Exploring Human Resource Management in Crowdsourcing Platforms*

Cristina Cabanillas

Vienna University of Economics and Business, Austria
cristina.cabanillas@wu.ac.at

Abstract. The correct execution of process activities is usually responsibility of the employees (i.e., human resources) of an organisation. In the last years, notable support has been developed to make resource management in business processes more efficient and customisable. Recently, a new way of working has emerged and caught significant attention in the market: crowdsourcing. Crowdsourcing consists of outsourcing activities in the form of an open call to an undefined network of people, i.e., the crowd. While in traditional resource management in business processes resources are known and task assignment is usually controlled, the workers in crowdsourcing platforms are unknown and are allowed to select the tasks they want to perform. These and other differences between resource management in business processes and in crowdsourcing platforms have not been explicitly investigated so far. Taking as reference the existing mature work on resource management in business processes, this paper presents the results of a study on the existing support for resource management in crowdsourcing platforms.

Keywords: business process management, crowdsourcing, empirical study, resource management

1 Introduction

Work is materialised in activities that must be completed, usually under temporal constraints. Nowadays, there are several ways to distribute the execution of activities. Business processes or workflows constitute a controlled definition (and execution) of the activities carried out in an organisation and are characterised as follows: (i) the workers are generally employees of the organisation and hence, easily accessible; (ii) the workers are typically offered or allocated the activities they can work on depending on their expertise; and (iii) each activity has one person responsible who can act individually or in collaboration with other workers for the completion of the job, being the outcome of an activity the result of a single execution (a.k.a. activity instance). On the other hand, in the last 10 years a new way of working known as *crowdsourcing* has become popular. Crowdsourcing consists of a web-based completion of publicly available

* This work is funded by the Austrian Research Promotion Agency (FFG) under grant 845638 (SHAPE).

activities ranging from simple tasks (e.g., picture tagging) to complex activities (e.g., software development). A crowdsourcing platform acts as an intermediary between a client (a company or an individual that needs an activity to be done) and the crowd (any person registered at the platform) in charge of executing the activities [1]. All the crowdsourcing platforms have in common that: (i) the workers are loosely coupled with the client as they do not have a contract with it but are paid for each activity "correctly" completed; (ii) usually, the workers can access any activity published on the platform and work on its execution; and (iii) due to the varied nature of activities and the high risk of cheating and misbehaviour, several instances of an activity are usually concurrently executed by different workers, and one or more results are taken into consideration for the final outcome of the job.

Despite their differences, similar steps must be carried out for work distribution in both the Business Process Management (BPM) and the crowdsourcing domains, such as the allocation of activities to suitable resources. In BPM, human resource[1] management has been widely investigated in the last years [2–4]. However, in the domain of crowdsourcing, the research efforts have been put on how to incentive workers [5] and how to assure quality of the results of the executions [6]. To the best of our knowledge, resource management in crowdsourcing platforms has not yet been investigated in a systematic way, so there might be room for improvement.

To address this gap, we have conducted a survey on the support for resource management in crowdsourcing platforms framed by the resource management concepts from BPM as well as quality assurance features described in the crowdsourcing literature and found in an exploration of crowdsourcing systems. This work contributes to understanding the current support and to discovering potential directions for future work in the crowdsourcing domain.

The paper is structured as follows. Section 2 introduces the vocabulary required to understand the study. Section 3 presents the hypotheses and the survey design. Section 4 analyses the results of the survey and outlines the limitations of the work. Finally, Section 5 draws conclusions and points out future work.

2 Background

In the following, we introduce the main concepts related to resource management in BPM and in crowdsourcing.

2.1 Resource Management in Business Processes

In BPM, resource management explores how resources are involved in the activities of the processes executed in an organisation. Three steps can be distinguished in resource management in BPM [7]. Fig. 1 illustrates them.

Resource assignment defines the set of conditions that resources must meet to be allowed to take part in an activity. These conditions are defined at design time

[1] From now on *resource* for the sake of brevity.

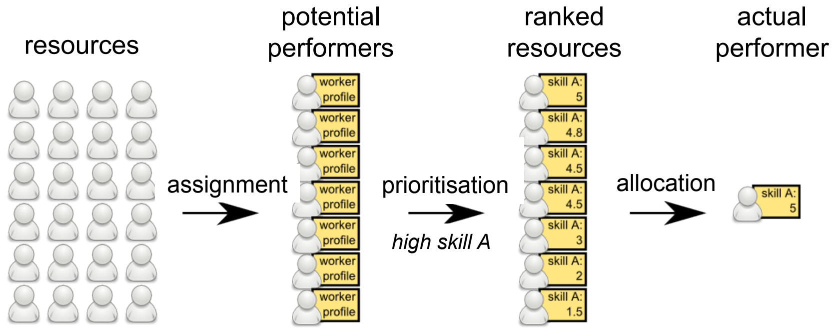

Fig. 1: Resource management in business processes

and are evaluated at run time when a process instance is executing, resulting in the set of *potential performers* of an activity instance. The languages for resource assignment rely on the concept of organisational model as a description of the part of the organisation involved in a business process. Common terms in organisational models are: person, role, position, organisational unit and the notion of capability or skill [8]. There are textual [4] as well as graphical [9] and hybrid [10, 3]) resource assignment languages, which differ in their expressiveness to define the selection conditions (e.g., based on organisational roles or on skills). A subset of the workflow resource patterns called creation patterns is typically used as evaluation framework of the expressiveness [2]. The most expressive languages support all of them [3, 4].

Resource allocation is the process of selecting one specific resource from the set of potential performers as *actual performer* of an activity instance. BPM systems usually perform resource allocation by offering an activity to one single resource or to several resources, or by allocating the activity directly to a specific resource. These and other techniques are collected in the subset of workflow resource patterns called push patterns [2]. Smarter ways of choosing the most appropriate resource for an activity instance to optimise, a.o., time or cost, are increasingly being investigated in the context of BPM [11].

Resource prioritisation is the definition of preferences to sort out the set of potential performers prior to resource allocation [12]. Properties that can be used for defining the preferences are, e.g., personal and organisational data, such as the value of predefined skills, the length of the worklist of resources at a specific point in time, or historic information that points out the ability of a resource for performing certain work. The outcome of the prioritisation is thus a ranking of potential performers that serves as input for the resource allocation technique.

These three steps apply not only to select the resource responsible for the execution of an activity but also for other responsibilities that may be associated with it. Responsibility is generally modelled in process-oriented organisations by using a so-called Responsibility Assignment Matrix (RAM) [13] that assigns one or more responsibilities to a specific organisational role for a process activity. For instance, in RASCI matrices [13] the available responsibilities are: responsible, accountable, support, consulted and informed.

2.2 Resource Management in Crowdsourcing Platforms

Crowdsourcing is technology that enables a large number of people contributing their knowledge and expertise to an activity[2] that would not be so valuable alone [14]. Crowdsourcing platforms play a crucial role between two types of registered users: the requesters (clients) and the crowd (workers). A complete crowdsourcing workflow is made up of four steps:

1. A requester submits an activity description to the platform defining, a.o., the due completion date and the associated remuneration. Ideally, on the platform the requester can specify if the task is available to the whole crowd or only to workers with specific characteristics as well as preferences for the performer of the activity.
2. All workers who are able to see the activity description can generally claim for its execution, except for activities restricted to a limited number of workers.
3. The workers that requested the activity may be ranked by the platform according to the criteria previously specified by the requester, and shown to the requester.
4. The requester can decide which worker(s) should perform the activity.
5. The selected worker(s) will then start their work and submit the results to the platform.
6. When the activity deadline is reached or all the requested activity instances are completed, the platform collects all the results from the workers and sends them to the requester, who proceeds to pay the workers for their job.

As can be observed, the resource management steps described for traditional BPM are also represented in the crowdsourcing domain, specifically: resource assignment maps to step 1, resource prioritisation maps to step 3 and resource allocation maps to step 4.

Unlike in traditional BPM, one of the biggest problems in crowdsourcing environments nowadays is the challenging mission of quality assurance. Wikis were the first crowdsourced applications run by non-profit organisations [14, 15] with the rise of the Web 2.0. Afterwards, with the emergence of commercial crowdsourcing platforms, like Amazon Mechanical Turk (AMT)[3], remuneration was a driving factor for people to join a crowdsourcing platform. Money-driven engagement implies that some workers try and cheat the system to maximize their earnings without delivering any useful contribution. Hence, every contribution of every worker may be incorrect and has to be checked against fraud and validity [15, 16]. Quality assurance can take place at several stages of the aforementioned workflow, e.g., by means of tests after a new worker is registered

[2] The term *task* is more common in the crowdsourcing domain but we will use *activity* for the sake of consistency.

[3] `https://www.mturk.com/mturk/welcome`

to the platform or before the results of an activity are sent to the requester for the subsequent invoicing, and it constitutes a critical matter to be considered in the management of resources in the crowdsourcing domain [6].

Two main classes of crowdsourcing platforms can be distinguished. *Marketplaces* are crowdsourcing platforms in a narrow sense, i.e., a platform where individuals and companies can post their activities and get them done by crowdworkers [17]. Steps 1 to 6 are performed as described above. Advanced mechanisms for quality assurance are not expected. Examples of marketplaces are oDesk, AMT and Microworkers. *Brokers* act as intermediaries and helpers between the requesters and the crowd, so that the requesters do not need to frame the task and post it to a marketplace. In addition, the broker typically offers complementary services such as quality control [17]. Examples of brokers are CrowdFlowers, CrowdControl and Microtask.

3 Research Design

Our aim is to explore resource management in marketplace and broker crowdsourcing platforms in terms of the support for resource assignment, resource prioritisation, resource allocation and quality assurance, as these are the characteristics that stand out regarding work distribution and completion.

We use an online questionnaire as research method because: (i) it supports geographical independence, i.e., any crowdsourcing company is accessible regardless of its location; (ii) it keeps confidentiality while providing insights, i.e., we can get precise information not publicly available while respecting privacy; and (iii) it keeps the balance between effort, results and drawbacks (low response rate and different perceptions [18]).

A typical workflow of a questionnaire is composed of six steps [19]. The first one is the selection of the sample (cf. Section 3.1). Afterwards, the research model with hypotheses must be defined (cf. Section 3.2). To make the research model measurable, an operationalidation step is crucial, whose resulting items must be arranged in a questionnaire (cf. Section 3.3). With the feedback collected from a pretest round, the questionnaire can be optimised and the data collection can be started (cf. Section 3.4). Afterwards, the analysis of the data and the evaluation uses the research model and tries to falsify the hypotheses (cf. Section 4).

3.1 Selection of the Sample

Our research on marketplaces and brokers resulted in a list of 55 companies, whose identities are kept confidential in this paper. For each of them, a contact person was identified through the websites of the companies, under the requirement of having a technical understanding and knowledge about their product. After the evaluation of the population the sample size was determined. We used the formula of Krejcie and Morgan [20] (cf. Equation 1) with the parameters shown in Table 1. This results in a number of 48 platforms that should respond in order to reach the given confidence and accuracy.

Parameter	*Value*	*Description*
N	55.00	Discovered platforms
X	1.95	Confidence level of 95%
P	0.5	Population proportion: 0.5 to get maximum sample size
d	0.05	Accuracy of 5% (margin of error)

Table 1: Equation parameters for estimating the sample size

ID	*Description*
H1	A broker platform will support more features than a marketplace platform.
H2	A marketplace platform will support more criteria for resource assignment and prioritisation than a broker platform.
H3	The supported criteria/features are rated as helpful.
H4	The higher the helpfulness of a supported criteria/feature, the higher the frequent usage.
H5	The unsupported criteria/features are rated as potentially unhelpful.
H6	The higher the necessity of supporting a criteria/feature in the future, the higher the perceived helpfulness.

Table 2: Hypotheses

$$s = \frac{X^2 N P(1-P)}{d^2(N-1) + X^2 P(1-P)} \tag{1}$$

3.2 Hypotheses

From now on we will differentiate between (i) *criteria*, which describe the conditions that can be defined for resource assignment and prioritisation, such as based on roles or skills (cf. Section 2.1); and (ii) *features*, which comprise the functionality that is provided, i.e., support for assignment, prioritisation, allocation and quality assurance functionalities. We want to discover the features implemented in the platforms and, for those platforms supporting assignment and prioritisation, the criteria used in them.

Our hypotheses are outlined in Table 2 and the derived research model is depicted in Fig. 2. The model is composed of seven constructs (boxes) connected through arrows that illustrate how the constructs are influenced by the hypotheses: (+) indicates a positive influence and (i) indicates a negative influence. The construct Task Type (TT) has been introduced for statistical purposes and thus has no influence on any other construct of a hypothesis.

In Section 2.2 we distinguished two types of crowdsourcing platforms: marketplaces and brokers. Brokers, by definition, provide additional features for requesters such as quality assurance. This leads to hypothesis H1. On the other hand, marketplaces are more specialised and hence, might have greater support of criteria for assignment and prioritisation. This leads to hypothesis H2. If a platform supports a criterion or feature it might be assumed that the criterion or feature is perceived as helpful for requesters. This leads to hypothesis H3.

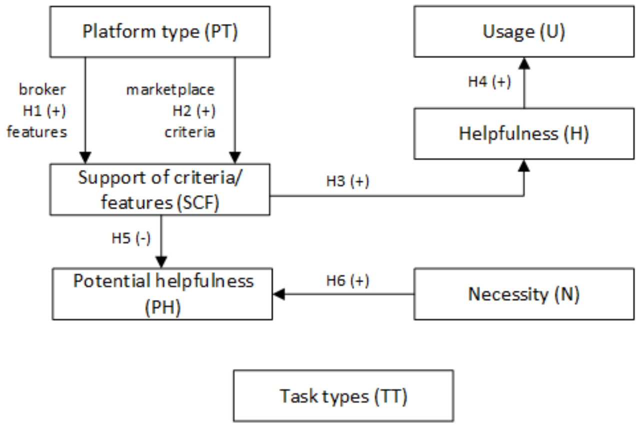

Fig. 2: Research model

Furthermore, if a criterion or feature is perceived as helpful there should be evidence of it in the form of a more frequent usage than other less helpful criteria or features. This leads to hypothesis H4. On the contrary, if a platform does not support a criterion or feature it might be assumed that the criterion or feature is perceived as potentially unhelpful for requesters. This leads to hypothesis H5. Finally, another construct which might influence the potential helpfulness of criteria and features is the necessity of supporting it in the future due to a high business competition. This leads to hypothesis H6.

3.3 Questionnaire

In order to measure the seven factors involved in the research model (cf. Fig. 2) we introduce measurement items in the form of questions in the questionnaire:

- Platform type (PT) is measured by a nominal scale with three items, only one of which can be selected by the user: (1) a definition of marketplace platform, (2) a definition of broker platform and (3) a free text field for platforms which would not classify themselves this way.
- Support of criteria/features (SCF) is measured by a multiple item choice table which contains various criteria or features. Due to space limitations, we refer to [21] for a complete description of the features and criteria used.
- Helpfulness (H) and usage (U) are measured by a 5-point Likert scale as it is a rating scale which measures the strength of agreement on a set of clearly defined statements [22]. The items are phrased so that a participant is requested to express his or her level of agreement from Strongly Disagree (1) to Strongly Agree (5). These questions are only asked if the participant has stated that their platform supports the corresponding criterion or feature. These two factors (H+U) are grouped by the corresponding criterion or feature and asked in a single multiple choice table.

Section name	*Questioned factors*	*Goals*
1. Introduction	-	Introduction to the survey, purpose and usage of collected data.
2. Worker selection and ranking	SCF, H, U, PH, N	Criteria supported for resource assignment and prioritisation
3. Support for the requester	SCF, H, U, PH, N	Features that help the requester decide upon resource allocation
4. Quality assurance	SCF, H, U, PH, N	Features to ensure quality of results
5. Advanced task assignment	SCF, H, U, PH, N	Support for functionality like, e.g., team composition
6. About the platform	PT, TT	Type of platform and tasks
7. About you	-	Basic corporate information

Table 3: Structure of the questionnaire

- Potential helpfulness (PH) and necessity (N) are also measured by a 5-point Likert scale ranging from Strongly Disagree (1) to Strongly Agree (5). In this case, the questions are only shown to participants who indicated that the feature or criterion is not supported by the platform. These two factors (PH+N) are grouped by the corresponding criterion or feature and asked in a single multiple choice table.
- Task types (TT) is a multiple choice scale where a participant can to make one or several choices on which task types their platform supports. The available task types, such as development, picture tagging, translation services and logo design, have been collected from the platform websites and generalised to reduce the number of types [21].

To improve the clarity of the questionnaire the items were structured and aggregated into topical groups, as depicted in Table 3. Several tools were evaluated for the generation of the questionnaire and Qualtrics[4] was selected to serve as online survey platform for this study, as it is open-source and it supports conditional table rows for multiple choice questions.

3.4 Data Collection

According to Beywl and Schepp-Winter, a pretest should be done by four to six participants [23]. We sent the questionnaire to four colleagues familiar with the topic, obtaining a positive final statement from all them accompanied with suggestions for minor language improvements, such as word order and typos. All the improvements were considered and integrated into the final version of the questionnaire, available in [21].

The questionnaire was sent to 55 companies that we had identified as marketplaces or brokers. We contacted associates of the management hierarchies (Chief/Head of Product, Product Manager, Chief Technology/Technical Officer (CTO), Chief Operating Officer (COO) or Chief Executive Officer (CEO)) but

[4] `http://qualtrics.com`

stated in the invitation e-mail that it could be forwarded to a capable person. On average, every 4 days for a 2-month period a reminder was sent to the companies that had not responded to the questionnaire so far. In the end, we received 14 valid questionnaires. This means a response rate of 25%, which is a quite good number for internet surveys w.r.t. the achievable rates of 20% to 30% [18]. Nevertheless, we did not achieve the required 48 responses to draw statistically reliable and accurate conclusions, for which we would have required a response rate higher than 89%. Due to the low response rate we achieved an accuracy of approximately 22% instead of desired 5%. However, the data analysis and evaluation already showed interesting results, which are summarised next.

4 Analysis

In the following we describe the result of the survey as well as the limitations and potential improvements discovered.

4.1 Result of the Survey

The evaluation of the results has been done with Microsoft Excel 2013 and with R [24]. The values are rounded to two decimals for better readability, except in cases where all values are too small to display anything (e.g., p-values). We use the significance level $\alpha = 0.05$ for all statistical tests, which indicates a 5% risk of concluding that a difference exists when there is no actual difference. Due to space limitations, the analysis of the data is summarised and grouped by hypothesis. For a detailed description of the evaluation, we refer to [21].

The first evaluation performed relates to hypothesis H1. Table 4 illustrates the number of platforms that support each feature (grouped according to the structure of the questionnaire: sections 3 to 5 - cf. Table 3) in both absolute and relative numbers. To evaluate whether H1 can be validated or not, we have to compare the means of supported features for both marketplaces and brokers. Broker platforms have on average 1.17 more features than marketplace platforms. We performed a statistical T-test to determine if the deviation is significant. The resulting p-value of 0.2496 indicates that the difference is not significant ($\alpha \leq$ p). Hence, we propose to reject hypothesis H1 and thus, we conclude that no significant difference in feature support exists between marketplaces and brokers.

The evaluation of hypothesis H2 brought up the findings outlined in Table 5. On average, 8.25 platforms support criteria for resource assignment and only 6 platforms support criteria for resource prioritisation. In summary, the criterion *familiarity with tasks* is the most supported one for marketplaces concerning assignment and for brokers concerning prioritisation. On the other hand, the criterion *skills* is the most supported one for broker platforms for resource assignment. For marketplaces covering resource prioritisation, there are 4 top supported criteria: *skills*, *familiarity with tasks*, *success rate* and *completion pace*. The statistical analysis shows that marketplaces support more criteria in the two categories (assignment = +1.13, prioritisation = +1.50) but none of the

Feature	*Marketplace* abs.	*Marketplace* %	*Broker* abs.	*Broker* %	*Total* abs.	*Total* %
Filter workers	5	62.50	4	66.67	9	64.29
Manual offer to workers	4	50.00	5	83.33	9	64.29
Team composition	4	50.00	4	66.67	8	57.14
Preferences	3	37.50	4	66.67	7	50.00
Redundant execution	7	87.50	4	66.67	11	78.57
Skill tests	6	75.00	5	83.33	11	78.57
Feedback on performance	5	62.50	5	83.33	10	71.43
Training tasks	3	37.50	3	50.00	6	42.86
Delegation	1	12.50	1	16.67	2	14.29
Accountable worker(s)	4	50.00	3	50.00	7	50.00
Supportive worker(s)	5	62.50	3	50.00	8	57.14
Consulted worker(s)	2	25.00	3	50.00	5	35.71
Informed worker(s)	3	37.50	2	33.33	5	35.71
Mean	6.50		7.67		7.54	
p-value	0.2496		-		-	

Table 4: Evaluation of the features supported by the platforms

probability values indicates a significant difference as no value was below 0.24. Since the difference between marketplaces and brokers for resource assignment is very small, a statistical T-test was performed on the data to calculate the significance of the differences. No category got a significant result. Therefore, we suggest that H2 is invalid, and thus, there is no significant difference between marketplaces and brokers regarding the criteria supported.

Hypothesis H3 states that the supported criteria are rated as helpful which means that the values have to be greater than 3. This hypothesis has been evaluated two times: for the supported criteria (section 2 of the questionnaire) and for the supported features (sections 3-5 of the questionnaire). Table 6 shows the mean values and the p-values of the factor H of all criteria for assignment and prioritisation. As no value is below 3, the respondents have not declined the helpfulness of the criteria. To validate this part of the hypothesis statistically a one-sided T-test was performed and the p-values were evaluated. 10 out of 16 possible criteria show a significant higher value than 3. Hence, we suggest that H3 is valid. As for the supported features, the situation is similar (cf. Table 7). No participant rated a supported feature as unhelpful. The only outlier is the feature *delegation*, which was rated as neutral. 8 of 13 features had also a significant positive rating. Hence, we can confirm hypothesis H3 as valid for features. Therefore, as both parts could validate H3 we suggest that H3 is valid.

Hypothesis H4 states that the higher the helpfulness of a supported criteria/feature, the higher the frequent usage. This hypothesis has also been evaluated twice. Table 6 shows an overview of the evaluation for the supported criteria, specifically, the Pearson correlation coefficient between the two factors H and U;

	Marketplace				*Broker*			
	Assignment		*Prioritisation*		*Assignment*		*Prioritisation*	
Criterion	abs.	%	abs.	%	abs.	%	abs.	%
Skills	6	75.00	5	62.50	5	83.33	2	33.33
Geographical position	6	75.00	3	37.50	4	66.67	2	33.33
Familiarity with tasks	7	87.50	5	62.50	4	66.67	3	50.00
Familiarity with requester	5	62.50	3	37.50	2	33.33	2	33.33
Expected salary	3	37.50	3	37.50	3	50.00	2	33.33
Success rate	5	62.50	5	62.50	3	50.00	2	33.33
Quality ranking	4	50.00	3	37.50	3	50.00	2	33.33
Completion pace	5	62.50	5	62.50	1	16.67	1	16.67
Mean	5.13		4.00		4.17		2.67	
p-value	0.2801		0.2410		-		-	

Table 5: Evaluation of the criteria supported by the platforms

	Assignment				*Prioritisation*			
	H		$H \rightarrow U$		*H*		$H \rightarrow U$	
Criterion	mean	p-value	cor	p-value	mean	p-value	cor	p-value
Skills	4.27	0.0002*	0.7717	0.0054*	4.00	0.0309*	0.9186	0.0035*
Geographical position	3.40	0.1717	0.8458	0.0020*	4.00	0.0171	-0.3953	0.5101
Familiarity with tasks	4.18	0.0002*	0.6119	0.0454*	4.00	0.0249*	0.9589	0.0002*
Familiarity with requester	3.71	0.1100	0.9626	0.0005*	3.40	0.2935	1.0000	0.0000*
Expected salary	3.67	0.0510	1.0000	0.0000*	3.60	0.0352	0.6124	0.2722
Success rate	4.00	0.0006*	1.0000	0.0000*	3.86	0.0226*	0.9262	0.0027*
Quality ranking	4.29	0.0021*	1.0000	0.0000*	4.60	0.0014*	1.0000	0.0000*
Completion pace	3.33	0.1816	0.9342	0.0064*	3.17	0.3054	0.8677	0.0251*

Table 6: Evaluation of factor H and of the correlations of factors H and U for supported criteria (* = significant value for $\alpha = 0.05$)

and the p-value, which indicates whether the correlation is significant or not. In short, all criteria for resource assignment has significant influence on the usage and only 2 criteria for resource prioritisation has no significant influence. The 2 non-validated criteria for ranking are *geographical position* and *expected salary*, whereas *geographical position* has a negative insignificant influence on the usage. Therefore, we suggest to confirm hypothesis H4 as valid for the criteria. Regarding the features, as shown in Table 7, 11 of 13 features have a strong positive influence on the usage and for 9 of them the influence is significant. Only 2 features (*team composition* and *delegation*) have no positive influence on the usage. Hence, we suggest that hypothesis H4 is also valid for features. Therefore, we could confirm the validity of hypothesis H4 for both criteria and features.

Hypothesis H5 states that the unsupported criteria/features are rated as potentially unhelpful, and has also been evaluated two times. As we are asking

Feature	*H* mean	*H* p-value	*H → U* cor	*H → U* p-value
Filter workers	3.78	0.664	0.9761	0.0000*
Manual offer to workers	3.78	0.0040*	0.8740	0.0021*
Team composition	3.63	0.1084	0.5130	0.1936
Preferences	3.57	0.0515	1.000	0.000*
Redundant execution	4.09	0.0030*	0.8503	0.0009*
Skill tests	3.64	0.0130*	0.8752	0.0004*
Feedback on performance	4.40	0.0003*	1.0000	0.0000*
Training tasks	4.50	0.0035*	0.8093	0.0511
Delegation	3.00	NA	NA	NA
Accountable worker(s)	4.14	0.0023*	0.7670	0.0442*
Supportive worker(s)	3.88	0.0105*	0.9078	0.0018*
Consulted worker(s)	3.80	0.0497*	0.8018	0.1027
Informed worker(s)	3.80	0.0889	1.0000	0.0000*

Table 7: Evaluation of factor H and of the correlations of factors H and U for supported features (* = significant value for $\alpha = 0.05$)

Criterion	*PH* mean	*PH* p-value	*PH → N* cor	*PH → N* p-value
Skills	3.67	0.0918	0.7559	0.4544
Geographical position	3.00	0.5000	0.9733	0.0267*
Familiarity with tasks	3.33	0.2113	NA	NA
Familiarity with requester	2.71	0.2285	0.9226	0.0031
Expected salary	2.71	0.2285	0.2475	0.5926
Success rate	3.20	0.3744	0.9609	0.0092*
Quality ranking	3.71	0.0041*	0.5916	0.1618
Completion pace	2.88	0.3813	0.8440	0.0084*

Table 8: Evaluation of factor PH and of the correlations of factors PH and N for supported criteria (* = significant value for $\alpha = 0.05$)

about factor potentially helpful (PH) we have to look for values below 3. The means and the p-values per criterion are listed in Table 8, where the means are showing a neutral picture: all values are more or less equal to 3, so are the overall mean and median. Only one criterion has a rating significantly different from neutral (*quality rating*) but this criterion is considered as potentially helpful and not unhelpful. Therefore, we can reject hypothesis H5 for criteria support. Regarding the supported features, the mean ratings of all features have been calculated and a statistical one-sided T-test has been performed. As shown in Table 9, only 4 of 13 features got significant p-values, where only the feature *manual offer to workers* is supporting the hypothesis, due the fact that the other three significant features have a positive rating. As most platforms rated most unsupported features as potentially helpful rather than unhelpful, we suggest to

	PH		*PH* → *N*	
Feature	mean	p-value	cor	p-value
Filter workers	4.00	0.0171*	0.8607	0.0611
Manual offer to workers	2.00	0.0171*	0.0000	1.0000
Team composition	3.00	0.5000	0.9576	0.0027*
Preferences	3.14	0.3679	0.9354	0.0020*
Redundant execution	3.33	0.2113	1.0000	0.0000*
Skill tests	3.67	0.0918	NA	NA
Feedback on performance	3.50	0.0908	1.0000	0.0000*
Training tasks	3.50	0.0518	0.3536	0.3903
Delegation	3.42	0.1049	0.0837	0.7958
Accountable worker(s)	3.57	0.0150*	-0.1667	0.7210
Supportive worker(s)	3.17	0.1816	0.5813	0.2262
Consulted worker(s)	3.44	0.0176*	0.6532	0.0565
Informed worker(s)	3.44	0.0845	0.4488	0.2256

Table 9: Evaluation of factor PH and of the correlations of factors PH and N for supported features (* = significant value for $\alpha = 0.05$)

reject the hypothesis H5 for features. Therefore, we could not confirm hypothesis H5 in any category, so we reject it.

Finally, hypothesis H6 states that the higher the necessity of supporting a criteria/feature in future, the higher the perceived helpfulness, and this hypothesis has also been evaluated two times. In the evaluation we check the correlation between the factors PH and N. Regarding the supported criteria, as shown in Table 8, half of the criteria have a significant correlation between necessity and potential helpfulness (*geographical position*, *familiarity with requester*, *success rate* and *completion pace*) and 2 more criteria have a strong influence but are not significant (*skills* and *quality rating*). The Pearson correlation coefficient for familiarity with tasks cannot be calculated due to a standard deviation of 0 (all respondents answered *neutral*) for the factor N. Therefore, hypothesis H6 cannot be generally validated or invalidated, it holds true for 4 criteria. As for the features supported, Table 9 shows that the evaluation has 4 significant values and hence, the hypothesis is valid for the following 4 features: *team composition*, *preferences*, *redundant execution* and *feedback on performance*. The feature *manual offers to workers*, which has been rated significantly negatively before, has a Pearson correlation coeficient of 0, which means that there is no relationship between the two factors. As only 4 of 13 features have a significant correlation we suggest that the hypothesis H6 is invalid. Therefore, since no category could verify that hypothesis H6 is fully valid, we reject the hypothesis.

Altogether, the results of our evaluations conclude that only hypothesis H3 and H4 are valid for the sample data. However, the invalidation of hypothesis H5 brings some light towards future extensions of the platforms to support the missing features and criteria. Moreover, the rejection of hypothesis H6 may be caused by the low response rate.

4.2 Limitations

This survey presents some limitations. For instance, one respondent had a problem with the question “Which statement describes your platform best?” and selected the option *marketplace*. However, from the description subsequently provided in the text field we could derive that the platform is actually a broker. Consequently, we assumed that the respondent selected *broker*. Therefore, this question should be a point for improvement as it should be easily understood by everyone. Furthermore, the low response rate suggests that the survey approach may not have been the best choice. Several companies did not respond at all, other companies gave harsh declinations. However, this fact may be due to a refusal of the companies to share information deemed confidential because of the increasing competition in this sector. In addition, the research performed was limited to one categorisation of crowdsourcing platforms (conceptual model) including only two types of platforms (marketplaces and brokers). However, there exist other types of platforms and many other distinction models, such as crowdsourcing objectives [25] or the four archetypes [26].

5 Conclusions and Future Work

This paper provides an overview of existing support for resource management in crowdsourcing platforms. The evaluation concludes that current crowdsourcing platforms focus their efforts on supporting features and criteria for resource assignment and prioritisation that have proved to be frequently used and hence, are deemed helpful; while it suggests that some unsupported features and criteria could be considered relevant in the future.

From the limitations of the survey we can conclude that, as a first attempt to extend this study, the questionnaire must be revised aiming at increasing the response rate. The rejected hypotheses can serve as a starting point for further investigations. In case of no success, a different research method should be explored. In addition, further potential extensions include taking into account a broader classification of the platforms.

References

1. D. Schall, B. Satzger, and H. Psaier, “Crowdsourcing Tasks to Social Networks in BPEL4People,” *WWW*, vol. 17, no. 1, pp. 1–32, 2014.
2. N. Russell, W. M. P. van der Aalst, A. H. M. ter Hofstede, and D. Edmond, “Workflow Resource Patterns: Identification, Representation and Tool Support,” in *CAiSE*, pp. 216–232, 2005.
3. L. J. R. Stroppi, O. Chiotti, and P. D. Villarreal, “A BPMN 2.0 Extension to Define the Resource Perspective of Business Process Models,” in *CIbS'11*, 2011.
4. C. Cabanillas, M. Resinas, A. del Río-Ortega, and A. Ruiz-Cortés, “Specification and Automated Design-Time Analysis of the Business Process Human Resource Perspective,” *Inf. Syst.*, vol. 52, pp. 55–82, 2015.

5. A. Singla and A. Krause, "Truthful Incentives in Crowdsourcing Tasks Using Regret Minimization Mechanisms," in *WWW*, pp. 1167–1178, 2013.
6. M. Allahbakhsh, B. Benatallah, A. Ignjatovic, H. R. Motahari-Nezhad, E. Bertino, and S. Dustdar, "Quality Control in Crowdsourcing Systems: Issues and Directions," *IEEE Internet Computing*, vol. 17, no. 2, pp. 76–81, 2013.
7. C. Cabanillas, "Enhancing the management of resource-aware business processes," *AI Communications*, vol. 29, no. 1, pp. 237–238, 2015.
8. O. Nicolae and G. Wagner, "Modeling and Simulating Organisations," in *EOMAS*, vol. 88, pp. 45–62, 2011.
9. C. Cabanillas, D. Knuplesch, M. Resinas, M. Reichert, J. Mendling, and A. Ruiz-Cortés, "RALph: A Graphical Notation for Resource Assignments in Business Processes," in *CAiSE*, vol. 9097, pp. 53–68, 2015.
10. W. M. P. van der Aalst and A. H. M. ter Hofstede, "YAWL: Yet Another Workflow Language," *Inf. Syst.*, vol. 30, no. 4, pp. 245–275, 2005.
11. G. Havur, C. Cabanillas, J. Mendling, and A. Polleres, "Automated Resource Allocation in Business Processes with Answer Set Programming," in *BPM Workshops (BPI)*, p. In press, 2015.
12. C. Cabanillas, J. M. García, M. Resinas, D. Ruiz, J. Mendling, and A. R. Cortés, "Priority-Based Human Resource Allocation in Business Processes," in *ICSOC*, vol. 8274, pp. 374–388, 2013.
13. Website, "Understanding Responsibility Assignment Matrix (RACI Matrix)." `http://project-management.com/understanding-responsibility-assignment-matrix-raci-matrix/`, Last accessed in March 2016.
14. S. Greengard, "Following the Crowd," *Commun. ACM*, vol. 54, no. 2, pp. 20–22, 2011.
15. M. Hirth, T. Hossfeld, and P. Tran-Gia, "Analyzing costs and accuracy of validation mechanisms for crowdsourcing platforms," *Mathematical and Computer Modelling*, vol. 57, pp. 2918–2932, 2013.
16. B. Satzger, H. Psaier, D. Schall, and S. Dustdar, "Auction-based crowdsourcing supporting skill management," *Inf. Syst.*, vol. 38, no. 4, pp. 547–560, 2013.
17. D. Schall, "Crowdsourcing Task Marketplaces," in *Service-Oriented Crowdsourcing* (S. N. York, ed.), ch. 2, pp. 7–30, SpringerBriefs in Computer Science, 2012.
18. K. Siau and M. Rossi, "Evaluation techniques for systems analysis and design modelling methods - a review and comparative analysis," *Inf. Syst.*, vol. 21, no. 3, pp. 249–268, 2011.
19. H. O. Mayer, *Interview und schriftliche Befragung: Entwicklung, Durchführung und Auswertung*. Mnchen: Oldenbourg Verlag, 4 ed., 2008.
20. R. V. Krejcie and D. W. Morgan, "Determining Sample Size for Research Activities," *Educational and Psychological Measurement*, vol. 30, pp. 607–610, 1970.
21. D. Anonymised, "Deliberately Anonymised," Master's thesis, Deliberately Anonymised, Deliberately Anonymised.
22. C. Gena, "Methods and Techniques for the Evaluation of User-adaptive Systems," *Knowl. Eng. Rev.*, vol. 20, pp. 1–37, Mar. 2005.
23. W. Beywl and E. Schepp-Winter, "Zielgeführte Evaluation von Programmen: ein Leitfaden." Bundesministerium fr Familie, Senioren, Frauen und Jugend, 2000.
24. R Development Core Team, *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria, 2008.
25. M. Vukovic, "Crowdsourcing for Enterprises," in *SERVICES*, pp. 686–692, 2009.
26. D. Geiger, M. Rosemann, E. Fielt, and M. Schader, "Crowdsourcing Information Systems - Definition, Typology, and Design," in *ICIS*, 2012.