Tampere University of Applied Sciences

Quality Control in Crowdsourcing:

approaches, evaluation and application.

Ghassan Abarbou

MASTER'S THESIS
November 2020

International Business Management

ABSTRACT

Tampereen ammattikorkeakoulu
Tampere University of Applied Sciences
International Business Management

GHASSAN ABARBOU:
Quality Control in Crowdsourcing Platforms: approaches, evaluation and application.

Master's thesis 52 pages, appendices 2 pages
November 2020

Crowdsourcing is a prominent paradigm within the modern digital business. More and more businesses are opting for this business model to drive their operations. With the emergence of this trend and the flourishing of its applications and activities, the need for customised quality control became evident. Currently, crowdsourcing platforms rely mainly on remote work over the internet. There exist a plethora of techniques to control the quality within crowdsourcing platforms. However, these techniques are situational and depend on the task at hand.

The presented work examines these quality control techniques to select the best result-yielding ones. The selected techniques are evaluated and their application is examined within the practical setting of the author's current employer. The evaluation is reflect on to conclude the outcomes of the research and propose the basis for integrated quality assurance strategy.

To examine the selected techniques, the research carried out a survey targeting a selected population representing different target groups involved in crowdsourcing-reliant projects. The survey's questionnaire included quantitative questions as well as open-ended qualitative ones. After the answers to the survey were received, the results were analysed quantitatively and qualitatively.

The results showed the importance of quality control in crowdsourcing projects; and that the techniques used are situational and depend on the stage of the crowdsourcing project they are applied to by the people involved. The results, however, provided insight into an integrated strategy that can build quality assurance within a whole crowdsourcing operation. The presented work sets the basis for this strategy.

Keywords: crowdsourcing, crowdsourcing platform, quality control, quality model, quality assurance, crowd, computational, compensation, review, feedback, effective communication, community management, task description.

CONTENTS

1 INTRODUCTION

Crowdsourcing platforms are becoming an omnipresent paradigm in today's business. With the rise of global businesses and the need of international talent for different projects; the need for crowdsourcing became evident and the practices within this paradigm flourished. The term crowdsourcing is a relatively young concept; it describes a web-based business model that relies on the involvement and innovation of different individuals in a distributed network (Souza et al., 2009). Crowdsourcing effectively takes advantage of the fact that some tasks are relatively easier to perform by humans rather than computers; profiting from the creative capabilities of the involved subjects as well as from today's abundant communication channels. It helps modern businesses cut on costs, improve productivity, widen the skill and talent pool, reduce training efforts and take advantage of customised skills for building task-targeted teams (Benatallah et al., 2013). However, the distributed and the open nature of such systems open the door for multiple challenges and risks. Dealing with these challenges is among the most active research fields within this paradigm. The most prominent challenge of the lot is quality control. With the diversity of the talent pool, the abundance of candidates and the variety of their skill levels; controlling the quality of what is produced in crowdsourcing platforms is a necessity. Project managers strive to set up effective quality control measures to ensure that the end product meets the requirements while still taking advantage of the perks of the crowdsourcing paradigm. Within this context, this research takes place. Working for the past couple of years on developing crowdsourcing systems and dealing closely with production teams responsible for running such systems; helped the author get familiar with the field and set up a solid background for the research.

The motivation behind this research is a personal one that comes from the author's interest in the subject, the continuous contact with crowdsourcing systems and the emergence of this trend within the modern business. In addition, the author's engineering background and previous experiences with artificial intelligence systems increased the interest in crowdsourcing as the two concepts are undoubtedly associated.

1.1 Problem definition

With the increase of interest in crowdsourcing systems and the benefits such systems provide, emerged the need for quality control measures. The need for ensuring that data, tasks and processes resulting from crowdsourcing collective creative effort made quality control one of the main focuses when designing and executing projects relying on this business model. The proposed research lies within this context and takes as basis the expertise and the familiarity of the author with developing and dealing with such systems. The author's current employer, Lionbridge Tampere Oy, within its Artificial Intelligence unit provides a portfolio of services relying primarily on the crowdsourcing paradigm. To drive projects, production teams rely on crowdsourcing and data outsourcing (Lionbridge AI, 2020) and on specifically designed tools developed in-house. One of the author's current tasks as a senior system engineer within this unit is to develop and maintain the later mentioned designed tools; providing project and task-specific systems that are aimed to be used by a globally distributed talent pool. The proposed research builds up on this basis. Combined with the technical knowledge, managerial skills were learnt and applied to evaluate and assess quality control measures within this portfolio.

The aim of the research is to examine the different approaches and the state-of-the-art techniques used for quality control in crowdsourcing platforms, evaluate the best result-yielding ones and examine the way they are applied within the author's employer's line of work and using the available resources from the company portfolio and experience.

The need of such research draws from the vitality of ensuring the desired quality in the services the company provides; which, in turns, ensures customer satisfaction i.e. is one of the most strived for and among the main defining characters for the company. The research questions the current work answers are formulated as follows:

- What are the best result-yielding quality control techniques in crowdsourcing platforms?
- How are these techniques evaluated and applied?

1.2 Method statement

The research starts by covering the state-of-the-art literature on the subject and examining the different approaches of quality control in crowdsourcing. This covers the basic aspect of the research and lays basis to the applied research aspects where the most used techniques are selected to be examined. The techniques are evaluated in terms of their application within the context of the company's portfolio, resources and general experience. The applied research aspect of the presented work relies on quantitative and qualitative research designs. The quantitative research is carried out by designing a custom survey that evaluates the use of a selected set of quality components by a selected population representing the employees of the company working on crowdsourcing projects. The data is then collected and reflect on to come up with a list of outcomes and possible improvements. The qualitative research is carried out next drawing conclusion on the non-numerical qualitative components of the survey. The results are then analysed quantitatively using statistical analysis combined with relational analysis to draw extra conclusions; and qualitatively using content analysis. The research concludes by synthesizing the outcomes and setting a basis upon which an integrated quality assurance strategy can be established.

1.3 Relevant company background

Briefly and covering the relevant aspect of the company the research is targeting, Lionbridge is a language service provider bringing "fast, scalable, high-quality language and AI services" to its customers (Lionbridge, 2020a). The company offers a wide range of services to customers, globally. Its services include: Content services (Technical writing, Training and e-learning, Financial Reporting, Multicultural marketing etc.), Translation services (natural language translation, Software localisation, Linguistic quality assurance etc.), Testing services (Functional, Compatibility, Interoperability, Performance and Accessibility testing), and last but not least Artificial Intelligence (AI) services, which are the most relevant services for this research (Lionbridge, 2020b). The company has adapted in recent years to the emerging Artificial intelligence and

language services market demand and mainly shifted its activities towards catering to this trend. In the Lionbridge Tampere branch, the dominant activities are related to providing AI services. The company relies on its strong community of resources to carry customer projects (Lionbridge, 2020a). In the context of Artificial Intelligence, projects rely heavily on human-created created data and carried tasks (Eskenazi, 2013). Data and tasks that depend on a variety of criteria and on the purpose of the system that is being developed for and on the project general requirements. Crowdsourcing is a strategic choice to accommodate for such a varied range of criteria.

The presented work relies on examining the quality assurance measures applied within this crowdsourcing strategy. The crowdsourcing operations of the company are based on a talent pool of over one million crowd workers who take part in different content processing, creation and evaluation tasks (Lionbridge, 2020b). The flow of the operations is simplified as follows (Lionbridge, 2020c).

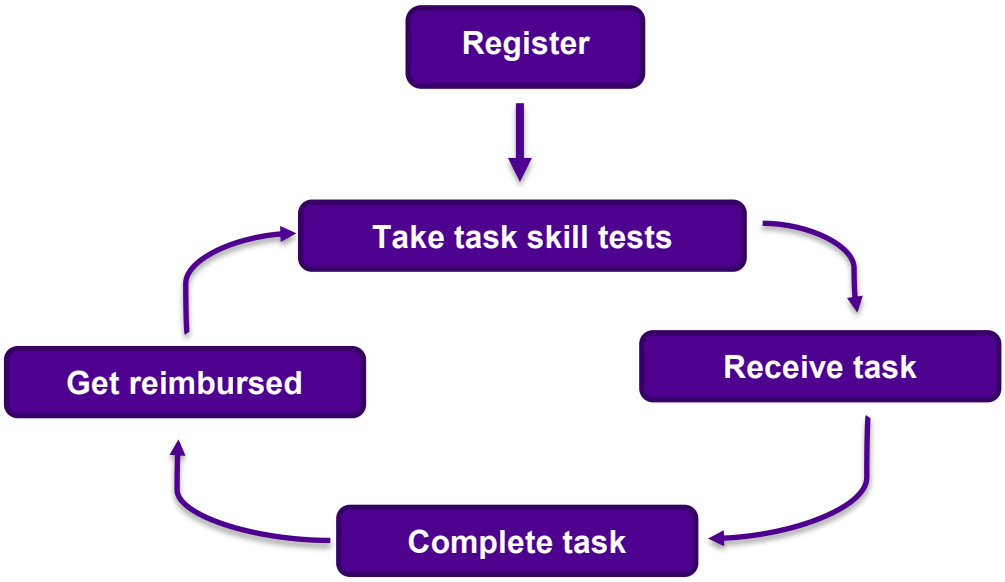


FIGURE 1. Lionbridge crowdsourcing flow

1.4 Research outline

The research shall start by examining the state-of-the art literature on crowdsourcing, quality in crowdsourcing and the quality control techniques. This sets the theoretical framework upon which the research is built and covers the basic research.

The next chapter shall explain the methodology of the research covering used the quantitative and qualitative designs. The chapter then shall cover the different aspects related to the design of the survey detailing the types of questions used and their aim, the sampling and target groups, the methods used to collect data, and the error correction methods. Then, I shall detail the methods used to analyse the data.

The research shall then analyse the results of the survey, reflect on the results in light of the theoretical framework and draw the main outcomes. Next, the outcomes shall be synthesised in the form of the basic components of an integrated quality assurance strategy.

The research shall conclude by a summary and general reflection as well as future work and research limitations.

2 QUALITY ASSURANCE IN CROWDSOURCING PLATFORMS

In the next section the research shall elicit literature relevant to the vision, general direction and scope of the topic at hand. The aim is to define the basic concepts related to quality assurance in crowdsourcing platforms as well as presenting the techniques most use the in discipline and how they contribute to making the practice successful and worthwhile. Ultimately setting up the theoretical framework upon which the research based its methodology. Literature is approached in a pragmatic manner to cover the relevant concepts without diving too much into details, respecting the scope of the work. Links are systematically established between the sections in light of the aim of the work. Given the nature of the work at hand, the technical aspect is used, and functional views are used as support to legitimise the choice of techniques and their feasibility. In order to use a certain technique its premise had to be examined first, evaluated second and its relevance to the project examined as a last step.

2.1 Crowdsourcing

2.1.1 Operational and functional definitions

Crowdsourcing as a word is formed by concatenating the two words crowd and outsourcing. The term was brought to the public by Howe (2006) to describe the behaviour of organisations or companies where they outsource a certain task to a large, mostly unknown crowd of workers through the medium of internet. Howe (2006) defines crowdsourcing as:

> “Crowdsourcing is the act of taking a job traditionally performed by a designated agent (usually an employee) and outsourcing it to an undefined, generally large group of people in the form of an open call”

Operationally, the foundation of crowdsourcing relies on the basic flow relying on the need for a task to be completed, the ability of non-experts to complete

the task, and the feasibility of issuing an open call to initiate the task. Building up on this foundation, the internet and modern communication media facilitate both the open call initiation, presentation, design, and workings of tasks (Eskenazi, 2013).

Form the above, we can conclude that crowdsourcing functionally can be defined by three main parts:

- The task: the actual work to be done.
- The open call: an available-to-the-masses audition/test or invitation to complete a task
- The crowd: a group of non-experts or in some cases experts that might have a specific set of skills to complete a task.

Building on these base concepts, the concept of the wisdom of the crowd emerges. Within the crowdsourcing paradigm the knowledge generated by the large non-expert participants or the “intelligent crowd” is considered to approach the knowledge of experts in the task or field at hand (Surowiecki, 2004). It, however, differs from task to task and from crowd to crowd (Eskenazi, 2013). For the sake of the current, the assumption that the quality of knowledge generated by an intelligent crowd approaches the one of experts (Surowiecki, 2004). With an intelligent crowd, as it is the case for the task the current work is targeting, it is also believed that tasks completed by an intelligent crowd (one that is selected using task specific criteria) are less erroneous than when completed by experts thanks to the aggregation of knowledge and the faster completion pace.

To better understand the operational components in crowdsourcing, the following example is used to simplify the different components. Consider a classification problem often encountered in fruit farms where participants are asked to sort fruits by size and are given a reference size for each of the sorting buckets. The employers let all the workers know that the fruits are to be sorted by size (open call), the workers are to sort the fruits having the defined sizes as a reference and sorting into separate buckets (task). Each worker is given an unsorted pile of fruit and different buckets with a picture of the reference size. At the end of the task the workers bring their buckets to a common area and empty

their buckets into corresponding containers with the same picture of the reference size. The employers check the containers at the end of the day and in each they might find bigger or smaller fruits compared to the reference size. However, the average of the sorting for each of the containers tends to be accurate and correspond to the reference size (aggregation).

In the context of the current work and the relevant tasks requiring knowledge and an intelligent crowd, Surowiecki (2004) defines the four main characteristics that make a crowd an intelligent crowd that is able to generate desirable knowledge as follows:

- Diversity: the crowd is diverse which adds an element of different perspective to the task making the generated knowledge of better quality.
- Independence: each member of the crowd has their own opinion that can be different from others within the specifications of the task.
- Decentralisation: knowledge is decentralised, and each member of the crowd has their own contribution that adds value to the task.
- Aggregation: ability to combine the independent, decentralised and diverse knowledge of the members of the crowd to find a consensus of high quality that serves the task.

Functionally, crowdsourcing is defined by the four characteristics that set the basis of the concept according to Surowiecki (2004). The paradigm is built upon the premise of mutual benefits (Wazny, 2017), where participants get a tangible or an intangible (personal satisfaction, social recognition etc.) and the issuer gets the knowledge needed to complete the task at hand. The four characteristics are further elaborated on as follows (Surowiecki, 2004):

- Diversity of opinion: members of the crowd will have different opinions about the task. The opinions might be similar and not differ that much from each other and might as well be either correct or incorrect.
- Independent opinion: each member of the crowd will have his or her individual opinion that is not influenced by others. The opinions would only be influenced by the task-specific instructions.
- Decentralised opinion: each member of the crowd would have local knowledge that they will apply to the task with the framework set by the compartmentalised task description (members of the crowd might not

have all information related to the task, the issuer of the open call controls what information is exposed).

- “Aggregable” opinion: the collective knowledge of the participant crowd can be aggregated and merged to form a collaborative opinion that will serve the purposes of the task at hand.

Given the constraints of cost effectiveness, crowdsourcing is a viable alternative to the tedious data collection, data annotations and data processing that influence most of modern systems and drive the information technology world in the current times and will continue to do so in the near future at least. Functionally and operationally, crowdsourcing offers a solid basis for data mining tasks while having a faster, more scalable and better knowledge compared to relying entirely on experts. (Wazny, 2017)

Throughout the literature, it is common to find a different definition of crowdsourcing, as it heavily depends on the tasks at hand and its specifications. Hosseini et al. (2015) showcase that there are contradictory definitions of crowdsourcing; however, they argue that it is heavily dependent on the task the crowdsourcing paradigm is used for. For example, classification tasks where the dataset processing only needs human involvement to classify data according to a certain characteristic (i.e. pictures having an item or not) (Li et al, 2013) do not require a particular skill; hence the element of an intelligent crowd is irrelevant and the members of the crowd do not need to have a specific set of skills. Whilst, for other tasks might need a certain set of skills that crowd workers are required to have to be able to take part in the task (Benatallah et al., 2013) (i.e. sentence annotation requires nativity of the language at hand, pattern recognition requires linguistic skills etc.). Nonetheless, the essence of the operational and the functional definitions from the above still holds as the main characteristics are independent of the task.

2.1.2 Crowdsourcing application and platforms

Applications of the crowdsourcing paradigm are multiple, and they cover a variety of fields that range from scientific applications to business-oriented

tasks. The main applicable classification that can be applied to crowdsourcing is the following (Wazny, 2017):

- Problem solving
- Data classification
- Producing a decision
- Knowledge discovery, management and generation

From the literature, the following are some examples of crowdsourcing applications in a variety of fields:

- Public health and health research (Prpic, 2017; Kamajian, 2015)
- Knowledge discovery and management (Brabham et al.,2014)
- Collaborative science projects (Wagy et al., 2018)
- Information sharing and voting systems (O'Leary, 2019)
- Selective sourcing with crowd assessment (Ghani et al. 2015)
- Data processing for machine learning (Daniel et al., 2018)

From the above, and throughout the literature the crowdsourcing paradigm is used in a variety of fields and takes on different forms of application. However, the essential premise of the said applications revolves around a common basis. The following diagram describes the flow of this basis (Hirth, Hoßfeld & Tran-Gia, 2013):

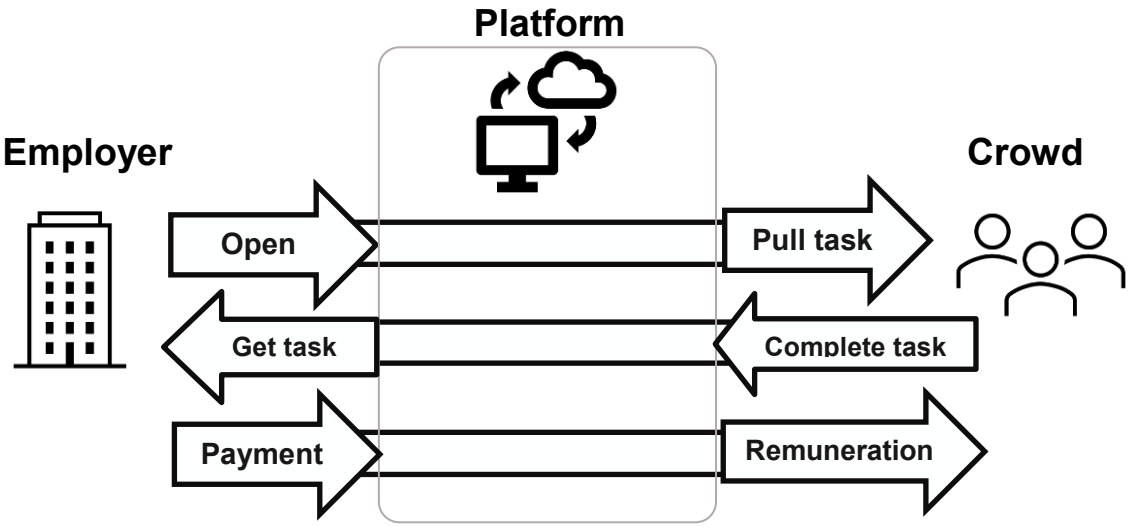


FIGURE 2. Basic structure of a crowdsourcing operation

In figure 2, the employer who is the issuer of the task, has a task to be completed. They issue an open call using the crowdsourcing platform to be used. The crowdsourcing platform is then used to design the task and present it to the crowd. The platform can also be used to select the crowd that would have access to the open call, as explained in section 2.1.1. The crowd workers pull the available tasks and submit their work using the platform. The crowdsourcing platform can then be used to process the submitted work (e.g. aggregation as explained in section 2.1.2) according to certain criteria defined by the employer who can then deliver the tasks. Once the work is completed, the employer issues payments to the members of the crowd using the platform and the crowd can redeem their remuneration according to their preferred method over the internet or through the provided payment details.

2.1.3 Benefits of using crowdsourcing

Benefits of using crowdsourcing differ from operation to operation and from application to application. However, the general consensus from the literature (Wazny, 2017; Zhao & Zhu, 2012; Daniel et al., 2018) showcases that crowdsourcing is particularly beneficial to drive operation relying on the advantages of the internet and modern communication media. Using the said communication media improves the connectivity and the collaboration ability of a distributed talent pool (Zhao & Zhu, 2012). This, in turns, allows for the optimisation of costs of carrying tasks, especially tasks most suitable for the crowdsourcing paradigm (section 2.1.2). Sourcing talents tailored for tasks is also a major benefit of crowdsourcing; where the right talents for the task are selected based on the settings of the open call (Wazny 2017), which improves the quality of the delivered tasks. Another aspect of crowdsourcing that also contributes to improved quality is the aggregation of the opinions that characterises such tasks (Lebraty & Lobre-Lebraty, 2013; Eskenazi, 2013). Crowdsourcing operations are also characterised by high mobility and scalability where large masses of participants can be targeted to complete a task that requires such settings (Lebraty & Lobre-Lebraty, 2013). Additionally, Kanhere (2011) argues that crowdsourcing takes advantage of the widespread of mobile devices and internet services to complete tasks that could not have

been completed before; making use of crowds that were not thought of as possible workers; especially for large data collection task. Wazny (2017) categorised the benefits of using crowdsourcing into two main categories that are in the essence of crowdsourcing operation: process-based and result-based, with some benefits belonging to both categories.

In essence, crowdsourcing indeed influences the process and the results of operations using it. The paradigm improves the process of the operations by improving the way the open call is designed, the scale of access to the task, robustness of task completion and the enriching of the talent pool. Results are improved by improving the selection process of participants, optimising the way the task is presented to the crowd, processing and aggregating the results.

These benefits certainly do not come without a considerable overhead (Daniel et al. 2018). To take full advantage of the said benefits quality control is a major component that needs to be carefully examined whenever the consideration to drive a crowdsourcing operation emerges. Quality control is considered the main challenge facing crowdsourcing platforms (Benatallah et al. 2013; Daniel et al. 2018; Brabham et al. 2013; Garcia-Molina et al. 2016)
In the next section the research shall cover the problematic of quality in crowdsourcing, cover a proposed taxonomy of quality control in crowdsourcing platforms, examine the associated quality model and elicit the most used techniques for quality control in crowdsourcing from the literature.

2.2 Quality in crowdsourcing

Quality control in crowdsourcing in the literature is quite scattered and there is a lack of centralised and standardised model that ensures quality in crowdsourcing platforms (Daniel et al., 2018). This is mainly due to the fact that most crowdsourcing tasks are highly customised, and the carried processes differ from task to task and from field to field. For example, within the data processing and collection field a task to collect textual data would require fundamentally different quality control compared to a task to collect audio samples; issues of data size, used software and technology, post processing

etc. have to be considered differently. From the author's own experience, it is highly challenging to ensure a high quality of the delivered task; each task defers in requirements, presentation, pre-processing and post-processing from the other. In the following section the research shall cover a proposed quality model applicable to the scope and elicit the related most used techniques. The research shall keep the examination of the quality control techniques focused on the ones that would be applicable within its scope.

2.2.1 Taxonomy

Daniel et al. (2018) propose a taxonomy that aims at covering the main components (as illustrated in figure 3) that constitute quality. Quality has to be controlled during the different stages the information flow goes through. Conceptually, the proposed taxonomy can be illustrated as follows (applied to figure 2):

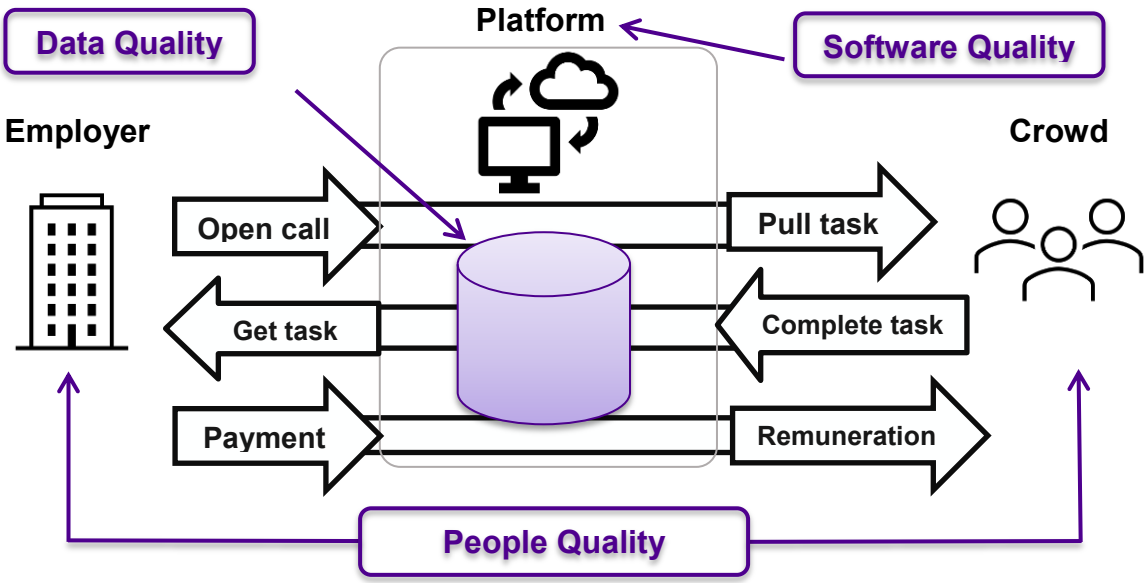


FIGURE 3. Quality taxonomy in a crowdsourcing operation

From this simplification of the taxonomy (figure 3), we can draw four major component that will lay basis to quality control in crowdsourcing platforms in general, and especially, will be adapted to lay foundation the theoretical framework of this work. The four components and their defining sub-elements are as follows (Daniel et al., 2018):

- Quality model: defined by the target quality attributes and the quality dimensions.
    - Quality attributes: define requirements and properties of the task. Include the selection criteria of the crowd workers, the format of the data processing, the target accuracy of data etc.; the attributes can be measurable (nativity of crowd workers, level of education, expertise) or abstract (aggregable opinion).
    - Quality dimensions: the components of the task that are derived from the above explained attributes.
- Quality control techniques and assessment: define the target attribute and the evaluation methods used to qualify it; as well as, who is responsible for the evaluation. They also define evaluation basis (e.g. computational, peer review, gold standard, human quality assurance, post processing etc.)
- Quality assurance: improve the overall process and result quality by examining the quality attributes and customising the quality model.
    - Strategies: high-level decisions that influence the vision of the operation in terms of quality. For example, rigorous crowd selection can be favoured to post processing and manual QA.
    - Actions: the concrete operations that aim at serving the adopted strategies and address quality issues.

The above taxonomy targets the different components of crowdsourcing operations. Consequently, its different aspects will be further detailed and examined.

2.2.2 Quality model

Dictated by the aim of this research and based on the literature, the following will be the target **quality dimensions** and their respective quality attributes or quality sub-dimensions for the quality taxonomy that will be examined. This is supported by the nature of the target projects and general consensus from the literature.

- **Data** (Daniel et al., 2018): Data is at the centre of focus in operations that are similar to the projects within the scope of the presented work. Input and output data need to maintain a certain level of quality that the origin, mediator and worker are required to satisfy. The target quality attributes for this dimension are:
  - Input completeness: refers to the percentage of completeness of data from the origin of the task. For example, for a task of image annotation, input completeness can the percentage of the images that render and are not corrupt within the original dataset, the percentage of duplicate images, or the format uniformity across images.
  - Output accuracy: refers to the percentage of completed task data that match the task description. Accuracy can be measured in different ways and generally is a requirement from the origin of the task. For example, given a sentence annotation task, the returned annotated sets have to match a certain accuracy when compared to the task standard (Polychronopoulos, 2017) (e.g. a QA post processing measuring algorithm, test cases passing using the annotated test etc.)
  - Output consistency: refers to the uniformity of task completion between the crowd workers given the same input (Daniel et al., 2018). For instance, given a sentence in a sentiment analysis task, output consistency can be the percentage of similarity in classification of a certain sentence or set of sentences among the crowd workers.
  - Reaction Time: refers to data integrity within real-time crowdsourcing platforms, especially for concurrent and dependent microtasks (building blocks of the major task that might depend on each other and might be completed by different crowd workers) (Yin, Chun & Sun, 2014). For example, this attribute can be the measured by latency in milliseconds of selecting an image to be annotated by a member of the crowd in different regions of the globe from a pool of images where a file is to annotate only once by one person.
- **Task** (Benatallah et al. 2013, Daniel et al. 2018) the work to be performed, it can take a multitude of forms and can cover any of the applications associated with crowdsourcing as well as combinations of applications:
  - Task description and definition: refers to the clarity and completeness of the instructions given to the members of the crowd to complete the task.

These instructions are pivotal in the smooth running of the work (Daniel et al. 2018). It can be, for example, a set of rules to follow when annotating an image with a list of objects to annotate and instructions of how to annotate them.

- The software and the interface (sub-dimension): rather abstract measure that can be categorised more as a quality sub-dimension. It refers to the general quality of the supporting software and the usability of the interface used to carry out the crowdsourcing operation.
- Granularity: also, a sub-dimension refereeing to how simple or complex the designed task is and how the crowd workers would carry the work given the complexity. Generally, the more complex the task is, the harder it is to achieve the desired quality for this sub-dimension (Benatallah et al. 2013).
- Incentive and compensation: refers the retention of members of the crowd and can be measured by the percentage of crowd workers staying in the platform or re-applying for tasks based on compensation and incentives offered by the employer to complete the task. Incentive is shown to also affect the overall quality of the operations (Benatallah et al. 2013).
- Performance and throughput: refers to the amount of work completed within a time period and can be measured for members of the crowd individually, per task or for the crowd in general. Finding the balance between the resources, compensation and load of work is the outcome of this attribute.

- **People** (Daniel et al. 2018): The people involved in the task, the crowd (people who perform the tasks) and the employer (people who design the task, present the task and select the crowd).
  - Employer: the issuer of the crowdsourcing operation including people who design the tasks, work on the platform to present and communicate the task to the crowd; and people who process, evaluate and deliver the output and people responsible for compensating the crowd workers. The main quality attributes for this sub-dimension are communication, fairness, and reputation.
  - Crowd: people who actively take part in completing the task. The crowd workers can be evaluated as individuals or as part of a group. Quality

attributes associated with this dimension are subjective and depend on the task at hand, the main ones are: the worker's profile, experience, credentials (Daniel et al. 2018), and reputation and expertise (Benatallah et al. 2013). To that we can add colluding, bots, and questionable workers (Checco et al., 2019)

2.2.3 Quality control techniques and assessment

Building up on the defined model, the next step within the adopted quality control taxonomy is to define the techniques used to measure the target quality dimensions and their corresponding quality attributes selected within the quality model. The research focuses on the best result-yielding techniques in terms of the general quality of crowdsourcing operations from the literature (Eskenazi, 2013; Daniel et al., 2018; Benatallah et al. 2013; Checco et al., 2019).
Benatallah et al. (2013) categorise the said techniques into two categories: design-time and run-time quality control measures.

- Design-time: techniques that are applied before the start of task and lay the foundation for the operating of the task; they include:
  - Effective preparation of the task: by defining the requirements properly without ambiguity and clearly stating the practicalities of task delivery, expected throughput and agreed on compensation.
  - Crowd workers selection: by selecting the right people for the right task (i.e. some tasks can be open to all, some tasks require nativity while some tasks require certain expertise that need to be tested beforehand)
- Run-time: focus on the processing of the task with the two involved parties (the employer and the crowd), they include:
  - Expert review: having a certain group of people either from the crowd, internal employees, or third-party people that are experts in the task at hand and are assigned to review the submitted tasks from the crowd.
  - Agreement among the crowd: if the majority of the answers of the crowd are the same either in terms of the output or input being the same, the data is accepted as correct and of the desired quality.
  - Ground truth: gold standard data that is accepted as correct by experts and used as reference and as a measure of quality.

- Majority consensus: the overall evaluation of a crowd worker by reviewers is accepted as general quality of the worker and their work.
- Direct evaluation: crowd workers are evaluated by the employers using the defined relevant quality attributes.
- Real-time support: providing real time guidance to crowd workers by experts in the task, designated crowd members, or people from the employer side.
- Optimise workflow: improve granularity of complex tasks and incrementally and iteratively improve the workflow for these tasks.

Daniel et al. (2018) in general terms agree with the above and build up on it. However, they had the defined techniques categorised into slightly different categories with more details and more techniques for each category. They then evaluate the extensive list of techniques for the state-of-the-art crowdsourcing platforms. Consequently, the following will elicit the most used techniques and ones that are relevant for the research. The three main categories are:

- Individual: techniques that are aimed at controlling and assessing the quality of members of the crowds on an individual level. The most relevant ones are:
  - Rating: this technique relies on assigning a task-dependant rating that might involve multiple quality attributes as an employer-defined formula to determine the quality of a crowd worker. The rating formula might for example include indices of how long the worker been registered in the platform, the number of tasks they participated in, the throughput of the said worker, their peer and direct evaluations and the expert reviews the worker acquired through working with the employer.
  - Skill tests: a technique relying on task-dependant functional tests or targeted questionnaires designed to determine how suitable the worker is for a task. Tests might be for nativity, for expertise or for pace of completion.
  - Expert review: similar to the run-time expert review of Benatallah et al (2013).
  - Usability checks: similar to design-time effective preparation of the task from Benatallah et al (2013).

- Group: techniques that assess the quality of a group of crowd workers based on their collective contribution. The most relevant ones are:
  - Voting: in this technique assessment of workers is obtained by a voting system where a group of reviewers assess the output of a group of crowd members and based on the votes the suitable workers with an acceptable quality assessment are identified.
  - Output agreement: similar to the agreement among the crowd of Benatallah et al (2013).
  - Peer review: similar to the expert review from Benatallah et al. (2013) with the difference being that the review is performed by multiple peers avoiding the bias of individual review. Draws some aspects from the majority consensus.
- Computational: automated techniques that rely on the algorithms to assess and control the quality of the crowd with minimal human involvement. The most prominent techniques of this category are:
  - Ground truth: similar to Benatallah et al. (2013)
  - Achievements: a technique based on assigning badges or certificates to crowd members based on their active historic of participation in tasks and aggregate of performance evaluation. This technique targets also the incentive quality sub-dimension where it plays two roles: assessing the quality of the crowd worker for the employer, and providing customised incentive and motivation for the worker.
  - Implicit feedback: pattern recognition is used to draw patterns on the behaviour of a worker while completing a task then providing implicit feedback to improve their workflow.
  - Task execution log analysis: this technique relies also on pattern recognition to determine the general interactions of the worker with the platform, the peak times of throughputs, the rejected tasks etc. Mining and analysis such attributes help in quality assessment as well as future planning.
  - Content analysis: this technique is also used to draw patterns on the complexity of tasks, how workers interact with such tasks and to what extent their quality of work is affected by the complexity of the task.

Checco et al. (2019) and add Eskenazi (2013) add couple of computational techniques that aim to prevent attacks on the quality of crowdsourcing operations by malicious programs or malicious crowd members.

- Bot prevention: bots are programs that simulate the behaviour of crowd workers often providing a high number of wrong answers and are mostly found in crowdsourcing platforms offering monetary remuneration. The most effective techniques to control this vulnerability is reCAPTCHA application that nowadays most crowdsourcing platforms implement. (Eskenazi, 2013)
- Collusion prevention: another malicious attack that affects crowdsourcing platforms; this one is related to malicious crowd members that share information, impersonate other workers or share information to gain advantage. Statistical models, identity checkers, tracking systems are used to identify suspicious behaviour within the crowd and prevent collusion.

2.2.4 Quality Assurance

After identifying and going over the most used and best result-yielding techniques applied in crowdsourcing platforms, the next section provides a customised list of quality attributes, quality dimensions as well as quality control and assessment techniques that lay the basis for the next chapters where a survey is designed to provide a custom approach to quality assurance. The list represents the target variables for a mixed (quantitative and qualitative) research which is carry out and is detailed in the next chapter.

- The importance of quality assurance for the general population.
- Quality and quality control perceptions for each of the target groups.
  - The target quality dimensions for the general population and per target group.
  - The most used techniques by each of the target groups.
- The most important quality attributes.
- The correlation between the applied quality control techniques and the target quality dimension.

- The experience of the target group dealing with quality control in their line of work.

The above list lays ground for the hypothesis the current work is formulating to answer the two research questions by extending the evaluation of the best result-yielding quality control techniques and how they are applied. The techniques are elicited from the literature then examined within the survey. In addition, the survey includes basis for gaining insight into what can be improved in the quality assurance practices within the practical setting.

3 METHODOLOGY

3.1 Quantitative and qualitative research

The presented work tackles two research questions:

- What are the best result-yielding quality control techniques in crowdsourcing platforms?
- How are these techniques evaluated and applied?

The two questions are of different nature. The first one is a descriptive research question (Business & IP centre, 2020); while the second is a relational research question with a case study component (Mass communication theory, 2011). To answer these questions adequately, different research methods were implemented.

Research is the systematic examination of a subject to gain insight into said subject. Research can be basic: expanding knowledge on a subject; or applied: improving the practices and aspects of a subject. Basic research often lays basis to applied research.

The applied research relies on two main distinct designs, qualitative and quantitative research methods. (Sharan et al. 2015)
Quantitative research aims at examining a subject by discovering new knowledge (or consolidating existing one) based on the simplifications of the complexities of the subject (O'Dwyer et al. 2015) and on determining a cause and effect relation, predict patterns, or describe a distribution (Sharan et al. 2015). It often answers questions that start by "what", "how many", "is there a correlation" etc. (Business & IP centre, 2020). Its main advantages are: replication, direct comparison of results, hypothesis testing, and uniformity of processing of the collected data.
Qualitative research aims at examining a subject by discovering new knowledge in light of the complexities of the subject in a natural setting (O'Dwyer et al. 2015). The questions qualitative research answers have a qualitative aspect to them such as: "how to", "examine", "discover", "explore" etc. Its advantages are

flexibility, gaining insight into the target group thinking and explaining unquantifiable research attributes (Rahman, 2017).
The two designs are often complementary rather than opposite to each other (O'Dwyer et al. 2015).

It is within the precise premise that the current work lies. Once the basic research was established by reviewing the literature and establishing the quality control model and theoretical framework; the foundation was set, and the practical aspect of the research was built upon it. The nature of the research questions dictated the choice of the practical research methods to use. To extend on the basic research established with the literature review, quantitative research was carried out to complete the answer for the first research question. In conjunction, the practical research was built upon using relational research to tackle the second research question; which was then consolidated with the qualitative research.

The quantitative research was carried out using a survey specially designed to support the formulated hypothesis. The hypothesis on a broad picture examines the importance of quality control in the practical setting and the correlation of the application of the best result-yielding quality control techniques and the quality of the product.

The qualitative research aspects were covered in a couple of questions within the survey where the target groups were asked to explain something in light of their experience. These questions are based on open-ended feedback from the participants where they share their narrative and experience regarding the subject (Carless & Douglas 2017); which is then analysed using content analysis.

3.2 Survey design

As stated above, the survey was designed to consolidate the findings of the literature review, examine the hypothesis and gain insight into the perceptions of the different target groups on the examined quality model.

3.2.1 Sampling

The general population that the survey targeted were employees within the company who are involved in crowdsourcing projects.

The target groups were:

- Project managers: managers within the company who deal with crowdsourcing projects
- Sourcing team: a team of employees taking care of sourcing needs for different projects.
- Production team: project coordinators, developers and production leads who drive the execution of the projects and the delivery of the final product.

3.2.2 Data collection method

To collect the data for the survey, a questionnaire was implemented in Microsoft orms with different questions targeting the list covered in section 2.2.4.
The survey was anonymous and was sent in separately to the different target groups. It was only accessible to people within the organisation with an emphasis on anonymity and abstinence of any personal details tracking.
At the beginning of the questionnaire, an overview of the survey's aim and practicalities was presented to the participants along with a disclaimer that the answers should not include any sensitive information or details on projects, clients or the concerned people.
The results of the survey are accessible only to the author and are used only within the context of this research (due to the nature of the research and the nature of the company)

3.2.3 Error reduction methods

To avoid the known errors associated with surveys, careful design and selection was applied before the survey was sent. To avoid the Nonresponse error (lack of response from all individuals) the survey was designed in a user-friendly way with clear and concise questions. To avoid the sampling error (individuals not

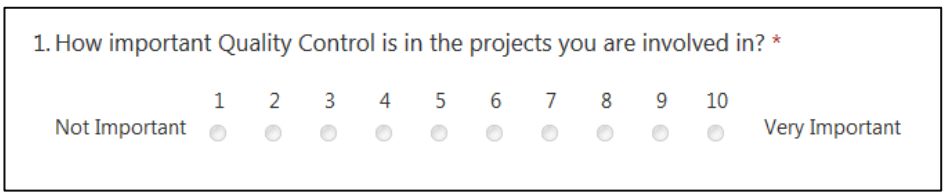
representing the target population), the survey was sent to selected individuals with prior experience in crowdsourcing projects, and the irrelevant answers were discarded. As for the measurement error (questions not targeting the topic of interest), the survey was designed carefully in light of the target quality dimensions, attributes, and techniques and was reviewed by the supervisor (from work side) before it was sent out; also branching was implementing where certain sentinel questions gave access to different branches depending on the answer. (Ponto, 2015)

3.2.4 Survey questions and evaluation

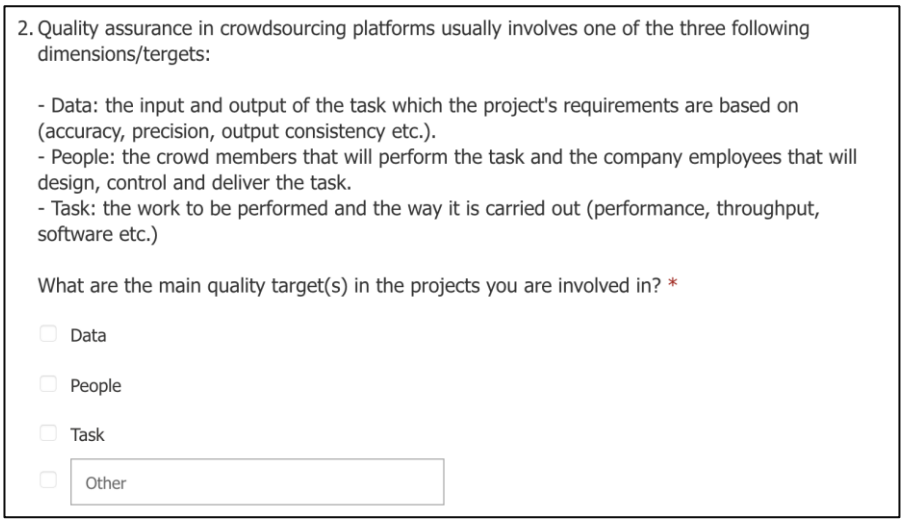
For the questions that used rating ranges, the choice of a 10 point Likert scale (from 1 to 10, Appendix 1) was due to the need to capture all different perceptions the respondent might have on the aspect being examined, the familiarity of the survey population with the aspects being examined, and ease of use of such a scale for the respondents. This scale also improves the variance of the answers to widen the range for the analysis and allocate for any answers that might provide variant rating which can be useful during the analysis. (Research gate, 2014)

Given the size of the target population and the number of qualitative questions, the evaluation of the survey was done using MS excel and the author's own analysis for the open-ended qualitative answers.

The following are some of main questions that were sent within the survey (not all questions are presented here for the conciseness of the report, all the questions will be presented in the results chapter).



➔ Numeric quantitative evaluation of the importance of quality control within the survey population and across target groups.

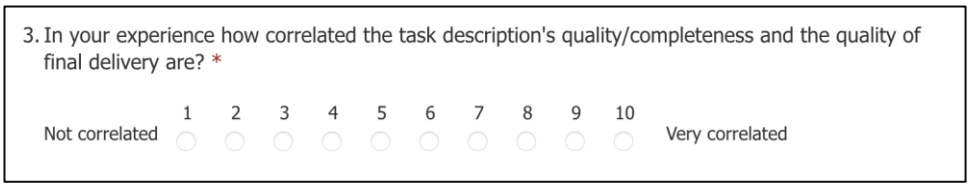

2. Quality assurance in crowdsourcing platforms usually involves one of the three following dimensions/tergets:

- Data: the input and output of the task which the project's requirements are based on (accuracy, precision, output consistency etc.).
- People: the crowd members that will perform the task and the company employees that will design, control and deliver the task.
- Task: the work to be performed and the way it is carried out (performance, throughput, software etc.)

What are the main quality target(s) in the projects you are involved in? *

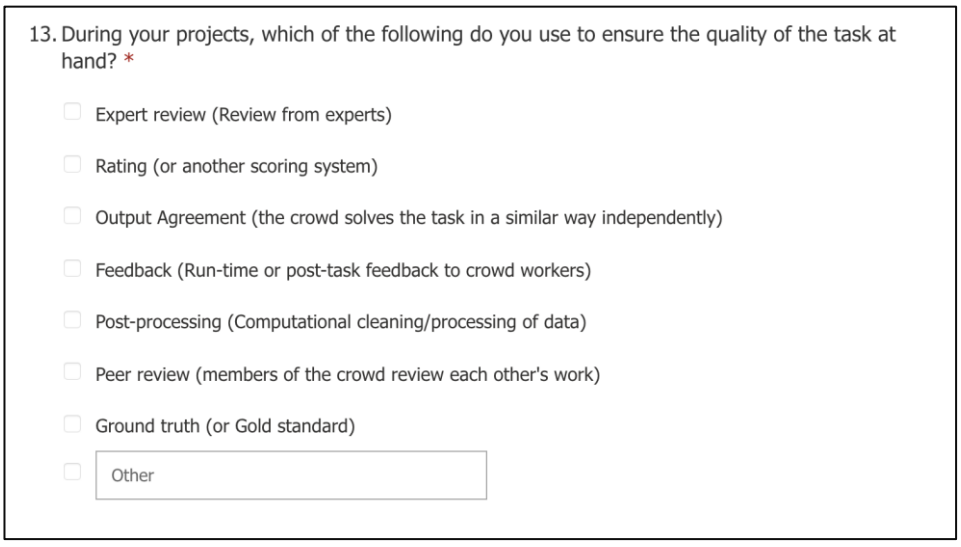
- [ ] Data
- [ ] People
- [ ] Task
- [ ] Other

➔ Numeric quantitative evaluation of the quality model dimensions within the survey population.


3. In your experience how correlated the task description's quality/completeness and the quality of final delivery are? *

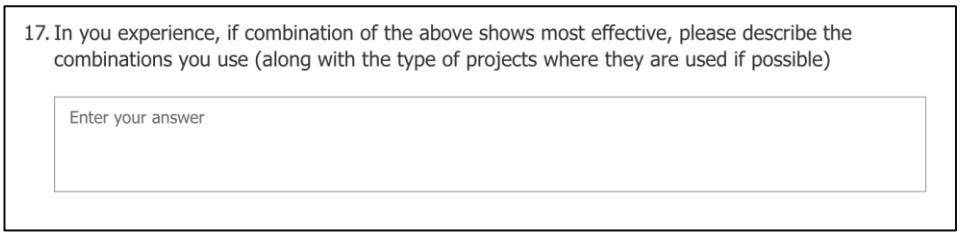
Not correlated 1 2 3 4 5 6 7 8 9 10 Very correlated

➔ Numeric quantitative evaluation of the task description quality attribute within the task quality dimension.

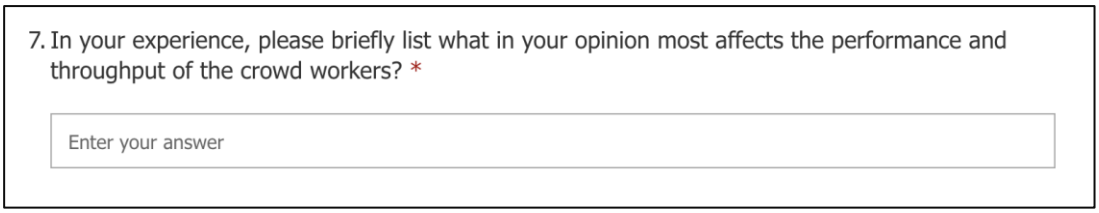

13. During your projects, which of the following do you use to ensure the quality of the task at hand? *

- [ ] Expert review (Review from experts)
- [ ] Rating (or another scoring system)
- [ ] Output Agreement (the crowd solves the task in a similar way independently)
- [ ] Feedback (Run-time or post-task feedback to crowd workers)
- [ ] Post-processing (Computational cleaning/processing of data)
- [ ] Peer review (members of the crowd review each other's work)
- [ ] Ground truth (or Gold standard)
- [ ] Other

➔ Numeric quantitative evaluation of the used quality control techniques used by the survey population and within the different target groups.



➔ Qualitative examination to gain insight into how the techniques are used within the actual work settings and when such combinations are used.



➔ Qualitative examination to gain insight into how a specific quality attributes (performance and throughput) is affected within the actual work settings.

3.2.5 Data analysis methods

Given the different research designs used, different methods were used to analyse the results.

For the quantitative research, statistical analysis was applied on the numeric quantitative data. Depending on the variable, an appropriate analysis method was used. For the descriptive questions, simple statistical analysis was used in the form of averages, percentages, ratings, and rankings. Simple statistical analysis was used to analyse the results in terms of the descriptive research aspect. As for the relational aspect of the research, the following are the methods used to analyse the data (Woodley, 2004; Statistics how to, 2020):

- Correlation analysis: used to examine correlations between variables. In this context correlation analysis was used to examine correlations (associations) between quality dimensions, quality attributes and used techniques.

- Cross-tabulation: used examine effects variables in terms of other variables. Used to examine relevance of certain quality attributes, dimensions and used techniques per target group.

For the qualitative research, the main relevant analysis method used to examine the qualitative open-ended question was text analysis (Question Pro, 2020). This was due to the number of respondents and the aim of the qualitative questions. The text analysis is a form of content analysis that is based on examining textual data in light of predefined themes or emerging themes (open coding). The data is coded based on the defined framework then analysed and presented. (Kuckartz 2019; Leavy, 2015; Lacity, 2015).

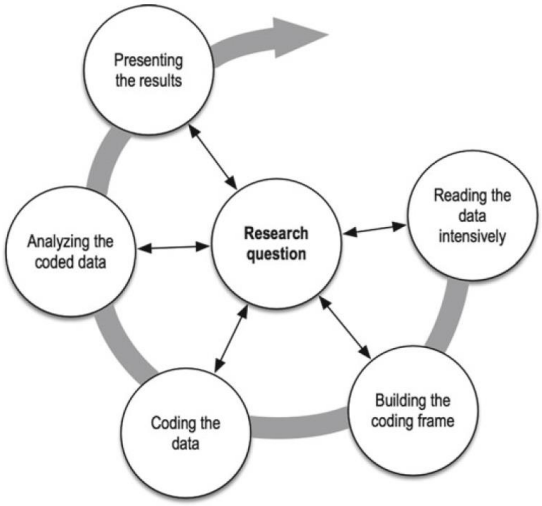


FIGURE 4. Stages of qualitative content analysis (Kuckartz 2019, 186)

Figure 4 simplifies the stages of qualitative analysis. For the couple of questions this research aimed at analysing qualitatively, the phases above were followed. In addition, dictated by the theoretical quality model; the themes were predefined based on the important quality dimensions and quality attributes that the open-ended questions are examining. In addition, emerging themes form analysing the answers were added and analysed as well.

4 RESULTS AND ANALYSIS

4.1 Descriptive quantitative results

The survey was taken by around 30 people. This was the intended target population which was dictated by the size of the local Tampere branch and the nature of the research targeting crowdsourcing projects and employees who are involved in this paradigm.

The suitability of the participants to the task was favoured to the number of participants in order to get relevant and targeted answers that are useful to the research. Some of the answers were omitted due to irrelevance. The following are the quantitative results of the research along with analysis of each question and its answers.

TABLE 1. Importance of Quality Control

	Rating value	%	
How important Quality Control is in the projects you are involved in?	10	58.06%	9.42 (mean)
	9	25.81%	
	8	16.13%	

Table 1 presents the results for a question that was aimed at identifying the importance of quality assurance to target population. The question was based on a rating from 1 (not important) to 10 (very important). The answers ranged from 8 to 10 with the vast majority rating quality assurance as very important in their line of work. This is expected as quality assurance is a defining characteristic of the company and focal point in its projects. With around 58% of answers rating the importance the highest possible, results from this question establish the premise that the survey population is aware of the importance of quality assurance and recognise the role it plays in the projects they are involved in.

TABLE 2. Quality dimensions

	Dimensions	%
What are the main quality target(s) in the projects you are involved in?	Data People	29.03%
	Data People Task	29.03%
	People	12.90%
	Data	9.68%
	People Task	9.68%
	Data Task	3.23%
	Process People Task	3.23%
	Task	3.23%

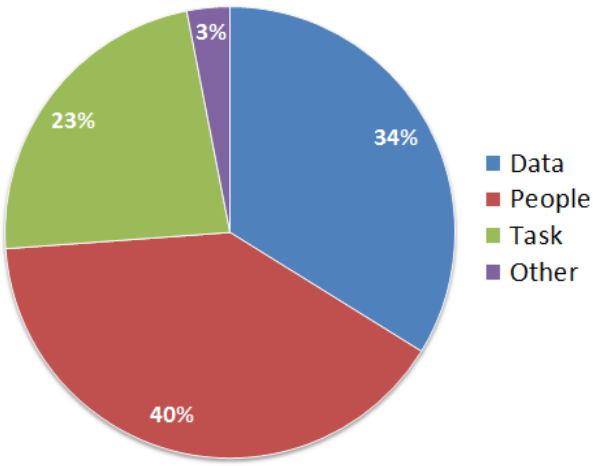


FIGURE 5. Quality dimensions

Table 2 presents the results of the question that aimed at examining the quality dimensions the survey population targets in their projects. The results showed that the majority of the population targets a set of quality dimensions in their projects with 74.2% of respondents targeting two or more quality dimensions that affect quality in their line of work. The main combinations were [Data, people, and task] and [Data and people], with the latter being embedded in the first. This shows that to assure quality, the survey population targets multiple quality dimensions. The question was designed to examine the defined quality model components. The results show that the survey population is aware that quality assurance relies on a quality model that is based on the defined quality dimensions. Figure 5 illustrates the distribution of the dimensions, where the most targeted dimensions were data and people.

TABLE 3. Data quality attributes

	Data quality attributes	%
Which of the following is (are) the most important aspect (s) of data in the projects you are involved in?	Complete Input Accurate Output Consistent Output Integral Input	48.39%
	Accurate Output Consistent Output	22.58%
	Accurate Output	12.90%
	Complete Input	6.45%
	Complete Input Accurate Output Integral Input	3.23%
	Consistent Output	3.23%
	Integral Input	3.23%

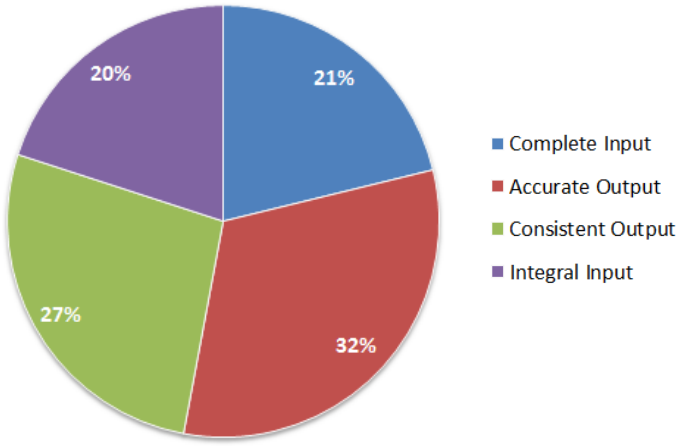


FIGURE 6. Data quality attributes distribution

Table 3 presents the results of the survey question that aimed at evaluating the familiarity and the most important quality attributes of the data quality dimension. The question was a multiple selection question with 4 options representing the defined quality attributes laying basis to the data quality dimension of the quality model set in the research theoretical framework. 48.4% of the population did chose all the quality attributes which shows a certain level of familiarity with what contributes to quality in data. However, some of the population went for a more straightforward approach where output accuracy was the main concern for them, with around 40% of the answers stated that output (either accuracy, consistency or both) is the main attribute that affects quality of data.

The distribution of the data quality attributes (figure 6), also consolidates this observation where most of the answers stated output accuracy as the most

important aspect (quality attribute) of data quality. Based on the theoretical framework, output accuracy and output consistency are closely related; which, brings the majority of the answers to the data quality question to be related mainly to the data output.

TABLE 4. Task description

	Rating value	%	
In your experience how correlated the task description's quality/completeness and the quality of final delivery are?	10	35.48%	8.58 (mean)
	9	29.03%	
	8	16.13%	
	7	6.45%	
	6	6.45%	
	5	3.23%	
	4	3.23%	

Table 4 presents the result of the question that aimed at evaluating the importance of task description for the survey population. Task description is a quality attribute of the task quality dimension from the quality model. The question was set up with a rating from 1 (not correlated) to 10 (very correlated). Around 80% of the answers ranked the correlation within the 4th quartile (close to the max). The results show that task description and the quality of the task are correlated but not strongly (8.58 mean).

TABLE 5. Software and user Interface quality

	Rating value	%	
How would you rate the importance of quality, user friendliness and performance of the software to the quality of the final delivery?	10	22.58%	8.19 (mean)
	9	25.81%	
	8	22.58%	
	7	16.13%	
	6	3.23%	
	5	9.68%	

Continuing on the task quality dimension, the next quality sub dimension to examine was the software quality importance and its significance to the survey population, presented in table 5. Similar to the question represented in table 4, the majority rating of the importance of the software and user interface performance fell within the 4th quartile (70%). However, compared to the task description, the quality of software was somewhat less important. The result

show that the survey population realises the role software plays in insuring task quality; but showed reserve in to how decisive software quality is to the quality of the task.

TABLE 6. Granularity

	Rating value	%	
How simple/complex are the tasks of the projects you are involved in?	10	22.58%	6.84 (mean)
	9	9.68%	
	8	22.58%	
	7	3.23%	
	6	16.13%	
	5	3.23%	

Table 6 presents the answers to the question that aimed at examining the granularity of the tasks the survey population is involved in. Granularity is a quality sub dimension within the task quality dimension. The rating of the question was from 1 (simple) to 10 (very complex). The reasoning behind the question was to examine how complex the tasks in the projects the survey population is involved are; and reflect on the premise of simplifying the tasks as a mean of improving quality. The results show that the a considerable portion of respondents (55%) qualified the task complexity in the projects they are involved in within the 4th quartile. 22.58% of the respondents ranked the complexity of the tasks as maximum. The results show that task complexity is an important issue that face the survey population. Hence, affect the quality of the task, since, as covered in the theoretical framework the more complex the tasks are in crowdsourcing project the more the quality is affected. The results also showed an extended variance in the granularity; which can be explained by the nature and variety of projects the survey population is involved in.

TABLE 7. Incentive and compensation

	Rating value	%	
How important are incentive and compensation to retain crowd workers for your projects and the quality of the final product?	10	35.48%	8.35 (mean)
	9	19.35%	
	8	22.58%	
	7	6.45%	
	6	6.45%	
	5	6.45%	

The question table 7 presents the results to aimed at examining the importance of incentive and compensation to retain crowd workers. Incentive and compensation is a quality attribute within the task quality dimension. The majority of the population recognised the importance of the incentive and compensation to retain workers and contribute to the quality of the task. 83% of the answers fell within the 4th quartile. The most picked rating was the maximum within the rating with more than 35% of the respondent asserting that incentive and compensation are very important factors to retain good workers and ultimately improve and maintain the quality of the task.

TABLE 8. People quality

	Rating value	%	
How correlated is the quality of people involved in the task (crowd workers and company employees) and the quality of the final product?	10	48.39%	9.19 (mean)
	9	32.26%	
	8	9.68%	
	7	9.68%	

Table 8 moves on to present the results of a question aimed at the next quality dimension, people. In the people quality dimension the quality of crowd workers and employees of the company are the relevant quality attributes. To this end, the question asked the respondent to determine how correlated the quality of these attributes is to the quality in their projects. The vast majority of the answers fell within the 4th quartile of the rating interval, with 90% of the answers. The lowest pick within the interval was the 7 rating. The results show that the population has a considerable belief in the quality of people who are involved in their project. The results also show that the quality of people is an important focus for the survey population.

The next question aimed at exploring this importance and what drives it. Based on the theoretical framework and the experience the author gained from involvement in crowdsourcing projects, the most used techniques to insure the quality of crowd workers were examined within the survey.

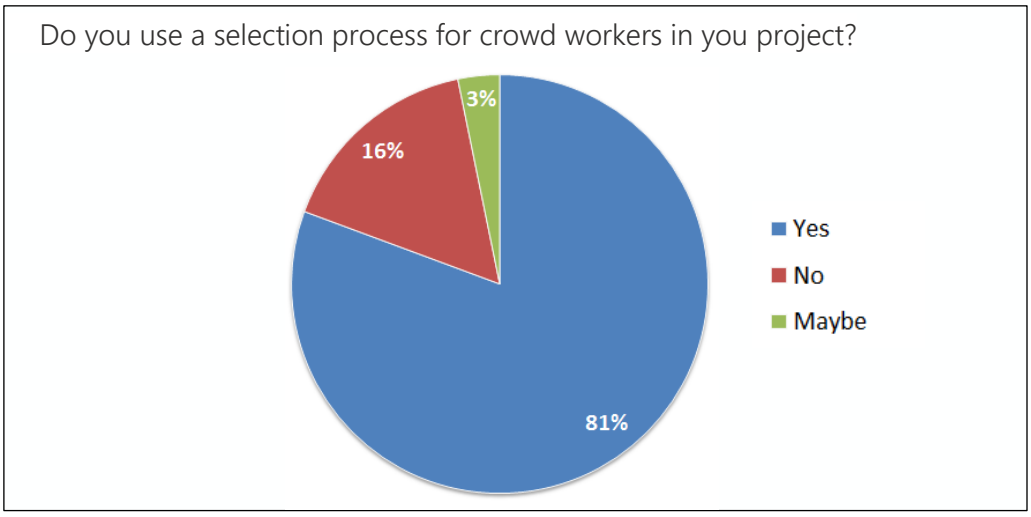


FIGURE 7. Use of crowd workers selection

Figure 7 presents the results to the question aiming at examining the use of crowd workers selection by the survey population. Crowd worker selection is a quality control technique used in design-time of projects. The results show that a considerable portion of the respondent (81%) uses a selection process in the projects they are involved in. Consequently, crowd workers selection is an essential factor to consider during the pre-production stage of crowdsourcing projects.

The next question was a branch question based on the answer the respondents picked. For the respondent that stated that they use a selection process, the next step was to examine the importance of the process to their projects and what it is based on.

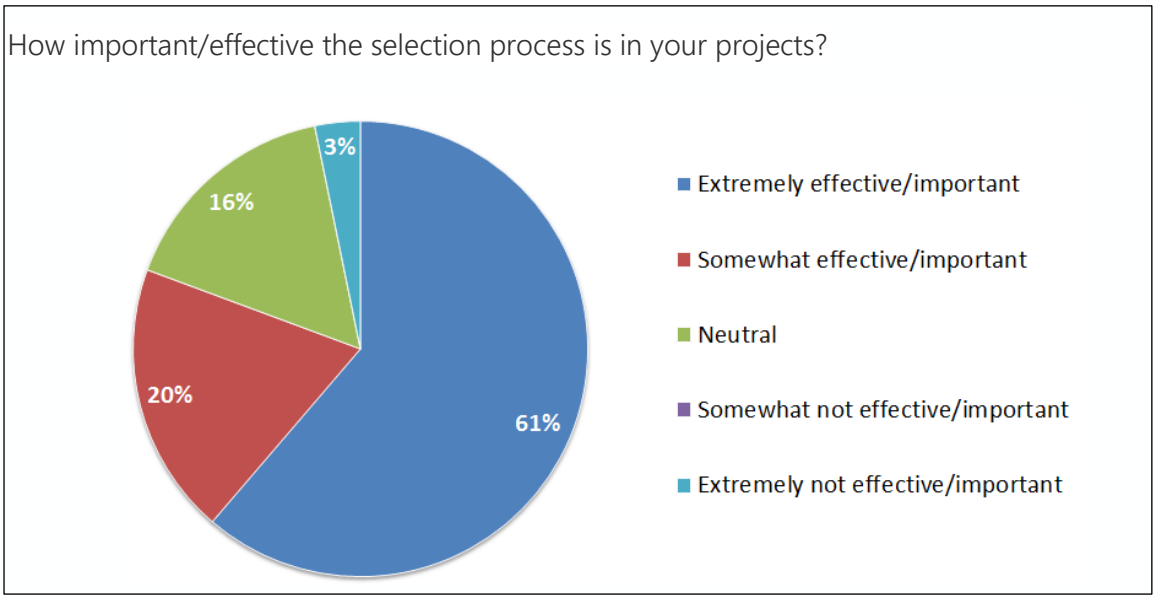


FIGURE 8. Effectiveness and importance of selection

Figure 8 presents the results the branch question for the respondents that said that they use a selection process. The question aimed as evaluating how the survey population perceives the effectiveness and importance of the selection process. 81% of the answers stated that the selection process was extremely to somewhat effective and important. The results show that for most of the respondent the selection process is vital within their projects. However, a considerable portion of the population showed reserve on the importance and effectiveness of the selection process.

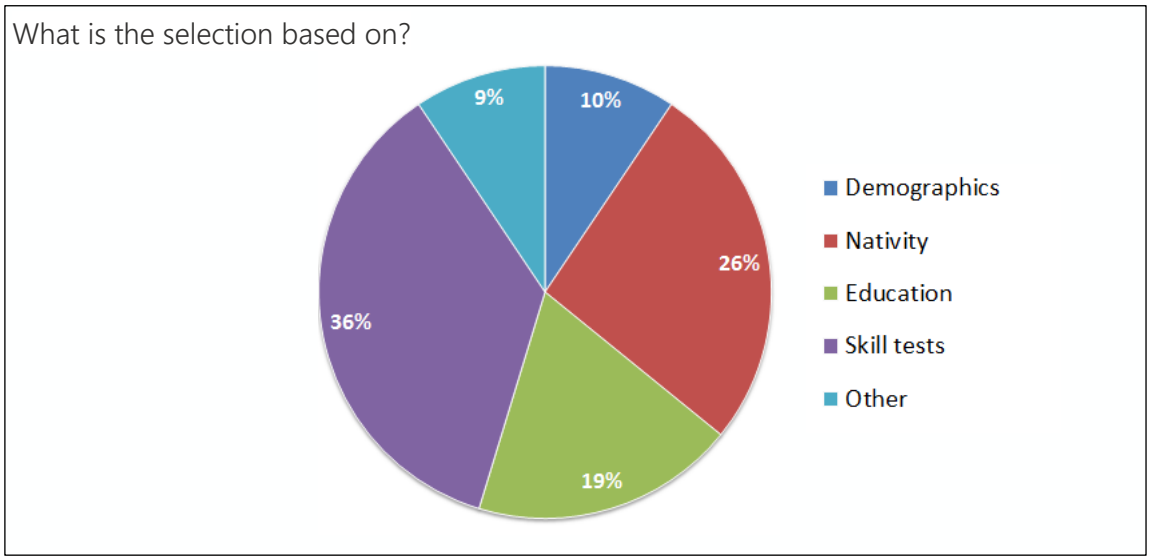


FIGURE 9. Selection criteria

Figure 9 shows the selection criteria that the respondents use. The criteria examination aims at identifying the main aspect that the survey population targets for the worker selection in their projects; which lays ground to further exploration and lays the foundation for the customisation of this crowd workers selection quality control technique. The results show a scattered distribution without clear trends; which is expected due the nature of the projects the survey population is involved in and their sub groups. Skill tests were; however, the most selected criterion with 36%.

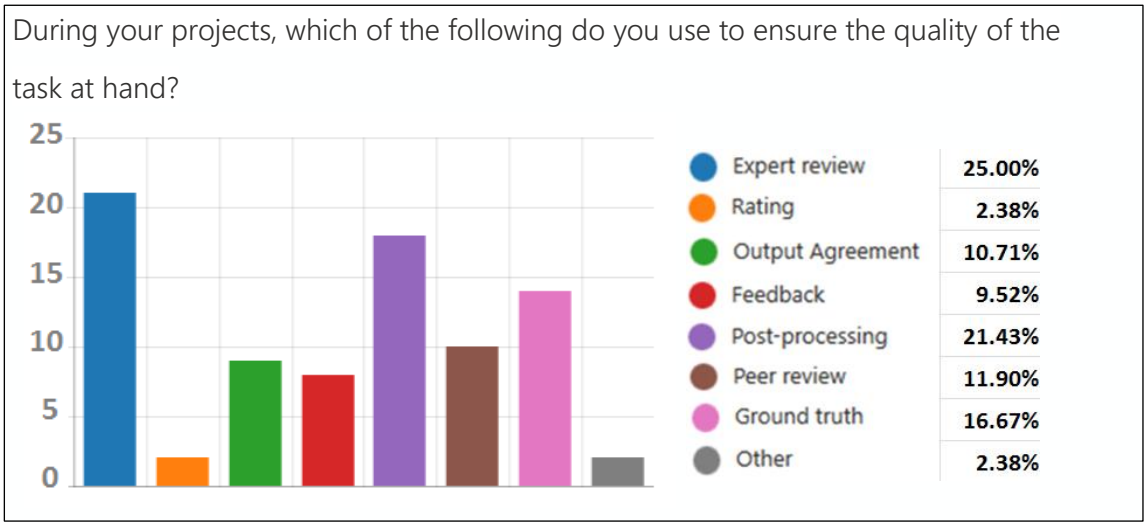


FIGURE 10. Quality control techniques

Moving on to evaluating other quality control techniques, figure 10 presents the results for the survey question that aimed at evaluating the most used techniques by the survey population. The results show that the most used quality control techniques were the expert review with 25% and post-processing with 21%, and to a lesser extent ground truth 16%, peer review 11% and output agreement with around 10%. The results show that the survey population uses a variety of techniques in the projects they are involved in, this in itself does not tell much but lays ground to further analysis and exploration that the qualitative analysis will cover.

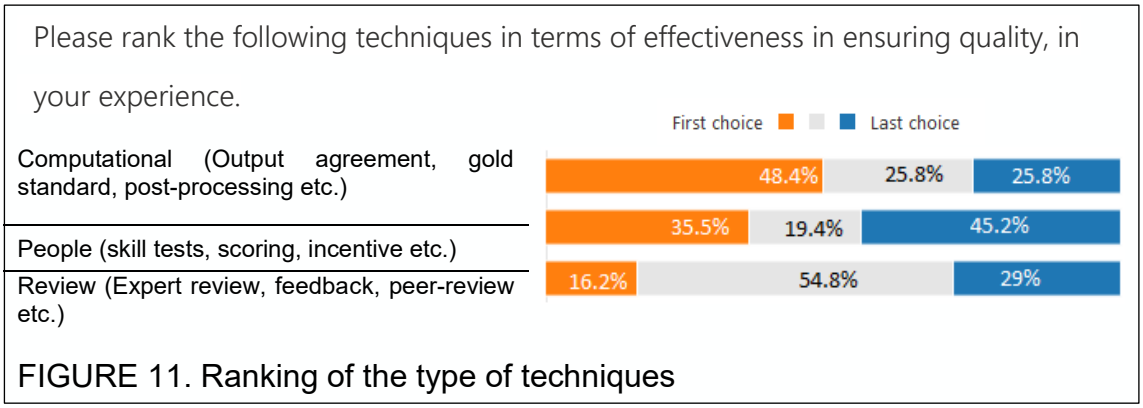


FIGURE 11. Ranking of the type of techniques

Figure 11 shows the results of a ranking the survey population was asked to perform to determine which type of quality control techniques is most effective. The quality control techniques were grouped and explained in terms of the quality attribute they target and the general aspect they represent. The aim for this question is to examine which paradigm of quality control techniques is most

relevant for the survey population, and to lay ground to further analysis about the customisation of these techniques. The results showed that the computational quality control techniques are considered the most effective by the survey population (48% 1st ranking). However, the other two paradigms are also quite prominent. People quality control techniques were second in terms of first ranking, 35% of the respondent ranked these techniques in the first ranking. Review based quality control techniques were ranked second by 55% of respondent, showing the place of these techniques as supporting to the other types of methods.

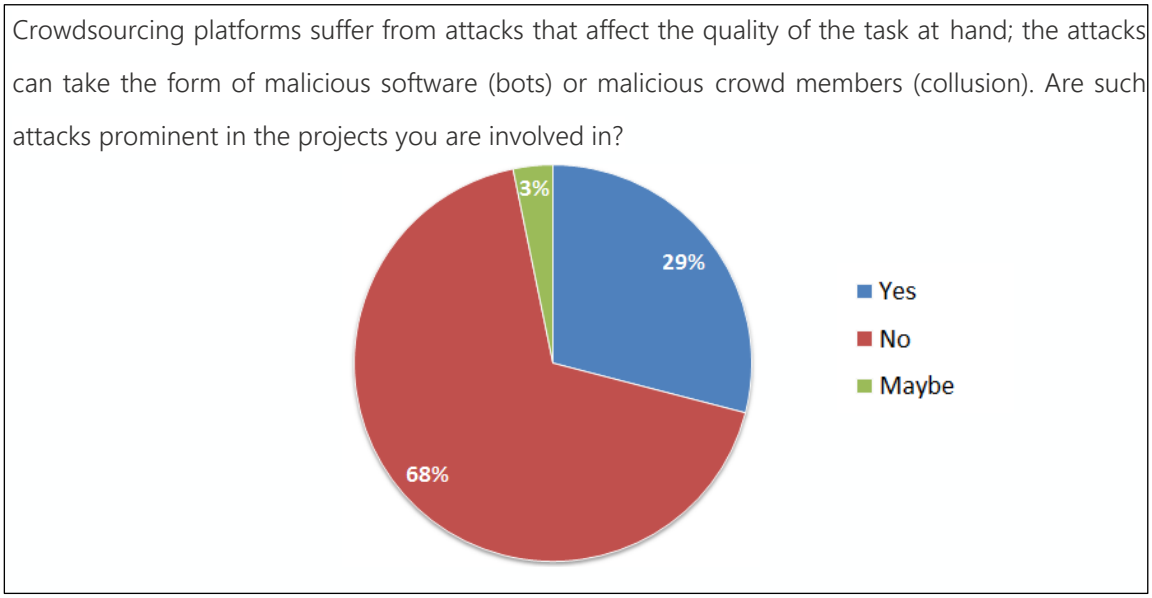


FIGURE 12. Crowdsourcing quality attacks

The question for which the results are shown in figure 12 was aimed at examining how affected the survey population is by the malicious quality attacks. Malicious quality attacks on crowdsourcing platforms are an emerging trend so it was particularly intriguing to determine if the issue was prominent for the survey population. The results showed that a relatively considerable portion of the respondent is aware of such attacks and that these attacks are prominent within their projects. However, the majority of the respondent stated that malicious quality attacks are not prominent within their project. This lowered the subset of respondent who answered a couple of questions that aimed at elaborating on this trend. Depending on the answer of this question the questionnaire branched to two other questions elaborating on the crowdsourcing quality attacks. The subset of respondents was significantly less

and their answers were distributed evenly, not offering much insight into the examined aspects. The first question was about the type of attacks (bots/bot like software or collusion) the answers were “50/50” for both types. The second question was checking if the attacks are handled or not and the majority of respondent stated that these attacks are handled. This portion of the survey did not give much insight into the examined aspect; hence, it is omitted from further analysis.

4.2 Relational results

To cover the relational aspect of the research questions, the relevant quantitative data from the above section is further analysed using cross tabulation and correlation analysis. The results further analysed are determined by the quality model and quality assurance framework defined in chapter 2.

To determine how each target group perceived quality control, the first aspect to examine was the quality dimensions that set the quality model and what are the target dimensions per target group of the survey population. Using cross tabulation, taking the responses for the question “What are the main quality target(s) in the projects you are involved in?” against the target groups of the survey population the following results were obtained.

TABLE 9. Quality dimensions by target group

Project Management		Production		Sourcing	
Data People	30.00%	Data People	23.08%	Data People	62.50%
Data People Task	30.00%	Data People Task	38.46%	Data People Task	12.50%
People Task	20.00%	People	30.77%	People Task	25.00%
Data Task	10.00%	Task	7.69%		
Data	10.00%				

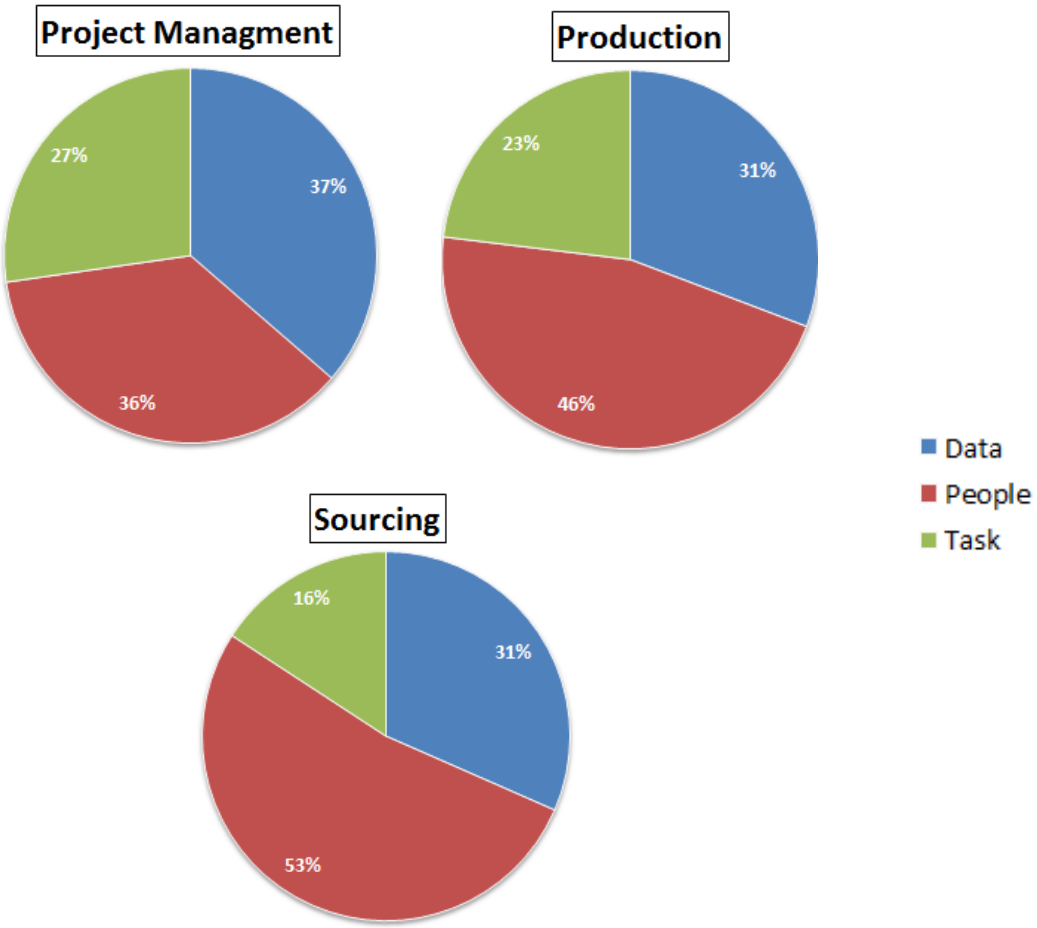


FIGURE 13. Quality dimensions distribution by target group

From table 9, we observe that the different target groups of the survey agreed predominantly on picking combinations of quality dimensions as targets in their line of work. [Data and People] and [Data, People, and Task] combinations were picked by the target groups; consolidating the results of the descriptive analysis of the same question. However, looking into the results by target groups, noticeable differences emerged.

- The project managers went for a 90% of answers picking some sort of combination of dimensions. This 90% was distributed mainly between two combinations of [Data and People] and [People and Task]; with the combination of the all three dimensions also included in the majority. The results for this group show that Project managers seems to be concerned with the different quality dimensions associated with quality control, with differences in which ones are more important to them. This can be explained by the nature of their line of work where they need to cover all, or the most important, aspects of quality in their projects depending on the type of project. 10% of the project manager picked data as the main quality

target for their project; this was not surprising given the nature of AI projects that rely heavily on data.

- The production team went also for a majority of answers stating a combination of dimensions as their quality targets. Similar to the project managers the production team went for a majority of [People and Data] and all dimensions combinations (61.5%). However, 30.7% of respondent choose [People] as the main quality target in their project. This was a surprising result and one that was not as clearly reflected in the descriptive results of the general population. The interpretation of this result can be drawn again from the nature of the production team work. Given that a large portion of their work revolves around coordination, the characteristics of people involved in the project holds a significant importance; the better the people you are coordinating the better your work is.
- The sourcing team followed the same trend of the descriptive analysis of the aspect at hand. The majority went for the same combinations as the two other target groups. However, this group went entirely for combinations of dimensions unlike the other two groups who had portions picking only one dimension as their target. 62.5% of the respondents of this group picked the combination of [Data and people] and 25% for the combination of [Data and task]. The interpretation of these peculiar results can be explained by the nature of the sourcing activity. Sourcing relies on a sourcing basis which can be either the data goals of the project or the task goals of project. For the data, an example of the [Data and People] combination can be that the sourcing request is based on output accuracy based on the input which is reflected on the quality of the desired people for the project (Benatallah et al., 2013). For the [People and Task] combination, the example would be that people required for the project need to meet a certain throughput threshold that determines their suitability for the task (Benatallah et al., 2013)

The relational analysis of the above aspect tackled (the quality dimensions perception per target group) and showed more details and more insight. All groups are aware of the importance of the different dimensions and their contribution to the quality of the operations. However, due the nature of the work focus of each group they favoured some dimensions to others.

The next aspect tackled was determining what the most used techniques were for each of the target groups. Cross-tabulation was again used to determine the distribution of the selected the quality control techniques by each group. Figure 14 presents the findings.

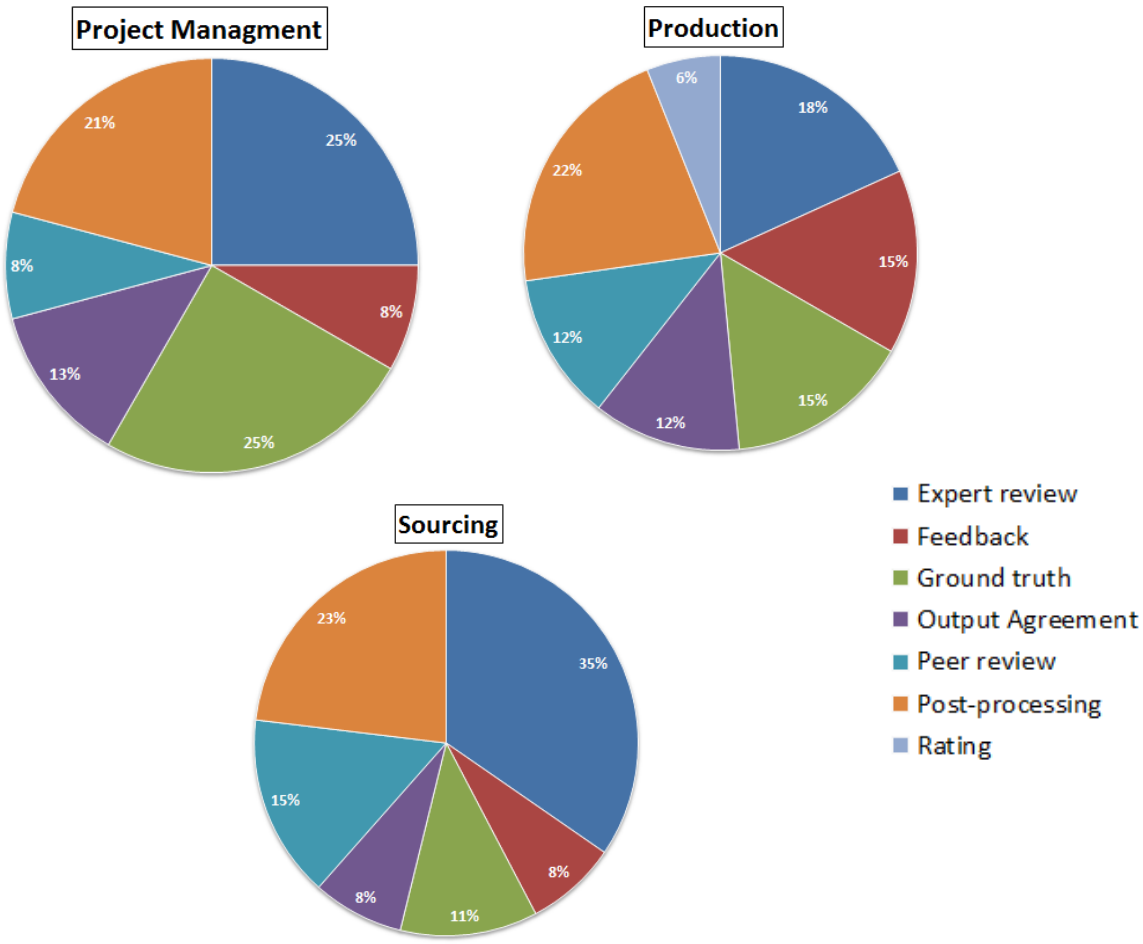


FIGURE 14. Quality control techniques per target group

The results show a varied distribution for each of groups. The patterns observed from the descriptive analysis for the same aspect for the general population still held, with the most used techniques being expert review, post-processing, and ground truth. However, the relational analysis showcased in figure 14 shows slight variance in the distribution per group. The project managers mostly picked expert review, ground truth and post-processing; the production team added to that feedback and introduced rating; while the sourcing team favoured expert review over the other techniques. This can be interpreted by the observation that the used techniques are circumstantial and situational. The teams use different techniques for different projects and at different stages of the project.

The next aspect the relational analysis targets is the correlation between quality attributes and the used techniques. To this end, and to focus the analysis, the correlations between the most selected best result-yielding quality control techniques and the corresponding quality attributes (quality dimensions) are examined.

Since the two variables are not increasing or decreasing and they are not numerical, data was simplified to binary values. The first variable represented the examined quality dimension or quality attribute and had binary values of 0 and 1 representing the choice (either picked or not) of the dimension or attribute by the respondent. Similarly, the selection of the quality techniques is represented by 0 or 1 for each of the respondent.

The most suitable technique to analyse such data was to use cross tabulation to generate a contingency table representing the variables (Woodley, 2004; Medium, 2018). To determine the strength of the association between the two variables (Medium, 2018), a Phi co-efficient was calculated from the contingency table with an online Phi calculator (Statology, 2020). The following are the results.

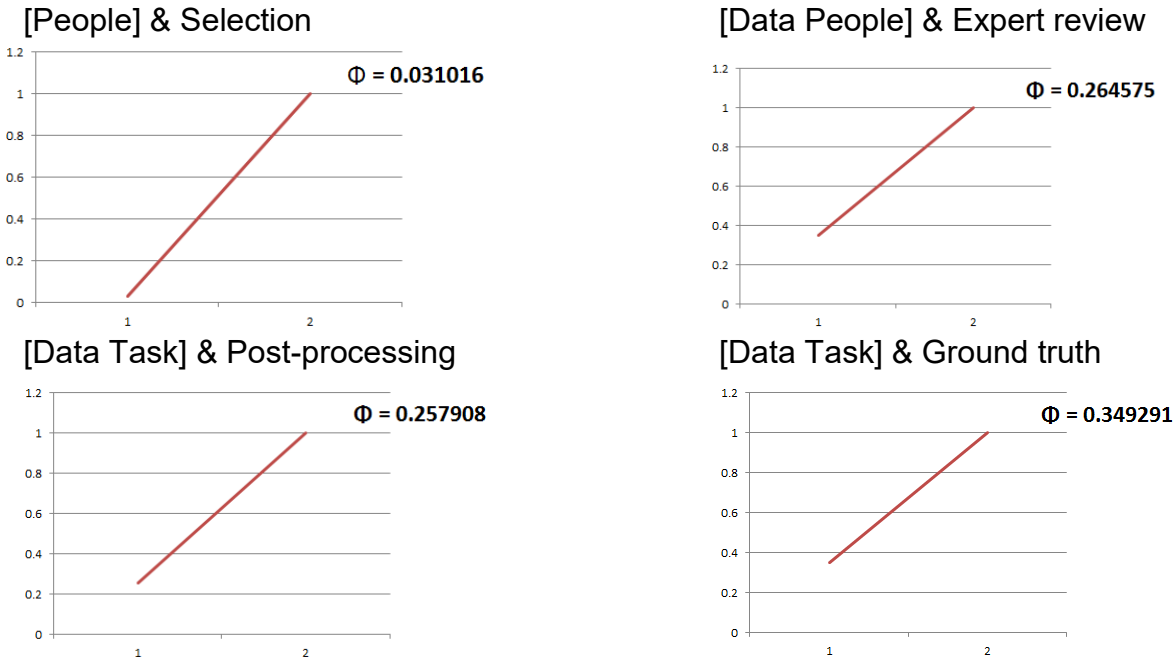


FIGURE 15. Correlation between quality control techniques and dimensions

Going over the results presented in figure 15, the first general impression is that there is a positive relationship between the quality dimensions and attributes and the used quality control techniques. All the examined techniques and quality attributes and dimensions showed a positive Phi co-efficient. The values, however, differed from examination to examination. The values of the Phi co-efficient are in a range between +1 (perfect positive relation) and -1(perfect negative relation) the values are further distributed on the range to draw interpretations (Statistics how to, 2020). The values obtained and showcased in figure 15, fell within three of the defined ranges (Full range in Appendix 2):

- [+.01 to +.19] -> no or negligible relation
- [+.20 to +.29] -> weak positive relationship
- [+.30 to +.39] -> Moderate positive relationship

The use of the selection process in correlation with the people quality dimension and attribute had a co-efficient of 0.03 indicating no or negligible relation. This can be interpreted by the fact that the selection process was a special case among the quality control techniques examined in the survey. Selection of crowd workers is one of the most used techniques; the majority of project use some sort of a selection process in one form or the other (81% from the descriptive analysis stated that they use a selection process). Hence, the association to a certain quality dimension produced a week relation.

The rest of the examinations fell within the weak to moderate positive relationships, consolidating the observation of the existence of a relationship between the quality dimensions and the used quality control techniques.

The high level aim of this analysis was to establish how associated are the line of work of the survey population and what they use to ensure quality. The quality dimensions and target quality attributes are influenced by the nature of the work and the type of projects, and now with perceived positive relation it seems that the choice of techniques is also dictated by the nature of the work and the type of projects.

4.3 Qualitative results

The survey included a couple of open-ended question that aimed at gaining more insight into the experience of the survey population with quality control.
The questions targeted different aspects of the quality model defined in the theoretical framework. The aim was at gaining more insight into how the respondents perceive these components, and what was their experience with such. In addition, the questions targeted the aspects that are not quantifiable and that were shown to be very important to quality control from the literature.

To conduct a qualitative research on the open-ended question, and as explained in section 3.2.5 a content analysis was carried out.
The themes were pre-defined from the literature review. The following is the list of themes that were used to code the textual data:

- Preparation
- Definition
- Selection
- Feedback
- Computational
- Software
- Review
- Compensation and Incentive
- Combination

After cleaning the data of noise and incomplete or irrelevant answers, the answers were coded with the above themes. This was achieved by author own reading of the answer, which was only possible due to the number of respondents; otherwise, more advanced textual analysis tools would have been needed. After reading each of the answers the corresponding theme or themes were marked in the corresponding column. Columns were added with emerging themes from the answers and marked as well. (Racer, 2017)
Ultimately, the above coding led to a better understanding of the examined aspects as well as to quantitative data that can be visualised and analysed statistically. The following are the results.

TABLE 10. Textual analysis for qualitative questions

Theme	%
Selection	39.02%
Preparation	36.59%
Feedback	34.15%
Definition	34.15%
Computational	31.71%
Compensation and incentive	29.27%
Review	26.83%
Software and tool	26.83%
Combination	19.51%
Motivation	17.07%
Effective communication	17.07%
Community management	14.63%
Team	14.63%
Training	12.20%
Customer involvement	7.32%
Complexity	7.32%
Amount of work	4.88%
Future opportunities	2.44%

Emerging trends

Table 10 represents the results of the coding applied on the answers of the respondents of the open-ended questions of the survey. The results are quantified and percentages are computed to give insight into what most of the answers referred to. The results consolidated the findings of the basic research and the quantitative results. The most used and the best results-yielding techniques and targeted attributes are significantly represented in the narratives from the survey population's experiences. The qualitative answers allowed the respondents also to emphasise the importance of combinations of quality attributes and quality control techniques to achieve and maintain overall quality.

The most intriguing part of the results was the emerging themes from the analysis of the qualitative question. A considerable portion of the answers brought more insight into quality components that are not covered or partially covered in the examined quality model; components originating from the experience of the survey population and the real world situation that they deal with within their line of work. The emerging themes were kept record of and were colour coded in table 10. Themes such as the motivation of the crowd workers, training the workers and the team, and effective communication are considerable additions to the outcomes of the research. Above all, community

management was mentioned by a couple of respondent using different terms. Community management includes all the aspects that this research is examining. It is the ultimate outcome of such study, and what this work is referring to as the quality assurance strategy. The mention of this within the open-ended answers did not come as a surprise knowing the experience of the survey population and their familiarity with the quality control in crowdsourcing platforms.

The qualitative results consolidated the findings from the descriptive quantitative and the relational results. It showed the importance of quality assurance within the practical setting of the presented work, the circumstantial and situational nature of the quality attributes and quality control techniques selection, and confirmed the familiarity with the best result-yielding techniques. Moreover, and most importantly, the analysis of the qualitative brought more insight into the real world application of quality control and confirmed the need for an integrated quality assurance strategy.

5 OUTCOMES AND SUGGESTIONS

The results of the quantitative and the qualitative research provided insight and gained new knowledge on the aspects listed as the basis of the hypothesis of the presented work. The literature review covered the basic aspect of the research and answered the descriptive research question this work tackled.
The findings were examined and consolidated by the practical aspect of the research. First by the quantitative results presentation and analysis; and the new knowledge gained using the relational analysis of the quantitative results. Then by the qualitative analysis which elaborated on the findings and gained more insight into the real world applications of the studied aspects. The main outcomes of the research are the following:

- Quality assurance is very important in crowdsourcing platforms
- The used quality control techniques are situational and circumstantial. They are dictated by the project requirements, the stage of the project they are applied on, and the people applying them. However, there are techniques that are common no matter the project or the setting.
- The different groups of targeted employees involved in crowdsourcing project perceive quality differently depending on their line of work. However, they are aware of the basis of quality in this type of projects.
- Often quality control techniques are used in combinations to cover project specific aspects.
- Real-life situations require the implementation of processes that caters to flexibility and adaptability.
- The need for an integrated quality assurance strategy that will cover all aspects of quality in crowdsourcing.

5.1 Quality assurance strategy

Developing a quality assurance strategy is a huge exercise and the scope of the current work cannot cover all of it. However, the author chose to present the basis for this strategy within the current work as it is one of the main and most interesting outcomes. The proposed strategy will rely on basic examined quality components and processes that support those components.

The strategy would involve three different stages of crowdsourcing projects.

- Design-time
  - Study the quality needs per project and align the needs with companywide quality requirements.
  - Involve all the different groups (Project management, production, sourcing and other) in the early stages of the project.
  - Establish monitoring and control plans for the quality needs of each project.
  - Establish a risk management plan to allocate for quality associated risks and define a contingency plan.
  - Study compensation plans for crowd workers based on available data of worker retention and worker feedback from other project (if not existent start recording it).
  - Design the selection process based on the quality needs.
  - Define and prepare the task thoroughly.
  - Lower the complexity of the tasks to be sent to the crowd when possible.
  - Design the computational processes to support the quality needs.
  - Evaluate the software and the tools to be used.
- Runtime
  - Execute the monitoring and control plans.
  - Teams to provide effective and frequent feedback to the crowd workers.
  - Re-evaluate quality contingency plans based on the running of the project.
  - Evaluate the effectiveness of the computational processes.
  - Manage the community and allow for process flexibility and adaptability to adjust for emerging quality needs or issues.

- Post-delivery
  - Record the quality needs.
  - Record metrics about the effectiveness of the computational processes.
  - Record metrics about the workers motivation, performance and retention.
  - Record the emerging trends concerning quality.

The above points are just a basis for a quality assurance strategy based on the analysis of the results and outcomes of the presented work. The actual strategy will have to account for a multitude of other factors such as cost and resource effectiveness, profitability etc.

6 CONCLUSIONS

6.1 Summary

The presented work tackled the question of quality control in crowdsourcing. Crowdsourcing is a business model based on outsourcing tasks that would traditionally be carried out by employees of the company to a larger group of crowd workers that are suitable to perform the task. The paradigm has a multitude of applications within the modern digital business and takes advantage of the modern communication mediums to enrich the talent pool, reduce cost and improve availability and reach. However, the main concern of this paradigm is quality control.

This research approached this concern and started by examining crowdsourcing and what constitutes quality in a crowdsourcing operation. The research then established a solid theoretical framework by covering the state-of-art literature on the subject. The quality components were defined in terms of a quality model that laid ground to quality dimensions and quality attributes targeted in crowdsourcing projects.

Based on the defined quality model, quality control techniques were examined and the most used best result-yielding ones selected to set basis for the hypothesis the practical aspect of the research examined. To this end, a quantitative research evaluated the selected components by examining them within a practical setting by conducting a survey on a selected population of the author's current employer. The quantitative research first examined the quantifiable aspects and explained the findings. Then, used relational research to further elaborate on the descriptive findings by evaluating associations between the components and the different perspectives different groups within the population had on the components. The survey also included qualitative measures aimed at gaining more insight into the real world applications and perceptions on the examined components. The research carried out a simplified qualitative analysis gaining additional insight.

The research outcomes answered the research questions and concluded on the components of the hypothesis. The outcomes confirmed the importance of quality control in crowdsourcing, confirmed the use of the best-result yielding quality control techniques and concluded on the correlation of component of the quality model of crowdsourcing operations. In addition, the outcomes concluded the circumstantiality of the used techniques and their association with the types of projects they are applied to. Ultimately, the research concluded on the need for a quality assurance strategy within crowdsourcing operations and set basis of said strategy with few theoretical points.

6.2 Research limitations

This research laid basis to quality assurance in crowdsourcing, which is a major exercise. The research focused on establishing a theoretical framework for the quality and evaluating some its components within a practical setting. However, the topic is broader than this and would require further effort.

As demonstrated, quality control techniques are varied and they depend on the task at hand. Given the nature of projects using the crowdsourcing paradigm, the techniques are varied and require customisation. In addition, within the scope of the current work, only the techniques relevant to a certain type of projects were considered. However, crowdsourcing projects can cover other types of project with different requirements and different needs. For example, voluntary crowd projects that do not involve monetary compensation would have a different quality model and different quality control techniques.

Another limitation was the number of participants in the survey due to the size of the target population; which limited to some extent the research. With a larger population and more inclusive target groups some of the quantitative and qualitative examinations would have been more accurate.

Some of the quantitative measures only provided observations and did not provide more detailed and rich analytical basis. This was due to the scope of the work and the number of answers obtained; another elaborative round of surveying would have shed more light into those aspects.

6.3 Future work and perspectives

As stated, this work is a basis for a much larger exercise. The next steps will start by presenting the findings to other concerned parties of the author's employer. Get their feedback on the initial findings and initiate discussions on the outcomes of the research. Then, conduct interviews with program managers to formulate the quality assurance strategy taking into consideration costs, resources and profitability of the proposed strategy. The work will carry on to applying the formulated strategy within a simulation project and evaluate the outcomes. Adjustments will have to be made and the organisational changes associated with such strategy will have to be evaluated and formulated.

Another perspective for the presented work is improving the examinations and extending the findings by conducting more in-depth interviews to make the outcomes more inclusive. Including more employees from higher management would provide a more strategic perspective that will help formulate a more complete quality assurance strategy. Reaching out to crowd worker would also be a possible lead providing a more practical perspective on the examined aspects and would enrich the research and improve the outcomes.

Crowdsourcing is still considered as young business model given the extent to which is it being used in modern business. Quality assurance within this paradigm is a very rich subject and one that is sought after. Consequently, the prospects for further development of such work are limitless. With the increasing interest in the topic and the advantages it offers in the transforming current digital business, the current work can be extended to cover more inclusively and more in-depth the approached concepts.

REFERENCES

Benatallah B., Allahbakhsh M., Ignjatovic, Motahari-Nezhad H. R., Bertino E. and Dustdar S. 2013. Quality Control in Crowdsourcing Systems: Issues and Directions. In: Proceedings of IEEE Internet Computing, 76-81.

Brabham, D. 2012. THE MYTH OF AMATEUR CROWDS: A critical discourse analysis of crowdsourcing coverage. Information, Communication & Society: AoIR Special Issue, 15(3), 394–410. https://doi.org/10.1080/1369118X.2011.641991

Brabham, D. C. 2011. Crowdsourcing: A model for leveraging online communities. InA. Delwiche & J. Henderson(Eds.), The Routledge Handbook of Participatory Culture.

Brabham, D., Ribisl, K., Kirchner, T., & Bernhardt, J. 2014. Crowdsourcing Applications for Public Health. American Journal of Preventive Medicine, 46(2), 179–187. https://doi.org/10.1016/j.amepre.2013.10.016

Business & IP Centre. 2016 .What Are Typical Quantitative Research Questions?. Read 25.09.2020 https://www.bl.uk/business-and-ip-centre/articles/what-are-typical-quantitative-research-questions

Carless, D., & Douglas, K. 2017. Narrative research. The Journal of Positive Psychology, 12(3), 307–308. https://doi.org/10.1080/17439760.2016.1262611

Checco, A., Bates, J., & Demartini, G. 2019. Quality control attack schemes in crowdsourcing. Twenty-Eighth International Joint Conference on Artificial Intelligence (IJCAI-19), Macao, China, 10-16 August 2019. Palo Alto, CA, United States: A A A I Press. https://doi.org/10.24963/ijcai.2019/8500

Daniel, F., Kucherbaev, P., Cappiello, C., Benatallah, B., & Allahbakhsh, M. 2018. Quality Control in Crowdsourcing: A Survey of Quality Attributes, Assessment Techniques, and Assurance Actions. ACM Computing Surveys (CSUR), 51(1), 1–40. https://doi.org/10.1145/3148148

Eskenazi, M. 2013. Crowdsourcing for speech processing applications to data collection, transcription and assessment . Wiley.

Garcia-Molina, H., Joglekar, M., Marcus, A., Parameswaran, A., & Verroios, V. 2016. Challenges in Data Crowdsourcing. IEEE Transactions on Knowledge and Data Engineering, 28(4), 901–911. https://doi.org/10.1109/tkde.2016.2518669

Ghani, K., Miller, D., Linsell, S., Brachulis, A., Lane, B., Sarle, R., Dalela, D., Menon, M., Comstock, B., Lendvay, T., Montie, J., & Peabody, J. 2015. Measuring to Improve: Peer and Crowd-sourced Assessments of Technical Skill with Robot-assisted Radical Prostatectomy. European Urology, 69(4), 547–550. https://doi.org/10.1016/j.eururo.2015.11.028

Hirth, M., Hoßfeld, T., & Tran-Gia, P. 2013. Analyzing costs and accuracy of validation mechanisms for crowdsourcing platforms. Mathematical and Computer Modelling, 57(11-12), 2918–2932. https://doi.org/10.1016/j.mcm.2012.01.006

Hosseini, M., Shahri, A., Phalp, K., Taylor, J., & Ali, R. 2015. Crowdsourcing: A taxonomy and systematic mapping study. Computer Science Review, 17, 43–69. https://doi.org/10.1016/j.cosrev.2015.05.001

Howe, J. 2006. The rise of crowdsourcin', Wired, vol. 14, no. 6, Read 06.06.2020 http://www.wired.com/wired/archive/14.06/crowds.html

Kamajian, S.D. 2015. How crowdsourcing & crowdfunding are fueling health care innovation. https://dx.doi.org/10.1016/ofp.v7i1.365

Kanhere, S. 2011. Participatory Sensing: Crowdsourcing Data from Mobile Smartphones in Urban Spaces, 2011 IEEE 12th International Conference on Mobile Data Management, Lulea, pp. 3-6, doi: 10.1109/MDM.2011.16.

Kuckartz U. 2019. Qualitative Text Analysis: A Systematic Approach. In: Kaiser G., Presmeg N. (eds) Compendium for Early Career Researchers in Mathematics Education. ICME-13 Monographs. Springer, Cham. https://doi.org/10.1007/978-3-030-15636-7_8

Lacity, J. 2015. Understanding Qualitative Data: A Framework of Text Analysis Methods. Journal of Management Information Systems, 11(2), 137–155. https://doi.org/10.1080/07421222.1994.11518043

Leavy, P. 2015. The Oxford handbook of qualitative research. Oxford University Press.

Lebraty, J., & Lobre-Lebraty, K. 2013. Crowdsourcing : one step beyond. Wiley.

Lewis, Sarah. 2015. Qualitative Inquiry and Research Design: Choosing Among Five Approaches. Health Promotion Practice 16.4 (2015): 473–475. Web.

Li, H., Yu, B., & Zhou, D. 2013. Error Rate Bounds in Crowdsourcing Models. ICML '13 Workshop: Machine Learning Meets Crowdsourcing

Lionbridge. 2020a. Read 2019. https://www.lionbridge.com/who-we-are/#about-us

Lionbridge. 2020b. Lionbridge What we do. Read 2019. https://www.lionbridge.com/

Lionbridge. 2020c. Careers. https://careers.lionbridge.com/careers-at-lionbridge-ai/?_ga=2.69196633.917816269.1606083639-1754292118.1538154183

Lionbridge AI . 2020. Read 2019. https://lionbridge.ai/services/

Mass Communication Theory. 2011. Writing Good Qualitative Research Questions Read 2020. https://masscommtheory.com/2011/05/05/writing-good-qualitative-research-questions/

O'Dwyer, Laura M., and James A. B. 2015. Quantitative Research for the Qualitative Researcher . Los Angeles: Sage. Print.

O'Leary, D. 2019. An empirical analysis of information search and information sharing in crowdsourcing data analytic contests. Decision Support Systems, 120, 1–13. https://doi.org/10.1016/j.dss.2019.03.003

Outside Two Standard Deviations. An overview of correlation measures between categorical and continuous variables. Read on 1.11.2020. https://medium.com/@outside2SDs/an-overview-of-correlation-measures-between-categorical-and-continuous-variables-4c7f85610365

Polychronopoulos, V. 2017. Techniques for Quality Control in Applications that Use Crowdsourced Input. UC Santa Cruz. https://escholarship.org/uc/item/57t1m99d

Ponto J. 2015. Understanding and Evaluating Survey Research. Journal of the advanced practitioner in oncology, 6(2), 168–171.

Prpic, J. 2017. Health Care Crowds: Collective Intelligence in Public Health.

Question Pro. 2020. Text Analysis. Read 15.10.2020. https://www.questionpro.com/tour/text-analysis.html

Rahman, M. S. 2017. The Advantages and Disadvantages of Using Qualitative and Quantitative Approaches and Methods in Language 'Testing and Assessment' Research: A Literature Review. Journal of Education and Learning 6.1 (2017): 102–. Print.

Ranard, B., Ha, Y., Meisel, Z., Asch, D., Hill, S., Becker, L., Seymour, A., & Merchant, R. 2013. Crowdsourcing—Harnessing the Masses to Advance Health and Medicine, a Systematic Review. Journal of General Internal Medicine : JGIM, 29(1), 187–203. https://doi.org/10.1007/s11606-013-2536-8

Saxton, G., Oh, O., & Kishore, R. 2013. Rules of Crowdsourcing: Models, Issues, and Systems of Control. Information Systems Management, 30(1), 2–20. https://doi.org/10.1080/10580530.2013.739883

Sharan M., and Elizabeth T. 2015. Qualitative Research: a Guide to Design and Implementation. 4th ed, Print.

Souza L., Ramos I. and Esteves J. 2009. Crowdsourcing Innovation: A Risk Management Approach. In: Proceedings of the of the 4th Mediterranean Conference on Information Systems, Greece.

Statistics how to. 2020. Correlation in Statistics: Correlation Analysis Explained. Read 15.10.2020. https://www.statisticshowto.com/probability-and-statistics/correlation-analysis/

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

Racer T. 2017. Qualitative analysis using Excel. Youtube. https://www.youtube.com/watch?v=P0gzlWNodKw

Research gate. 2014. When should one switch to the 10-item Likert scale?. Read on 01.11.2020. https://www.researchgate.net/post/When-should-one-switch-to-the-10-item-Likert-scale

Wagy, M., Bongard, J., Bagrow, J., & Hines, P. 2018. Crowdsourcing Predictors of Residential Electric Energy Usage. IEEE Systems Journal, 12(4), 3151–3160. https://doi.org/10.1109/JSYST.2017.2778144

Wazny, K. 2017. “Crowdsourcing” ten years in: A review. Journal of Global Health, 7(2), 020602–. https://doi.org/10.7189/jogh.07.020601

Woodley, A. 2004. PREST in open and distance learning: Getting and analysing quantitative data (A3 module). Commonwealth of Learning (COL).

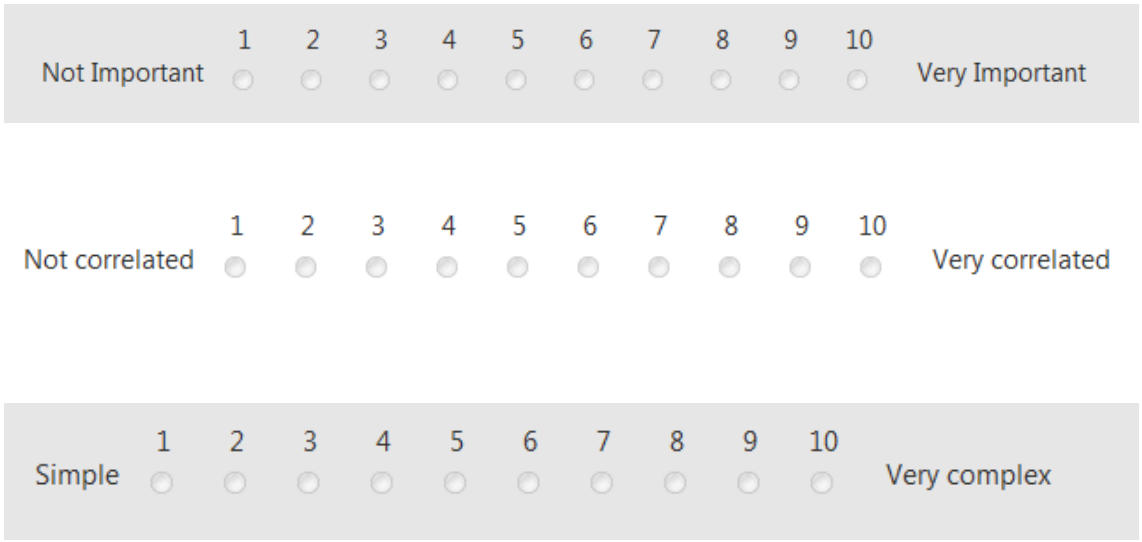
Yin, M., Chen, Y., & Sun, Y. 2014. Monetary Interventions in Crowdsourcing Task Switching. HCOMP.

Zack, n.d 2020. Statology: Phi Coefficient Calculator. Read 01.11.2020. https://www.statology.org/phi-coefficient-calculator/

Zhao, Y., & Zhu, Q. 2012. Evaluation on crowdsourcing research: Current status and future direction. Information Systems Frontiers, 16(3), 417–434. https://doi.org/10.1007/s10796-012-9350-4

APPENDICES

Appendix 1. Used rating ranges



Appendix 2. Phi co-efficient interpretation scale

+.70 or higher	Very strong positive relationship
+.40 to +.69	Strong positive relationship
+.30 to +.39	Moderate positive relationship
+.20 to +.29	weak positive relationship
+.01 to +.19	No or negligible relationship
0	No relationship
-.01 to -.19	No or negligible relationship
-.20 to -.29	weak negative relationship
-.30 to -.39	Moderate negative relationship
-.40 to -.69	Strong negative relationship
-.70 or higher	Very strong negative relationship