

# Community Heuristics for Crowdsourcing Platforms Interface Evaluation

Shazia Zaheer  
Faculty of Science and Arts Sharurah  
Najran University  
Kingdom of Saudi Arabia  
shaziazaheer\_17@yahoo.com

Vera Anatol'evna Gnevasheva  
MGIMO University  
Moscow, Russia  
vera\_cos@rambler.ru

Brunilda Subashi  
Faculty of Public Health  
University "Ismail Qemali", Vlora, Albania  
brunasubashi@gmail.com

Dr Mustapha Bin Danquah  
Asalat2007@gmail.com  
Atlantic International University, USA

---

## Abstract

Crowdsourcing is growing rapidly in both industry and academia, introducing new ways of conducting work and improving our understanding of how to utilize the potential of crowds. Related research has emphasized on how to improve crowdsourcing platforms and related practices to foster collaboration, motivation, trust, quality and creativity. However these challenges don't seem to be as apparent in vibrant online communities. Research in how to make online communities work, provides insights in how to address the challenges crowdsourcing is facing right now. For this work we have gathered from literature relevant design guidelines (heuristics) for online communities and have applied them to 20 crowdsourcing platforms to evaluate how those platforms conform to the heuristics. The heuristics can be used as a tool for designers of crowdsourcing platforms, to evaluate how to improve these platforms and to compare them against their competition. The paper highlights the current challenges of crowdsourcing platforms to acquire positive aspects of online communities.

*Keywords: Crowdsourcing; design methods; online community analysis and support*

---

## 1. Introduction

Crowdsourcing is growing in both industry and academia. At the time of writing this paper, Amazon's Mechanical Turk (AMT) has more than 800 thousand tasks available for completion by workers; Upwork, claims to have more than twelve million workforce and \$1 billion+ worth of work done annually [1]; 85% of the largest corporates have already used crowdsourcing in the last ten years [2]; and one crowdsourcing platform alone (Samasource) has lifted almost fifty thousand people out of poverty in developing countries [3].

These are impressive developments especially when we consider how recent crowdsourcing is as a phenomenon -

the term itself was only coined in 2006. Nevertheless, the potential population of requesters as well as of workers is way larger than the aforementioned figures. There are currently approximately 3.6 billion people connected to the Internet [4] many of whom would benefit from the employment and income both in developing as well as in more developed economies. As a result an increasing number of new crowdsourcing platforms are launched [5]. The design of such platforms appears as a key component for their success in enabling more people to become part of this new way of working and to support different kinds of work. Related research has explored how to design such platforms [6] but currently, designers of such platforms

lack a systematic way to evaluate their platforms against good practices and compare to their competitors.

Researchers highlight that crowdsourcing platforms must integrate both the technical and social needs of the workers in their platform [7] and have to recognize the sociality of work and the shared identities produced through paid collaboration [8]. Some researchers even claim that crowdsourcing platforms are ideally similar to open-source communities [9]. The social element seems to be an important intrinsic motivation to contribute in crowdsourcing platforms [10–12]. It is telling that in the case that platforms fail to provide collaborative tools, workers have shown to create their own tools for collaboration [7]. The worker's perspective can be overlooked with designers focusing primarily on the requester's side with the objective to acquire more projects and increase revenue. Researchers have raised in the past the ethical issues [13] and the improvement of the relationship between workers and requesters [14,15].

Many key challenges crowdsourcing is facing right now, have already been tackled in the field of online communities. For example, in crowdsourcing, research has been concerned with enhancing among crowd workers motivation [11,12,16–19], collaboration [7,8,20,21], creativity [8,20,22] and trust [23]. Our research proposes that these challenges can be addressed by approaching the design of crowdsourcing platforms as designing an online community. For example, the principles of moderation in online communities have been used to improve the quality of work within crowdsourcing using self-assessment and feedback [24]. In another example, a community of crowd workers took collective action to improve their own working environment [25]. Thus, while crowdsourcing platforms often do not provide the tools to connect workers [7] there is a need for evaluation tools to identify such shortcomings for designers and platform owners to take action to enhance the development of their "crowd" into a "community".

By having a vibrant, active community, crowdsourcing platforms can potentially benefit in many ways such as increasing loyalty to the platform, collaboration, and trust. Fortunately, there is already substantial literature that presents guidelines for designing and developing successful online communities [26–30]. However, to the best of our knowledge, it has not been attempted till now to apply such guidelines to evaluate crowdsourcing platforms. Already a recent study investigated common design features in citizen science projects using online community principles [31]. Our research goes on to exploring crowdsourcing platforms with monetary rewards.

With this paper we wish to contribute to the literature in the following ways: 1) to gather the most relevant to crowdsourcing guidelines for the design of online communities; 2) to present a comprehensive way to use those guidelines so that these can be utilized by a greater public; 3) to apply these guidelines in different crowdsourcing platforms; 4) to discuss whether existing guidelines for online communities need to be extended for covering crowdsourcing platforms. In the following

sections we present the literature on which we base our guidelines -which we will name heuristics from now on; the application of those heuristics to twenty existing crowdsourcing platforms; a reflection upon the extent to which known community heuristics are applicable to crowdsourcing platforms.

## 2. Related Work

Crowdsourcing has been used for evaluation purposes on a plethora of domains from search systems [32] to graphic perception experiments [33] to privacy filters [34], just to mention a few. Nevertheless, there have been no attempts, to the extent of our knowledge, to develop methodology to evaluate interaction design aspects of crowdsourcing platforms themselves. By "design aspects" we mean the intersection between the user interface (UI) and the community around the platform. This is not to say that there is no prior knowledge that can guide the design of crowdsourcing platforms. For example, recent research has shown that an increase in participation from the members of a platform can provide individuals with more chances to get noticed, sharpen their creative skills, and strengthen a sense of community [35]. Such general findings though have not yet been compiled into methodology leaving designers of crowdsourcing platforms with the responsibility to locate relevant research and interpret it into their own application context.

To address this apparent gap in evaluation methodology, we set off to develop heuristics to guide the design and expert evaluation of crowdsourcing platforms, in with the widely used approaches for the design and evaluation of user interfaces and web pages [36,37]. The difference with classical heuristic evaluation is that we specifically focus on crowdsourcing platforms. In this effort we draw on research in online communities. For example, Kraut et al. put forward a number of "design claims" that translate theory to design alternatives that may help achieve community goals [29]. Kim proposes nine timeless design strategies that characterize successful, sustainable communities derived from experience in designing many online environments for large corporates in the USA [28]. Gurzick and Lutters present eight design guidelines for online community design, which they illustrate in the case of the online community called "Fieldtrip" [27]. Finally, Preece et al. [30] applies Nielsen's [38] usability heuristics for the design of an online health community, combining them with their own developed "sociability heuristics" which provide a step wise iterative process for improving communities from a member's perspective. Drawing on these sources, we compiled a set of heuristics that compiles design advice relevant to crowdsourcing, removing duplication and adopting a consistent phrasing and abstraction level. These are discussed below, together with an explanation of their foundations upon related work.

For the purpose of our research, we adopt the definition of an online community by Kraut et al. [29] as any virtual space where people come together with others to converse,

exchange information or other resources, learn, play or just be with each other. We adopt the following encompassing definition of crowdsourcing, by [6] : “crowdsourcing is an umbrella term for a variety of approaches that harness the potential of large crowds of people by issuing open calls for contribution to particular tasks”. This definition includes paid crowdsourcing but does not exclude the use of non-dedicated platforms such as social networking systems or other computer supported cooperative systems. Since it is debatable if crowdsourcing in general can be seen as an online community, we distinguish the two terms in discussion of our research.

## 2.1 Heuristics

In this section we will describe each heuristic shortly. We categorize the heuristics in two levels: the general heuristic and its subcategories. We first shortly explain in one to two sentences the general idea of the Level-1 heuristics followed by their subcategories’ description. With each heuristic we cite relevant sources we base it on.

### 2.1.1 Purpose

The platform's purpose identifies the members’ needs and identifies the owner’s goals.

**Clarity:** A clear purpose should describe how it identifies the members’ needs and the owner’s goals [27,30].

**Visibility:** The purpose should be adequately visible, so newcomers see and understand what needs the platform can fulfill [27,29]. Stated by Gurzick & Lutter (2009) it is important both to orient and entice newcomers to the community as well as to provide a common frame of reference for more seasoned members [27].

**Idealism:** The platform has a purpose that contributes to society, justice or altruism; and thus reaches further than the platform alone [39].

### 2.1.2 Moderation

Moderators and other users should monitor that the platform stays a pleasant working environment. Regulation will help to make the expected behaviour clear and can be referred to when violated.

**Monitoring:** The platform should provide the possibility for users to report undesired behaviour thus allowing the community to monitor itself. Users should be able to control their contribution on the platform and allowed to revise it [29].

**Regulations:** Regulation or policies should be available on the platform and easy to find [29,30], allowing users to refer to them when discussions are derailed or when posts do not align with the purpose of the platform. The regulations should be open for debate for further improvement. However, displaying regulations too prominently, may convey the impression that they are not always followed [29]

for which the platform may be negatively perceived.

**Moderators:** Moderators should be consistent. The platform should train or inform them about their role. Moderators can make mistakes, however members should be able to contact them and make an appeal if they don’t agree with the moderator’s modifications [29]. Moderators should have a place to share difficult situations, so they can advise each other and act consistently throughout the platform [28–30].

### 2.1.3 Members

Members should be able to build up an identity on the platform using a profile. For long lasting communities, the platform should recruit and trigger new members to contribute and stimulate the current members’ involvement.

**Self-presentation:** Members need a profile where they can present themselves [26,28,30]. The profile could display elements such as: profile picture, biography and topics that the user is interested in [27]. Personalizing

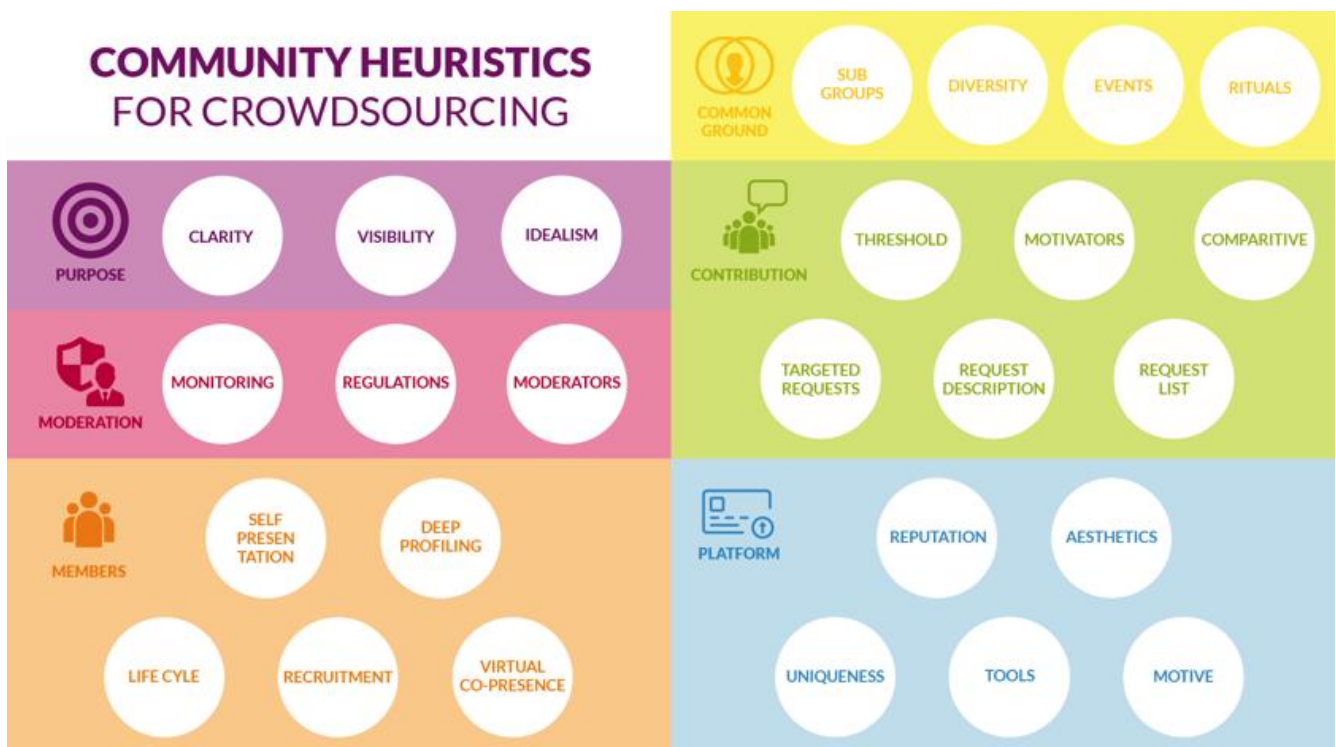


Fig. 1. Based on existing literature we gather six major heuristics, and define their subcategories, for evaluating the state of a community in a crowdsourcing platform

features and activities satisfy people's need to develop individual style and create a social statement through the design of their profile [26].

*Deep profiling:* The platform should support deep profiling capabilities which it can achieve with: reputation or ranking systems [29,40], interaction archives and tools that provide an indication of who did what [26]. The platform can provide a perceived fit between a focal person's belief of his or her identity and the recognition and verification of this identity by other community members [41]. Although both deep profiling and self-presentation, can both be part of a member's profile, deep profiling pertains to information the platform provides about members' activity and reputation, whereas self-presentation pertains to information the members provide themselves.

*Lifecycle:* The platform should be able to facilitate the membership lifecycle [27] consisting of: Welcoming its visitors, by instructing its novices, rewarding its regulars, empowering its leaders and honouring its elders [28].

*Recruitment:* A community should continuously seek new members [30], to grow and sustain an active amount of members. The platform should actively recruit new members by external communication and promotion. Seeing which friends already use the platform (e.g., through social media channels), will raise the likelihood to join the platform as well [29]. Present members should be aware of the importance of newcomers, by inviting members to the platform and by interacting in a friendly and stimulating way [29].

*Virtual co-presence:* Finding an inactive online community will yield little motivation to interact in it [27,41]. The platform should give the impression that is a populated space and needs a critical mass to do so [30,42]. This can be done by a list of the platform's (online) members [28], adding time marks to posts and showcasing the latest contributions [29].

#### 2.1.4 Common ground

The platform should offer mechanisms that support members to find common ground. Members should be able to subdivide from the community to form intimate subgroups in order to accommodate growth and prevent becoming too diverse.

*Subgroups:* The ability for members to partly separate themselves from the community, will maintain a sense of intimacy as the community expands [28]. A subgroup will raise the identity-based commitment to the community as a whole, if it is in line with the general purpose of the platform [29]. The platform should facilitate mechanisms that increase the likelihood that members will encounter similar people to themselves [28], which can be achieved by creating subgroups.

*Diversity:* If the members of the platform have too diverse interests in the platform, it can lower the commitment to the platform and drive members away [29]. The platform should be aware of the diversity and when necessary create subgroups [28,29].

*Events:* The platform should organize events to reinforce the purpose and values of the community. Events will help

to define the community, remind members what they have in common and what their community is all about [28].

*Rituals:* Incorporating community rituals into the platform will make the members feel at home. Having certain rituals will lay the foundation for a true online culture [28].

#### 2.1.5 Contribution

Stimulating members to contribute to the platform, can be one of the toughest tasks of the platform [29]. The platform should show what other members have contributed, make a certain appeal to the members by targeted requests and have a clear description of the request. It is important that the threshold before contribution is adequate and that there are enough intrinsic and extrinsic motivators for the members to contribute.

*Threshold:* The effort that a member has to make to contribute, such as making an account [27]. If it takes too much effort, it is less likely they will contribute, but at the same time, the quality of contributions will be higher [29]. For example, providing credit card information before being able to contribute would create a higher threshold. When a lot of spam or non-relevant contributions are made, the threshold probably should be increased. When nobody, or too few, are contributing, one of the factors can be that the threshold for contribution is too high. Members can play an active role in familiarizing the newcomers to the platform and thus lowering that threshold [28,29].

*Motivators:* What kind of motivations does the member have in order to contribute? Platform owners should be aware of what drives their members, since this can differ per platform [17,43]. A distinction is made between intrinsic motivators (inherently interesting or enjoyable) and extrinsic motivators (outcomes) [44]. Providing rewards and other extrinsic motivators for requests that are otherwise intrinsically motivating, could undermine the intrinsic interest in the task and thus should be treated with care [29,44].

*Comparative:* Members should be able to compare what others have contributed and thus being able to learn the normative behaviour. Members will be more likely to have a more divergent set of contributions when being able to compare themselves [17,29].

*Request list:* A list of the requests should be present, with sorting and tracking mechanisms [29] so members can find tasks which fits their needs and capabilities.

*Request description:* In the description of the required contribution, it should be clear what impact the fulfilment of the contribution will have; is it complementary or substitute [29]? Emphasizing that a member has a unique position or capability will make people more willing to contribute [29]. When workers perceive a certain task as meaningful they increase their productivity [19].

*Targeted requests:* Members should be invited to contribute [27–29], which can be done by targeted requests that match the interest and capabilities of the member [29]. The status, likeability and familiarity of requester improve the change of contribution [29].



### 2.1.6 Platform

The platform should present itself to its members with a unique position compared to competitors, having a visual professional appearance, a trustworthy reputation and motive. The platform should offer tools that help fulfil its purpose.

**Reputation:** Articles of the platform in the news, support by celebrities, awards: all contribute to a positive reputation of the platform [29]. Showcasing the achievements of the platform, helps to understand the value the platform offers and can raise expectations about future success [29]. Platform owners could show the growth of the platform, amount of contributions made and the amount of years it has been established [27,29].

**Aesthetics:** A better looking platform means that people expect it to be better [29]. The platform should provide a professional user experience, and shouldn't encounter any technical difficulties [27,30].

**Uniqueness:** With ever rising amount of crowdsourcing platforms, the platform should serve a unique purpose that other platforms do not offer [29].

**Tools:** The platform offers tools that contribute to fulfilling its purpose. Those tools can cover different areas, such as making a contribution, communicating, collaborating etc. Supportive tools can be the reason that the members will become part of the community and add to the uniqueness of the platform [29].

**Motive:** The motive of the creators of the platform has to be clear to the members. An "about" page of its initial creators and their motivation of creating the platform, will help for members to understand this motive [27].

## 2.2 Heuristic Evaluation

Although there already are well-established logging tools to measure the success of social networks, for example by tracking the responsiveness and interaction between members (e.g. Lithium\*, Philips†), there are no attempts to establish an inspection method for crowdsourcing platforms. Social networks are clearly a different ilk of systems when compared to crowdsourcing platforms. Social networks do not have calls to specific tasks and their main purpose is for their users to exchange their everyday life moments. Consequently, the inspection methods for such platforms would significantly differ.

Inspection methods, such as heuristics, are complementary to logging. Platform owners can have their users' logs, nevertheless they do not have user logs of their competitors. Heuristics can benchmark how they are doing in comparison with their competition. Furthermore, when designing a new platform, heuristics can assist the design,

instead of acting as an evaluation tool. Designers could use heuristics as a checklist to find out whether their initial designs comply with good practices. Furthermore, prior work has shown that inexperienced evaluators' perception of –in that case usability- heuristics is quite similar to experienced ones [45]. We do not claim that user logs are of little use; after all, user logs would be of great salience to evaluate design interventions –whether based on heuristics or any other method. But what we do want to emphasize is that both user logs and inspection methods have their own place in design and evaluation since they serve a different perspective for the same goal.

The heuristics described in this paper follow the same process as the UI usability evaluation (heuristic evaluation) [37], however, the process and description have been modified and re-elaborated with the aim to explore and get results about existing and applicable principles of communities in crowdsourcing platforms.

The approach of using heuristics has been applied in other domains such as: ambient displays [46] games [47] and groupware [48], just to name a few. Heuristic evaluation is one of the main inexpensive usability engineering methods and easy to apply compared to other evaluation methods. In crowdsourcing and communities a lot of information cannot be quantified. In this case, a scored evaluation is the solution to get some quantitative data from the analysis of qualitative aspects related with the communities of their users and their offline and online interaction.

## 3. Methodology

### 3.1 Evaluated Platforms

To assess how well the heuristics support the evaluation of different crowdsourcing platforms we have applied them to a diverse set of twenty platforms shown in Table 1, which were chosen to match the eight characteristics of a crowdsourcing platform [49]. We grouped the twenty platforms in five different categories: *digital work*, *design*, *ideation*, *microwork* and *research*. The reason we did that was to be able to compare not just a platform with another one, but also between platforms of the same category as well as compare categories of platforms. Rather than a principled classification based on the characteristics of the platform (e.g. [6]), we classified platforms according to their purpose as platform owners are more interested in how their platform compares to their competition.

Table 1  
 List of the twenty surveyed platforms

Platform	Description
<i>Digital work</i> Upwork	A marketplace for freelancers
Topcoder	Online computer programming and design competitions.

\* <https://www.lithium.com/>

† <https://www.linkedin.com/pulse/20141120100755-62042713-how-to-measure-the-success-of-your-internal-social-network>

Freelancer	A marketplace for freelancers
Crowdsource	Managed crowdsourcing platform, providing trained Upwork workers.
<b>Design</b>	
Jovoto	Crowdsourcing innovative ideas for the challenges of big brands
99Designs	Graphic design marketplace.
Battle of Concepts	Crowdsourcing solutions to mostly local societal issues.
Cadcrowd	3D model marketplace and competitions.
<b>Ideation</b>	
Tricider	Crowdsourcing the collection of and voting for ideas.
Synthetron	Collective brainstorming/discussions with voting mechanisms
OpenIDEO	Global community working together to design solutions for the world's biggest challenges.
Innocentive	Innovation market with solutions to business, social, policy, scientific, or technical problems.
<b>Microwork</b>	
Microworkers	Crowdsourcing microwork
Crowdflower	Data analysis combining machine learning and microwork
Samasource	Providing microwork and training to workers in poverty countries.
MTurk	Crowdsourcing microwork
<b>Research</b>	
Prolific.ac	Surveys for academic research
Roamler	Real-life microwork for companies
Usability Hub	Design feedback from fellow designers
AYTM	Various market research methods

### 3.2 Process and Data Gathering

We recruited four evaluators among students in our department that had at least half-a-year experience in the field of crowdsourcing: a 2nd year bachelor student, a 2nd year master student and two PhD students. The evaluators

were given a training session consisting of a 45 minute presentation explaining the heuristics in detail and a 15 minute demonstration of how to use the evaluation form (see next section). Then they evaluated five platforms, one from each category. The evaluators could choose the platform they wanted to evaluate, but each platform was evaluated only once. We are aware that different evaluators may find different problems when evaluating a single platform [36] yet since this is the first time we apply such heuristics, we were more interested in their applicability and understanding the broader design issues of crowdsourcing platforms rather than being exhaustive in uncovering all points of improvement for a specific platform. In any case, we would expect that our evaluators would identify at least approximately one third of the issues [50]. We advised our evaluators to choose platforms that they were more familiar with. Each participant received as a reward a €50 gift voucher.

We created an online evaluation form using Google Forms\*, which our evaluators used to fill in once for each platform. Additionally, a short interview with each evaluator was held after the evaluations to collect general impressions. During the interview they reported to have needed between half-an-hour up to two hours to evaluate each platform using the heuristics depending on the platform's complexity. Nevertheless, for all evaluators the first evaluation took longer, between one and three hours, since they were still getting familiar with the heuristics and with some of the platforms. To give a concrete example of what our evaluators were asked, we present the evaluation questions, of the three subcategories of the heuristic "Moderation":

#### 1. Moderation

##### 1.1. Monitoring

1.1.1. Monitoring Q1: Is it possible for users to report undesired behaviour?

1.1.2. Monitoring Q2: Can users control the output of their contribution?

##### 1.2. Regulations

1.2.1. Regulations Q1: Are there regulations present on the platform?

1.2.2. Regulations Q2: Are they placed on a correct place on the platform?

1.2.3. Regulations Q3: Are the regulations up for debate by users?

In total, the heuristic evaluation (operationalized with the form) has 51 evaluation questions spread along the heuristics and their subcategories. Each of the 51 questions is accompanied with an extra question to provide evidence (more about the "evidence" is explained later on). With this amount of questions, the evaluator has the capacity to go in depth and explore the details of the community heuristics on a certain platform.

To quantify the community heuristics, we used a simple three-point scale: "No": the platform does not adhere to the

\* Form can be accessed at: <http://goo.gl/bDSXfM>

heuristic; “Semi”: there is some evidence of an effort to adhere to the heuristic but this effort is not sufficient; and “Yes”: the platform fully adheres to the heuristic. For every heuristic a score is given from 0 (“No”) to 2 (“Yes”), allowing us to add all heuristics together to a total score called “t”. Since crowdsourcing platforms can differ in their setup, we expected that not all of the community heuristics from literature would be applicable. Thus, in addition to the aforementioned three-point scale, the evaluators also had the choice of Not Applicable (NA). Having this option would indicate if any of the heuristics were not applicable for crowdsourcing.

The evaluators were asked to provide evidence for every evaluation question. They could either submit a link to a specific page of the crowdsourcing platform, a screenshot or a short description to substantiate their score. Providing evidence not only helps understand how the heuristics are interpreted, but also stimulates the evaluators to be more thorough with their choices and explain their reasoning.

Before starting the evaluation, evaluators had to indicate how familiar they were with the platform they evaluated, choosing between the four options shown in Table 2.

Table 2  
Levels of expertise in a certain crowdsourcing platform that our evaluators had to choose from.

Familiarity	Description
Somewhat	Browsed the website but have neither worker nor requester account.
Quite	Have browsed the website and have either worker or requester account and have browsed their view.
Comfortable	In addition to above have completed tasks or have posted tasks and received worker input.
Expert	In addition to above - I frequently complete or post tasks.

Besides that, we asked them as what kind of user they had used the platform before (more than one options selectable): *visitor, worker, requester, moderator or platform owner*. We requested this information for identifying biases caused by familiarity with the platform.

#### 4. Results

In this section we first analyse the platforms based on the heuristics. We examine three heuristics for which the reviewed platforms score well and three heuristics that help identify required improvements for the reviewed platforms. Further, we analyse how well the five different types of crowdsourcing platforms identified above fare against the heuristics. Finally, we compare two platforms side by side, to highlight the differences at a finer level of detail.

#### 4.1 Evaluation per heuristic

In this section we first present our analysis by grouping the results according to the heuristics. First three examples are described in which crowdsourcing platforms perform well (i.e. adhere to the heuristic) and three in which they can improve.

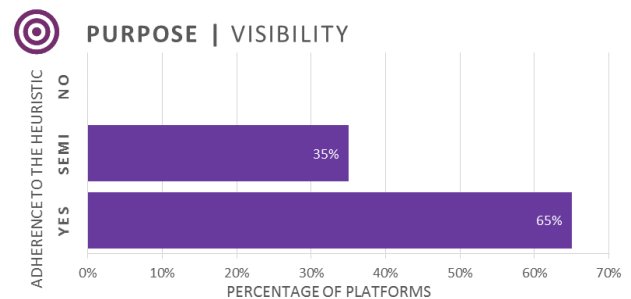


Fig. 2. 65% of the 20 surveyed platforms prominently display their purpose

##### 4.1.1 Heuristics in which platforms perform well

The platforms assessed were rated highly with regards to making their purpose visible (Figure 2) –which is important for newcomers to the community and to provide a common frame of reference for regular members. In most cases the purpose was displayed in a slogan underneath the platform logo or on the homepage as sentence in a large font followed by a call-to-action. For example, Cadcrowd, displays the following sentence spanning a whole screen “Freelance 3D design, 3D modelling and CAD drafting” followed by the sub sentence “Hire a 3D modeller, 3D designer or CAD drafting freelancer for your project on demand”. An improvement that we can suggest is that after the user has logged in the homepage stating the purpose will not be accessible anymore. An example of this approach is UsabilityHub, where clicking on the platform logo at the top left corner will redirect users to their personal dashboard, instead of the platform’ homepage.

The evaluators rated the platforms highly regarding the “reputation” heuristic, as can be seen in Figure 3. Examples of external reputation that our evaluators found were Facebook ratings, blog posts on Quora, Reddit forums, independent sites reviews and testimonials presented on the site itself. Internally, platforms showcase the achievements

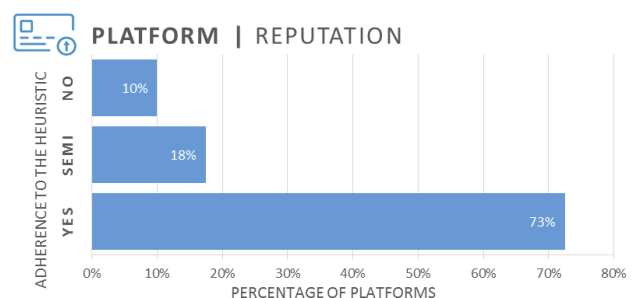


Fig. 3. High percentage of the 20 surveyed platforms has a positive reputation among different media channels and they make sure to showcase their achievements back in their own platforms.

on the platform itself quite well, with 73% of the platforms doing so. Platforms can showcase the achievements by displaying the amount of money received by the crowd, amount of workers, amount of competition etc. An example is found in Jovoto, displaying these achievements on their homepage as can be seen in Figure 4.



Fig. 4. Jovoto displaying the achievements of the platform, contributing to the reputation of the platform.

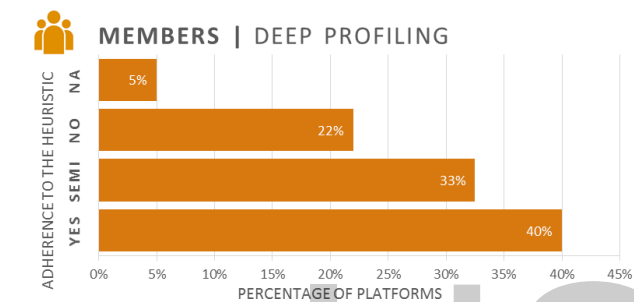


Fig. 5. Already 73% of the 20 surveyed platforms offer deep profiling possibilities to the members.

Already 73% of the platforms were found to support deep profiling possibilities to their members (Figure 5). The platform *Freelancer* uses a lot of different deep profiling options, such as the number of stars, written reviews by requesters, number of jobs completed, number of deliverables that were submitted on time and on budget and finally the repeat hire rate. Furthermore, a worker can also be recommended by a requester that is represented in *Freelancer*'s user interface by a heart symbol. Workers can prove the skills they claim to have by taking a paid test provided by the platform. When successfully completing the test, a certification will be show on the worker's profile. The platform states "By completing a certification you can expect to earn 50% more per year than your competitors"\*. In that way workers can distinguish themselves with certificates that are valuable within the platform itself.

Other examples include 99Designs and Crowdfunder that allow deep profiling by providing badges, but these cannot be seen by other members. Although this is a step in the right direction, these platforms could improve their deep profiling by also making these badges publicly visible on the user's profile.

#### 4.1.2 Heuristics in which platforms need improvement

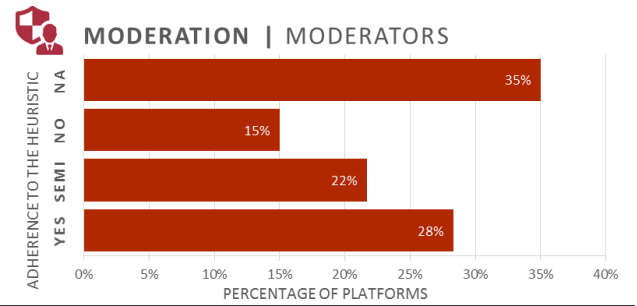


Fig. 6. 35% of the 20 surveyed platforms are NA in terms applying to the heuristic moderators.

Evaluators considered that moderation was inapplicable in many cases. The moderators present on the platform need to be trained and need to be easily available for contact by workers. Furthermore they should have a place to gather and discuss certain problems. The high proportion of Not Applicable (NA) responses that can be seen in Figure 6 has two explanations. First, our evaluators could not have the access to the part of the platform that was necessary to evaluate the heuristic or moderators were not being present on the platform at all. For two of the three evaluation questions, our evaluators filled in NA 50% of the time, meaning that half of the assessed platforms do not have any form of moderators present other than contacting the owners of the platform directly, which can be considered as a high threshold. A specific comment from one of our evaluators when evaluating *Microworkers*: "Moderators are invisible. You can come in contact with a moderator when you complain your task is not revised properly, but otherwise, I could not find any moderator." Some platforms like *Topcoder* behave like an online community where the experienced workers have actually taken the role of moderators. This result matched prior research that has found that platforms underestimate the role that their own workers can have in the managing the community on the platform [21]. *OpenIDEO* has a page called "community" where one can apply for one of three roles a good moderator could also fulfil†. These roles are: community cross-pollinator, community prototype and social media ambassador.

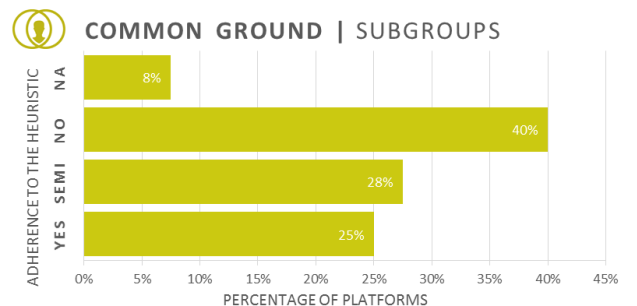


Fig. 7. Few of the 20 surveyed platforms we evaluated support the creation of subgroups for their members

\* <https://www.freelancer.com/exam/exams/buy.php?id=2>

† <https://challenges.openideo.com/content/community>



The creation and presence of subgroups is currently problematic, with only 25% of the platforms that fully support them (Figure 7). A good example in our list is the platform OpenIDEO that for every task (challenge in their case) creates a subgroup that becomes a small community within the platform. Within the challenge workers can interact and support each other, but that still complies with the general purpose of the platform, which is working together to design solutions for the world's biggest challenges.

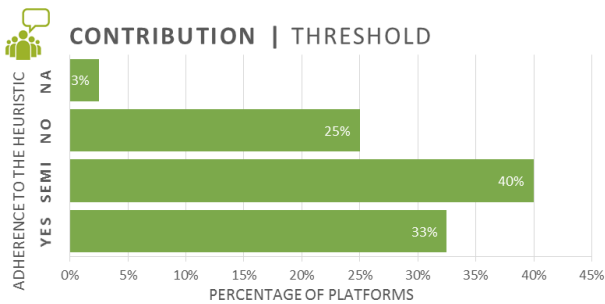


Fig. 8. Only 33% of the 20 surveyed platforms have a threshold for contribution that is just right

The threshold to contribute seems to differ a lot among platforms (Figure 8). In the platforms we evaluated the threshold was mostly manifested in the creation of an account for the platform. Our evaluators reported that some platforms require elaborate personal information, or even training to reach a certain level before being able to contribute. Even worse, they may not even be able to contribute at all after the training as, for example, in the case of Roamler where a code is required to activate the account that is not actually available. A good example for this heuristic is Tricider, requiring only one click to vote for the best idea. However one can contribute more by adding new ideas or suggesting the pros and cons of the ideas given all without having to make an account. Only when one wants to stay up to date or moderate the contest, an account is required.

#### 4.2 Comparison among different kinds of platforms

Among the different types of platforms, the inclusion of communities differs. As already explained in the methodology, for every heuristic a score is given from 0 to 2, allowing us to add all heuristics together to a total score called "t". The maximum score would be calculated as: would be 4 platforms x 51 heuristics x maximum score of 2 which equals max t = 408. If we count all the scores from the same types of platforms together, we find a division of two groups occurring. The types of *Microwork* (t = 166) and *Research* (t = 188) are the platforms that score the lowest, meaning that for these two types most improvements can be made for the design of communities. Three other types of platforms were given similar total scores : *Ideation* (t = 275), *Design* (t = 288), and *Digital*

*Work* (t = 286). These kinds of platforms require more skills and creativity than the research and microwork platforms, which could explain the difference. The differences between the two groups were found in all the different heuristics categories.

Two heuristics will be presented in which a certain type of platform performs well and three heuristics in which they need improvement.

##### 4.2.1 Heuristics in which types of platforms perform well

Already in the general comparison among the different types of categories, the gap between research, microwork and the other three types is the biggest. That is why we only describe when the platforms differ from this pattern in a remarkable way.

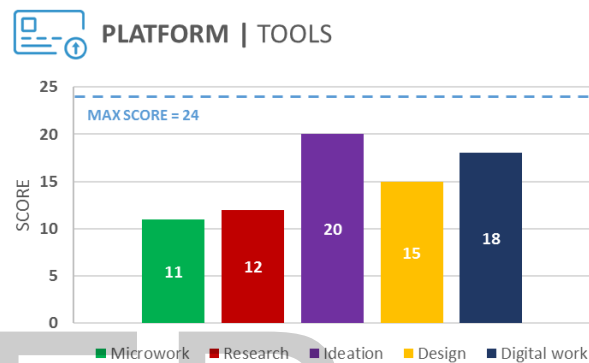


Fig. 9. Ideation type of platforms provide the best tools for contribution on the platform.

The ideation platforms seem to flourish in the tools heuristics (Figure 9). These platforms would need the most creative input for users and thus provide the best tools. Since the platforms can be considered as a tool for requester to find and chose the best worker of contribution, the platform should also be open for improvement. An example is OpenIDEO that provides a lot of resources that could benefit their workers to make better contributions such as brainstorm or interview toolkits. They are also open to feedback for the improvement of the platform, since in the footer they state "Please give us your feedback".

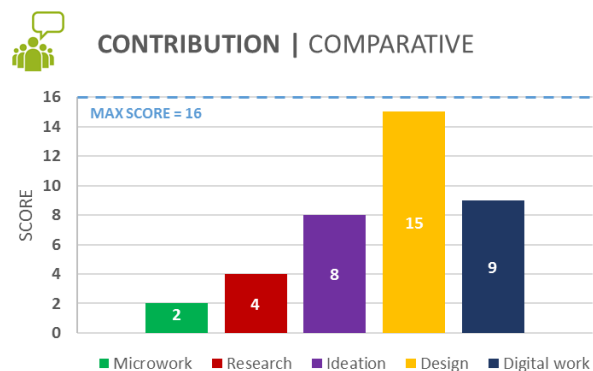


Fig. 10. Design type of crowdsourcing platforms perform best in allowing their members to compare themselves towards to other members.

The heuristics “comparative” is most present for the Design type of platforms, with the near perfect score of 15 (Figure 10). On platforms in the category “Design”, one can easily see which contribution have been made for competition. For the worker it is important to see these contributions in order for them to judge if they have a chance of winning a certain competition. The workers could also see which contribution would have won in the past and thus learn what the expected level is. In these kinds of platforms it is also normal to communicate directly to the requester during the competition. This helps workers to learn the normative behaviour of the platform thus help them understand how to become a better worker. The design platform 99Designs allows comparing the features described. Within microwork and research type of platforms, the possibility for comparison seems to be completely absent.

4.2.2 Heuristics in which types of platforms need improvement

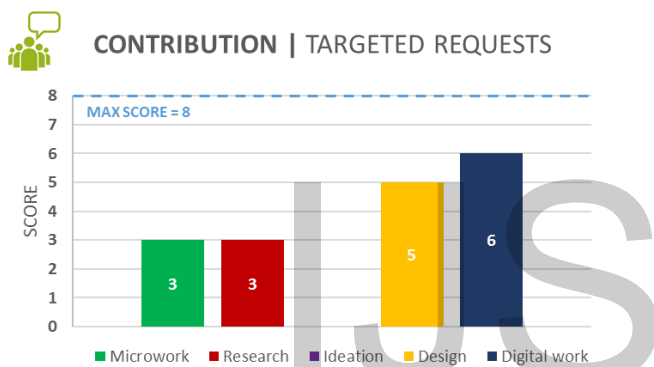


Fig. 11. The Ideation type of platform do not make targeted requests

None of the evaluated ideation platforms make targeted requests towards their participants (Figure 11). An explanation could be that the nature of ideation platforms are be more open than other platforms, not wanting to push people to contribute. The evaluators describe two ways in which targeted requests are made. The first one are the targeted request from the requester that can filter workers to appeal to a more specific group of workers. For example, the platform Microworkers makes a distinction between “Basic” microtasks which all workers can perform and the possibility to “Hire Group”, to target a specific type of workers, such as only the English-speaking workers. The second type of targeted requests can be made by the platform towards its workers. Based on the data that the platform has gathered from its workers, it can learn and directly target a specific group by sending a notification to those workers of the new request.

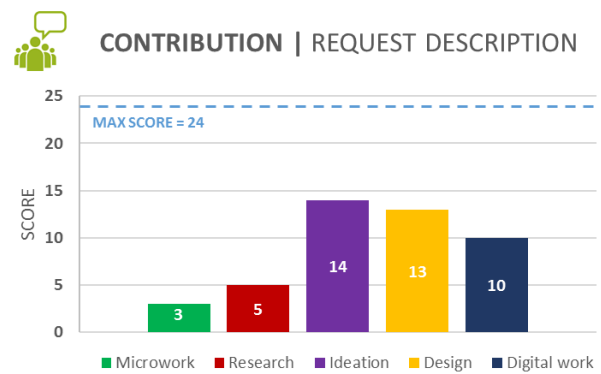


Fig. 12. Microwork and research platforms can improve their request description a lot compared to other types of platforms.

Mainly Microwork and Research platforms could support a request description to raise a feeling of complementary contribution and appeal to unique capabilities of a worker (Figure 12). As already described in the previous heuristic, Microworkers allows to select certain workers that can fulfil the task. The feeling of a complementary contribution may be harder for microwork type of platforms, given the individual nature of the contributions, but not impossible. A requester that would have many workers perform a certain tasks for them, could send an update towards all the Microworkers thanking them and sharing the results of impact they their contributions had.

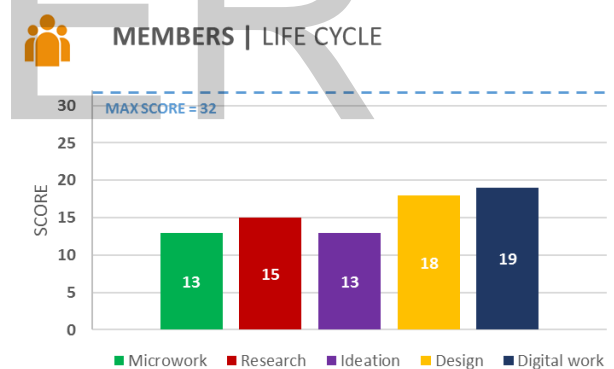


Fig. 13. Ideation and microwork platforms support the member’s life cycle the least

Surprisingly, Ideation and Microwork platforms seem to not fare the worst concerning the appreciation of members’ life cycle (Figure 13). Especially the first step of welcoming your visitors, where the platform explains how it is used, is often found on the Homepage, FAQ; only one of the ideation platforms fulfil this heuristics whereas the three other ideation platforms do not. The platforms should be clearer on what it is that they offer and require from their users. However, once the step is made to make an account on the platform, the ideation type of platform performs the best. At this point the platform instructs their novices, which seems a learning point for almost all platforms. Evaluators have found example of tours, resources and video training in order to stimulate the novice users to make a first contribution. The reward for

regular use seems quite clear in crowdsourcing, since there is almost always a monetary reward to work for. Microwork platforms need to honour their elderly better. For example Microworkers has a special page showcasing the best workers where one could filter on most paid, most stars, most tasks etc.

#### 4.3 Comparing two platforms using the heuristics

Radar diagrams (Figure 14) are well suited for a heuristic evaluation comparison of two platforms, because they give a recognizable shape based on the score. The more circular the radar, the more balanced the scores; the spikier the radar, the more variation in the scores. The size of the radar plot on the axes indicates the total score percentage itself, while the shape shows good and bad areas. These results show that the evaluation procedure and heuristics proposal can be tools to analyse the community features integrated in a platform as a whole. The radar diagrams are not a result to determine if a certain platform is successful or not, but rather to illustrate how the platforms use and integrate their online communities. When the evaluators rated a heuristic with NA, the score could be from 0 to 2. The graphs represent the maximum (orange) and minimum (light orange) scores of the platform. The platforms that embrace their community the most seem to be OpenIDEO and Topcoder each having hexagon shape without spikes that fill the graph almost completely. The radar graphs show that three of the four research platforms do not fulfil any of the common ground heuristics.

Another use of the heuristics is for competitor analysis. Within the same type of platforms, the platforms Cadcrowd and 99Designs are compared. The platforms seem similar at first sight, both supporting their community by a forum. The difference is only that the contributions made on 99Designs are graphic designs and Cadcrowd 3D model designs. Can the heuristics reveal what these platforms can learn from each other? Although the demand for 3D modellers versus graphic designers would differ, the competition within the graphic design contest platforms is a lot bigger than in 3D model competitions. Cadcrowd claims to have over 11.000 workers, whereas 99Designs has over 360.000 workers.

The platforms are compared using the order of the heuristics, starting with the main heuristics purpose. Only the heuristics that reveal interesting differences are described.

**Purpose:** Cadcrowd does not support viewing the home page when logged-in, which decreases the visibility of the purpose of the platform. 99Designs hosts competitions without asking for a fee for non-profit organizations\*, showcasing that even monetary driven platforms can live up to the heuristic idealism.

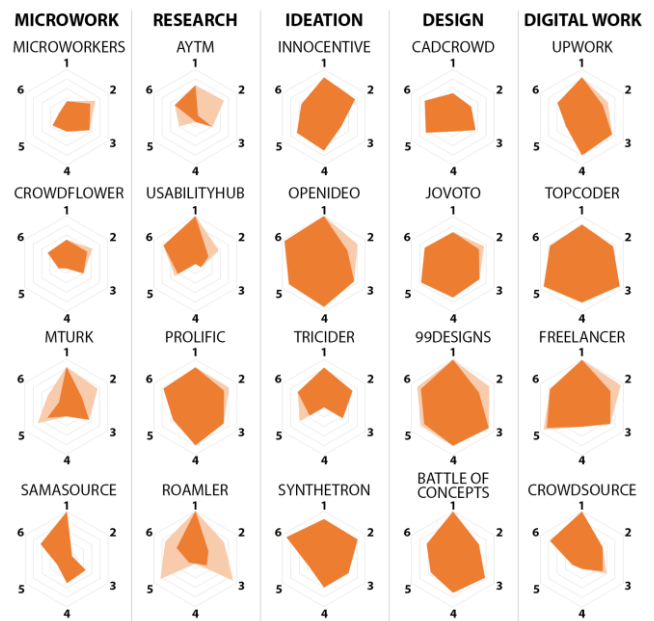


Fig. 14. Radar diagrams based on the heuristics evaluation of various crowdsourcing platforms. The numbers resemble the heuristic: 1 = Purpose; 2 = Moderation; 3 = Members; 4 = Common ground; 5 = Contribution; 6 = Platform

**Moderation:** In terms of monitoring, Cadcrowd allows members to report competitions that do not follow the platform guidelines, but not individual contributions made by the workers. 99Designs does allow members to flag worker’s contributions and the platform even rewards you with a badge called “First Flag” explaining the importance of flagging inappropriate messages. Besides that, the badge also explains that one could also directly send a personal message when seeing an issue with a user’s post.

**Members:** In terms of deep profiling, 99Designs allows requesters to give a review to the designer, where Cadcrowd doesn’t support this feature. 99Designs has a whole page with supportive tools and resources for their designers to improve their designs. Cadcrowd provides a blog, which is mostly targeted at the requesters instead of workers by showcasing the 3D models their workers have made and giving general advice in how to create successful products. A unique way of honouring elderly members within the lifecycle heuristic is provided by 99Designs, which launched their first community eBook showcasing the “compilation of epic designs from our oh-so-talented community”†. Cadcrowd has a more dynamic way of honouring their members by giving all their members a rank that is displayed on the profile based on points you gather for contributing and winning contests. The rank allows them to provide a list of the best designers. 99Designs allows viewing designs, but isn’t possible to filter on the best designers, making it harder to find the best designer for your task.

**Contribution:** The threshold for 99Designs is quite high, since it is required that the platform should validate

\* <https://99designs.nl/nonprofits/>

† <https://99designs.nl/blog/portraits/check-out-our-first-ever-designer-ebook/>

member ID cards before allowing them to make a contribution, a measure probably intended to limit the amount of spam to the platform. A good example on how to lower the threshold for requesters is by Cadcrowd, which adds a “Post similar contest” button to every competition. 99Designs allows you to make valuable request descriptions, one can choose from a list of logos and choose the ones that fit their taste the most. In the next step the requester can move a slider chosen between contrasts such as Classic and Modern, helping the workers to understand what the requester is looking for. The counter effect can be that the contributions are not diverse from each other and thus limiting the creativity of the designers.

## 5. Discussion

In this section we would like to first address methodological challenges, then discuss based on the experience of conducting this study a classification of the incorporation of communities in crowdsourcing and finally discuss the definition issues that arise from our work.

### 5.1 Methodological Challenges

The initial challenge when conducting an evaluation that our evaluators faced, was to make an account on the platform that they were asked to evaluate. Although in most platforms this is a workable and quick action, there are platforms that restrict that or require extra checks. For example, in the platform Roamler one can only make an account if they have received an invite by an existing member or from the platform administrator. In our case, our evaluator was not able to get this access and thus had to use “not applicable” (NA) for several heuristics such as the “members” and “contributions” heuristics. Other platforms might have geographical restrictions. For example, MTurk currently only accepts workers and requesters from the United States and India. The access to the request list is still possible without having an accepted account by MTurk, and thus our evaluator was able to evaluate most of the platform.

Other evaluators would probably face the same issue for other platforms. Having mentioned the previous challenges, a worker account is not strictly necessary to perform the heuristic evaluation we presented in this paper. For example, for the heuristics of self-presentation or deep-profiling it might help the evaluator to review the full extent of the platform, nevertheless by inspecting other worker’s public profiles an evaluation can be certainly performed. Another option to address this issue is to create a requester account. In most platforms that is as easy, if not easier. Through a requester account one can then inspect worker’s profiles and report their findings relating to the worker’s profile.

Another issue our evaluators reported was fatigue when having to conduct the evaluation in one go. We expected and had informed our evaluators that an evaluation session

would last around one and a half hour nevertheless, for most of our evaluators, it took more than that. Our evaluators for the first platform reported up to three hours with a minimum of one and a half hours. Subsequent sessions took around the time we expected, i.e. one and a half hour. Nevertheless, even that amount of time was perceived as laborious and difficult, even though our evaluators did not perform the evaluation in one go. We conducted a frequent analysis to check whether the fatigue we would expect affected our evaluators. Figure 15 clearly shows that this is not the case.

This finding raises the issue of making the evaluation not just easier, but also making it a better experience, since we do expect a typical session to last at least one and a half hours. There are a few ways to address this. Firstly, one could break the session down and better guide evaluators to manage their expectations. For example, one can introduce regular breaks and mention for each step what would the estimated time be to complete a particular heuristic. Secondly, we are currently exploring the use of cards to guide the evaluation (Figure 16). Each card briefly describes one heuristic. Moreover, a short paper guide is included with the deck of cards. In this way we hope that the evaluation becomes accessible and reduces the need to study the method, or read documentation and provides a low threshold way to start applying the heuristics. The process of an evaluation with the cards could open up to even a collaborative effort that would also be perceived as more pleasurable. Lastly, the heuristics could be restructured in a more efficient way. For example, one of the first heuristics “visibility of the purpose”, is checked on the homepage of the platform. At later stage, the reputation of the platform also has to be checked mostly on the homepage. These kind of heuristics could be grouped together, based on the current input from the evaluators and thus limit the searching time.





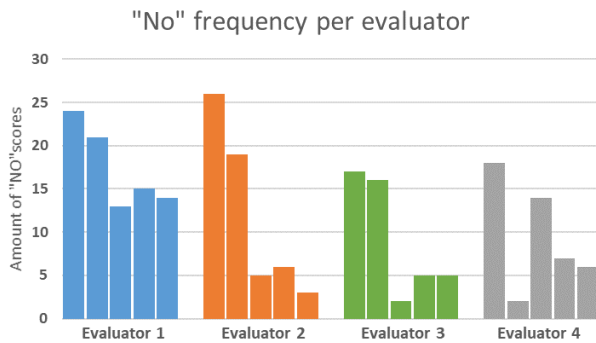


Fig. 15. Clustered bar chart for “No” frequency per evaluator. If our evaluators would experience fatigue, we would expect an upward trend in the bar charts since answering “No” carries the least effort. We do not observe the upward trend in the bar chart.

Fig. 16. An idea to make the process of evaluation more accessible and fun is with the use of cards, which we have already piloted in workshop settings.

It should be kept in mind that applying a certain heuristic, finding and reporting relevant evidence, requires more time and effort from the part of the evaluators. This could introduce a bias in favour of answering “no” that the specific heuristics is not covered in a certain platform or “NA” that the heuristic is not applicable. To guard against such behaviours an extra check was included for the cases where they had checked with “no”. After our evaluators had concluded their evaluation we asked them, after a period of few days, to re-check the cases that they had rated a heuristic with “No” or “NA”. Future application of our heuristics also need to be aware of this shortcoming and need to take care that an extra check is planned to ensure that there was enough effort put for each heuristic.

## 5.2 Community incorporation

In inspecting the twenty crowdsourcing platforms we found that they have three different ways in how they use communities. We dub those as: platforms that have an *unregulated community*, an *externally controlled community* or an *integrated community*. We further discuss the three types.

### 5.2.1 Unregulated community

Some platforms do not support their participants at all, i.e. they do not provide their members any form of venue to encounter each other or discuss. In some of these, a community can be created by the workers themselves [33]. The platform’s employees do not control or participate in the community making workers dependent on each other. Without the facilitation by the platform there is not designated place for workers to gather, which can result in multiple fragmented communities emerging. For example

there are at least 12 external dedicated communities\* for MTurk. This phenomenon leaves substantial untapped.

There is an opportunity for MTurk to create a strong community, which would provide one place to manage and gather all its members. Another example is Microworkers that has a forum type community on Reddit† created by the workers of the platform. Nevertheless, there are several platforms that do not have even this type of community, such as: Tricider, Battle of Concepts, UsabilityHub, AYTM and Prolific. These platforms seem to have taken upon them to address all potential questions and comments from workers and requesters, whereas a community ideally could support the platform in this function.

### 5.2.2 Controlled community

A controlled community is created and controlled by the platform. The community isn’t integrated in the platform since a link is used within the platform to redirect to the community environment. The contributions made in the community are not reflected in the worker’s profile back in the platform. Examples of this kind of platforms display the link to the community as a subcategory of the main menu or in the footer. Specific examples of this type that we found are: Upwork, Freelancer, 99Designs, Cadcrowd and Crowdfunder.

Crowdfunder is an interesting borderline case since it has an external community in the form of a Tumblr page‡. The latest post in September 2016 is from the community manager conveying to the workers “Goodbye, I’ll miss you” –by that announcing she will not be the community manager anymore. In a Twitter message on the page, the workers are requesting for a new community manager. The community set-up by Crowdfunder depends on their community manager who acts as a form of helpdesk for workers. With the departure of the community manager at the time of writing this manuscript, the platform community can transform into the uncontrolled community type. This is even more so, as the link to the external community cannot be found in the platform, so at this moment we could better classify the platform in the uncontrolled community type.

### 5.2.3 Integrated community

The last type of platforms integrates the community completely. Here all types of contributions are reflected and sorted in the members’ profiles. Examples of such platforms are: Topcoder, OpenIDEO and Jovoto.

Then there are platforms that have an integrated community, in which case they act as community sourcing platforms§. Instead of crowdsourcing the tasks in a form of an open call enabling everyone to contribute, the call is specified towards a certain community that the platform controls, hence the term “community sourcing”. The

\* <https://www.reddit.com/r/mturk/wiki/communities>

† <https://www.reddit.com/r/Microworkers/>

‡ <http://crowdfundercommunity.tumblr.com/>

§ <http://amysampleward.org/2011/05/18/crowdsourcing-vs-community-sourcing-whats-the-difference-and-the-opportunity/>

threshold to become part of these communities is high since the platform decides whether one is allowed to become part of the community. Examples of these kind of platforms are: Roamler, Samasource, Crowdsourcing and Synthetron. The platform InnoCentive also offers community sourcing, but as a separate service that is called InnoCentive@Work<sup>\*</sup>.

### 5.3 Crowdsourcing Platform or Online Community? The issue of definition

One of the main questions we had when starting this endeavour, was to what extent the heuristics we gathered from literature on communities would apply to crowdsourcing platforms. After having the experience of performing several evaluations ourselves and the interviews of our evaluators we feel that all heuristics for online communities do apply to crowdsourcing platforms.

However, there is one sub-item in the heuristic of "Regulations" and two subcategories - "Recruitment" and "Idealism" - that we feel are not applicable or at least need some special consideration. The sub-item in "Regulations" is: "the regulations should be open for debate for further improvement". In an online community it is only expected that each member, or at least all prominent members would debate regulations with the aim to improve them. Nevertheless, when it comes to crowdsourcing platforms, specific companies with for-profit objectives own these platforms. While without a vibrant community the platforms would not exist, debating the regulations could have adverse implications for the business.

Another specific heuristic that raised a discussion among us was the one of "recruitment". This might be controversial in crowdsourcing, at least from the point of view of some crowd-workers: if one would recruit more workers she would create more competition for herself in the platform. Most crowdsourcing platforms are competition-based. This means that if one would invite and actively recruit more members one is automatically lowering her chances to win a certain competition. Thus, to actively promote this behavior crowdsourcing platforms need to carefully think of rewards and clearly outline the benefits for existing members. For example, a potential benefit might be that if there are more people in the platform that by itself might attract more requesters and in turn more competitions so the competition-to-worker-ratio might be potentially equal or lower. Furthermore, if the platform would be able to visualize the ideal balance between the amount of requests and workers, it could issue more informed requests on its own workers to invite new ones.

Finally, in the case of *Idealism*, one could be inclined to question its applicability to crowdsourcing since most platforms are for-profit. Yet idealism is not as such incompatible with profit. One might link idealism to sustainability and social responsibility. This is maybe a

"blind spot" for current platforms. Samasource is a good example of a platform that prominently highlights its link to social responsibility by employing and training crowd workers from developing countries mainly in Africa. We find that idealism is also a good example of a heuristic originating from communities that can help crowdsourcing platforms to further develop. The role of communities transcends financial matters; ideals could potentially help in addressing challenges that crowdsourcing faces, such as low quality contributions or engaging crowd-workers.

Furthermore, with the way we structured our evaluation, we had planned to gather empirical evidence of which community heuristics would be less applicable to crowdsourcing platforms. We had expected that evaluators would choose the "NA" option when doubting the applicability of a certain heuristic. Nevertheless, we found out that this option was only chosen when it was extremely difficult for our evaluators to evaluate a certain heuristic. For example, an item such as: "Do moderators have a place to gather and discuss certain problems?" is almost impossible to evaluate from a workers' perspective - which is how the evaluators were viewing the platform (they were not moderators in this case). Rather than heuristics being more or less applicable we feel that it is more challenging to identify those. For example, when it comes to the heuristic of *diversity*, the question that is raised is not whether this is applicable to crowdsourcing but rather how can a platform be aware of this? This question raises the issue of the need to invent mechanisms that would scale our approach or make it easier and more enjoyable.

As mentioned above we have adopted the definition by Estellés-Arolas and González-Ladrón-de-Guevara [49] which provides eight criteria, that crowdsourcing platforms need to fulfil to be practical and operational. Nevertheless, when checking platforms against these criteria our assessment diverged from those reported in [49]. For example, according to these authors YouTube does not qualify as a crowdsourcing platform since it only satisfies two of their eight defining criteria. For example, according to the authors YouTube members cannot earn a living by being a professional "YouTuber", an assertion that is no more true. Here we note temporal effects: platforms and practices shift in time, and these days one could consider YouTube as meeting all these criteria. Furthermore, YouTube visitors can be considered as the requesters of the platform, requesting more videos from their favourite YouTuber.

Another example that challenges Estellés-Arolas and González-Ladrón-de-Guevara [49] definition is Wikipedia. In their article, Wikipedia lacks three criteria to be considered a crowdsourcing platform. The first according to the authors is that the "crowdsourcer" isn't clearly identified. We would challenge this argument since the Wikipedia itself can be considered the crowdsourcer just like Threadless -again an example they consider a crowdsourcing platform. Another argument is that Wikipedia does not "use an open call of variable extent". Nevertheless, throughout the website Wikipedia makes clearly open calls to contribute in the form of articles. For

<sup>\*</sup> <https://www.innocentive.com/offering-overview/innocentivework>

the aforementioned reasons, we would like to revisit the definition of what crowdsourcing is and as an extension how does it differ from online communities. We propose four criteria:

1. There is an open call for contribution that means that is uncertain who will contribute
2. The open call is by a requester in a certain need that can vary (e.g. ideas, knowledge, work, creativity etc.)
3. The compensation can be variable from the requester: money, knowledge, experience, recognition, etc.
4. There is a platform that facilitates the process without contributing to the open call itself

When considering the aforementioned criteria, we can make a distinction between crowdsourcing platforms and online communities, as illustrated in Figure 17. The first circle illustrates crowdsourcing platforms such as MTurk, which is mostly a marketplace for workers and requesters, supported by the platform. The second circle is more unique, where the platform is the requester at the same time, as in the case of Quirky and Threadless. The third circle describes communities where the roles of requester and worker are unclear and perhaps dynamic. The work that we have conducted in this paper raises even more questions that could be addressed in future research. It is interesting to explore ways to guide the evolution of crowdsourcing platforms, the changing roles of workers and members, introducing stronger community elements to crowds, or allowing communities to spin-off a crowdsourcing platform.

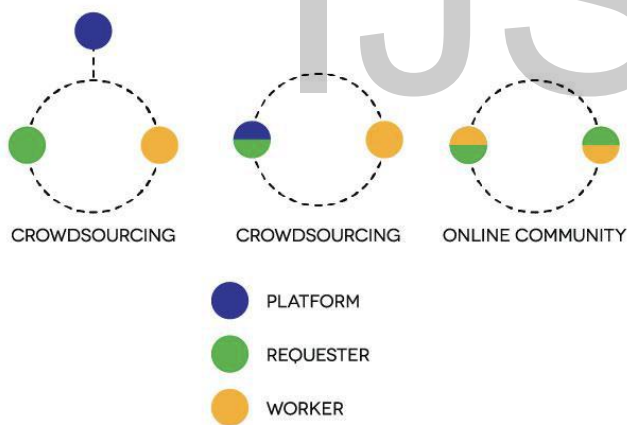


Fig. 17. Proposed distinction between crowdsourcing platforms and online communities.

## 6. Conclusion

This paper contributes a set of heuristics to support the expert review of crowdsourcing platforms. These criteria derive from the domain of online communities with some adaptations that pertain primarily to crowdsourcing platforms.

The heuristics consist of six major categories: 1. *Purpose*: identify the members' needs and the owner's goals on the platform and make the purpose visible. 2.

*Moderation*: monitor the platform, have easily accessible regulations and trained moderators who can be directly contacted. 3. *Members*: allow members to present themselves on their profile and offer deep profiling options, while giving a feeling of virtual co-presence. Guide members based on their life cycle and recruit new ones using your own members. 4. *Common ground*: organize events, introduce rituals and create sub groups while taking care to have a balanced diversity of members. 5. *Contribution*: provide the right threshold, motivators and targeted requests for members to contribute on the platform and manage expectations by request lists, request descriptions and showcasing previous contributions. 6. *Platform*: offer the right tools to facilitate your members, showcase their reputation and motivation of the platform and offer a unique and aesthetically professional platform.

We applied those heuristics to 20 mainstream crowdsourcing platforms with four evaluators. Our results show that these were largely applicable, supporting the argument that to a large extent crowdsourcing platforms are a special case of online communities. The heuristics helped identify and substantiate a number of possible improvements for different platforms, but also drew out some common attributes that characterize different types of platforms. Two specific elements we can generalize based on our survey that crowdsourcing platforms can improve are: guiding workers based on the specific stage in their life cycle, allowing workers to themselves become moderators of the platform and making targeted requests to workers.

In future work we aim to improve the efficiency of applying such heuristics. Potential directions include shortening the lists of heuristics, experimenting with different forms of presenting them and identifying potential redundancies between heuristics or prioritizing between them, e.g., based on the severity of the issues they help identify. For these directions we are currently experimenting in splitting the evaluation itself into microtasks to speed it up [51]. Having an evaluation of all crowdsourcing platforms enables us to compare them. Currently we are testing and talking with crowdworkers to make a comparison platform, enabling workers and requesters alike to support their decision making process in finding and requesting work done through existing platforms. Lastly, we are organizing workshops, based on the heuristics we presented in this paper, to designers who are building communities and crowdsourcing platforms. Our tools such as the heuristic cards and workshops details can be found at [www.acamponie.nl](http://www.acamponie.nl)\*

## References

- [1] Upwork Global Inc., About Us, (n.d.). <https://www.upwork.com/about/> (accessed November 18, 2016).
- [2] Y. Roth, The state of crowdsourcing in 2015, (2015) 1–18. <https://en.eyeka.com/resources/analyst-reports>.
- [3] Samasource, Q2 2016 IMPACT SCORECARD, 2016.

\* <http://www.acamponie.nl>

- [4] Miniwatts Marketing Group, World Internet Users Statistics and 2016 World Population Stats, (n.d.). <http://www.internetworldstats.com/stats.htm> (accessed September 4, 2016).
- [5] Crowdsourcing LLC, Directory of Sites, (n.d.). <http://reports.crowdsourcing.org/> (accessed May 27, 2016).
- [6] D. Geiger, E. Fieft, M. Rosemann, M. Schader, Crowdsourcing Information Systems –Definition, Typology, and Design, Proc. 33rd Int. Conf. Inf. Syst. 2012. Assoc. Inf. Syst. Electron. Libr. (AISEL). (2012) 1–11. <http://aisel.aisnet.org/icis2012/proceedings/ResearchInProgress/53> (accessed September 12, 2016).
- [7] M.L. Gray, S. Suri, S.S. Ali, D. Kulkarni, The Crowd is a Collaborative Network, in: Proc. 19th ACM Conf. Comput. Coop. Work Soc. Comput. - CSCW '16, ACM Press, New York, New York, USA, 2016: pp. 134–147. doi:10.1145/2818048.2819942.
- [8] A. Kittur, J. Nickerson, M. Bernstein, The Future of Crowd Work, Proc. CSCW '13. (2013) 1–17. doi:10.1145/2441776.2441923.
- [9] O. Stewart, J.M. Huerta, M. Sader, Designing crowdsourcing community for the enterprise, Proc. ACM SIGKDD Work. Hum. Comput. - HCOMP '09. (2009) 50. doi:10.1145/1600150.1600168.
- [10] W. Soliman, V.K. Tuunainen, Understanding continued use of crowdsourcing systems: An interpretive study, J. Theor. Appl. Electron. Commer. Res. 10 (2015) 1–18. doi:10.4067/S0718-18762015000100002.
- [11] D.C. Brabham, Moving the Crowd at Threadless: Motivations for participation in a crowdsourcing application, Information, Commun. Soc. 13 (2010) 1122–1145. doi:10.1080/13691181003624090.
- [12] H. Zheng, D. Li, W. Hou, Task Design, Motivation, and Participation in Crowdsourcing Contests, Int. J. Electron. Commer. J. Int. J. Electron. Commer. 15 (2011) 1086–4415. doi:10.2753/JEC1086-4415150402.
- [13] M.S. Silberman, L. Irani, J. Ross, Ethics and Tactics of Professional Crowdwork, Xrds. 17 (2010) 39–43. doi:10.1145/1869086.1869100.
- [14] B.B. Bederson, A.J. Quinn, Web workers unite! addressing challenges of online laborers, in: Proc. 2011 Annu. Conf. Ext. Abstr. Hum. Factors Comput. Syst. - CHI EA '11, ACM Press, New York, New York, USA, 2011: p. 97. doi:10.1145/1979742.1979606.
- [15] J. O'Neill, D. Martin, Relationship-based business process crowdsourcing?, in: Lect. Notes Comput. Sci. (Including Subser. Lect. Notes Artif. Intell. Lect. Notes Bioinformatics), Springer Berlin Heidelberg, 2013: pp. 429–446. doi:10.1007/978-3-642-40498-6\_33.
- [16] J. Rogstadius, V. Kostakos, A. Kittur, B. Smus, J. Laredo, M. Vukovic, An Assessment of Intrinsic and Extrinsic Motivation on Task Performance in Crowdsourcing Markets, ICWSM. 11 (2011) 17–21.
- [17] D. Schultheiss, A. Blieske, A. Solf, S. Staudtner, How to encourage the crowd? a study about user typologies and motivations on crowdsourcing platforms, in: Proc. - 2013 IEEE/ACM 6th Int. Conf. Util. Cloud Comput. UCC 2013, 2013: pp. 506–509. doi:10.1109/UCC.2013.98.
- [18] O. Stewart, D. Lubensky, J.M. Huerta, Crowdsourcing participation inequality, Proc. ACM SIGKDD Work. Hum. Comput. - HCOMP '10. (2010) 30. doi:10.1145/1837885.1837895.
- [19] D. Chandler, A. Kapelner, Breaking monotony with meaning: Motivation in crowdsourcing markets, J. Econ. Behav. Organ. 90 (2013) 123–133. doi:10.1016/j.jebo.2013.03.003.
- [20] A. Kittur, Crowdsourcing, collaboration and creativity, XRDS Crossroads, ACM Mag. Students. 17 (2010) 22. doi:10.1145/1869086.1869096.
- [21] M. Fuge, K. Tee, A. Agogino, N. Maton, Analysis of Collaborative Design Networks: A Case Study of OpenIDEO, J. Comput. Inf. Sci. Eng. 14 (2014) 21009. doi:10.1115/1.4026510.
- [22] L. Yu, J. V. Nickerson, Cooks or Cobblers? Crowd Creativity through Combination, Proc. 29th Annu. ACM SIGCHI Conf. Hum. Factors Comput. Syst. (2011) 1393–1402. doi:10.1145/1978942.1979147.
- [23] R. Laplante, M.S. Silberman, Building Trust in Crowd Worker Forums : Worker Ownership , Governance , and Work Outcomes, in: 2016.
- [24] S. Dow, A. Kulkarni, Shepherding the crowd yields better work, Comput. Support. Coop. Work. (2012) 1013–1022. doi:10.1145/2145204.2145355.
- [25] N. Salehi, L.C. Irani, M.S. Bernstein, A. Alkhatib, E. Ogbe, K. Milland, Clickhappier, We Are Dynamo: Overcoming Stalling and Friction in Collective Action for Crowd Workers, Proc. ACM CHI '15 Conf. Hum. Factors Comput. Syst. 1 (2015) 1621–1630. doi:10.1145/2702123.2702508.
- [26] L.M. Gallant, G.M. Boone, A. Heap, Five heuristics for designing and evaluating Web-based communities, First Monday. 12 (2007). <http://firstmonday.org/ojs/index.php/fm/article/view/1626/1541> (accessed September 10, 2016).
- [27] D. Gurzick, W.G. Lutters, Towards a design theory for online communities, in: Proc. 4th Int. Conf. Des. Sci. Res. Inf. Syst. Technol. DESRIST 09, ACM Press, New York, New York, USA, 2009: p. 1. doi:10.1145/1555619.1555634.
- [28] A. Kim, Jo, Community building on the Web, Peachpit Press, 2000.
- [29] R.E. Kraut, P. Resnick, S. Kiesler, Building successful online communities : evidence-based social design, MIT Press, 2011.
- [30] J. Preece, C. Abras, D. Maloney-Krichmar, Designing and evaluating online communities: research speaks to emerging practice, Int. J. Web Based Communities. 1 (2004) 2–18. doi:10.1504/IJWBC.2004.004795.
- [31] N. Reeves, R. Tinati, S. Zerr, M.G. Van Kleek, E. Simperl, From Crowd to Community, in: Proc. 2017 ACM Conf. Comput. Support. Coop. Work Soc. Comput. - CSCW '17, ACM Press, New York, New York, USA, 2017: pp. 2137–2152. doi:10.1145/2998181.2998302.
- [32] V.R. Carvalho, M. Lease, E. Yilmaz, Crowdsourcing for search evaluation, ACM SIGIR Forum. 44 (2011) 17. doi:10.1145/1924475.1924481.
- [33] J. Heer, M. Bostock, Crowdsourcing graphical perception, in: Proc. 28th Int. Conf. Hum. Factors Comput. Syst. - CHI '10, ACM Press, New York, New York, USA, 2010: p. 203. doi:10.1145/1753326.1753357.
- [34] P. Korshunov, S. Cai, T. Ebrahimi, Crowdsourcing approach for evaluation of privacy filters in video surveillance, in: CrowdMM 2012 - Proc. 2012 ACM Work. Crowdsourcing Multimedia, Co-located with ACM Multimed. 2012, ACM Press, New York, New York, USA, 2012: pp. 35–40. doi:10.1145/2390803.2390817.



- [35] Y. Zhao, Q. Zhu, Evaluation on crowdsourcing research: Current status and future direction, *Inf. Syst. Front.* 16 (2014) 417–434. doi:10.1007/s10796-012-9350-4.
- [36] J. Nielsen, R.L. Mack, *Heuristic evaluation*, Wiley, 1994.
- [37] J. Nielsen, R. Molich, *Heuristic Evaluation of user interfaces*, CHI '90 Proc. SIGCHI Conf. Hum. Factors Comput. Syst. (1990) 249–256. doi:10.1145/97243.97281.
- [38] J. Nielsen, 10 Heuristics for User Interface Design: Article by Jakob Nielsen, Jakob Nielsen's Alertbox. (1995) 1–2. doi:10.1016/B978-0-12-385241-0.00003-8.
- [39] S. Reiss, Multifaceted Nature of Intrinsic Motivation: The Theory of 16 Basic Desires., *Rev. Gen. Psychol.* 8 (2004) 179–193. doi:10.1037/1089-2680.8.3.179.
- [40] Y. Xu, Literature Review on Web Application Gamification and Analytics, 2011. <http://csdl.ics.hawaii.edu/techreports/11-05/11-05.pdf> (accessed November 18, 2016).
- [41] M. Ma, An identity-based theory of information technology design for sustaining virtual communities, in: *Int. Conf. Inf. Syst. (ICIS '04)*, 2004: pp. 559–570. <http://aisel.aisnet.org/icis2004/44/>.
- [42] D.R. Raban, M. Moldovan, Q. Jones, An empirical study of critical mass and online community survival, in: *Proc. 2010 ACM Conf. Comput. Support. Coop. Work - CSCW '10*, ACM Press, New York, New York, USA, 2010: p. 71. doi:10.1145/1718918.1718932.
- [43] W. Mason, D.J. Watts, Financial incentives and the "performance of crowds", in: *Proc. ACM SIGKDD Work. Hum. Comput. - HCOMP '09*, ACM Press, New York, New York, USA, 2009: p. 77. doi:10.1145/1600150.1600175.
- [44] R.M. Ryan, E.L. Deci, Intrinsic and Extrinsic Motivations: Classic Definitions and New Directions, *Contemp. Educ. Psychol.* 25 (2000) 54–67. doi:10.1006/ceps.1999.1020.
- [45] C. Rusu, V. Rusu, S. Roncagliolo, D. Quinones, V.Z. Rusu, H.M. Fardoun, D.M. Alghazzawi, C.A. Collazos, Usability Heuristics: Reinventing the Wheel?, in: *Springer, Cham*, 2016: pp. 59–70. doi:10.1007/978-3-319-39910-2\_6.
- [46] J. Mankoff, A.K. a. K. Dey, G. Hsieh, J. Kientz, S. Lederer, M. Ames, Heuristic Evaluation of Ambient Displays, *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.* (2003) 169–176. doi:10.1145/642611.642642.
- [47] D. Pinelle, N. Wong, T. Stach, Heuristic evaluation for games: usability principles for video game design, *Proc. SIGCHI Conf. Hum. Factors Comput. Syst.* (2008) 1453–1462. doi:10.1145/1357054.1357282.
- [48] K. Baker, S. Greenberg, C. Gutwin, Heuristic Evaluation of Groupware Based on the Mechanics of Collaboration, *Eng. Human-Computer Interact.* 2254 (2001) 123–139. doi:10.1007/3-540-45348-2\_14.
- [49] E. Estellés-Arolas, F. González-Ladrón-de-Guevara, Towards an integrated crowdsourcing definition, *J. Inf. Sci.* 38 (2012) 189–200. doi:10.1177/0165551512437638.
- [50] J. Nielsen, T.K. Landauer, A mathematical model of the finding of usability problems, in: *Proc. INTERCHI '93 Conf. Hum. Factors Comput. Syst.*, IOS Press, 1993: p. xviii, 547 .
- [51] S. à Campo, V.J. Khan, Crowdsourcing community heuristic evaluations, (2017). doi:10.18420/ecscw2017\_p6.