests were done, the results should be interpreted as an indication of internal consistency and should not be taken by heart, as the size of the sample is too low for these measures to be reliable. The items were measured using a 7-point Likert scale ranging from 1 Strongly Disagree to 7 Strongly Agree.

The calibration was done based on both survey and interview data. The scores ranged from a minimum of 4,00 to a maximum of 6,80, with a mean score on the scale of 5,96. The descriptive statistics for the survey conditions are also depicted in table 12. Although the results might imply that all teams had a high level of psychological safety (mean score 5,96 higher than the middle value of 4,00), this was not the case. The interview data revealed there was a sample bias, as there were only two instances of teams that reported low psychological safety when asked to elaborate on survey answers. On the basis of these qualitative insights, the breakpoint for this scale was set at a value of 5,00. This resulted in 17 out of 19 teams scoring high on psychological safety.

Table 12. Survey items for psychological safety

Condition	No.	Question	Reference
Psychological Safety	Q2.1	The members of my group had a hard time listening to an opposing point or perspective.	Xu & Yang (2010)
Psychological Safety	Q2.2	If you made a mistake on this team, it was often held against you	Edmondson (1999)

Psychological Safety	*Q2.3*	*Members of this team were able to bring up problems and tough issues* *(deleted for analysis)*	*Edmondson (1999)*
Psychological Safety	Q2.4	People on this team sometimes rejected others for being different	Edmondson (1999); Xu & Yang (2010)
Psychological Safety	Q2.5	It was safe to take a risk on this team	Edmondson (1999)
Psychological Safety	Q2.6	It was difficult to ask other members of this team for help	Edmondson (1999)
Psychological Safety	*Q2.7*	*No one on this team would deliberately have acted in a way that undermined my efforts* *(deleted for analysis)*	*Edmondson (1999)*

Teamwork

To measure the condition of teamwork, the scale from Lohmann, Pratt, Benckendorff, Stricktland, Reynolds and Whitelaw (2018) was slightly adapted and used. One item was deleted from the scale, as it was not relevant to the cases studied. Other items were reformulated to fit the research context

The reliability tests showed that the scale scored high Cronbach's Alpha values, with values of 0,967 (Lohmann et al., 2018) and a value of 0,840 measured on the basis of this study's data. As these values are greater than 0,80, this concerned a high internal consistency of the items and therefore also proved this scale reliable. The final scale used consists of 11 items, all of which are related to teamwork. The items were measured using a 7-point Likert scale ranging from 1 Strongly Disagree to 7 Strongly Agree. The scale scores ranged from a minimum of 3,60 to a maximum of 7,00, with the mean score for the scale being 5,32. The calibration was done on the basis of the survey data and the interview data in the same way as for the other conditions. The breakpoints were chosen to be 5,00, but also for these conditions there were teams with survey scores above the breakpoint, but for which their reports clearly showed the level of teamwork could be considered low. A good example of this is the team of D, with a survey score of 5,60. The interviewee reports that one of her team members decided to do things how they want, and they and one other team member could not handle that very well. They say it did not have a very positive effect on the collaboration and also they said the commitment was much lower in the end, as it turned out to be difficult to meet in person because not everyone lived in the city. In such two instances the scoring was based on the interview data.

Table 13. Survey items for teamwork

Condition	No.	Statement	Reference
Teamwork	Q3.1	Key decisions about our company were made by the entire team	Lohmann et al. (2018)
Teamwork	Q3.2	Most of the time, members of my team asked each other for feedback on their work	Lohmann et al. (2018)
Teamwork	Q3.3	Team members acknowledged the points of view of others	Lohmann et al. (2018)
Teamwork	Q3.4	The contributions of members of the team assisted our shared understanding of the project	Lohmann et al. (2018)
Teamwork	Q3.5	My team was dedicated to the task	Lohmann et al. (2018)
Teamwork	Q3.6	My team worked well together	Lohmann et al. (2018)
Teamwork	Q3.7	The unique skills and talents of each team member was fully valued and utilised	Lohmann et al. (2018)
Teamwork	Q3.8	The interaction within the team helped my team members to understand each others' point of view	Lohmann et al. (2018)
Teamwork	Q3.9	Working as a team allowed my team members to work smarter, not harder.	Lohmann et al. (2018)
Teamwork	Q3.10	My team members were able to learn new skills and knowledge from one another.	Lohmann et al. (2018)

Team Interaction

Team interaction consists of two subdimensions, namely *inbound interaction* and *outbound interaction*. A scale was developed that combines both subdimensions. *Inbound interaction* was measured by a developed scale that consisted of 6 items and *Outbound interaction* was also measured by a 6-item scale. The Cronbach's Alpha for the scale of team interaction is good with a value of 0,781, indicating a high reliability of the developed scale. As removing any items from this scale would not result in a larger Cronbach's Alpha, no items were removed. The items were measured using a 7-point Likert scale ranging from 1 Strongly Disagree to 7 Strongly Agree. This scale has a mean score of 4,95, and it ranged from a minimum of 3,67 to a maximum of 6,67. Based on interview data, the breakpoint has been set to a value of 5,00. The developed items can be seen in table 14 below.

Table 14. Survey items for team interaction

Condition	No.	Question	Reference
Inbound Interaction	Q5.1	Other teams did not pay attention to our ideas	Schemmann, Chappin and Herrmann (2017)
Inbound Interaction	Q5.2	Other teams used every possibility to provide my team with feedback	Zhu, Kock, Wentker & Leker (2019)

Inbound Interaction	Q5.3	Other teams discussed our ideas with us	Zhu, Kock, Wentker & Leker (2019)
Inbound Interaction	Q5.4	The feedback my team received from other teams was diverse in terms of themes	Zhu, Kock, Wentker & Leker (2019)
Inbound Interaction	Q5.5	The feedback my team received from other teams was constructive.	Zhu, Kock, Wentker & Leker (2019)
Inbound Interaction	Q5.6	The feedback my team received from other teams was useful.	None
Outbound Interaction	Q4.1	My team did not pay attention to the ideas of other teams	Schemann, Chappin and Herrmann (2017)
Outbound Interaction	Q4.2	My team used every possibility to provide other teams with feedback on their ideas	Schemann, Chappin and Herrmann (2017); Javadi Khasraghi & Hirschheim (2021); Chan, Li and Zhu (2015)
Outbound Interaction	Q4.3	My team discussed the ideas of other teams with them	Javadi Khasraghi & Hirschheim (2021); Bullinger, Neyer, Rass & Moeslein (2010); Fuger, Schimpf, Füller & Hutter (2017)
Outbound Interaction	Q4.4	My team provided constructive feedback to other teams.	Zhu, Kock, Wentker & Leker (2019)
Outbound Interaction	Q4.5	The feedback my team provided to other teams was diverse in terms of themes.	Zhu, Kock, Wentker & Leker (2019)
Outbound Interaction	Q4.6	My team provided useful feedback to other teams.	None

Client Interaction

Client interaction was measured by a new scale that was developed for this study. The scale consists of 8 items. The Cronbach's Alpha for the scale showed high internal consistency with a value of 0,789. As removing one item (*Q5.1: Our client did not pay attention to our idea)* from this scale would only result in a slightly larger Cronbach's Alpha (0,804), none of the items was removed. The items were measured using a 7-point Likert scale ranging from 1 Strongly Disagree to 7 Strongly Agree. The scale scores ranged from a minimum of 4,13 to a maximum of 7,00, with the mean score for the scale being 5,59. For the calibration, the survey data and the interview data were used. The breakpoint was set on 5,30. This was set higher than the other conditions, based on the following rationale. The data on client interaction was often difficult to dichotomise, as the survey scores and the interview data combined showed that the survey scores were not very consistent with the interviewees' answers. More specifically, some interviewees reported information that was considered to show a low level of client interaction, while their survey score was above the breakpoint of 5,30. A good example is the following team, which has a survey score of 5,63, seemingly indicating a high level of client interaction, while the opposite was reported by the interviewee. It was reported that their client was very busy and responses to the teams' questions took a long time. It was also said the clients' feedback was very technical by nature and not diverse. Moreover, the client did

not provide all the relevant information about the problem context at the start of the project. To deal with this dichotomisation issue, on the one hand, a higher calibration threshold was set so that the cases in the middle could be included. This regarded the 5,30 threshold, which could be interpreted as a more ''strict'' interpretation of the quality required for client interaction to score high. For cases that deviated even more strongly - the survey data completely contradicted their interview data - it was chosen to follow the qualitative insights instead of taking the survey score. Three cases were treated this way. Robustness tests were done to check for the sensitivity of the results when a more lenient threshold was set (4,70). As the findings showed largely similar patterns, these were considered robust. Appendix E provides the findings of these tests.

Table 15. Survey items for client interaction

Condition	**No.**	**Question**	**Reference**
Client Interaction	Q6.1	Our client did not pay attention to our idea.	Schemmann, Chappin and Herrmann (2017)
Client Interaction	Q6.2	Our client used every possibility to provide my team with feedback on our idea.	Zhu, Kock, Wentker & Leker (2019)
Client Interaction	Q6.3	Our client discussed our ideas with us.	Zhu, Kock, Wentker & Leker (2019)
Client Interaction	Q6.4	The feedback our client provided on our idea was diverse in terms of themes	Zhu, Kock, Wentker & Leker (2019)
Client Interaction	Q6.5	Our client provided my team with constructive feedback.	Zhu, Kock, Wentker & Leker (2019)
Client Interaction	Q6.6	Our client provided my team with suggestions and recommendations on how to continue.	Zhu (2018)
Client Interaction	Q6.7	Our client did not provide my team with enough information about the problem context.	Brunswicker, Bilgram and Fueller (2017)
Client Interaction	Q6.8	The feedback my team received from the client was useful.	None

4. Results

This section presents the results of this study. The section is divided into two main parts, each presenting the results for a different operationalisation of the outcome condition: *team success.* The first part revolves around *idea quality* as a measure of team success, whereas the second part revolves around *project continuation* as a measure of team success. For both outcome conditions, the results from the analyses for necessary conditions and for the truth table analyses are presented and explained.

4.1 Idea Quality

This part revolves around identifying the configurations of conditions that lead teams to achieve a high level of idea quality in their innovation challenge. To recap, this regards teams that won the edition of the circular challenge they participated in.

Analysis for necessary conditions

The analysis for necessary conditions shows that psychological safety with a consistency value of 1 can be considered a necessary condition, as it supersedes the threshold of 0,90 (Ragin, 2006). In practical terms, all teams that exhibit team success also score high on psychological safety. This means that psychological safety cannot explain the differences across the outcomes of the cases (Ragin, 2006), and, therefore, this condition will not be used in the truth table analysis for this outcome condition. Table 16 shows the consistency and coverage values for the conditions analysed. Whenever a tilde, ~, is added to a condition, it means the *absence* of that condition.

Table 16. Analysis for necessary conditions for idea quality

	Outcome variable: Idea Quality (Winning a contest)		
Condition	**Abbreviation**	**Consistency**	**Coverage**
Psychological Safety	PS	1.00	0.41
	~PS	0.00	0.00
Teamwork	TW	0.57	0.36
	~TW	0.43	0.38
Client Interaction	CI	0.71	0.50
	~CI	0.29	0.22
Gender Diversity	GD	0.43	0.33
	~GD	0.57	0.40

Cognitive Diversity	CD	0.57	0.36
	~CD	0.43	0.38
Team Interaction	TI	0.71	0.45
	~TI	0.29	0.25

Truth table analysis

The next step in the analysis is the core of QCA: revealing the configuration of causal conditions that leads to a high level of team success. Table 17 shows an overview of these different solutions. The table uses the notation devised by Ragin & Fiss (2008) to present the solutions of the truth table analysis. This representation makes a distinction between ‘’core conditions’’ and ‘’peripheral conditions’’. Core conditions are presented by large symbols (● for presence; ⊗ for absence) and come out of the analysis for parsimonious results, meaning these relations are strongest. Peripheral conditions are presented by smaller symbols and come out of the analysis for intermediate results (● for presence; ⊗ for absence), meaning these relationships are less strong, although these are still relevant. Whenever cells for conditions are empty, these can be interpreted as ‘’don't care’’ conditions, meaning these conditions can be both present or absent for the outcome to occur (Ragin & Fiss, 2008).

The truth table analysis revealed three configurations that led teams to achieve high levels of idea quality. The set of solutions is very consistent across all cases, as the solution consistency of ‘1’ is above the threshold of 0,80 (Fiss, 2011). The coverage of the configurations is 0.71. For the separate configurations, the raw coverage varies between 0.14 and 0.29. More important is the unique coverage equals the raw coverage for all solution terms, indicating that the three configurations are the only possible solutions that can explain the corresponding teams (Rihoux & Ragin, 2008).

The latter shows that any duality or ambiguity is eliminated. The results are explained through interview data. The interviews generally provided deep qualitative insights about the teams, the conditions, and the wider context. Quotes of interviewees are provided whenever suitable.

Table 17. Results of the analysis for idea quality

Configurations for Idea Quality			
Configuration	**1) Collaborative internal knowledge integration**	**2) Collaborative expert knowledge integration**	**3) Non-collaborative External knowledge integration**
Conditions			
Psychological Safety	-	-	-
Teamwork	●	●	⊗
Cognitive Diversity	●	⊗	●
Gender Diversity	●	⊗	⊗
Team Interaction			●
Client Interaction		●	●
Raw Coverage	0.29	0.29	0.14
Unique Coverage	0.29	0.29	0.14
Consistency	1	1	1
Solution Coverage	0.71		
Solution Consistency	1		

Configuration 1: Collaborative internal knowledge integration

The first solution is labelled **‘’collaborative knowledge integration’’**, as the high level of teamwork combined with the high level of cognitive diversity played a very important role for teams within this configuration.

The solution shows that teams that scored high on teamwork, high on cognitive diversity and high on gender diversity may result in a high level of idea quality. The mentioned solution is independent of the other conditions that are shown in the table, team interaction and client interaction, meaning that regardless of these conditions the solution is sufficient for the outcome to occur.

Team P is an excellent example of the teams that fall into this solution. The team is extremely diverse, both in terms of cognitive diversity and gender diversity. Indeed, it scores very high on cognitive diversity, concretely meaning that the team exhibited a broad variety of nationalities, educational backgrounds and work experience. This is also shown by the following quote when asking about whether or not their group was diverse:

Interviewee Team P

Yes, definitely . . . I was really surprised by how well we all complemented each other

The team also scored high on teamwork and this is explained well in detail by the interviewee. Amongst the points said are that individual members were very committed to the tasks, each coming fully prepared to the meetings, already having done much research within their specific field of expertise. This is shown well by the following quote:

Interviewee Team P

Yeah, uhm, we were very overwhelmed with to-do's, and the team did very, very well in terms of swarming whatever the workload was.. where it's like, I can't do this. What can you do? Okay, cool. I'll do it. You know, I'll pick that up. I'll run with it . . . And so that happened a lot where we would sort of loosely have an idea of what we wanted to do and we would set up a check-in with the group and the amount of deep research that had already been completed by each person sort of based on what their speciality was, they would have an overwhelming amount of very useful information during the check-in moments . . . Everyone would always come to the table very prepared and going beyond the call of duty . . . I think this created a really strong and positive team dynamic.

For the other team, the different cultural backgrounds were said to bring up several issues, relating to communication and the variety of perspectives and problems brought up. Concerning the communication issues, the interviewee reports that more elaborate communication helped to resolve these issues, as the quote below shows:

Interviewee Team S

Yeah, so I think since we're all from different backgrounds, sometimes we didn't really speak exactly the same language. And on top of that also language barriers between people who speak English from different language backgrounds sometimes made communication a bit difficult, I think. I remember a couple of times a chemical issue was raised, for example, and the rest of the team didn't really see the problem, and then you're just trying to explain to each other what you mean by using different words. So, that's kind of the hard part about listening to an opposing point of perspective, is that sometimes we lacked kind of the terminology in common to tackle it together. But it all worked out after more elaborate communication.

Concerning the variety of perspectives and problems that were brought up, the high dedication of the team members helped them to solve all the problems resulting from the new perspectives. This is shown well by the following quote:

Interviewee Team S

> I think what happened was that since you have so little time and in the Circular Challenge you discover a lot. And you all have different backgrounds, you just come up with more and more questions about your own projects, and you're raising like different types of problems that could happen and then you're all working together to kind of solve all these problems. And since we all have diverse backgrounds and work together, I don't think that necessarily made us pick the most important things. We didn't really prioritise that. So, we just made sure we tried to research everything. So we worked very hard.

As the configuration also shows, both team interaction and client interaction score ‘’don’t care’’, meaning that the level of these conditions was not found to be of importance for the teams that fall within this solution.

Configuration 2: Collaborative expert knowledge integration

The second solution is labelled ‘’**collaborative expert knowledge integration**’’, as the high level of client interaction in combination with a high level of teamwork played a very important role for teams with this configuration.

This configuration showed that teams that scored low on cognitive diversity and low on gender diversity, but high on teamwork and client interaction might achieve high idea quality. The results are independent of the level of team interaction.

In contrast to the previous solution, the teams that fall within this solution show a lower level of both gender diversity and cognitive diversity. Although the teams scored high on the dimension of educational diversity, they scored low on the other two dimensions: nationality diversity and job diversity. This fact was acknowledged and reflected upon by one of the teams’ interviewees:

> Interviewee Team F
>
> So, I think, we had different backgrounds. I think [team member 1] came from it with a design background, [team member 2] was a bit more technical and then [team member 3] and I had a very similar background. So, like, somewhere in between economics and politics.. But, in regards to other types of diversity, all four of us were Dutch, and we had a similar, like, there weren't big discrepancies in the way we grew up, I think, and we all had a student life. In personality sense, of course, there was diversity, but in, like, cultural sense there wasn't much diversity.

The interview data revealed that the level of client interaction was very high among these teams. Both interviewees reported that not only did their client pay much attention to their idea, but they also provided very relevant feedback. When asked about the feedback the client gave them, the interviewees reported that their client provided them with diverse perspectives which were also very useful for the development of their idea. The following quotes illustrate this well.

Interviewee Team F

. . . they approach it from different perspectives, and even though those perspectives were quite annoying at some point, because it was quite useful, they were different perspectives

Interviewee Team K

We had a very good penultimate presentation, very good feedback, and at that point, we felt like we could win if we were adding basically the questions that we couldn't answer during that round.

The teams in this solution also scored high on teamwork. When asked about it, a well-functioning task division and the high dedication of the group members helped the teams to collect and process much relevant information. The following quote from team K illustrates this well:

Interviewee Team K

I think we had a clear vision on what we had to do and we were motivated to go there and I think we were collecting a lot of information and then sort of getting those clear. So, I think just looking back I felt like we were working well together.

As the configuration also shows, team interaction scores ‘’don’t care’’, meaning that the level of this condition was not found to be of importance.

Configuration 3: Non-collaborative full knowledge integration

The third solution is labelled ‘’**non-collaborative external knowledge integration**’’, as the low level of teamwork in combination with a high level of team interaction and a high level of client interaction played a very important role for teams with this configuration.

This solution consists of teams that score low on teamwork, high on cognitive diversity, low on gender diversity, and high on both team interaction and client interaction. Also, this solutions’ configuration of conditions is sufficient for the outcome to occur.

There is one team that fits this solution and that is team B. This team shows a low level of teamwork and that is, because, according to the interviewee, two team members showed low commitment by doing little work, and by often being absent. One of these two was relatively more involved in the project than the other and did do some work, but did not stand out. Interestingly, the interviewee does elaborate on the great collaboration with the remaining teammate. From the interview data, it became clear that most of the work was accomplished through the fruitful collaboration between these two. This is well illustrated by the following quote:

Interviewee Team B

…but on the other hand, I think, especially [my teammate] and I were, like, super focused and really wanted to win and deliver a great product.

The team scored high on cognitive diversity, but the interviewee reported that it was not very clear how these fields were interconnected, as is seen by the following quote:

Interviewee Team B

Um, yeah, I think it was very different, but it didn't really, like, connect all the time.

This group shows both high scores on team interaction and client interaction. With regards to team interaction, the interviewee reports having had regular conversations with other participants during lab time, mostly about their ideas. This is illustrated well by the following quote:

Interviewee Team B

Yeah. So, in the beginning, there was not much attention but since we came up with interesting material samples, suddenly there was more attention on our team. I think there were, like, five or six people who really came to us and gave a lot of feedback, but the rest of the teams didn't really come up to us.

Also, the interviewee reports having a client that was very engaged with their project and was providing a lot of information. This is illustrated by the following quote:

Interviewee Team B

So, any moment of the day we could come to her office and talk to her or call her and then she would really open up to us, like, I want to help you in any way possible . . .
the only thing was that they were a little messy in what they already knew . . . so, we got a lot of information, but we had to figure out a lot ourselves as well.

Robustness Test

A robustness test was done to check for the sensitivity of the results. The robustness test was done by performing the analyses with an alternative outcome condition for idea quality which included the quality of the prototype as a measure. As this analysis found configurations that were highly similar compared to the original analysis, and thereby hold water with respect to necessity and sufficiency, the findings can be considered robust (Schneider & Wagemann, 2012). An overview of the results of the robustness analysis with a more detailed description can be found in appendix E.

4.2 Project Continuation

This part revolves around identifying the configurations of conditions that lead teams to achieve a high level of project continuation. To recap, this regards teams that continued their project three months after the end of the challenge.

Analysis of necessary conditions

Also for project continuation, the analysis of necessary conditions shows that having high psychological safety can be considered a necessary condition. With a consistency score of '1', this is well above the threshold of 0.90 (Ragin, 2006). This means that this condition is also omitted in the consecutive truth table analysis for project continuation.

Table 18. Analysis of necessary conditions for project continuation

	Outcome variable: Project Continuation (Continuation after three months)		
Condition	**Abbreviation**	**Consistency**	**Coverage**
Psychological Safety	PS	1.00	0.41
	~PS	0.00	0.00
Teamwork	TW	0.71	0.45
	~TW	0.29	0.25
Client Interaction	CI	0.71	0.50
	~CI	0.29	0.22
Gender Diversity	GD	0.14	0.11
	~GD	0.86	0.60
Cognitive Diversity	CD	0.57	0.36
	~CD	0.43	0.38
Team Interaction	TI	0.71	0.45
	~TI	0.29	0.25

Truth table analysis

Table 19 shows the results for project continuation. The truth table analysis revealed three solutions that led teams to achieve high levels of project continuation. The set of solutions is very consistent across all cases, as the solution consistency of '1' is well above the threshold of 0,80 (Fiss, 2011). The solution is 0.86. For the subset solutions, the raw coverage varies between 0.14 and 0.57. The unique coverage varied between 0.14 and 0.29. The raw coverage and unique coverage were identical for solution 3, indicating that the teams are uniquely explained by the corresponding solutions.

The raw coverage and unique coverage differ from one another for configurations 1 and 2, with 0.57 and 0.29 for the former and 0.14 and 0.43 for the latter. This means that there are some teams that can be explained either by the first or the second configuration. Two teams fell into this category. As the interview data provided rich qualitative insights into the cases, these were used for clarification. Quotes of interviewees are provided whenever suitable.

Table 19. Results of the analysis for project continuation

Configurations for Project Continuation			
Configuration:	**1) Collaborative internal knowledge integration**	**2) Collaborative expert knowledge integration**	**3) External knowledge integration**
Conditions			
Psychological Safety	-	-	-
Teamwork	●	•	
Cognitive Diversity	●		•
Gender Diversity	●	⊗	⊗
Team Interaction	⊗		•
Client Interaction		●	●
Raw Coverage	0.14	0.57	0.43
Unique Coverage	0.14	0.29	0.14
Consistency	1	1	1
Solution Coverage		0.86	
Solution Consistency		1	

Configuration 1: Collaborative internal knowledge integration

The first solution is labelled ‘’**collaborative internal knowledge integration**’’, which equals the name of the first solution of the analysis revolving around idea quality. This was done as also in this analysis the combination of teamwork and cognitive diversity proved to be very important for the outcome to occur.
This configuration shows that teams that score low on team interaction, but high on teamwork, high on cognitive diversity and high on gender diversity might lead to project continuation, irrespective of the level of client interaction.
There is one team in this configuration and that is team S. This team was diverse in terms of gender diversity (2 women and 2 men), and in terms of cognitive diversity, with multiple team members from different countries and also having complementary educational backgrounds. The team reports how well they were collaborating, even considering the unforeseen circumstances of the COVID-pandemic they had to deal with. This becomes clear from the following quote:

> Interviewee Team S
> *Yeah, so I remember we had a conversation about this, how happy we were about collaborating together, and also with the team, especially because back then COVID happened. It was really the first time for a lot of us to work together online. Yeah, it was really the first time doing this. So we were all kind of surprised as well, how easy that went. We thought you really would need to be together to work together well, and we all were very happy that the online situation didn't change anything about how we worked together, and that it just worked out really well. Yeah, I remember, everybody was very, very happy about the interactions and about how it went.*

The level of team interaction for this team was low. The team did have some contact with other teams, but they kept it to a minimum as they were very competitive. The following quote illustrates this well.

> Interviewee Team S
> *Yeah. So on a basic level, we were really looking at what other people were doing and mostly in terms of how prepared they were and what kind of solutions they were coming up with, but mostly from a competitive attitude. So, we were just trying to win. So, in that sense, we kept track of the competition.*

As the configuration also shows, client interaction scores ‘’don’t care’’, meaning that the level of this condition was not found to be of importance for this team.

Configuration 2: Collaborative expert knowledge integration

The second solution is labelled ‘’**collaborative expert knowledge integration**’’, which equals the name of the second solution of the analysis revolving around idea quality. This was done as also here the high level of client interaction in combination with a high level of teamwork played a very important role for teams with this configuration.

This configuration showed that teams that scored low on gender diversity, but high on teamwork and high on client interaction might lead to project continuation. The results are independent of the level of team interaction and cognitive diversity. Team O is a good example of this solution. The team scored low on gender diversity, with a percentage of 75% women. When asked to elaborate on the interaction with their client, it is mentioned that they were very engaged with the project during the Circular Challenge, providing the required information and other support whenever needed. This is clearly shown in the following quote.

> Interviewee Team O
> *Yeah, because they were very, very attentive to the ideas that we came up with. There was complete participation from their team. They were very organised in terms of all the questions we asked them. They tried to be on time, replied with responsibility and on exact topics we were keen on. I think they were very honest. If they didn't know anything, they would clearly tell us, but at the same time help us understand how we could get that information which was very positive for me. And yeah, so that's what I felt. I think that's why I felt that they were very attentive.*

Besides the engagement during the Circular Challenge, the high level of client interaction also proved very important after the event. Multiple teams reported that their client was interested in their idea after the event, through the stimulation of planning follow-up meetings to discuss the ideas or to plan further steps ahead. The following quotes illustrate this well:

> Interviewee Team O
> *We are thankful for the municipality because they are now arranging a meeting with us where we can present the product to them.*

> Interviewee Team F
> *. . . and then we called together and we said, okay, let's continue, we want to do a feasibility study. That seemed like the most logical step, and our client had said that they would like to continue with the idea.*

As the configuration also shows, cognitive diversity and team interaction score ‘’don’t care’’, meaning that the level of these conditions was not found to be of importance for these teams.

Configuration 3: Full knowledge integration

The third solution is labelled ‘’**external knowledge integration**’’. This name is partially similar to the name of the third configuration in the analysis revolving around idea quality, the only difference being that the former concerned a low level of teamwork, while the level of teamwork is not important in this configuration. This configuration showed that teams that scored low on gender diversity, but high on cognitive diversity, high on team interaction and high on client interaction might lead to project continuation. The results are independent of the level of teamwork. Team N is a good example of this solution. This team reported having close contact with other teams during the event which proved very valuable. They would have discussions with other teams during structured feedback sessions and during lab time, often providing new insights about the ideas. This becomes clear from the next quote.

> Interviewee Team N
> *Yeah, so, the best example was maybe [team Z], because they were in the lab so often. So we would come by and then we were like, Oh yeah, did you think of this and this and this already? Yeah, we did, but we don't know if it's going to be possible.. And then, yeah, especially [my teammate], because he had such a background in the wood recycling industry, he was, like, oh yeah, actually, if you're gonna make RVS plates with it, maybe you can sell it for this and this and this, because he knew things from that industry that they didn't know from at [team Z] themselves . . . and it was also very, very often the other way around, that we gave a presentation and people were really saying, like, oh yeah, so you're just going to separate? Where's your product? Is there actually a market for it? So, in that way you will just keep in contact.*

The team also reports having a good relationship with the client, whereby the client would support them in learning about the waste stream at hand and the most important solution requirements.

> Interviewee Team N
> *Yeah. So in the beginning, we had a lot of loose ideas about which they were very... I would say, critical. So, we always had a very big discussion about what would be actually a nice thing to do, what will be innovative, but what will also be, you know, there were a lot of parameters that were important for them and we learned about those on the way, so it was very helpful to have these discussions and get their insights and, yeah, I mean, they had a stake in that as well. So, it was very, very good to have those discussions.*

The teams in this solution scored high on cognitive diversity, which was also the case for this team. This team actually reported that they were very diverse in terms of educational backgrounds and interests, although the interviewee had not expected that at the start of the challenge. The diverse insights show that the team members very well complemented one another, as is shown by the following quote when asked about the diversity in their group:

Interviewee Team N

So, at first I actually thought no, because we had this weird situation where we just met everyone, and we realised we all came from [the same] University... thinking like, Oh, that's so typical, like, that everyone here is from the same university and is trying to do something with the circular economy, but if you think about it in hindsight, now that we've evolved a bit further, I think that actually, our backgrounds are really, really quite diverse, despite the fact that it's all very, very much in the same region.. because the business development side for example, some things that I have really no experience with... and then the technological side, which I'm quite experienced with and [my team member A] as well is something that, for example, [my other two team members B and C] really don't have, and [my other team member D] really, at least that's the way I feel about it, really has a very out of the box way of thinking, because he did an interdisciplinary study. So, in that way, yeah, there was actually diversity, at least in interests, but also, I think, therefore, educational background.

As the configuration also shows, teamwork scores ‘’don’t care’’, meaning that the level of this condition was not found to be of importance for these teams.

5. Discussion

The aim of this study was to shed light on the interplay of team-related and interaction-related conditions that lead innovation teams to be successful within the context of offline team-based crowdsourcing for innovation.

Necessary conditions

The tests for necessary conditions revealed that psychological safety is a necessary condition for both variations of team success: idea quality and project continuation. Psychological safety being a necessary condition can be explained due to the sample of the study. Due to a high occurrence of psychological safety among the teams (17 out of 19), this almost automatically turned into a necessary condition. This can be explained due to the following. The people that participate in the Circular Challenge are mostly young professionals and (master)students, seeking to broaden their horizons through learning about circular entrepreneurship and building a network. This also shows clearly in the communication on the event's channels. Moreover, participants are selected through an interview on the basis of their motivation and skills, as personal communication with the organisers revealed. These observations suggest that motivation and openness are not only common among participants but rather are requirements for participation in this challenge. This is in line with research stating that the motivations of contributors that participate in crowdsourcing for innovation events stem from the aim of learning and gaining new expertise (Seidel & Langner, 2015) or from the joy of participation in the ideation process (Aitamurto, Landemore & Saldivar Galli, 2017), and contact with peers (Bretschneider, Leimeister & Mathiasen, 2015; Leimeister Huber, Bretschneider & Krcmar, 2009). Despite the sample bias, the data shows that psychological safety is of high importance. Multiple interviewees, both from successful and unsuccessful teams mentioned that it was very important to allow each other to give out-of-the-box ideas, without rejecting those or judging one's capabilities, as these ideas often lead to better ideas. These observations are perfectly in line with scientific research showing that psychological safety is positively related to team performance (e.g. Edmundson, 1999; Kostopoulos & Bozionelos, 2011; Newman, Donohue & Eva, 2017), and, moreover, it also strengthens the growing scientific evidence that psychological safety is in fact very important for knowledge development (Xu & Yang, 2010; Mu & Gnyawali, 2003; Choo, Linderman, & Schroeder, 2007) and innovation processes (Edmondson & Mogelof, 2006), as it ensures a climate in which it is normalised to make mistakes and learn, which are an inherent part of the innovation process (Edmondson & Mogelof, 2006).

Configurations

The results revealed that there are three types of pathways that lead teams to achieve team success, one of which is internal and the other two being external. In the internal pathway, the team-related conditions (teamwork, cognitive diversity, gender diversity) were found to be important, whereas the external pathways showed a high prominence for the interaction-related conditions (team interaction and client interaction).

For a clear overview of these pathways and their labels, please see table 20 below. Besides the labels, the specific solutions are also given codes to distinguish them. IQ stand for Idea Quality and PC stands for Project Continuation. The number indicates the number of the solution as referred to in the results section. The pathways that could lead teams to achieve a high level of team success are:

1. *Collaborative internal knowledge integration*: teams that exhibited a high level of cognitive diversity and a high level of gender diversity, in combination with a high level of teamwork. The solution for project continuation also includes a low level of team interaction.
2. *Collaborative expert knowledge integration:* teams that exhibited a low level of gender diversity, in combination with a high level of teamwork and a high level of client interaction. The solution for idea quality also includes a low level of cognitive diversity.
3. *(Non-collaborative) external knowledge integration:* teams that exhibited a high level of cognitive diversity, a low level of gender diversity, and a high level of team interaction and client interaction. The solution for idea quality also includes a low level of teamwork.

Table 20. Configurations for high levels of team success.

Outcome Condition	**Idea Quality**	**Project Continuation**	**Idea Quality**	**Project Continuation**	**Idea Quality**	**Project Continuation**
Solutions	**IQ 1**	**PC 1**	**IQ 2**	**PC 2**	**IQ 3**	**PC 3**
Teamwork	●	●	•	•	⊗	
Cognitive Diversity	●	●	⊗		●	•
Gender Diversity	●	●	⊗	⊗	⊗	⊗
Team Interaction		⊗			•	•
Client Interaction			●	●	•	●
	Collaborative Internal Knowledge integration		**Collaborative Expert Knowledge integration**		**Non-Collaborative External Knowledge integration**	**External Knowledge integration**

The different nature of the pathways shows the relative prominence the team-related conditions in some pathways to team success and the prominence of interaction-related conditions in other pathways to team success. Therefore, the data seems to suggest that there is a substitutive relationship between these types of conditions. These findings complement previous research showing the importance of team-related factors or interaction-related factors. Multiple scholars have investigated the effect of team-related (e.g. Lykourentzou, Antoniou, Naudet, Dow, 2016; Dissanayake, Zhang, Gu, 2015a) and interaction-related factors on team performance (e.g. Javadi Khasraghi & Hirschheim, 2021; Fuger, Schimpf, Füller & Hutter (2017). However, these studies mainly study these relationships in isolation, e.g. by performing regression analyses (e.g. Javadi Khasraghi & Hirschheim, 2021; Schemmann, Chappin, Herrmann, 2017). That makes that currently; there is very little known on the interdependence between team-related and interaction-related conditions within crowdsourcing for innovation. The current findings could be in line with the research done by Javadi Khasraghi & Hirschheim (2021), which show that cross-team collaboration has a large effect on team success in crowdsourcing contests. Moreover, this is in line with Zhu, Kock, Wentker & Leker (2019), who found that experts play a crucial role in idea competitions as they integrate the knowledge of the teams that participate. The substitutive relationship could also be seen in light of the open innovation paradigm. Central in this paradigm is the idea that firms can and should use both external and internal ideas to advance their technology (Chesbrough, 2003). By means of purposive inflows and outflows of knowledge, access to a wider range of ideas and perspectives is created and thereby, internal innovation can be accelerated (Chesbrough, 2006). Most research on open innovation is from the firm's perspective, but there are also studies that regard it from the perspective of start-ups (Usman, van Haverbeke, 2016). Start-ups can benefit from open innovation practices by using external knowledge flows (Brunswicker and Vanhaverbeke, 2015), to overcome the scarcity of resources (Gassmann, Enkel, Chesbrough, 2010). This same idea is arguably also shown within this study, as the crowdsourcing teams falling within the external pathways compensated for their low level of internal team-related conditions by collaborating closely with external parties like the client that provided them with valuable resources. As this study does shed light on these interrelations, it arguably complements our current understanding of the circumstances under which teams can be successful within crowdsourcing for innovation.

6. Theoretical Implications

The vast majority of studies on team success within crowdsourcing for innovation have focused on online instances of such events. Also, the emphasis lies strongly on the investigation of the success of individual contributors and not on teams. Research about offline team-based crowdsourcing for innovation was therefore lacking. This study contributes to the field by shedding light on those forms of crowdsourcing for innovation. Moreover, most studies done in the field were focused on the investigation of relationships between deemed to be important conditions and team success in isolation. There is a very small number of studies that focuses on investigating the interrelationships between these factors, due to which insights into these interrelationships are largely lacking. This study applies a configurational design using crisp set qualitative comparative analysis (csQCA) with the main aim of providing insights about these interrelationships. The findings of this study suggest that the methodological choice for this study was well-founded, as new interdependencies have been found. The findings reveal the existence of three pathways to team success, one of which is internal and the other two external. The internal pathway is characterised by the importance of team-related factors such as teamwork, cognitive diversity, and gender diversity, while the external pathways highlight the significance of interaction-related factors such as team interaction and client interaction. These findings imply the existence of a substitutive relationship between team-related and interaction-related conditions. As this has previously been underexplored, the results of this study have the potential to inform future research and practitioners in understanding the circumstances under which teams can be successful within crowdsourcing for innovation.

7. Practical Recommendations

Based on the findings of this study, there are several practical implications for organisers of crowdsourcing for innovation events, contributing teams, and knowledge-seeking organisations. First of all, organisers of crowdsourcing for innovation events should be aware of the large role knowledge-seeking organisations, or clients, play with respect to the quality of the ideas generated throughout the competition. Underlying the nature of the client interaction, there are several mechanisms that are paramount to either idea quality or project continuation. Clients were shown to provide diverse perspectives and put the pieces together for successful teams. The nature and quality of the feedback and information flow is, therefore, of utmost importance. Event organisers could make efforts in the introduction of measures to ensure the high quality of this feedback and information flows throughout the contest and also ensure these are consistent across teams. As the findings show that a high level of client interaction was extra important for teams in which the internal conditions were not ideal with respect to either teamwork or cognitive diversity, this is even more important. Moreover, event organisers could be aware that clients influence on whether or not teams continue after the project. If a goal is to generate social impact through such events, organisers could select their partnerships on the basis of whether or not the partners actually are open to the implementation of solutions or rather on exploring a new field of knowledge. Secondly, organisers should be aware of the role that teamwork plays in regard to overall team diversity. Diverse teams are a double-edged sword: on the one hand, the variety with regards to perspectives, knowledge and problem-solving mechanisms can provide a wealth of knowledge and contribute to high idea quality. On the other hand, the differences regarding communication preferences, norms and values can provide friction and knowledge exchange and thereby hindering knowledge development. The data suggest teams that scored very high in terms of teamwork were able to navigate these issues. Therefore, during the challenge, the organisers could facilitate teamwork, especially in extra diverse teams.

8. Limitations and Future Research

The study provided valuable insights into the interplay between contributor-related and interaction-related conditions and its effects on team success within offline team-based crowdsourcing for innovation. However, there are some limitations to this study that make that the results of this study should be interpreted with caution. Suggestions for future research are included whenever relevant.

First of all, the findings of this research cannot be easily generalised to other settings, as they are dependent on the context of the Circular Challenge. This means that the results are specific to the particular conditions, environment, and circumstances of this particular study and may, therefore, not be applicable or relevant to other settings. Despite not being able to generalise the findings to other settings, the study still provides a novel perspective on the interplay between team-related and interaction-related conditions, which might serve as a venue for further research to be conducted. In terms of future research, this highlights the importance of conducting further studies on offline team-based crowdsourcing for innovation on other platforms to validate and generalise the findings. In addition, there was a high occurrence of psychological safety among the sample. It would be interesting to investigate a sample in which there is more variation to see whether the current findings hold water.

Secondly, ideally, the number of cases would have been higher to gain more reliable results. However, the study's findings have a reliability of 90% so the relationships found could not have been based on random data. This is for 5 conditions and 19 cases on the basis of the benchmark tables by (Marx & Dusa, 2011). As this study still exceeds 90% reliability threshold (Marx et al., 2013) and as the models in this study are theory-based, the results are expected to be reliable.

Thirdly, although the conditions included in this study were based on literature, some important conditions might not have been included. There is one condition that stands out among the conditions that might be included and this relates to the properties of the waste streams the teams had to work with. The data revealed that in some exceptional cases, the waste streams were more difficult than others. The Circular Challenge is structured in such a way that every team gets assigned a client with a waste stream. Every team's project is thereby unique. According to the organisers, this is done to stimulate a positive culture during the project and to stimulate cooperation among teams so that they can learn from one another and build meaningful relationships. However, this also means that other conditions could contribute to teams achieving team success. A condition that was explicitly mentioned to influence team success was that of waste stream difficulty and more specifically, the level of waste stream homogeneity. Some waste streams are more difficult than others. Waste streams that are heterogenous by nature were mentioned to be more difficult to deal with. An example of this could be municipal organic waste (Dutch: GFT afval). About 30% of this waste stream normally consists of other waste than organic waste, according to an interviewee[1]. The expensive separation process and the resulting heterogenous waste stream make it extra difficult to make a product of. Another factor that makes waste streams more difficult than others is the chemical complexity. Among the cases, there was a team, for example, dealing with PFAS-contaminated dredge.

PFAS, or Per- and PolyfluoroAlkyl Substances, is the umbrella term for a family of manmade complex chemical compounds that are used for resisting heat, oil, stains, grease and water. It is commonly used in products like cooking ware, textile coatings, and many other appliances in industries like aerospace, photographic imaging, and electronics (Wang et al., 2017). The team dealing with this waste stream reported having a knowledge gap with respect to chemistry, despite this team actively seeking help for solving their PFAS problem while also having someone in their team with a master's level educational background in chemical engineering. It is not known whether the jury evaluations took into account the difficulty of the waste streams. However, this is not expected. The implication for this research specifically is that these teams were probably less likely to become successful within the contest, even though they might have had the potential concerning their team-related and interaction-related conditions. Among the many types of factors researched in relation to team success are also factors related to the nature of the task (Seidel & Langner, 2015) or the level of the challenge (Shah, 2006). As both waste stream heterogeneity and chemical complexity relate to the task, and these seem to be important for eventual team success, it is proposed to include task difficulty as a factor in further studies to shed light on the pathways that are important with respect to regular tasks and difficult tasks. Other factors could be taken into consideration in follow-up research. Having a shared vision was mentioned as an important factor for teams that continued. This is in line with the literature on innovation teams and start-ups, which emphasises the importance of having a shared purpose. When team members are working towards the same goal, it helps to prevent confusion and misunderstandings that can lead to delays and setbacks (Pearce & Ensly, 2004). Task-related backgrounds and skills were thought to be important for team success. When asked about the knowledge and skills gaps, a design background, a chemistry background, and a business background were often mentioned as important. When asking about factors driving team success, firstly, having a design background and secondly, having a chemistry background were mentioned to be most important. The rationale behind this was that skills like experience with the often messy design process, the ability to think out-of-the-box, and having the right skill set to make a physical product were all considered very important. With respect to the chemistry backgrounds, the rationale was that these skills would allow the teams to learn about the inherent properties of their waste stream faster and more efficiently through experimentation in the lab. As the Circular Challenge mainly revolves around commercialising excess waste streams and the vast majority of the innovations are physical products, these are very understandable rationales. With respect to theory, there are several studies that confirm that project-specific skills are important for contributor success. For example, Dissanayake, Zhang, Gu (2015b) found that both skill level and skill alignment are important predictors of team performance. Accordingly, Nolte, Chounta, Herbsleb (2020) found that the fit of technical capabilities with respect to the project is important for teams to continue their project. The match between project-required skills and actual skills has not been researched in the context of offline team-based crowdsourcing for innovation yet. A potential for future research could therefore be to include this condition in the analysis.

Thirdly, a suggestion for further research would be to apply the method of fuzzy set qualitative analysis. In some cases, it proved difficult to decide on a proper membership score for the respective condition, and this dichotomisation might have resulted in the loss of valuable information. This was mentioned in the methodology for one of the teams. On the one hand, they were very frustrated by their client, as they only got their physical waste stream in the penultimate week of the challenge, thereby having significantly less time than other teams to experiment with the waste stream in the laboratories, which was mentioned by other interviewees to be of vital importance for the development of their ideas. However, on the other hand, they reported that they could never have achieved as much as they did (they became runner-up) as a direct consequence of the close collaboration with one of their client's employees, who basically joined them during every team session. Assigning such a team a score of '0' or '1' does not do justice to the situation, as the data richness for such ordinal or scale variables is then not utilised, resulting in a loss of information (Herrmann & Cronqvist, 2009). A suggestion for future research would be to use the method of fs/QCA, as the allowance of in-between decimal scores enables ordinal or scale variables to retain their information richness (Herrmann & Cronqvist, 2009). Nevertheless, these issues minimally impact the reliability of the findings, as multiple robustness tests have been conducted for idea quality as well as the level of client interaction, as was shown in the methodology section. It was found that the main findings stayed intact. The outcome of this analysis can be found in Appendix E.

9. Conclusion

The aim of this study was to gain an understanding of the combinations of team-related and interaction-based conditions that lead to team success in offline team-based crowdsourcing for innovation. This study used a configurational approach to study these interrelationships. The studied sample consisted of 19 teams that participated in an edition of an innovation challenge called the “Circular Challenge” in the Netherlands. The data was collected through structured interviews and self-completion questionnaires with individual team members and crisp-set qualitative comparative analysis (csQCA) was used to analyse the data. The findings show that both team-related and interaction-related factors are important for achieving success in this context. Moreover, the findings show that there were several types of combinations (pathways) that could lead to team success. More specifically, this could be achieved through an internal pathway in which team-related conditions are prominent, or through external pathways in which interaction-related conditions are more important. Also, the study suggests that these types of conditions have a substitutive effect. It is theorised that external pathways occur for teams that lack important internal team-related factors, but that these can be compensated by external interaction-related factors. Client interaction plays a large role in this substantive relationship through the provision of valuable resources for the crowdsourcing teams. This study contributes to the existing literature in several ways. Within the domain of crowdsourcing for innovation, it provides a novel view of the interplay between team-related and interaction-related conditions that are important for team success. It complements the underexplored field of research within offline team-based crowdsourcing, and it makes a methodological contribution by confirming the importance of studying relationships together as these capture the complex dynamics of real-life situations better.

References

Afuah, A., & Tucci, C. L. (2012). Crowdsourcing as a solution to distant search. Academy of management review, 37(3), 355-375.

Aitamurto, T., Landemore, H., & Saldivar Galli, J. (2017). Unmasking the crowd: participants' motivation factors, expectations, and profile in a crowdsourced law reform. Information, Communication & Society, 20(8), 1239-1260.

Ali-Hassan, H., & Allam, H. (2016). Comparing crowdsourcing initiatives: Toward a typology development. Canadian Journal of Administrative Sciences/Revue Canadienne des Sciences de l'Administration, 33(4), 318-331.

Armisen, A., & Majchrzak, A. (2015). Tapping the innovative business potential of innovation contests. Business Horizons, 58(4): 389-399.

Axel, M. Cambré, B. Rihoux, B. (2013). *Chapter 2 Crisp-Set Qualitative Comparative Analysis in Organizational Studies*. In Fiss, P. Cambré, B. Marx, A. (Eds). Configurational Theory and Methods in Organizational Research (pp.23-47). Emerald Group.

Baumgartner, S.: (2006), 'Measuring the Diversity of What? And for What Purpose? A Conceptual Comparison of Ecological and Economic Biodiversity Indices', Working Paper, University of Heidelberg.

Baker, D. P., & Salas, E. (1992). Principles for measuring teamwork skills. Human factors, 34(4), 469-475.

Bell, S. T., Villado, A. J., Lukasik, M. A., Belau, L., & Briggs, A. L. (2011). Getting specific about demographic diversity variable and team performance relationships: A meta-analysis. Journal of Management, 37(3), 709–743. https://doi.org/10.1177/0149206310365001

Benefield, G. A., Shen, C., & Leavitt, A. (2016). Virtual team networks: How group social capital affects team success in a massively multiplayer online game. In Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing (pp. 679-690).

Beretta, M., J. Björk, and M. Magnusson. (2017). Moderating ideation in web-enabled ideation systems. Journal of Product Innovation Management. https://doi.org/10.1111/jpim.12413

Berg-Schlosser, D., De Meur, G., Rihoux, B., & Ragin, C. C. (2009). Qualitative comparative analysis (QCA) as an approach. Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques, 1, 18.

Bilgram, V., Brem, A., & Voigt, K. I. (2008). User-centric innovations in new product development—Systematic identification of lead users harnessing interactive and collaborative online-tools. *International journal of innovation management*, *12*(03), 419-458.

Blohm, I., Zogaj, S., Bretschneider, U., & Leimeister, J. M. (2018). How to manage crowdsourcing platforms effectively?. California Management Review, 60(2), 122-149.

Bogers, M., Foss, N. J., & Lyngsie, J. (2018). The "human side" of open innovation: The role of employee diversity in firm-level openness. Research Policy, 47(1), 218-231.

Boons, M., Stam, D., & Barkema, H. G. (2015). Feelings of Pride and Respect as Drivers of Ongoing Member Activity on Crowdsourcing Platforms. Journal of Management Studies, 52(6): 717-741.

Boons, M., & Stam, D. (2019). Crowdsourcing for innovation: How related and unrelated perspectives interact to increase creative performance. Research Policy, 48(7), 1758-1770.

Boss, V., Kleer, R., & Vossen, A. (2019). Walking Parallel Paths or Taking the Same Road? The Effect of Collaborative Incentives in Innovation Contests. In Managing Innovation: Understanding and Motivating Crowds (pp. 111-144).

Boudreau, K. J., & Lakhani, K. R. (2013). Using the crowd as an innovation partner. Harvard business review, 91(4), 60-9.

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

Brabham, D. C. (2010). Moving the crowd at threadless: motivations for participation in a crowdsourcing application. Information, Communication & Society, 13(8), 1122–1145.

Bretschneider, U., Leimeister, J. M., & Mathiassen, L. (2015). IT-enabled product innovation: Customer motivation for participating in virtual idea communities. Int. J. Product Development, 20(2), 126-141

Brito, J. (2008). Hack, Mash, & Peer: Crowdsourcing government transparency. The Columbia Science and Technology Law Review, 9, 119–157.

Brunswicker, S. and Vanhaverbeke, W. (2015), “Open innovation in small and medium-sized enterprises (SMEs): external knowledge sourcing strategies and internal organizational facilitators”, Journal of Small Business Management, Vol. 53 No. 4, pp. 1241-1263

Brunswicker, S., Bilgram, V., & Fueller, J. (2017). Taming wicked civic challenges with an innovative crowd. Business Horizons, 60(2), 167-177.

Bryman, A. (2012). Social research methods. Oxford University Press.

Bullinger, A. C., Neyer, A.-K., Rass, M., & Moeslein, K. M. (2010). Community-based innovation contests: Where competition meets cooperation. Creativity and Innovation Management, 19, 290–303.

Campbell, K., & Mínguez-Vera, A. (2008). Gender diversity in the boardroom and firm financial performance. Journal of business ethics, 83(3), 435-451.

Carmeli, A., Reiter-Palmon, R., & Ziv, E. (2010). Inclusive leadership and employee involvement in creative tasks in the workplace: The mediating role of psychological safety. Creativity Research Journal, 22(3), 250-260.

Chanal, V. and Caron-Fasan, M.L. (2010). The difficulties involved in developing business models open to innovation communities: the case of a crowdsourcing platform. M@n@gement, 13, pp. 318–340.

Chan, K. W., Li, S. Y., & Zhu, J. J. (2015). Fostering customer ideation in crowdsourcing community: The role of peer-to-peer and peer-to-firm interactions. Journal of Interactive Marketing, 31(1), 42-62.

Chappin, M. M., Cambre, B., Vermeulen, P. A., & Lozano, R. (2015). Internalizing sustainable practices: a configurational approach on sustainable forest management of the Dutch wood trade and timber industry. *Journal of Cleaner Production*, *107*, 760-774.

Chesbrough, H.W., (2003). Open Innovation: The New Imperative for Creating and Profiting from Technology. Harvard Business School Press, Boston

Chesbrough, H. W. (2006). Open Innovation: The New Imperative for Creating and Profiting From Technology. Boston, Massachusetts: Harvard Business School Press

Cho, Y. (2022). Team diversity, perspective taking, and employee creativity: The importance of psychological safety. Social Behavior and Personality: an international journal, 50(2), 1-11.

Choo, A. S., Linderman, K. W., & Schroeder, R. G. (2007). Method and context perspectives on learning and knowledge creation in quality management. *Journal of operations management*, *25*(4), 918-931.

Chui, M., Manyika, J., Bughin, J. and Dobbs, R. (2012). The social economy: unlocking value and productivity through social technologies. McKinsey Global Institute.

Chung, J., & Monroe, G. (1998). Gender differences in information processing: an empirical test of the hypothesis-confirming strategy in an audit context. *Accounting & Finance*, *38*(2), 265-279.

Cohen, W. M., & Levinthal, D. A. (1990). Absorptive capacity: A new perspective on learning and innovation. *Administrative science quarterly*, *35*(1), 128-152.

Colón-Emeric, C. S., Ammarell, N., Bailey, D., Corazzini, K., Lekan-Rutledge, D., Utley-Smith, Q., . . . Piven, M. L. (2006). Patterns of medical and nursing staff communication in nursing homes: Implications and insights from complexity science. Qualitative Health Research, 16, 173-188. doi:10.1177/1049732305284734

Cooper, R. G. (2008). Perspective: The stage-gate® idea-to-launch process—update, what's new, and nexgen systems. *Journal of product innovation management*, *25*(3), 213-232.

Dahlander, L., & Gann, D. M. (2010). How open is innovation?. Research policy, 39(6), 699-709.

Dahlander, L., & Piezunka, H. (2014). Open to suggestions: How organizations elicit suggestions through proactive and reactive attention. Research Policy, 43(5): 812–827.

Dahlin, K. B., Weingart, L. R., & Hinds, P. J. (2005). Team diversity and information use. Academy of management journal, 48(6), 1107-1123.

Dai, Y., Byun, G., & Ding, F. (2019). The direct and indirect impact of gender diversity in new venture teams on innovation performance. Entrepreneurship Theory and Practice, 43(3), 505-528.

Darley, W. K., & Smith, R. E. (1995). Gender differences in information processing strategies: An empirical test of the selectivity model in advertising response. *Journal of advertising*, *24*(1), 41-56.

Díaz-García, C., González-Moreno, A., & Jose Sáez-Martínez, F. (2013). Gender diversity within R&D teams: Its impact on radicalness of innovation. Innovation, 15(2), 149-160.

Dissanayake, I., Zhang, J., & Gu, B. (2015a). Task division for team success in crowdsourcing contests: Resource allocation and alignment effects. Journal of Management Information Systems, 32(2), 8-39.

Dissanayake, I., Zhang, J., & Gu, B. (2015b). Virtual team performance in crowdsourcing contest: a social network perspective. In 2015 48th Hawaii International Conference on System Sciences (pp. 4894-4897). IEEE.

Edmondson, A. (1999). Psychological safety and learning behavior in work teams. *Administrative science quarterly*, *44*(2), 350-383.

Edmondson, A. C., & Mogelof, J. P. (2006). Explaining psychological safety in innovation teams: Organizational culture, team dynamics, or personality. Creativity and innovation in organizational teams, 21, 28.

Emmenegger, P., Kvist, J., & Skaaning, S. E. (2013). Making the most of configurational comparative analysis: An assessment of QCA applications in comparative welfare-state research. Political research quarterly, 185-190.

Estellés-Arolas, E., & González-Ladrón-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. Journal of Information science, 38(2), 189-200.

Fiss, P. C. (2007). A set-theoretic approach to organizational configurations. Academy of Management Review, 32, 1180–1198.

Fiss, P. C. (2011). Building better causal theories: A fuzzy set approach to typologies in organization research. Academy of Management Journal, 54, 393–420.

Forza, C. (2002). Survey research in operations management: a process-based perspective. *International journal of operations & production management*.

Foxall, G., & Tierney, J. (1984). From CAP 1 to CAP 2: User-Initiated Innovation from the User's Point of View. *Management Decision*, *22*(5), 3-15.

Franke, N., & Shah, S. (2003). How communities support innovative activities: an exploration of assistance and sharing among end-users. *Research policy*, *32*(1), 157-178.

Fuger, S., Schimpf, R., Füller, J., & Hutter, K. (2017). User roles and team structures in a crowdsourcing community for international development–a social network perspective. Information Technology for Development, 23(3), 438-462.

Füller, J., & Matzler, K. (2007). Virtual product experience and customer participation—A chance for customer-centred, really new products. Technovation, 27(6-7), 378-387.

Gassmann, O., Enkel, E. and Chesbrough, H. (2010), "The future of open innovation", R&D Management, Vol. 40 No. 3, pp. 213-221.

Geiger, D., Rosemann, M., Fielt, E. and Schader, M. (2012). Crowdsourcing information system–definition, typology, and design. ICIS 2012 Proceedings. Available at: https://pdfs.semanticscholar.org/e8c6/f13256b9423c67c59599b283cca868a5649e.pdf (accessed 15 November 2016).

Ghezzi, A., Gabelloni, D., Martini, A., & Natalicchio, A. (2018). Crowdsourcing: a review and suggestions for future research. International Journal of Management Reviews, 20(2), 343-363.

Gibson, C. B., & Gibbs, J. L. (2006). Unpacking the concept of virtuality: The effects of geographic dispersion, electronic dependence, dynamic structure, and national diversity on team innovation. Administrative science quarterly, 51(3), 451-495.

Greckhamer, T., Misangyi, V. F., & Fiss, P. C. (2013). The two QCAs: From a small-N to a large-N set theoretic approach. In Configurational theory and methods in organizational research (Vol. 38, pp. 49-75). Emerald Group Publishing Limited.

Gu, Q., Wang, G. G., & Wang, L. (2013). Social capital and innovation in R&D teams: the mediating roles of psychological safety and learning from mistakes. R&d Management, 43(2), 89-102.

Harrison, D. A., Price, K. H., & Bell, M. P. (1998). Beyond relational demography: Time and the effects of surface- and deep-level diversity on work group cohesion. Academy of Management Journal, 41, 96-107. doi:10.2307/256901

Harrison, D. A., Price, K. H., Gavin, J. H., & Florey, A. T. (2002). Time, teams, and task performance: Changing effects of surface-and deep-level diversity on group functioning. *Academy of management journal*, *45*(5), 1029-1045.

Herrmann, A. M., & Cronqvist, L. (2009). When dichotomisation becomes a problem for the analysis of middle-sized datasets. International Journal of Social Research Methodology, 12(1), 33-50.

Hienerth, C. (2006). The commercialization of user innovations: the development of the rodeo kayak industry. R&d Management, 36(3), 273-294.

Hoegl, M., & Gemuenden, H. G. (2001). Teamwork quality and the success of innovative projects: A theoretical concept and empirical evidence. Organization science, 12(4), 435-449.

Hopkins, R. (2011), “What is crowdsourcing”, in Sloane, P. (Ed.), A Guide to Open Innovation and Crowdsourcing: Advice from Leading Experts in the Field, Kogan Page Publishers, London.

Horwitz, S. K., & Horwitz, I. B. (2007). The effects of team diversity on team outcomes: A meta-analytic review of team demography. *Journal of management*, *33*(6), 987-1015.

Hossain, M., & Kauranen, I. (2014). Competition-based innovation: The case of the X Prize Foundation. *Journal of Organization Design*, *3*(3), 46-52.

Howe, J. (2006) ‘Crowdsourcing: A Definition’, Crowdsourcing: Tracking the Rise of the Amateur (weblog, 2 June), URL (accessed 24 November 2006): http://crowdsourcing.typepad.com/cs/2006/06/crowdsourcing_a.html

Hülsheger, U. R., Anderson, N., & Salgado, J. F. (2009). Team-level predictors of innovation at work: a comprehensive meta-analysis spanning three decades of research. Journal of Applied psychology, 94(5), 1128.

Jain, R. (2010). Investigation of Governance Mechanisms for Crowdsourcing Initiatives. in Proceedings of Americas Conference on Information Systems, pp. 557–563

Jaribion, A., Khajavi, S. H., Järvihaavisto, U., Nurmi, I., Gustafsson, R., & Holmström, J. (2021). Crowdsourcing properties and mechanisms of mega hackathons: The case of junction. IEEE Transactions on Engineering Management.

Javadi, E., & Gebauer, J. (2019). An Experimental Study of the Attention-based View of Idea Integration: The Need for a Multi-level Dependent Variable. Journal of Information Technology Theory and Application (JITTA), 20(1), 1.

Javadi Khasraghi, H., Wang, X., & Li, Y. (2020). Crowdsourcing Contests: Understanding the Effect of Environment and Organization Specific Factors on Sustained Participation. In Proceedings of the 53rd Hawaii International Conference on System Sciences.

Javadi Khasraghi, H., & Hirschheim, R. (2021). Collaboration in crowdsourcing contests: how different levels of collaboration affect team performance. Behaviour & Information Technology, 1-17.

Jeppesen, L. B., & Lakhani, K. R. (2010). Marginality and problem-solving effectiveness in broadcast search. *Organization science*, *21*(5), 1016-1033.

Joecks, J., Pull, K. and Vetter, K. (2013). Gender diversity in the boardroom and firm performance: what exactly constitutes a “critical mass?, Journal of Business Ethics, Vol. 118 No. 1, pp. 61-72.

Jones, G., Chace, B. C., & Wright, J. (2020). Cultural diversity drives innovation: empowering teams for success. International Journal of Innovation Science.

Karachiwalla, R., & Pinkow, F. (2021). Understanding crowdsourcing projects: A review on the key design elements of a crowdsourcing initiative. Creativity and Innovation Management, 30(3), 563-584.

Kark, R., & Carmeli, A. (2009). Alive and creating: The mediating role of vitality and aliveness in the relationship between psychological safety and creative work involvement. Journal of Organizational Behavior: The International Journal of Industrial, Occupational and Organizational Psychology and Behavior, 30(6), 785-804.

Kazman, R., & Chen, H. M. (2009). The metropolis model a new logic for development of crowdsourced systems. Communications of the ACM, 52(7), 76-84.

Kessel, M., Kratzer, J., & Schultz, C. (2012). Psychological safety, knowledge sharing, and creative performance in healthcare teams. Creativity and innovation management, 21(2), 147-157.

Kirkman, B. L., Cordery, J. L., Mathieu, J., Rosen, B., & Kukenberger, M. (2013). Global organizational communities of practice: The effects of nationality diversity, psychological safety, and media richness on community performance. Human relations, 66(3), 333-362.

Kitsios, F., & Kamariotou, M. (2018). Open data hackathons: an innovative strategy to enhance entrepreneurial intention. International Journal of Innovation Science, 10(4), 519-538.

Klein, H. A., & McHugh, A. P. (2005). National differences in teamwork. Organizational simulation, 229-251.

Kostopoulos, K. S., & Bozionelos, N. (2011). Team exploratory and exploitative learning: Psychological safety, task conflict, and team performance. Group & Organization
Management, 36, 385–415

Krogslund, C., D. D. Choi, and M. Poertner. (2014). “Fuzzy Sets on Shaky Ground: Parameter Sensitivity and Confirmation Bias in FsQCA.” Political Analysis 23: 21-41

Kurtzberg, T. (2005). Feeling creative, being creative: An empirical study of diversity and creativity in teams. Creativity Research Journal, 17, 51-65. doi:10.1207/s15326934crj1701_5

Langner, B., & Seidel, V. P. (2015). Sustaining the flow of external ideas: The role of dual social identity across communities and organizations. Journal of Product Innovation Management, 32(4), 522-538.

Leimeister, J. M., Huber, M., Bretschneider, U., & Krcmar, H. (2009). Leveraging crowdsourcing: activation-supporting components for IT-based ideas competition. Journal of management information systems, 26(1), 197-224.

Levitt, T., (1963). Creativity is not enough. Harv. Bus. Rev. 41 (3), 72–83.

Lohmann, G., Pratt, M. A., Benckendorff, P., Strickland, P., Reynolds, P., & Whitelaw, P. A. (2019). Online business simulations: authentic teamwork, learning outcomes, and satisfaction. Higher Education, 77(3), 455-472.

Lorenzo, R. Voigt, N. Tsusaka, M. Krentz, M. and Abouzahr, K. (2018), “How diverse leadership teams boost innovation”, Boston Consulting Group, available at: www.bcg.com/publications/2018/how-diverse-leadership-teams-boost-innovation.aspx.

Lüthje, C., Herstatt, C., & Von Hippel, E. (2005). User-innovators and "local" information: The case of mountain biking. Research policy, 34(6), 951-965.

Lykourentzou, I., Antoniou, A., Naudet, Y., & Dow, S. P. (2016). Personality matters: Balancing for personality types leads to better outcomes for crowd teams. In *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing* (pp. 260-273).

Majchrzak, A., & Malhotra, A. (2016). Effect of Knowledge-sharing Trajectories on Innovative Outcomes in Temporary Online Crowds. Information Systems Research, 27(4): 685-703.

Martins, L. L., Schilpzand, M. C., Kirkman, B. L., Ivanaj, S., & Ivanaj, V. (2013). A contingency view of the effects of cognitive diversity on team performance: The moderating roles of team psychological safety and relationship conflict. Small Group Research, 44(2), 96-126.

Marx, A., Cambré, B., & Rihoux, B. (2013). Crisp-set qualitative comparative analysis in organizational studies. In *Configurational theory and methods in organizational research*. Emerald Group Publishing Limited.

Mateos de Cabo, R., Gimeno, R., & Nieto, M. J. (2012). Gender diversity on European banks' boards of directors. *Journal of business ethics*, *109*, 145-162.

Mathieu, J., Maynard, M. T., Rapp, T., & Gilson, L. (2008). Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future. *Journal of management*, *34*(3), 410-476.

McEwan D, Ruissen GR, Eys MA, Zumbo BD, Beauchamp MR. (2017). The Effectiveness of Teamwork Training on Teamwork Behaviors and Team Performance: A Systematic Review and Meta-Analysis of Controlled Interventions. PLOS ONE 12(1): e0169604. https://doi.org/10.1371/journal.pone.0169604

Mello, A. L., & Rentsch, J. R. (2015). Cognitive diversity in teams: A multidisciplinary review. Small Group Research, 46(6), 623-658.

Mohammadi, A., Broström, A., & Franzoni, C. (2017). Workforce composition and innovation: How diversity in employees' ethnic and educational backgrounds facilitates firm-level innovativeness. Journal of Product Innovation Management, 34(4), 406-426.

Mu, S., & Gnyawali, D. R. (2016). Developing Synergistic Knowledge in Student Groups. The Journal of Higher Education. https://www.tandfonline.com/doi/abs/10.1080/00221546.2003.11780864?journalCode=uhej20

Natalicchio, A., Petruzzelli, A. M., & Garavelli, A. C. (2014). A literature review on markets for ideas: Emerging characteristics and unanswered questions. Technovation, 34(2), 65-76.

Newman, A., Donohue, R., & Eva, N. (2017). Psychological safety: A systematic review of the literature. Human resource management review, 27(3), 521-535.

Nolte, A., Chounta, I. A., & Herbsleb, J. D. (2020). What Happens to All These Hackathon Projects? Identifying Factors to Promote Hackathon Project Continuation. Proceedings of the ACM on Human-Computer Interaction, 4(CSCW2), 1-26.

Nunnally, J. C., & Bernstein, I. H. (1979). This week's citation classic. Psychom. theory, 34, 1982.

Nussbaum, M. C. (2020). Legal reasoning. In Tasioulas, J. (Ed.), The Cambridge companion to the philosophy of law (pp. 59–77). Cambridge University Press.

O'reilly, T. (2005). What is web 2.0.

Østergaard, C. R. & Timmermans, B., & Kristinsson, K. (2011). Does a different view create something new? The effect of employee diversity on innovation.Research Policy, Elsevier, 40(3), pp. 500-509.

Parameswaran, M., & Whinston, A. B. (2007). Social Computing: An Overview. Communications of the Association for Information Systems , 762-780.

Pearce, C. L., & Ensley, M. D. (2004). A reciprocal and longitudinal investigation of the innovation process: The central role of shared vision in product and process innovation teams (PPITs). Journal of organizational Behavior, 25(2), 259-278.

Pedersen, J., Kocsis, D., Tripathi, A., Tarrell, A., Weerakoon, A., Tahmasbi, N., ... & De Vreede, G. J. (2013, January). Conceptual foundations of crowdsourcing: A review of IS research. In 2013 46th Hawaii International Conference on System Sciences (pp. 579-588). IEEE.

Penin, J. and Burger-Helmchen, T. (2011). Crowdsourcing ´ of inventive activities: definition and limits. International Journal of Innovation and Sustainable Development, 5, pp. 246–263.

Pe-Than, E. P. P., Nolte, A., Filippova, A., Bird, C., Scallen, S., & Herbsleb, J. D. (2018). Designing corporate hackathons with a purpose: The future of software development. *Ieee Software*, *36*(1), 15-22.

Poetz, M.K. and Schreier, M. (2012). The value of crowdsourcing: can users really compete with professionals in generating new product ideas? Journal of Product Innovation Management, 29, pp. 245–256.

Post, C. (2012). Deep-level team composition and innovation: The mediating roles of psychological safety and cooperative learning. Group & Organization Management, 37(5), 555-588.

Quinn, J. B. (1985). Managing innovation: controlled chaos. University of Illinois at Urbana-Champaign's Academy for Entrepreneurial Leadership Historical Research Reference in Entrepreneurship.

Ragin, C. C. (1987). The comparative method: Moving beyond qualitative and quantitative strategies. Berkeley1987.

Ragin, C. C. (2000). *Fuzzy-set social science*. University of Chicago Press.

Ragin, C. C. (2006). Set relations in social research: evaluating their consistency and courage. Polit. Anal. 14, 291e310.

Ragin, C. C. (2008). Redesigning Social Inquiry: Fuzzy Sets and beyond. University of Chicago Press, Chicago.

Ragin, C. C., & Fiss, P. C. (2008). Net effects analysis versus configurational analysis: an empirical demonstration. In C. C. Ragin (Ed.), Redesigning social inquiry: Fuzzy sets and beyond (pp. 190–212). University of Chicago Press.

Ragin, C. C. & Davey, S. (2016). *Fuzzy-Set/Qualitative Comparative Analysis 3.0*. Irvine, California: Department of Sociology, University of California. Retrieved from: https://www.socsci.uci.edu/~cragin/fsQCA/software.shtml

Ragin, C. C. (2018). *User's Guide to Fuzzy-Set/Qualitative Comparative Analysis 3.0*. Irvine, California: Department of Sociology, University of California. Retrieved from: https://www.socsci.uci.edu/~cragin/fsQCA/software.shtml

Reynolds, A., & Lewis, D. (2018). The two traits of the best problem-solving teams. Harvard Business Review, 2.

Riedl, C., & Woolley, A. W. (2017). Teams vs. crowds: A field test of the relative contribution of incentives, member ability, and emergent collaboration to crowd-based problem solving performance. Academy of Management Discoveries, 3(4), 382-403.

Rihoux, B., & De Meur, G. (2009). Crisp-set qualitative comparative analysis (csQCA). *Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques*, *51*, 33-68.

Rihoux, B., & Ragin, C. C. (2004). Qualitative comparative analysis (QCA): State of the art and prospects. In *APSA 2004 Annual Meeting, Panel 47-9*.

Rihoux, B., & Ragin, C. C. (2008). Configurational comparative methods: Qualitative comparative analysis (QCA) and related techniques. Sage Publications.

Rudd, J., Stern, K., & Isensee, S. (1996). Low vs. high-fidelity prototyping debate. interactions, 3(1), 76-85.

Salas, E., Cooke, N. J., & Rosen, M. A. (2008). On teams, teamwork, and team performance: Discoveries and developments. Human factors, 50(3), 540-547.

Saka-Helmhout, A., Chappin, M. M., & Rodrigues, S. B. (2021). Corporate Social Innovation in Developing Countries. Journal of Business Ethics, 1-17.

Schemmann, B. (2018). Crowdsourcing innovation: Understanding online idea generation and knowledge production (Doctoral dissertation, University Utrecht).

Schemmann, B., Chappin, M. M., & Herrmann, A. M. (2017). The right kind of people: Characteristics of successful ideators' online behaviour. Creativity and Innovation Management, 26(3), 277-290.

Schemmann, B., Herrmann, A. M., Chappin, M. M., & Heimeriks, G. J. (2016). Crowdsourcing ideas: Involving ordinary users in the ideation phase of new product development. Research Policy, 45(6), 1145-1154.

Schenk, E., & Guittard, C. (2009). Crowdsourcing: What can be outsourced to the crowd, and why? Paper provided by HAL in its series of working papers with numbers halshs-00439256_v1

Schenk, E., & Guittard, C. (2011). Towards a characterization of crowdsourcing practices. Journal of Innovation Economics, 7, pp. 93–107.

Schermer, B.W. Hagenauw, D. Falot, N. (2018). Handleiding Algemene verordening gegevensbescherming en Uitvoeringswet Algemene verordening gegevensbescherming. Ministerie van Justitie en Veiligheid. Retrieved from (19-01-2023): https://open.overheid.nl/repository/ronl-dd12795b-eaa8-4e23-b552-96ef285cb9ad/1/pdf/Handleiding%20Algemene%20verordening%20gegevensbescherming.pdf

Schneider, C. Q., & Wagemann, C. (2007). Qualitative Comparative Analysis (QCA) und Fuzzy Sets: Ein Lehrbuch für Anwender und jene, die es werden wollen. Barbara Budrich.

Schneider, C. Q., & Wagemann, C. (2010a). Qualitative comparative analysis (QCA) and fuzzy-sets: Agenda for a research approach and a data analysis technique. Comparative Sociology, 9(3), 376-396.

Schneider, C. Q., & Wagemann, C. (2010b). Standards of good practice in qualitative comparative analysis (QCA) and fuzzy-sets. Comparative sociology, 9(3), 397-418.

Schneider, C. Q. and C. Wagemann. (2012). Set-theoretic Methods for the Social Sciences. A Guide to Qualitative Comparative Analysis. New York, NY: Cambridge University Press.

Seidel, V. P., & Langner, B. (2015). Using an online community for vehicle design: project variety and motivations to participate. Industrial and Corporate Change, 24(3), 635-653.

Shah, S. K. (2006). Motivation, governance, and the viability of hybrid forms in open source software development. *Management science*, *52*(7), 1000-1014.

Sharma, A. (2010). Crowdsourcing critical success factor model: Strategies to harness the collective intelligence of the crowd. Retrieved January 27, 2014, from http://irevolution.files.wordpress.com/2010/05/working-paper1.pdf

Simons, T. (1995). TOP MANAGEMENT TEAM CONSENSUS, HETEROGENEITY, AND DEBATE AS CONTINGENT PREDICTORS OF COMPANY PERFORMANCE: THE COMPLIMENTARITY OF GROUP STRUCTURE AND PROCESS. In Academy of Management Proceedings (Vol. 1995, No. 1, pp. 62-66). Briarcliff Manor, NY 10510: Academy of Management.

Singh, V., Terjesen, S., & Vinnicombe, S. (2008). Newly appointed directors in the boardroom:: How do women and men differ?. *European management journal*, *26*(1), 48-58.

Smeets, B., Chappin, M., & Kaashoek, B. (2012). Getting More out of the Crowd: How to Organize Crowdsourcing for Innovation. In Academy of Management Proceedings (Vol. 2012, No. 1, p. 13816). Briarcliff Manor, NY 10510: Academy of Management.

Soltani, P. M., Pessi, K., Ahlin, K., & Wernered, I. (2014). Hackathon: A method for digital innovative success: A comparative descriptive study. In Proceedings of the 8th European Conference on IS Management and Evaluation (pp. 367-373).

Spellerberg, I. F., & Fedor, P. J. (2003). A tribute to Claude Shannon (1916–2001) and a plea for more rigorous use of species richness, species diversity and the 'Shannon–Wiener'Index. *Global ecology and biogeography*, *12*(3), 177-179.

Steffen, W., Rockström, J., Richardson, K., Lenton, T. M., Folke, C., Liverman, D., ... & Schellnhuber, H. J. (2018). Trajectories of the Earth System in the Anthropocene. Proceedings of the National Academy of Sciences, 115(33), 8252-8259.

Tegarden, D. P., Tegarden, L. F., & Sheetz, S. D. (2009). Cognitive factions in a top management team: Surfacing and analyzing cognitive diversity using causal maps. Group Decision and Negotiation, 18, 537-566. doi:10.1007/s10726-007-9099-1

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

Thomann, E., & Maggetti, M. (2020). Designing research with qualitative comparative analysis (QCA): Approaches, challenges, and tools. Sociological Methods & Research, 49(2), 356-386.

Trainer, E. H., Kalyanasundaram, A., Chaihirunkarn, C., & Herbsleb, J. D. (2016). How to hackathon: Sociotechnical tradeoffs in brief, intensive collocation. In Proc. 19th ACM Conf. Comput.-Supported Cooperative Work & Social Computing (CSCW 2016), 2016, pp. 1118–1130.

UNESCO Institute for Statistics. (2015). INTERNATIONAL STANDARD CLASSIFICATION OF EDUCATION. Fields of education and training 2013 (ISCED-F 2013) – Detailed field descriptions Retrieved on January 7, 2023 from: http://uis.unesco.org/sites/default/files/documents/international-

standard-classification-of-education-fields-of-education-and-training-2013-detailed-field-descriptions-2015-en.pdf

Van Dijk, H., Van Engen, M. L., & Van Knippenberg, D. (2012). Defying conventional wisdom: A meta-analytical examination of the differences between demographic and job-related diversity relationships with performance. Organizational Behavior and Human Decision Processes, 119(1), 38-53.

Viswanathan, V., Atilola, O., Goodman, J., & Linsey, J. (2014). Prototyping: A key skill for innovation and life-long learning. In 2014 IEEE Frontiers in Education Conference (FIE) Proceedings (pp. 1-8). IEEE.

Von Hippel, E. (1978). A customer-active paradigm for industrial product idea generation. *Research Policy*, *7*(3), 240-266.

Vukovic, M. (2009). Crowdsourcing for enterprises. In 2009 congress on services-I (pp. 686-692). IEEE.

Walter, T., & Back, A. (2011). Towards measuring crowdsourcing success: An empirical study on effects of external factors in online idea contest.

Wang, R. (2021). Team Diversity and Team Success in Collaborative Crowdsourcing. Communication Studies, 1-17.

West, J., & Bogers, M. (2014). Leveraging External Sources of Innovation: A Review of Research on Open Innovation. Journal of Product Innovation Management, 31(4): 814-831

Whitla, P. (2012). Towards an integrated crowdsourcing definition. Journal of Information Science, 38, pp. 189–200.

Whitley, R. (2000). The Intellectual and Social Organization of the Sciences (2nd ed.). New York: Oxford University Press.

Xu, H., Wu, Y., & Hamari, J. (2022). What determines the successfulness of a crowdsourcing campaign: A study on the relationships between indicators of trustworthiness, popularity, and success. Journal of Business Research, 139, 484-495.

Xu, Y., & Yang, Y. (2010). Student learning in business simulation: An empirical investigation. Journal of Education for Business, 85(4), 223-228.

Zaamout, K., & Barker, K. (2018). Towards quantifying behaviour in social crowdsourcing communities. In Proceedings of the AAAI Conference on Human Computation and Crowdsourcing (Vol. 6, pp. 203-212).

Zhao, Y., & Zhu, Q. (2014). Evaluation on crowdsourcing research: Current status and future direction. Information systems frontiers, 16, 417-434.

Zheng, H., Xie, Z., Hou, W., & Li, D. (2014). Antecedents of solution quality in crowdsourcing: The sponsor's perspective. Journal of Electronic Commerce Research, 15(3), 212.

Zhu, H., Kock, A., Wentker, M., & Leker, J. (2019). How does online interaction affect idea quality? The effect of feedback in firm-internal idea competitions. Journal of Product Innovation Management, 36(1), 24-40.

Appendix A - Consent form template

INFORMED CONSENT FORM for participation in:

Crowdsourcing for Innovation

Investigating the interaction- and contributor-related characteristics that lead to team success for offline crowdsourcing for innovation

To be completed by the participant:

I confirm that:

- I am satisfied with the received information about the research;
- I have been given opportunity to ask questions about the research and that any questions that have been risen have been answered satisfactorily;
- I had the opportunity to think carefully about participating in the study;
- I will give an honest answer to the questions asked.

I agree that:

- the data to be collected will be obtained and stored for scientific purposes;
- the collected, completely anonymous, research data can be shared and re-used by scientists to answer other research questions;
- video and/or audio recordings may also be used for scientific purposes.

I understand that:

- I have the right to withdraw my consent to use the data;
- I have the right to see the research report afterwards.

Name of participant : _____

Signature: _____ Date, place: ___ / ___ / ____, _____

Utrecht University

Appendix B - Data table

Team Name	CD	GD	PS	TW	TI	CI	W	W+PR	PC
Team P	1	1	1	1	1	1	1	1	0
Team F	0	0	1	1	1	1	1	1	1
Team K	0	0	1	1	0	1	1	1	1
Team S	1	1	1	1	0	0	1	1	1
Team B	1	0	1	0	1	1	1	1	1
Team E	1	1	1	0	1	1	1	1	0
Team L	0	0	1	0	1	0	1	1	0
Team R	1	0	1	1	1	0	0	0	0
Team O	1	0	1	1	1	1	0	1	1
Team N	1	0	1	1	1	1	0	0	1
Team A	0	1	1	1	0	1	0	0	0
Team H	0	1	1	1	0	1	0	0	0
Team M	1	0	1	1	0	0	0	0	0
Team I	0	0	1	1	0	0	0	0	0
Team C	1	1	0	0	1	1	0	0	0
Team J	1	1	1	0	0	0	0	0	0
Team D	0	0	1	0	1	0	0	1	1
Team Q	1	1	1	0	1	0	0	0	0
Team G	0	1	0	0	0	0	0	0	0

Appendix C - Truth tables

Truth table representation for Team Success (Idea Quality = Position Challenge)

Teams	Cognitive Diversity	Gender Diversity	Teamwork	Team Interaction	Client Interaction	Team Success
F, K	0	0	1	0	1	1
D, L	0	0	0	1	0	1
S	1	1	1	0	0	1
B	1	0	0	1	1	1
P	1	1	1	1	1	1
N, O	1	0	1	1	1	0
C, E	1	1	0	1	1	0
H, A	0	1	1	0	1	0
I	0	0	1	0	0	0
M	0	1	0	0	0	0
G	1	0	1	0	0	0
J	1	1	0	0	0	0
R	1	0	1	1	0	0
Q	1	1	0	1	0	0

Truth table representation for Team Success (Project Continuation)

Teams	TW	CI	GD	CD	TI	PC
O, N	1	1	0	1	1	1
S	1	1	0	0	0	1
K	1	0	1	1	0	1
B	1	1	0	0	1	1
F	0	1	0	1	1	1
D, L	0	0	0	0	1	0
A, H	1	1	1	0	0	0
E, C	0	1	1	1	1	0
G	1	0	0	0	0	0
J	0	0	1	0	0	0
I	1	0	0	1	0	0
M	0	0	1	1	0	0
Q	1	0	0	1	1	0
R	0	0	1	1	1	0
P	1	1	1	1	1	0

Appendix D - Robustness test Idea Quality

Instead of taking solely the position in the challenge to determine the level of idea quality, now also prototype quality was taken into consideration . Concretely, this meant that any team that scored high on prototype quality, regardless of whether or not they won the challenge, now also scored a high level of idea quality. The reasoning behind this is that although the jury was able to rank the top two teams, there could be variety in the level of idea quality across editions, so that in some editions the average level of idea quality could be higher than in other editions. As position within the challenge is a relative measure, taking solely this measure means that variation across the editions potentially could not be accounted for. The prototype quality can be evaluated to the same standard across competitions and could in that way also reveal other teams that, in fact, scored high on idea quality and thus team success. Following that line, additionally integrating the measurement of prototype quality into the measurement, idea quality is measured in a more reliable way To measure the *prototype quality* of the contributors, the level of fidelity of a prototype was used. Low-fidelity prototypes have limited functionality and they are made to present general ideas to get a feel of the product or service at hand, examples include visual models, sketches, mock-ups. High-fidelity prototypes have complete functionality and can be used as a final product (Rudd, Stern, Isensee, 1996). Therefore, prototypes that resembled an end-product instead of models, sketches, or mock-ups, were regarded as high fidelity. Both the runners-up, as well as teams that neither were first nor second in the challenge still could achieve a high level of idea quality, regardless of their position in the challenge. This resulted in two teams being added to the list of successful teams, compared to the first analysis.

Calibration Robustness Test Team Success (Idea Quality = Position Challenge + Prototype Quality)

Position in the challenge		**Prototype Quality**		**Team Success**	**Team Success**
				Idea Quality	**Idea Quality (Position + Prototype Quality) - Robustness Test**
				Only winning counts	*Equal evaluation of conditions*
Position	**Membership**	**Fidelity**	**Membership**	**Membership**	**Membership**
Winner	1	High	1	1	1
Winner	1	Low	0	1	1
Runner-up	0	High	1	0	1
Runner-up	0	Low	0	0	0
Other	0	High	1	0	1
Other	0	Low	0	0	0

Truth table representation for Team Success (Idea Quality = Position Challenge + Prototype Quality)

Teams	TW	CI	GD	CD	TI	IQ
D, L	0	0	0	0	1	1
K	1	1	0	0	0	1
S	1	0	1	1	0	1
F	1	1	0	0	1	1
B	0	1	0	1	1	1
P	1	1	1	1	1	1
N, O	1	1	0	1	1	0
E, C	0	1	1	1	1	0
A, H	1	1	1	0	0	0
I	1	0	0	0	0	0
G	0	0	1	0	0	0
M	1	0	0	1	0	0
J	0	0	1	1	0	0
R	1	0	0	1	1	0
Q	0	0	1	1	1	0

The results of the robustness test are presented in the table. The results are highly similar to the results before, as configuration 1 and 2 are both identical to configuration 1 and 2 in the original analysis as shown in section 4.1. Configuration 3 is slightly different, with two conditions now omitted. This new configuration shows that teams that scored low on teamwork, low on gender diversity, and high on team interaction may result in a high level of idea quality, regardless of the level of cognitive diversity, or client interaction. As the main patterns are similar, the analyses are regarded robust.

Table. Results of the robustness for team success (idea quality)

Configurations for achieving high idea quality (position challenge + prototype quality)			
Configuration:	1 (=1)	2 (=2)	3
Conditions			
Psychological Safety	-	-	-
Teamwork	●	•	⊗
Cognitive Diversity	●	⊗	
Gender Diversity	●	⊗	⊗
Team Interaction			•
Client Interaction		●	
Raw Coverage	0.222222	0.222222	0.333333
Unique Coverage	0.222222	0.222222	0.333333
Consistency	1	1	1
Solution Coverage	0.777778		
Solution Consistency	1		

Appendix E - Robustness test Client Interaction

Table. Results of the robustness for high idea quality (lenient version of client interaction)

Configurations for Idea Quality			
Configuration	**1) Collaborative internal knowledge integration**	**2) Collaborative expert knowledge integration**	**3) Non-collaborative External knowledge integration**
Conditions			
Psychological Safety	-	-	-
Teamwork	●	•	⊗
Cognitive Diversity	●	⊗	●
Gender Diversity	●	⊗	⊗
Team Interaction			•
Client Interaction	•	●	•
Raw Coverage	0.29	0.29	0.14
Unique Coverage	0.29	0.29	0.14
Consistency	1	1	1
Solution Coverage	0.71		
Solution Consistency	1		