

# Understanding the complementary nature of paid and volunteer crowds for content creation

Claudia Flores-Saviaga, Ricardo Granados, Liliana Savage, Lizbeth Escobedo, Saiph Savage

Published: 30 November 2020

## Abstract

Crowdsourced content creation like articles or slogans can be powered by crowds of volunteers or workers from paid task markets. Volunteers often have expertise and are intrinsically motivated, but are a limited resource, and are not always reliably available. On the other hand, paid crowd workers are reliably available, can be guided to produce high-quality content, but cost money. How can these different populations of crowd workers be leveraged together to power cost-effective yet high-quality crowd-powered content-creation systems? To answer this question, we need to understand the strengths and weaknesses of each. We conducted an online study where we hired paid crowd workers and recruited volunteers from social media to complete three content creation tasks for three real-world non-profit organizations that focus on empowering women. These tasks ranged in complexity from simply generating keywords or slogans to creating a draft biographical article. Our results show that paid crowds completed work and structured content following editorial guidelines more effectively. However, volunteer crowds provide content that is more original. Based on the findings, we suggest that crowd-powered content-creation systems could gain the best of both worlds by leveraging volunteers to scaffold the direction that original content should take; while having paid crowd workers structure content and prepare it for real world use.

## Keywords:

Crowdsourcing; Social Good; Wikipedia; Content Creation.

## 1 Introduction

Organizations often need to create and communicate content, either online or printed, to promote their work. However, content creation involves not only financial, but human resources as well. Nonprofit organizations have limited budgets [21].

This means that they need to be smart about how to efficiently use the available resources to achieve their goals. One increasingly

attractive option is to crowdsource their content creation needs [17][23]. This can involve the creation of a catchphrase or idea generation for new products for the organization [3][16]. There are two main ways organizations with a limited budget can crowdsource content creation tasks: Volunteers and crowd workers.

Volunteers could help out with or without a small remuneration. Many of these volunteers are intrinsically motivated to assist in the agenda that the organization is pushing forth; consequently, they may be motivated to engage in a great amount of work for the organization for little benefit [20]. One potential source of volunteers is social media [12] and can be asked for simple tasks such as gathering information [19] or content creation [6]. The content creation can be more diverse than face-to-face discussion [9]. The disadvantages are that it can take on extremely varied response rates [18], and deflates the degree of social capital. Social capital means all the resources an individual has at their command, all of which can be extracted from his or her social network [4].

Although, researchers have explored the implementation of social media as a way to gather and use volunteer participation [15], past research has focused on studying contributions from people's social capital [2][5][26]; this is a problem as some organizations, especially new ones, might not have a large network to rely on. Previous work has also explored the possibility of asking strangers on social media questions [15][26]. Though intrinsic motivation should follow on to higher quality production results [1], previous studies have found that the design content creation, such as creating a slogan or a full Wikipedia biography, taken from online communities can be of both a lower quality and quantity than organizations expect [13][25].

The other way an organization can crowdsource content is to post the work to a crowd marketplace which provides a more stable source of workers, but the quality of their work may also be questionable. This type of workers tend to complete their work primarily based on extrinsic motivations [14]. A financially motivated audience can be accessed through analysis platforms such as Voyant, or commercial platforms such as Upwork or Amazon Mechanical Turk (AMT) [14]. Potential drawbacks include financial costs: although a single instance of content creation

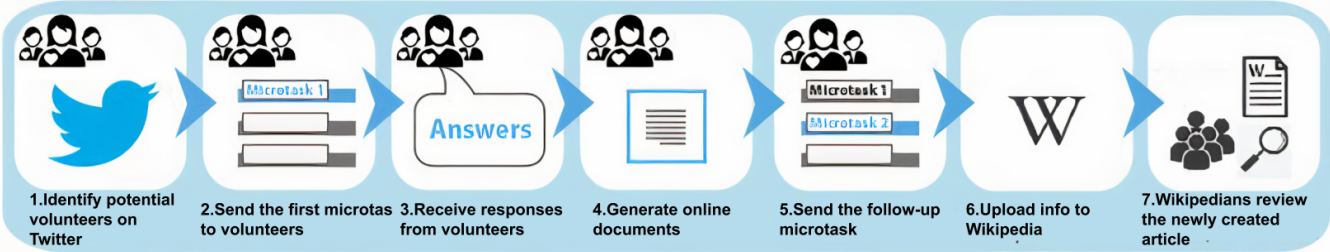
---

Flores-Saviaga, C.  
West Virginia University  
cif0001@mix.wvu.edu

Granados, R., Savage, L., Savage, S.  
Universidad Nacional Autónoma de México  
disavage@gmail.com, saiph.savage@mail.wvu.edu

Escobedo, L.  
CETYS  
lizbeth.escobedo@cetys.mx

## a) Volunteer Crowds



## b) Paid Crowds

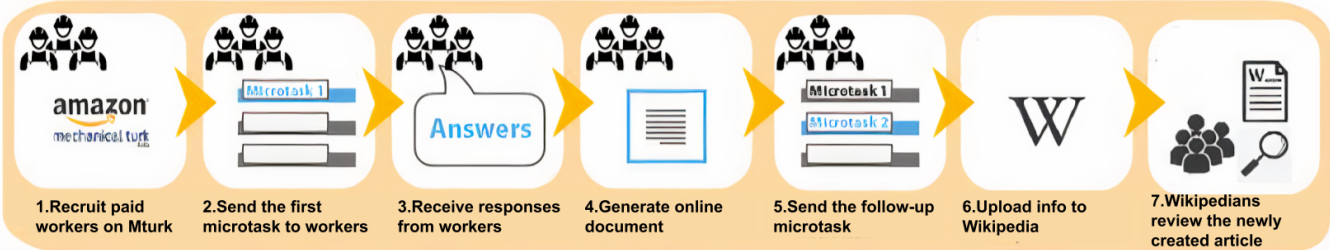


Fig 1. Workflow followed by online volunteers and crowd workers to produce a biography for Wikipedia

generation is typically affordable (e.g. \$10 U.S. dollars), generating content creation for many other instances could get expensive quite quickly.

Organizations with limited budgets may wish to mix and match these different populations of workers, depending on the precise nature of their priorities (e.g., time to completion, level of originality, reliability of content quality) and available resources (e.g., access to social capital, monetary constraints). To make these decisions in a principled way, requires an understanding of the precise strengths and weaknesses of each population of crowd workers. While previous work has studied how crowd workers compare with volunteers for more simple jobs, including marking related pictures or providing micro-content creation [6][10], there is a lack of empirical research on the strengths and weaknesses of each population of workers for content creation tasks. This knowledge could be very useful for organizations deciding whether and how to leverage crowdsourcing for their content creation needs. In this paper, we investigate the strengths and weaknesses of paid and volunteer crowds for content creation tasks. We focused on the following research questions:

- How do different crowds compare in terms of the originality of the content task generated results?
- How do different crowds compare in terms of usefulness of the answers?
- How do crowds compare in terms of accuracy following instructions?

To answer these research questions, we conducted an online study, in which volunteers and crowd workers were recruited to execute content creation tasks for three real world NGOs focused on empowering women. Volunteers were recruited and orchestrated directly from social media; crowd workers were recruited in AMT and orchestrated through Chorus<sup>1</sup> [8], a crowd-powered dialogue system. Our content creation tasks ranged in complexity from simply providing keywords, generating a slogan for the NGO to create a biography on Wikipedia. To gain a nuanced

understanding of the strengths and weaknesses of volunteers and crowd workers, we measured key subcomponents of quality: originality, usefulness, and accuracy to follow instructions.

Our work, to our understanding, is the first to directly compare crowd workers and volunteers, and explore a range of indicators of work quality (e.g. quantity, quality, and degree of content) to discover possible trade-offs. Our methodology also demonstrates how social media volunteers can be orchestrated to execute complex content creation tasks directly on social media via simple text messages

## 2 Crowds

In the following section we present the crowds we considered in our study: volunteer crowds from social media and crowd workers. All the participants for this study were anglophone speakers. Due to privacy we only gather information about the job done but not about the person that participated in the activities. All the responses received from a participant that lead to a completed task were stored by a human agent into the database.

### 2.1 Volunteer Crowds

Volunteer crowds were recruited on social media using Twitter streaming API. Based on the previous work [6] we developed a workflow that monitors Twitter activity real-time and listens for words related to feminism, identifying suitable volunteers in order to send questions to them.

The experimental workflow, Figure 1.a shows the process followed to collect information for each task on Twitter. The first step was to identify potential volunteers by listening in on a real-time filtered feed of publicly posted Tweets. Filtering was primarily done by keywords related to feminism, such as "#women", "#genderequality", "#feminism", "#womenrights", among others. When a potential volunteer was identified, the first microtask was sent. Microtasks were sent as @replies to one of the recent relevant tweets, thus giving some context for why the microtask was directed to them. Next, a human operator would monitor Twitter for responses to the microtask. If the volunteer responded to the

<sup>1</sup> <http://talkingtothecrowd.org/>

first microtask with a relevant response, then the follow-up microtask would be sent.

## 2.2 Paid Crowds

In this study, the paid crowd was recruited from AMT. We used a third-party software, Chorus [8], which is a conversational agent powered by crowdsourcing, to communicate with the worker. One of the key design goals of Chorus is to allow end-users to naturally develop open dialogs with crowd workers without being aware of the boundaries of conversational sessions. This type of open dialog is very similar to the kinds of conversational threads that occur on Twitter, which also do not have clear communication boundaries. Therefore, it was useful to compare with the social media volunteers recruited on Twitter. Workers in Chorus get paid with \$0.2 per conversational session lasting 11 minutes on average. On top of the base rate, to incentive workers, they were granted extra bonus money according to their contributions to the conversation.

The workflow used to dispatch microtasks to crowd workers can be seen in Figure 1.b. To recruit workers, a human agent initiated a conversation on Chorus, to send them the first microtask, the human agent waited for the feedback from the workers and kept sending the follow-up microtasks. The agent stored the responses from crowd workers for further analysis.

## 3 Content creation tasks and metrics

Our goal was to understand the type of content that volunteers and paid crowd workers produced on demand. For this purpose, we sourced to these two groups different content creation tasks, and studied how: (1) original and (2) useful the content created was; and (3) how accurate the people followed instructions. Each of these metrics and the methodology to evaluate them will be explained in detail in the following sections.

We studied these three metrics because they are some of the most important points organizations and communities consider when deciding what content to incorporate. We were particularly interested in creating a body of knowledge that organizations could use to decide what crowd to select for their content creation needs. The focus of our investigation is on the case study of non-profits centered on feminism. We selected this area because two of the co-authors had direct involvement with NGOs working on feminism. This facilitated getting real world feedback about the content that volunteers and crowd workers were generating. As a side note, one critical point of several organizations is the lack of coverage of women. Therefore, understanding how crowd workers or volunteers could be best leveraged to fight this information bias is important.

We asked crowd workers and volunteers to participate in creating content for three different tasks. These tasks were common to most NGOs, and had a growing level of complexity. Thus allowing us to better understand the strengths and weaknesses of content creation in each type of crowd. We measured task complexity based on the number of micro-tasks involved. The tasks we considered were: keywords, slogan, biography of a famous woman. Each of these tasks was conducted for a real world non-profit organization.

### 3.1 Keywords Task

Many NGOs struggle with identifying the best keywords to be more searchable online, or to better describe themselves to their potential volunteers and sponsors. Therefore, the objective of this task was to get five words that best described an NGO after visiting their

website and reading about their goals. We consider this task the simplest one in our study. For this task, people from each crowd were given an image that contained a brief description of the non-profit organization along with the website. For this task we focused into creating content for an NGO focused on empowering women in technology<sup>2</sup>.

### 3.2 Slogan Task

The objective of this task was to create a slogan, a "catchy phrase", that describes the organization, as well as providing an image to go along with the phrase. For this task we focus into creating content for an NGO focused on empowering women in digital journalism<sup>3</sup>. Note that this task involved two microtasks:

- 1) **Slogan Microtask.** Consist in writing a phrase or motto, describing the NGO.
- 2) **Image Microtask.** Consist in providing an image related to the phrase.

### 3.3 Biography Task

The objective of this task was to create a short biography of a noteworthy woman. We focus this particular task on one of the largest knowledge sources of the world Wikipedia. We chose Wikipedia because coverage of important women is a critical problem faced by the organization, where gender-oriented disparity in articles exist throughout the entire encyclopedia [24]. The Wikipedia Biography task consists of three microtasks:

- 1) **Request Name Microtask.** Consists in providing the name of a woman who did not have a biography on Wikipedia and that they considered was noteworthy enough.
- 2) **Request Information Microtask.** Consists in giving basic biographical information about the suggested woman, as well as references that support the information.
- 3) **Structure Information Microtask.** Consists on structuring and refining the data that was provided about the woman, to produce a final text of the biography that could be directly incorporated into Wikipedia.

For the latter microtask we provided crowds with a link where we had collected all data from the previews microtasks that they had contributed collectively. This link was an online document that they could edit together to prepare their final biography. Note that paid crowds did not have access to the online documents from volunteers, and vice versa. The biographies that each crowd generated were directly uploaded by a human agent to Wikipedia.

## 4 Evaluation

We requested tasks from volunteers and crowd workers from July 15th until September 7th, 2016. We guided crowds to generate content via a series of questions. Each task thus involved a series of questions, and each microtask was tied to one question. Table 1 presents the different questions we used for each microtask. We alternated between tasks for which we requested content, to minimize any sequence effect.

For the keyword and slogan tasks real world NGOs helped us to validate the data. And for the biography task we submit the biography to Wikipedia and wait for the moderators comments, to evaluate the work. As said before we study originality, usefulness and accuracy to follow instructions. We analyzed only the final content that crowds created for each microtask. We did not consider

<sup>2</sup> <http://www.eniac.org.es/english/home/>

<sup>3</sup> <http://www.chicaspoderosas.org/>

off-topic responses in our analysis. The results for keywords and slogan are presented together as the methodology to evaluate originality is the same, afterwards we present the results for the biography task that uses a different one.

### 4.1 Keyword and Slogan Tasks

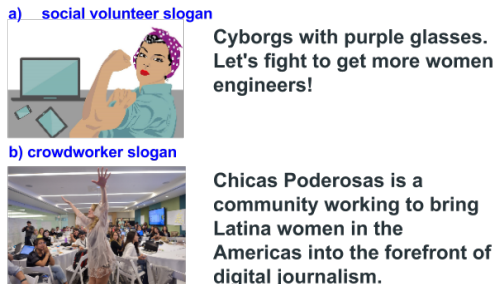


Figure 2. Example of the Slogan content that a) volunteers and b) paid crowd workers provided.

Volunteers and paid workers generated 20 slogans, and provided 20 sets of keywords. Figure 2 shows an example of the content that a) volunteers produced; and b) paid crowds produced for the Slogan task.

#### 4.1.1 Method: originality

To measure originality we considered that a piece of content was original when it did not appear on the NGO's official website. We use this method to be able to qualify the active participation and interest of the people, considering that copy-paste action would mean non involvement or interest at all. For the case of keywords we verified on the "ENIAC" website and for slogan we checked on "Chicas Poderosas" website. For both keyword and slogan tasks, we calculated the percentage of words that did not appear on the site over the number of words that were provided. And lastly, for the slogan we also measured the originality of the images provided (did the image appear or not on the page's official site).

Table 1. Text description used in each microtask.

Task	Micro-Task	Question Text
Keywords	1	Can you give me five words that you would use to describe the work of [non-profit name]?
Slogan	1	Can you give us a slogan that we could use to promote the work of [non-profit name]?
Slogan	1	Can you send us an image that you think could complement the slogan you just created?
Wikipedia Bio	1	Can you give me the name of a noteworthy woman who you think should have a biography on Wikipedia but currently does not?
Wikipedia Bio	2	Can you give me facts about her?

Table 2. Originality statistics per crowd on each task.

	Keywords	Slogan
Volunteer Crowd	51%	100%
Paid Crowd	15%	73%

#### 4.1.2 Results: originality

Table 2 summarizes the originality ratings of volunteers and paid crowds. We can observe that volunteers gave more original responses than crowd workers. It is interesting to observe that even for complex tasks that involve generating a new motto and an image, volunteers were able to produce original content. However, it is important to notice that this creativity should not be confused with usefulness. An answer could be original, meaning, it is not a copy/paste from the non-profit web page, but it could end up not being useful for the NGO. For the Keywords task, volunteers had significantly more original responses (51%) than paid workers (15%), z test of difference in proportions = 5.53, p<.001. The same was true for the Slogan task, where volunteers produced significantly more original on-topic answers (100%) than paid workers (73%), z = 2.94, p<.01.

#### 4.1.3 Method: usefulness

We implemented usefulness as a measure to assess how helpful the content produced is for an NGO and if the NGO would use it. In order to measure usefulness, we asked two members of the NGOs participating in the study to categorize the answers given by each crowd for the Keyword and Slogan Tasks. Only two categories were used: Useful or Trash. Useful was content that the organization felt that they could use, and Trash was content that they would likely never use. Each piece of content was reviewed by the NGOs members independently, and decided whether the organization would likely use it or not. In case there was a tie, a third volunteer was called to undo the tie and reach an agreement about the categorization. The two coders agreed on 97% of the answers with a Cohens kappa = .94 for the Keywords task, and Cohens kappa = .94 for the Slogan task.

Table 3. Usefulness statistics per crowd to complete each task.

	Keywords	Slogan
Volunteer Crowd	46%	44%
Paid Crowd	100%	89%

#### 4.1.4 Results: usefulness

From Table 3 we can note that across tasks the paid crowd tended to give more useful content. For the Keyword task 100%, z = -3.99, p < .001, of the paid workers delivered only useful words while the social media volunteers gave less useful content for the NGO. The same effect was observed in the Slogan task where almost 90%, z = -2.80, p < .01, of paid workers delivered useful content while many volunteers tended to provide content that seemed to be more to support the general cause of empowering women, but were not necessarily related to the NGO. For instance, one person gave to the NGO of "chicas poderosas" (an organization focused on digital journalism and women) the slogan of: "cyborgs with purple glasses. Let's fight to get more women engineers!".

#### 4.1.5 Method: accuracy to follow instructions

We were also interested in studying the accuracy of each crowd to follow instructions. In this case we use the word accuracy to the ability of the social media volunteer or crowdsourcing worker to

follow the instructions exactly as they were provided, no more, no less that was asked for. In many cases organizations need content that has a certain format or follows certain standards to use it. Therefore, we evaluate the compliance on the instructions, penalizing any deviation.

For the Keywords task, the instruction was to provide only five words describing the given NGO. We penalized if they provided more words, or if they provided fewer words. The capability to follow instructions was calculated on a scale from 0 to 100, and each word counted 20 points, therefore, in order to get the maximum score, five words should have been given. We penalized people for each extra keyword that was given by removing 20 points.

For the Slogan task, we measured how much crowds followed instructions per micro-task. For the first microtask we gave people 100 points if they gave us some phrase and 0 if not. For the second microtask we gave people again 100 if they provided an image for the slogan and 0 if not.

#### 4.1.6 Results: accuracy to follow instructions

In Figure 3 we observe that the paid crowd tended to follow guidelines more than volunteers. For the Keyword task we observed that only 38% of all submissions followed the specified instructions of providing five words. Meanwhile, 86% of the submissions from crowd workers followed the established rule. This difference was statistically significant,  $z = -3.71$ ,  $p < .001$ .

Volunteers tended in general to provide more than what was requested from them. But consequently ended up not following the specifics of the task. This result is aligned with the findings of Yen et al. [9], which found that people on online forums tended to give the most responses to a small feedback related task.

From Figure 3 we see that volunteers followed instructions more accurately for the particular micro-task of providing a motto for the NGO, where over 70% of the content that crowds created followed the specifics. For this microtask we did not observe a statistically significant difference in how many volunteers and paid workers followed instructions,  $z = -1.13$ ,  $p = 0.13$ .

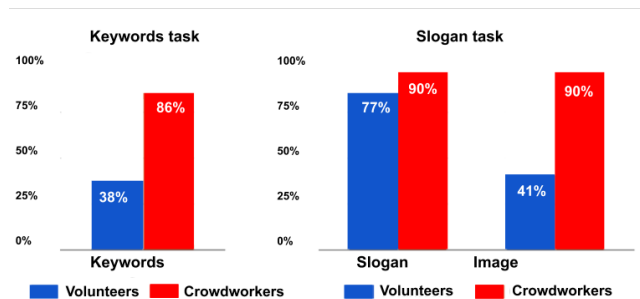


Figure 3. Capability to follow instructions.

For the microtask of providing an image, less than 45% of social media volunteers provided an image. Paid workers, on the other hand, 90% of the time followed the instructions exactly. For this microtask, we see a difference between how accurate volunteers and crowd workers followed instructions,  $z = -3.60$ ,  $p < .001$ .

We believe that volunteers are not consistent with how much they follow instructions likely due to their intrinsic motivations [20]. Since they are not paid, volunteers value having flexibility in

how they complete work [7]. It is thus likely that volunteers are focusing on the aspect of the task that they most care about, or find the most fulfilling.

## 4.2 Biography Task

To provide a more externally valid test of our research questions, we decided to carry on a content creation task deployment on the popular online encyclopedia, Wikipedia<sup>4</sup>. Wikipedia is special in this endeavor as its articles can be edited by anyone. We decided to test out both crowds in this environment.

Figure 1 shows the main question-asking flow used for collecting answers for each microtask from both crowds. For the “Request Name Microtask”, the human agent would dispatch the request to both crowds, and wait until an original name was received. Once this task was accomplished the “Request Information Microtask” was dispatched. The human agent waited for the biographical information given by many of the social media volunteers and paid workers and stored it in a sandbox. Once this small work assignment was accomplished, the “Structure Information Microtask” was dispatched.

### 4.2.1 Method: originality

For the “Request Name Microtask”, we evaluated if volunteers and crowdworkers provided a biography that was not on Wikipedia. This shows the active participation of the volunteers and crowdworkers to do a brief research of the existent biographies on Wikipedia as well their commitment to the cause to search for a name that is missing.

For the “Request Information Microtask”, we evaluated whether they provided references or not.

### 4.2.2 Results: originality

On the “Request Name Microtask”, volunteers tended to give more original names of women than the paid crowd had. We observed that the paid crowd had a hard time coming up with women that did not have a biography on Wikipedia. In general, they tend to give push-back answers such as “Everyone is on Wikipedia already”, “Personally I can't think of any”. When this happened, we suggested a name to the crowd worker to keep up testing the next work assignments.

For the “Request Name Microtask”, 82% of crowd workers gave us a name, while 86% of social media volunteers did. This percentage changed for the “Request Information Microtask”, where references are asked. Volunteers only 54% gave references, while 100% of the crowd workers asked gave references.

From the “Biography Task”, the data that we obtained was that 43% of the people who made it until the second work assignment, completed structured the information to create the biography, while 100% of crowd workers did.

### 4.2.3 Method: usefulness

We tested the usefulness of the structured biography on Wikipedia using ORES<sup>5</sup>, (Objective Revision Evaluation Service), a web service designed by Wikimedia Foundation to help detect and remove vandalism and the survival rate in order to check how many articles were not deleted by the community. We considered a biography useful if it had not been marked for deletion by the Wikipedia community.

ORES uses three quality models to rate editions to Wikipedia articles: “damaging”, “good faith” and “reverted”. With probabilities from 0 to 1. We considered an observation period of 1

<sup>4</sup> <https://www.wikipedia.org>

<sup>5</sup> <https://ores.wikimedia.org/>



week to wait for the feedback of the community on Wikipedia as the patrolling process that the community uses to flag articles or delete them ensures that 80% of all new articles are patrolled within an hour of creation, and 95% within a day [11]. The mechanism by which articles can be flagged, starts right after the article has been created. Once the article is on Wikipedia, it appears on a special page that is monitored by trusted, verified Wikipedians who review the new article and can flag it for deletion.

**4.2.4 Results: usefulness**

We used Wikipedia to test out in the wild what happened with the information given by both crowds in the work assignment to structure the biography. Only the users from each crowd that completed the last work assignment from the Biography task pipeline were suitable for usefulness testing from the Wikipedia community. This is, they gave us full biography information about a woman, and the information was structured enough to upload to Wikipedia.

From the social media volunteers, 3 biographies were created on Wikipedia, 1 of them was marked for deletion in less than 24 hours, and two are still unreviewed. From the crowd workers, 4 stubs were created on Wikipedia, and none of them were marked as deleted. The articles that were not deleted, received feedback mainly related to the necessity to add citations, some editing and grammar correction.

**Table 4. ORES results. The only bio that survived from the volunteers was for Magaly Pineda Tejada.**

Bio	Damaging (true)	Good faith (true)	Reverted (true)
Magaly Pineda Tejada	0.026	0.986	0.041
Laura J. Esserman	0.008	0.993	0.009
Virginia Hubbell	0.008	0.993	0.014
Women of the Apollo Program	0.004	0.996	0.009
Suzanne RD Tata	0.008	0.993	0.013
Pamela Palenciano	0.013	0.993	0.017

ORES results for the quality models (see Table 4.) show that the information uploaded to Wikipedia was not considered damaging (a probability closer to 0 for the damaging model indicates that the edit was not considered to cause damage), they were saved in good faith, (a score closer to 1 indicates that the edit was saved in good faith), and had a low probability of being reverted (a score closer to 0 indicates that a low probability exists that it will be reverted).

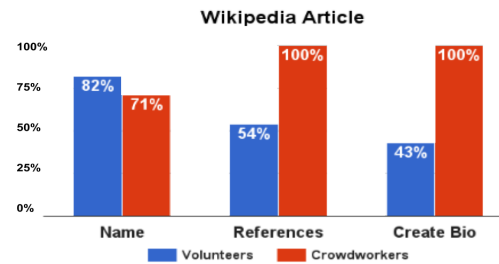
**4.2.5 Method: accuracy following instructions**

We calculated per microtask, on a binary scale of 0 or 100. For the “Request Name Microtask” a 100 was given if the woman was not already on Wikipedia, and 0 if the woman was already on Wikipedia, or if it was a spam response (meaning it was any woman name). For the “Request Information Microtask” 100 was given if the user provided at least one piece of information about the requested name. In case of the “Structure Information Microtask”,

100 was given if the user structured the information given, or at least did some structuring in a sandbox.

**4.2.6 Results: accuracy following instructions**

Figure 4 illustrates differences in reliability between the volunteer and paid workers. While a majority (82%) of the volunteers followed instructions for the “Request Name Microtask”, only approximately half of them followed instructions in the subsequent microtasks (54% for “Request Information Microtask”, and 43% for the “Structure Information Microtask”). In contrast, paid workers followed instructions at similar rates to volunteers for the first microtask (71% of workers,  $z = -1.37, p = .17$ ), but were significantly more willing to complete subsequent microtasks, once they were given a name to work on; 100% of workers completed the “Request Information Microtask”,  $z = 1.68, p = .09$ ; and 100% of workers completed the “Structure Information Microtask”  $z = 1.89, p = .06$ .



**Figure 4. Capability to follow instructions in real world deployment.**

**5 Discussion**

In this study, our objective was to investigate the strengths and weaknesses of volunteers and paid crowd workers for crowdsourced content creation. Our overall goal was to generate knowledge that can be used by organizations to make appropriate decisions regarding which combinations of workers they should use depending on their priorities (e.g., time to completion, originality) and available resources (e.g., social capital, budget).

**Table 5. Summary of results.**

Bio	Damaging (true)	Good faith (true)	Reverted (true)
Keyword	Originality	51%	15%
	Usefulness	46%	100%
	Accuracy following instructions	38%	86%
Slogan	Originality	100%	73%
	Usefulness	44%	89%
	Accuracy following instructions	77%	90%
Biography	Originality	More names	Less names
	Usefulness	More rejected	More accepted
	Accuracy following instructions		

Our experiments found that volunteers typically generate ideas with more originality. Meaning that this crowd gets more involved and invested in the task, as they are more dedicated to try to provide original and no copied information. As we can see from Table 5, The volunteer crowds were better than paid workers for all the originality tests that we evaluated. However, the main drawback that we found is that they tend to generate more spam responses than crowdworkers. This could be explained as this is a volunteer work they tend to put more effort into the tasks that they found more interesting and putting aside the instructions.

Paid crowd workers are better at following instructions and producing high-quality content. From Table 5 we can observe that usefulness and accuracy following instructions was totally dominated by the paid workers. Crowd workers tend to follow instructions and avoid working more than required, providing exactly the information and activities that were asked. This makes their work much more useful in terms that their results are ready to use by the NGOs. Nevertheless, they struggled in terms of original creation, this was the case despite the fact that crowd workers who regularly contribute to Chorus may be more creative than average crowd workers, since participation in Chorus exposes them to many diverse experiences and problems [8].

An organization needs to specify the type of content creation to decide which type of worker is needed. If the task needs originality the best option is to work with volunteers on social media. This comes with the limitation that the organization will have to deal with a larger number of responses and maybe not all of them tailored in the way they expect. This will insert possible delays of data cleaning and organization that need to be taken into account. On the other hand, if the task requires accuracy following instructions and with specific structure to be useful, it is better to use crowd workers. In this case the organization has to organize and plan the activities and tasks with great detail, as these workers will tend to avoid overdoing or providing more than what is asked. This way requires that the organization is aware of the exact needs and steps to tackle their content needs.

## 6 Limitations and future work

An important limitation of our study is that we did not request or track the identity of the social media volunteers or crowdworkers due to the privacy implications. This is important if other elements that could also contribute to the quality of results, such as the participant's domain experience, their content creation expertise, and their demographics are needed to be tested. Future work is needed to analyze how these factors influence the results obtained from content creation requests online and to study the generalizability of our results for different instances of the crowd categories tested in this work. We look forward to testing the results obtained by tackling social media volunteers with more domain expertise, different content creation tasks, different sizes of social networks, and offer visualization of the task created [22].

Another important issue not considered in this work are the potential ethical considerations of engaging paid crowd workers for certain kinds of NGO work, such as advocacy. For some nonprofits, this may mean that paid crowd workers are a non-starter. Yet, our findings still have useful implications for such settings, as they provide such nonprofits with information about what gaps they may need to fill in terms of their content creation needs if they exclusively rely on volunteers (e.g., disciplined, high-quality work). These gaps might be filled with dedicated, highly-trained staff, or perhaps also with research on guidance mechanisms for improving volunteer work.

## 7 Conclusion

Organizations are more and more often accessing social networks and paid task markets in order to recruit personnel to help out accomplish their goals. To promote their work, organizations like nonprofits need to recruit volunteers or crowdworkers for some of their content creation needs. In this study, we presented a comparison of Twitter volunteers and crowdworkers from Amazon Mechanical Turk to execute content creation tasks with different levels of complexity such as suggesting keywords to describe a non-profit organization, creating a slogan for a non-profit organization, and a real world deployment to create a Wikipedia biography. We discovered that volunteers should be leveraged if the organization is in need of original answers as they tend to generate more original ideas than crowdworkers; however, if the task requires that precise instructions to be followed or a certain level of structuring such as the one where text requires to be referenced, then a paid crowd market could be the best option. Alternatively, one organization could leverage volunteers to scaffold the direction that original novel content should take, while having paid crowd workers structure content and prepare it for real world use. We hope our results will enable and encourage organizations to more effectively harness the immense potential of both crowds to execute content creation tasks, to get the best of both worlds and generate content that best suits their needs.

## 8 References

- [1] Amabile, Teresa M. "Motivation and creativity: Effects of motivational orientation on creative writers." *Journal of personality and social psychology* 48.2 (1985): 393.
- [2] Brady, Erin, Meredith Ringel Morris, and Jeffrey P. Bigham. "Social microvolunteering: Donating access to your friends for charitable microwork." *HCOMP*. 2014.
- [3] Boudreau, Kevin J., and Karim R. Lakhani. "Using the crowd as an innovation partner." *Harvard business review* 91.4 (2013): 60-9.
- [4] Burke, Moira, Robert Kraut, and Cameron Marlow. "Social capital on Facebook: Differentiating uses and users." *Proceedings of the SIGCHI conference on human factors in computing systems*. 2011.
- [5] Farzan, Rosta, Saiph Savage, and Claudia Flores Saviaga. "Bring on board new enthusiasts! A case study of impact of Wikipedia art+ feminism edit-a-thon events on newcomers." *SocInfo2016*.
- [6] Flores-Saviaga, Claudia, Saiph Savage, and Dario Taraborelli. "Leadwise: using online bots to recruit and guide expert volunteers." *CSCW*. 2016.
- [7] Harp, Elizabeth R., Lisa L. Scherer, and Joseph A. Allen. "Volunteer engagement and retention: Their relationship to community service self-efficacy." *Nonprofit and voluntary sector quarterly* 46.2 (2017): 442-458.
- [8] Huang, Ting-Hao Kenneth, et al. "Is There Anything Else I Can Help You With?" Challenges in Deploying an On-Demand Crowd-Powered Conversational Agent." *HCOMP*. 2016.
- [9] Hui, Julie S., Elizabeth M. Gerber, and Steven P. Dow. "Crowd-based design activities: helping students connect with users online." *DIS*. 2014.

- [10] Kim, Joy, and Andres Monroy-Hernandez. "Storia: Summarizing social media content based on narrative theory using crowdsourcing." *CSCW*. 2016.
- [11] Kumar, Srijan, Robert West, and Jure Leskovec. "Disinformation on the web: Impact, characteristics, and detection of wikipedia hoaxes." *WWW*. 2016.
- [12] Lovejoy, Kristen, and Gregory D. Saxton. "Information, community, and action: How nonprofit organizations use social media." *Journal of computer-mediated communication* 17.3 (2012): 337-353.
- [13] Marlow, Jennifer, and Laura Dabbish. "From rookie to all-star: professional development in a graphic design social networking site." *CSCW*. 2014.
- [14] Mason, Winter, and Duncan J. Watts. "Financial incentives and the performance of crowds." *Proceedings of the ACM SIGKDD workshop on human computation*. 2009.
- [15] Nichols, Jeffrey, et al. "Analyzing the quality of information solicited from targeted strangers on social media". *CSCW* 2013.
- [16] Poetz, Marion K., and Martin Schreier. "The value of crowdsourcing: can users really compete with professionals in generating new product ideas?." *Journal of product innovation management* 29.2 (2012): 245-256.
- [17] Rosen, Peter A. "Crowdsourcing lessons for organizations." *Journal of Decision Systems* 20.3 (2011): 309-324.
- [18] Rzeszotarski, Jeffrey M., and Meredith Ringel Morris. "Estimating the social costs of friendsourcing." *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 2014.
- [19] Savage, Saiph, Andres Monroy-Hernandez, and Tobias Höllerer. "Botivist: Calling volunteers to action using online bots." *CSCW*. 2016.
- [20] Silva, Filipa, Teresa Proença, and Marisa R. Ferreira. "Volunteers' perspective on online volunteering-a qualitative approach." *International Review on Public and Nonprofit Marketing* 15.4 (2018): 531-552
- [21] Stukas, Arthur A., et al. "Motivations to volunteer and their associations with volunteers' well-being." *Nonprofit and Voluntary Sector Quarterly* 45.1 (2016): 112-132.
- [22] Toxtli, C., Flores-Saviaga, C., Aguilar, F., Monroy, A., Flores, J., Herrejon, J., ... & Savage, S. "BotViz: Data Visualizations for Collaborations With Bots and Volunteers". *CSCW* 2016.
- [23] Vukovic, Maja. "Crowdsourcing for enterprises." *2009 congress on services-I*. IEEE, 2009.
- [24] Wagner, Claudia, et al. "It's a Man's Wikipedia? Assessing Gender Inequality in an Online Encyclopedia." *ICWSM*. 2015.
- [25] Xu, Anbang, and Brian Bailey. "What do you think? A case study of benefit, expectation, and interaction in a large online critique community." *CSCW*. 2012.
- [26] Yen, Yu-Chun, et al. "Social network, web forum, or task market? Comparing different crowd genres for design feedback exchange." *DIS*. 2016.



© 2020 by the authors. This work is licensed under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License. To view a copy of this license, visit <http://creativecommons.org/licenses/by-nc-nd/4.0/> or send a letter to Creative Commons, PO Box 1866, Mountain View, CA 94042, USA.