

# **Towards the Development of a DSS Supporting the Integration of Crowdsourcing in Theory Testing: Analytical Framework Design**

*Completed Research Full Papers*

**Ijeoma Enwereuzo**

Victoria University of Wellington  
Ijeoma.enwereuo@vuw.ac.nz

**Pedro Antunes**

Victoria University of Wellington  
Pedro.antunes@vuw.ac.nz

**David Johnstone**

Victoria University of Wellington  
David.johnstone@vuw.ac.nz

## **Abstract**

Crowdsourcing can be an adopted strategy for researchers where tasks are distributed to internet users to harness different forms of data, which adds to the reliability and validity of the research process. As theory testing is an essential part of the research process, involving activities that most times needs input from different and diverse participants, the need to consider adopting the crowdsourcing strategy for theory testing is paramount. Adopting a design science paradigm to manage this challenge, we design an analytic framework which comprises of important attributes that need to be considered if crowdsourcing is to be used for any of the theory testing activities. The framework which was justified using sample cases gives us an insight into what attributes make such activity crowdsourcing-able. The value of this artefact lies in its capacity to help researchers utilize crowdsourcing to their advantage.

## **Keywords**

Crowdsourcing, theory testing, design science, framework.

## **Introduction**

In academia, a discipline will not have much leverage, if it does not have useful theories to contribute (Niederman & March, 2015). Theory helps scientists to describe and explain a process or sequence of events. It also helps scholars to explain the complexities of the empirical world by providing a linguistic tool for organizing it (Colquitt & Zapata-Phelan, 2007). Theory testing which is an integral part of theory building is very important to the development of Information Systems (IS) discipline, just as Bacharach (1989) said, “if it is not testable, no matter how profound or aesthetically pleasing it may be, it is not theory”. We are therefore interested in supporting theory testing in the IS field and how crowdsourcing (CS) can be used to support the theory testing process. CS is a strategy in which a single task can be fragmented into multiple tasks which can then be delivered to a large group of people in the form of an open call (Howe, 2006).

In the previous paper (Enwereuzo et al., 2017, 2019), we analyzed how theories have been tested in IS and we proposed a conceptual framework and a model. The framework and model were evaluated using card sorting, the outcome showed that the framework is sound and that students (Ph.D.) couldn't comprehend the theory testing process and the activities involved. The outcome also indicates that even though Ph.D. students were knowledgeable about theory building, they were not familiar with theory testing. The experiment conducted showed that Ph.D. students were not aware of the variety of pathways that can be used to test theory. They also seemed to be unaware of the diversity of theory testing activities that can be crowdsourced. Therefore, it is one thing to know the steps to take; it is another to know what types of activities are involved in those steps, especially during the validation stage, which comprises of data collection and analysis. Having identified those activities, there are different pathways to choose when carrying out the testing process. A pathway is a set of linked activities that instantiate an intent from the beginning to end, while an Intent summarizes the general purpose of a study taking a theory testing

perspective, which does not concern the specific problem under investigation (Penker, 2000). Having studied the conceptual framework and model which constitute a solid foundation for this study, the next logical step in our research is the development of a support system which aims at assisting researchers to know the theory testing steps, theory testing pathways and activities, crowdsourcing requirements and how to make decisions as to which theory testing activity can be crowdsourced. This type of support is usually associated with Decision Support Systems (DSSs) (Arnott & Pervan, 2005; Hosack et al., 2012). This research objective of this paper is to *analyze and design a CS analytic framework for supporting the integration of CS in theory*. This is a step towards developing the primary artefact - a DSS, which is a future study.

The remainder of this paper is organized as follows. Section two describes the research background of the study. Section three describes the method adopted for the study. In Section four, we look at the tool architecture, development and evaluation. Finally, Section five highlights the research contributions.

## Research Background

Several researchers have considered the adoption of CS in theory testing. For instance, Lowry et al. (2016) advocated CS as a way to increase the quality of data collection. The advantages brought by CS concerning demographics, psychometrics, and structural properties of data samples were highlighted by Steelman et al. (2014). We note that using crowdsourcing for data collection has been around for some time. In particular, crowdsourcing is becoming a very common way to deliver questionnaires in behavioral research (Bates and Lanza 2013; Behrend et al. 2011; Jarmolowicz et al. 2012; Peer et al. 2017). In user studies, it has also started to be used as an instrument to collect user data (Kittur et al. 2008; Stewart et al. 2017). And it is also regarded as an important component of citizen science, where it supports the distributed collection of research data (Bonney et al. 2009; Gura 2013). However, the relationship between crowdsourcing and theory testing with human participation has not yet been systematically explored and established. Establishing this relationship is important from both theoretical and practical perspectives. From a theoretical viewpoint, finding innovative ways to research humans is paramount to rigorously test existing theories and advance new theories, while from a more practical viewpoint, existing crowdsourcing platforms open up opportunities to develop relatively easy-to-implement methods to conduct theory testing. The possibilities brought by crowdsourcing seem even more relevant in a research landscape where problems are becoming increasingly complex and the time available for completing doctoral studies remains constant, if not shrinking (Blagojević et al. 2017). This is especially true in the Information Systems (IS) field where theory building, and therefore theory testing, have a fundamental role in shaping the identity of the field (Gregor 2006). Furthermore, it has been reported that participation rates in scientific studies have been declining (Galea and Tracy 2007; Van Gelder et al. 2010). Finally, as alternative dissertation formats, such as the three-paper dissertation, are becoming increasingly popular, the need for multiple theory testing efforts is likely to increase (Fong 2017; Jump 2015). The crowdsourcing strategy offers a new way to incentivize humans to participate in research.

### ***Crowdsourcing Requirements with Respect to Theory Testing***

Requirements are important and should be clearly defined because it is essential for both the technical designers and the users of the new system or developed tool to be absolutely clear on what they want and expect the system to do (Mumford, 1985). Requirements here tend to lean more to the non-functional requirements, which deals more on the constraints, qualities, characteristics, and properties that a system or tool should possess (Malan & Bredemeyer, 2001). For crowdsourcing to be used for theory testing, these set of requirements must be considered both by the user and for the tool design. The requirements can be categorized into two: Mandatory requirements – are requirements that must be met before CS can be used, they are based on platform conditions and limitations and are what CS is made up of. Desirable requirements – are requirements that are flexible depending on what aspect the researcher is exploring. They don't have to be met for CS to be considered.

#### **Mandatory Requirements**

- **Horizontal decomposability:** This is a requirement that states that tasks must be decomposable in several simple, independent tasks that can be executed in parallel by the crowd. Task decomposition

leads to faster task turn-around time, which is the main advantage of CS (Heer & Bostock, 2010; Vukovic et al., 2010).

- **One-off:** This requirement entails that the crowd task must be executed only once. Repeated tasks are usually not allowed (platform limitation) (Mason & Suri, 2012; Yang et al., 2008).
- **Simplicity:** The crowd task should be meaningful, easy to delineate and have clear inputs and outputs. Simplicity is considered one of the most important conditions for the success of CS (Geerts, 2009).
- **Limited interaction:** The crowdsourcer and the crowd have a limited channel for interaction, this implies that communication loops are usually not possible (platform limitation) (Wexler, 2011; Wu et al., 2013).
- **Computer mediation:** The crowd task should be remotely executed, using the Internet for coordination and communication. CS is mainly carried out on different platforms, and most CS platforms are web-based (Andriole, 2010; Kleemann et al., 2008).

### **Desirable Requirements**

- **Time-boundness:** It should be possible to complete the crowd task in a bounded period of time. One of the advantages of using CS is having and obtaining a quick turn around on task, therefore, timing is crucial to any CS task (Bernstein et al., 2011; Aniket Kittur et al., 2008).
- **Skills matching:** It should be possible to match the domain knowledge required by the crowd task with the skills of the selected members of the crowd. Mismatch of skills and tasks leads to low-quality results (Geiger et al., 2011; Schenk & Guittard, 2011).
- **Training:** Training, if required, should not be complex or prolonged. In current CS platforms, training is still considered challenging (Park et al., 2014).

## **Research Method**

The design science paradigm was adopted for this research. We adopted this paradigm because design science is a problem-solving paradigm that seeks to create and evaluate IS artefacts to solve identified organizational problems (Gregor & Hevner, 2013). Design science involves the designing of artefacts, making research contributions, evaluating the designed artefacts, and communicating the results to appropriate audiences (Peppers et al., 2007). By artefacts we mean anything that can be transformed into material existence as an artificially made object (e.g., model, instantiation) or process (e.g., method, software) (Goldkuhl, 2002; Gregor & Hevner, 2013).

Design science is focussed on the iterative design of IS artefacts. In this study, our iterative development started with the development of the conceptual framework, then we proceeded to the model development. We consider these artefacts as secondary artefacts, because they aid in the development of the instantiation artefact, which is the primary artefact. The analytic framework which is considered as a secondary artefact evolved from the development of previous secondary artefacts and serves to provide users' support and guidance (Hevner & Chatterjee, 2010; Von Alan et al., 2004). The design science paradigm which views the to-be-built tool as an artefact distinguishes two primary research activities: build and evaluate (March & Smith, 1995). The build activity involves generating an initial artefact based on a preliminary problem frame and a set of requirements, which then lead to the artefact design. The build activity is then followed by an evaluation activity, which confronts the design against the objectives and suggests a new problem frame, which may then lead towards another design cycle, until a satisficing solution is obtained (Sein et al., 2011). The evaluation for this artefact will be done using example cases from existing literature (Sonnenberg & vom Brocke, 2011). This evaluation serves the purpose of showing that the artefact design incorporates the solution to the stated problem (Sonnenberg & vom Brocke, 2011).

## **Decision Support Crowdsourcing Analytical Framework**

The Decision Support Crowdsourcing Analytic Framework Development (DSCAFD) is based on the two secondary artefacts, which are the conceptual framework and the model (Enwereuzo et al., 2017, 2019). The model feeds the DSCAFD and serves as a knowledge base to the DSCAFD. The DSCAFD basically allows users to play with the model using a set of restrictions and requirements.

## Crowdsourcing Attributes

The analysis of which activities can be crowdsourced or not needs to be systematic and anchored on a transparent procedure. To increase systematicity and transparency, we first define the set of crowdsourcing attributes considered in this study. These attributes are generated from the requirements mentioned above. The attributes are organized into four categories related to the task, medium, crowd, and quality. We consider these to be the core attributes characterizing crowdsourcing.

**Task.** The task corresponds to “what the crowd has to do” (Estelles-Arolas & Gonzalez-Ladron-de-Guevara, 2012). Some tasks are completely independent of the context in which they are executed, e.g. classifying photos according to a set of given categories, which means they are *context-free*. Other tasks depend on a specific context to be executed, such as a physical location, operational environment, or virtual place. Examples include gathering data on public libraries, while driving a car, or from commonly used social media. In these cases, tasks are *context-dependent*. Finally, some tasks may have to be performed in restricted contexts, which may not be easily accessible or reproducible. Examples include gathering data from hospitals and selecting records from a private database. In these cases, we designate the tasks as *context-restricted*. We consider that context-free tasks are easy to crowdsource, while context-dependent tasks are difficult to crowdsource, and context-restricted tasks cannot be crowdsourced. The rationale for considering context-dependent tasks difficult to crowdsource is that certain constraints have to be enforced, which make tasks more difficult to specify, control and execute. In principle, we consider context-restricted tasks as not crowdsource-able. This is because it would be possible to circumvent the restrictions, then they would be dependencies. Consider as an example that you would like to crowdsource data collection in a hospital. Sending the crowd to the hospital without consent would be unethical, and therefore the task would be context-restricted. However, if you can get consent from the hospital, then the task becomes context-dependent. Tasks can also be classified according to complexity, structure, and interdependence. Regarding complexity, a task can be classified as *micro* if it can be done quickly, uses repetition, or requires little cognitive effort; or it can be classified as *macro* if it requires considerable cognitive effort (e.g. planning what to do and making decisions (Kittur et al., 2011)). Structure classifies the task’s problem and solution as either *well-defined* (e.g. classifying photos using a set of tags) or *ill-defined* (e.g. generating new research ideas) (Nakatsu et al., 2014). Interdependence classifies how the task outputs are generated. The task outputs can be *pooled* (e.g. individual ideas can be pooled together), *coordinated* (e.g. gather data first, and then analyze data), or *shared* (e.g. collaborative writing) (Crowston, 1997). We consider that well-defined, pooled and micro-tasks can be easily crowdsourced, while ill-defined, coordinated or shared tasks can be difficult to crowdsource.

**Support.** Firstly, we consider communication support, which allows sending the inputs and getting the outputs from the crowd. Regarding this attribute, we consider the following values: *channel available* (e.g. using a crowdsourcing platform or social media), *channel must be developed* (e.g. a mobile application must be created to communicate with the crowd), or *channel is unavailable* (e.g. the function must be communicated within proprietary software, which cannot be accessed by the crowd). Secondly, we consider process support, which allows the crowd to perform the function. Regarding this attribute, we consider the following values: *process support not needed* when the function can be carried out by the crowd with no additional support (e.g. counting birds in a forest), *process support exists* when there is an existing technology that can be used by the crowd to execute the function (e.g. a crowdsourcing platform such as MTurk, Threadless, and iStockphoto), *develop process support*, when such support must be developed by the researcher (e.g. a website must be created to classify an ecosystem), and *process support unavailable*, when process support is needed but cannot be developed. Finally, we have to consider what type of coordination support is required to collect the outputs generated by the crowd. Regarding this attribute, we consider the following values (Crowston, 2012;): *pooled outputs* (e.g. individual ideas can be pooled together), *coordinated outputs* (e.g. gather data first, and then analyze data), or *shared outputs* (e.g. collaborative writing). We consider that crowdsourcing cannot be done if no channel is available or process support is unavailable. If the communication channel is available, the outputs can be pooled, and process support exists or is not needed, then it is easy to crowdsource. Also, if the communication channel is available, process support exists, and outputs are coordinated or shared, then it is easy to crowdsource. All other combinations of attributes should be seen as difficult to crowdsource. In particular, if the communication channel must be developed or process support must be developed, it is difficult to crowdsource.

**Crowd.** The crowd is the collection of people who execute the task. The participation of the crowd is either *voluntary* (e.g. open software development) or *rewarded* (e.g. prizes and money). We should also consider if either an *incentive mechanism exists* (e.g. most crowdsourcing platforms offer reward mechanisms) or an *incentive mechanism is missing*. We consider it is easy to crowdsource if an incentive mechanism exists and the participation is rewarded, and also if an incentive mechanism is missing but participation is voluntary. An example of the latter case is the successful use of crowd mapping in disaster relief actions, which depend on voluntary contributions (Gao et al., 2011). The crowd may have to be either *individually selected* or *profiled* according to a set of criteria (e.g. generic skills, geographical locations, or answers to a questionnaire (Amazon, 2011)). We consider that it is easy to crowdsource when the crowd can be profiled, while individually selecting the participants makes it difficult to crowdsource.

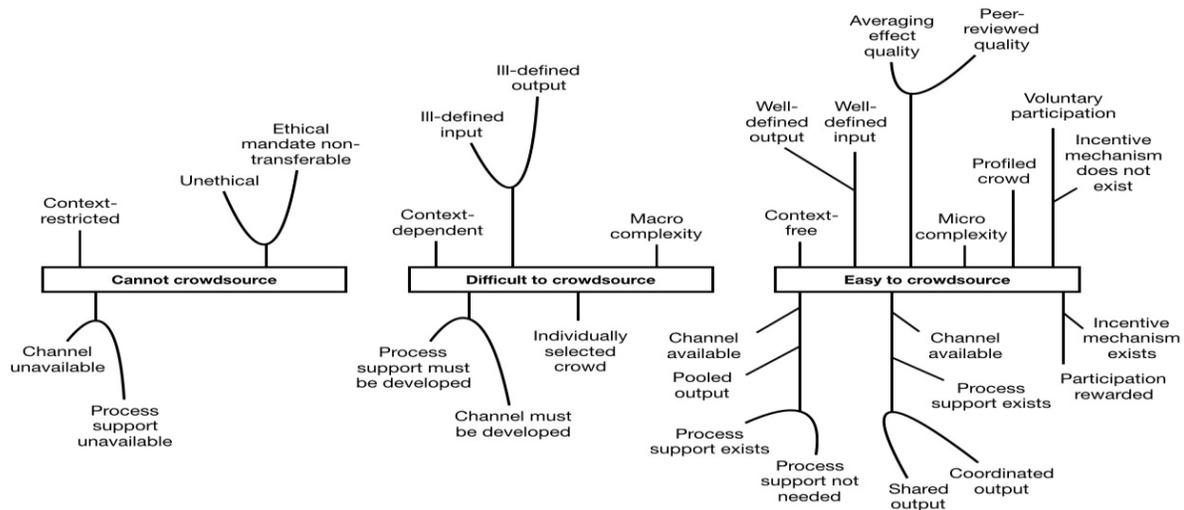
**Quality.** Considering the loose characteristics of crowdsourcing, it seems natural that the task outputs should be subject to quality assessment. Quality assurance may rely on the *averaging effect*, which uses multiple data sources to increase accuracy and trust (Brunt & Meidell, 2018). It can also be *peer-reviewed*, where quality assurance is actually done by the crowd. And finally, it may also rely on *quality monitoring*, where either a sample or the entire data set is checked for accuracy by the researcher. We consider that, if the averaging effect or peer-reviews are viable, then the task is easy to crowdsource. Monitoring has a neutral impact in relation to crowdsourcing, as it is a traditional approach to quality assessment.

**Ethics.** We should also consider if it is either *ethical* or *unethical* to execute a task. We suggest that an unethical task cannot be crowdsourced (e.g., gathering data from inside a hospital without consent). If the task is ethical, we need to further consider if the *ethical mandate is transferable* (i.e., if the mandate obtained by the researcher can be transferred to the crowd) or the *ethical mandate is non-transferable* (i.e., only the researcher has the mandate, which cannot be transferred to anyone else). We suggest that neither unethical tasks nor non-transferable mandates should be crowdsourced.

### **Analytic Framework Development**

Based on the set of crowdsourcing features discussed above, we now propose an analytic procedure to check if a task can be crowdsourced or not. The procedure is applied to each individual task that one wishes to crowdsource and checks systematically how the attributes related to the task, structure, crowd, quality, and ethics determine the feasibility to crowdsource. As illustrated in Figure 1, the feasibility test may have three results: 1) we cannot crowdsource; 2) we can crowdsource, but it may be difficult to accomplish; and 3) it is easy to crowdsource. The first check to do regards the impossibility to crowdsource (Figure 1, left). We discussed five conditions under which it is not possible to crowdsource. They are related to the task context, channel availability, process support, and ethics. The next feasibility check determines if crowdsourcing is difficult to accomplish (Figure 1, centre). We found seven conditions making a task difficult to crowdsource. They are related to the task context, type of task, types of input and output, channel and support development, and crowd selection. The final feasibility check determines if crowdsourcing is easy to accomplish (Figure 1, right). Here, various combinations of conditions have to be fulfilled to determine that a task is easy to crowdsource.

Overall, they consider 19 different conditions. We observe that some combinations of conditions may lead to conflicting results. For instance, if the task is classified as macro, then it is difficult to crowdsource. And if the crowd has to be profiled, then it is easy to crowdsource. In these cases, we adopt a pessimistic approach and suggest that any conditions leading towards the impossibility to crowdsource should take precedence over any other conditions; and any conditions leading towards the difficulty to crowdsource should take precedence over any other conditions indicating it is easy to crowdsource. Finally, in the eventual cases where a combination of conditions would lead towards none of the above, we suggest the default should be that it is difficult to crowdsource. We note that the procedure shown in Figure 1 could have been more formally presented as a decision tree. However, because of the high number of attributes and conditions to consider, such a tree would be too large to be actually useful. The adopted graphical scheme seeks to summarize the various decisions that have to be considered in a more compact way, which is done at the cost of not showing every possible combination of attributes and conditions in the whole decision path. To better understand the adopted graphical scheme, we need to define the meaning of the various graphical elements used. Boxes represent the assessment, which determines if a particular activity can be crowdsourced or not and if yes, if it is easy or difficult. V-branches represent a logical



**Figure 1. Feasibility check that determines if a function can be crowdsourced (A V-branch corresponds to a logical AND while a U-branch corresponds to an OR).**

“AND”, while U-branches represent a logical “OR”. For instance, if we have well-defined outputs AND well-defined inputs, then it is easy to crowdsource. If we have ill-defined inputs OR ill-defined outputs, then it is difficult to crowdsource. Next, we use this procedure in some example cases.

## Justification by Applying the Procedure: Example Cases

The examples were selected to illustrate different approaches to theory testing (see figure 2). This is the first step adopted in justifying the procedure.

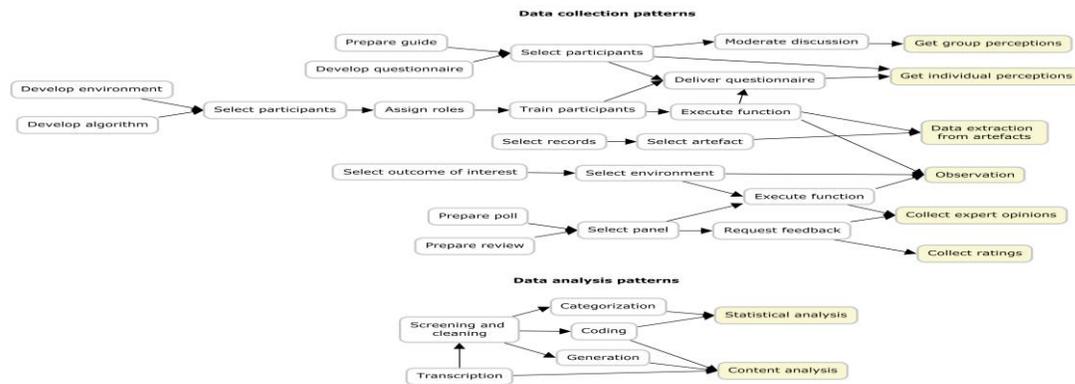
### Example 1

This example uses an article written by Burton-Jones and Volkoff (2017), which was drawn from the pool of reviewed papers. The study adopted an inductive approach to theory building and the case study approach to theory testing. Theory testing comprised the demonstration of the proposed theoretical constructs in a specific case dealing with a community-based healthcare system. The authors obtained their data through focus groups, interviews, observations of system use by shadowing workers, and document analysis (Section 3.3). Matching these specific theory testing activities to the pattern system shown in Figure 2, we can infer which of the patterns best suit what was done. We first note that an environment was selected (a community care unit of a regional health authority in Canada). The paper does not provide details about other activities done between selecting the environment and selecting the participants. After the participant selection, the authors report they got a group and individual perceptions. Regarding document analysis, the paper provides a brief account of the types of artefacts that were selected for data extraction (e.g. business cases and project descriptions). Finally, the paper provides evidence that specific outcomes of interest and environments were selected for observing workers (shadowing). We can, therefore, identify the following data collection patterns that have been used:

- Select environment - [No other activity reported] - Select participants - Moderate discussion - Get group perceptions
- Select environment - [No other activity reported] - Select participants - Get individual perceptions
- Select records - Data extraction from artefacts
- Select outcome of interest - Select environment – Observation

Burton-Jones and Volkoff (2017) analyzed data by transcribing the outputs from focus groups, interviews and shadowing, coding all data, and finally performing content analysis. We can, therefore, identify the two data analysis patterns that have been used:

- Transcribe - Code - Content analysis
- Code - Content analysis



**Figure 2. Theory testing patterns (Adapted from (Enwereuzo et al.,2017, 2019))**

Crowdsourcing was not used in this article. However, we can discuss how it could have been applied using the schema presented in Figure 1.

**Select environment:** This is a context-restricted activity because the researcher alone has to determine the environment which best suits the study. In this example, the environment was a regional health authority. Considering the nature of the activity and the research goals, we recognize that it cannot be crowdsourced, which is aligned with the results from our procedure.

**Select participants:** Burton-Jones and Volkoff (2017) note the selection of participants had to be carefully done. They targeted specific sites belonging to the same community care unit. Within those sites, they targeted specific types of participants, including front-line staff with access to the system (clerks and clinicians), middle managers, directors, and executives. Furthermore, as noted earlier, the focus groups participants were carefully selected to avoid having supervisors and subordinates in the same group. Based on these constraints, we suggest the activity has macro complexity and therefore is difficult to crowdsource. The activity also seems context dependent, since the selection of participants takes into consideration the characteristics of the target organization. Therefore, we regard this activity as context-dependent and therefore difficult to crowdsource. Finally, considering all requirements and constraints, it also seems that it would be difficult to profile the crowd. So, the crowd would have to be individually selected, which again makes the activity difficult to crowdsource.

**Moderate discussion:** Burton-Jones and Volkoff (2017) do not provide many details about the focus groups, which suggests that a typical approach was adopted, with the researchers conducting the discussion of topics of interest (an observation supported by the statements reported in the paper, which cover a variety of issues with the system). We suggest this is a macro, context-dependent activity. It has ill-defined outputs, because of the complexity associated with moderating a group of people while adapting the script to what the participants report and what the researchers find interesting. This suggests the activity is difficult to crowdsource. However, another concern is about ethics. Reflecting on the sophistication of the topic, i.e. use of patient health records by health professionals, the diversity of stakeholders, and the complexity of data the researchers aim to collect, i.e. data about effective use, do we think it would be ethical to crowdsource such an activity? We suggest the answer should be no and consequently, the activity should not be crowdsourced.

**Get group perceptions:** From what Burton-Jones and Volkoff (2017) report, this activity had macro complexity, was context-dependent, and had ill-defined inputs and outputs, all contributing to being difficult to crowdsource.

**Get individual perceptions:** Details were not given by Burton-Jones and Volkoff (2017) about the interviews, and how they were conducted, so we assume that a typical approach was adopted using semi-structured interviews. We suggest this activity would be difficult to crowdsource because it has macro complexity and is context-dependent. Burton-Jones and Volkoff (2017) highlight the interviewers had to participate in several preparation sessions to understand the work context, which supports the classification as context-dependent.

**Select records:** The same argument we used regarding the “select environment” activity applies here. This activity cannot be crowdsourced.

**Data extraction form artefacts:** Burton-Jones and Volkoff (2017) do not report any specific procedure to extract data from artefacts, which suggests they moved immediately to data analysis.

**Select outcome of interest:** This is a context-restricted activity because only the researchers knew the outcome of interest, which was to develop and test a theory of effective use. Considering the nature of this activity, we recognize it cannot be crowdsourced.

**Observation:** As reported by Burton-Jones and Volkoff (2017), observation was done by shadowing front line workers. We suggest that this activity is difficult to crowdsource because it is context-dependent: it must be done in a specific environment, shadowing health workers like nurses, physicians, etc. However, a mandate to gather data using the crowd would have to be secured by the researchers, which seems difficult to obtain in this case, as an unknown crowd would have to wander in a patient care environment. Consequently, we suggest this activity cannot be crowdsourced.

**Transcribe:** We suggest that this activity is easy to crowdsource because it has micro complexity, is context-free and has a clear output. As reported by Burton-Jones and Volkoff (2017), the gathered data was about the use of a new software system, which suggests the data does not raise privacy issues. Therefore, the transcription could have been crowdsourced.

**Code:** Burton-Jones and Volkoff (2017) report that coding was done by identifying emerging themes. This activity has micro complexity, which suggests that it could have been crowdsourced. However, the output is ill-defined, which suggests that instead, it is difficult to crowdsource.

**Content analysis:** We suggest this activity would be difficult to crowdsource because it has an ill-defined output, which is based on the analyzed data. Furthermore, it has macro complexity, since it requires particular skills to interpret the data using a specific framework.

All in all, the research reported by Burton-Jones and Volkoff (2017) had 17 theory testing activities, of which six were primary and eleven secondary. Of the primary activities, one could not have been crowdsourced, and five would be difficult to crowdsource. Regarding secondary activities, four could not have been crowdsourced, four would have been difficult to crowdsource, and one could have been easy to crowdsource.

## **Example 2**

In this example, we consider a study by Sanyal (2016), which was also drawn from the pool of reviewed papers. The author studied the effects of bidding strategies on the economic performance of auctions and used an experimental approach to theory testing. As discussed in detail in Section 4 of the paper, theory testing consisted of the demonstration of the proposed theoretical constructs in the laboratory using different auction designs. Experimental data was obtained by the author in the laboratory (Section 4.2). A particular environment was developed for the experiment (a bidding interface). The paper mentions the participants were undergraduate university students. The data gathering procedure adopted the following pattern:

- Develop environment - Select participants - Run function - Data extraction from artefacts

In this particular case, no specific roles had to be assigned to the participants, since they were all bidders. Training was also not considered in the experiment. Regarding data analysis, Sanyal (2016) report that some activities we identify in the pattern system were automated by the bidding interface. In particular, the interface automatically removed dead bids from the data set, which is an automated form of screening and cleaning. The interface also automatically generated data in two groups, from symmetric and asymmetric bidders, which is a form of categorization. Considering that these two activities were automated, the data analysis pattern ended up having one single activity to consider:

- [Automated screening and cleaning] - [Automated categorization] - Statistical analysis

Crowdsourcing was not used in this article. However, we can discuss how it could have been applied using our procedure.

**Develop environment:** According to Sanyal (2016), a Web site with a bidding user-interface had to be developed for the experiment. Therefore, this is a context-dependent and macro activity because it requires specific skills to develop the software. Considering the nature of the activity, we suggest it would be difficult to crowdsource.

**Select participants:** This activity has micro complexity. According to Sanyal (2016), there were no criteria for selecting the participants. For that reason, this is also a context-free activity. Therefore, we suggest this activity would be easy to crowdsource.

**Run function:** This activity falls into micro complexity because it is repetitive and requires a single skill, which is to place bids. Being only focused on bidding, it is also context-free. Process support would have to be developed to run the experiment. However, since we have already considered that decision element

in the “develop environment”, here we have to consider that process support exists. Regarding the communication channel, the researcher used email to recruit the participants on campus, which means a communication channel is available. Sanyal (2016) noted the participants were rewarded through cash payments, which suggests an incentive mechanism would be needed. However, the same argument regarding process support applies here: since the environment has to be developed, the incentive mechanism can be integrated into the software. Therefore, we regard this activity as easy to crowdsource.

**Data extraction from artefacts:** Specific details were not given by Sanyal (2016) about how the data was extracted, but we can infer from the paper that the Web site stored the user data generated during the experiments. Therefore, this activity has micro complexity because it only involves retrieving stored data. It also has well-defined inputs and outputs. We suggest that this activity would be easy to crowdsource.

**Statistical analysis:** Sanyal (2016) compared the effects of price in jump bidding using statistical techniques. We suggest this activity has macro complexity and therefore is difficult to crowdsource.

## Discussion, Recommendations, and Conclusion

This section is structured around the artefact developed by this research and its contributions to research and practice. The artefact we have developed is the mechanism to check if a theory testing activity can be crowdsourced or not. The value of this artefact lies in its capacity to help researchers making complex design decisions, in this particular case regarding how they can utilize crowdsourcing to their advantage. Gathering and analyzing data to prove a theory can be a daunting endeavor. The complexity of today’s world, as well as the sophistication of the research undertaken today, suggests researchers need to collect very large amounts of data pertaining to complex environments, considering a multitude of factors, contextual elements, and stakeholders, which require new, innovative approaches.

The proposed mechanism helps researchers systematically checking if crowdsourcing can be applied to a wide range of theory testing activities organized as a pattern system. So, our proposition is not just to check if data collection per se can be crowdsourced or not. Our proposition is to check if every activity related to theory testing can be crowdsourced or not. We applied the proposed mechanism to two examples illustrating very different research endeavors. In the two examples, we could make a rational recommendation regarding the decision to crowdsource or not using simple questions, which ask about specific characteristics of the research project in consideration to attributes required by crowdsourcing. In spite of the potential values described above, we should also recognize some limitations of this research. One limitation to consider is that the decision to crowdsource may extend beyond the intrinsic characteristics of theory testing activities, for instance, taking into consideration the specific characteristics of the research undertaken. From a research point of view, our work extends existing research by showing that crowdsourcing can be used for different activities apart from surveys, which most of the existing literature that used crowdsourcing focused on. Through the sample cases, we were able to operationalize and demonstrate how some of the activities gotten from those papers which were not CS could be CS creating awareness and enlightenment.

Future work that has to be done involves transforming this framework into a tool which will operationalize the framework and provide further support to the researcher. This development will contribute to a better understanding of the framework. After this is done, the utility of the tool needs to be confirmed. Thus, the future research should provide further details about the framework regarding information, data and technical structures.

## References

- Aken, J. E. v. (2004). Management research based on the paradigm of the design sciences: the quest for field-tested and grounded technological rules. *Journal of management studies*, 41(2), 219-246.
- Arnott, D., & Pervan, G. (2005). A critical analysis of decision support systems research. *Journal of information technology*, 20(2), 67-87.
- Bacharach, S. B. (1989). Organizational theories: Some criteria for evaluation. *Academy of management review*, 14(4), 496-515.
- Burton-Jones, A., & Volkoff, O. (2017). How can we develop contextualized theories of effective use? A demonstration in the context of community-care electronic health records. *Information Systems Research*, 28(3), 468-489.

- Colquitt, J. A., & Zapata-Phelan, C. P. (2007). Trends in theory building and theory testing: A five-decade study of the Academy of Management Journal. *Academy of Management Journal*, 50(6), 1281-1303.
- Crowston, K. (2012). Amazon mechanical turk: A research tool for organizations and information systems scholars *Shaping the Future of ICT Research. Methods and Approaches* (pp. 210-221): Springer.
- Enwereuzo, I., Antunes, P., & Johnstone, D. (2017). Towards the Development of a DSS Supporting the Integration of Crowdsourcing in Theory Testing: Conceptual Framework and Model. Paper presented at ECIS.
- Estelles-Arolas, E., & Gonzalez-Ladron-de-Guevara, F. (2012). Towards an integrated crowdsourcing definition. *Journal of Information Science*, 38(2), 189-200. doi:10.1177/0165551512437638
- Geiger, D., Seedorf, S., Schulze, T., Nickerson, R. C., & Schader, M. (2011). *Managing the Crowd: Towards a Taxonomy of Crowdsourcing Processes*. Paper presented at the AMCIS.
- Gregor, S., & Hevner, A. R. (2013). Positioning and presenting design science research for maximum impact. *MIS quarterly*, 37(2), 337-355.
- Hevner, A., & Chatterjee, S. (2010). *Design research in information systems: theory and practice* (Vol. 22): Springer Science & Business Media.
- Hosack, B., Hall, D., Paradice, D., & Courtney, J. F. (2012). A look toward the future: decision support systems research is alive and well. *Journal of the Association for Information Systems*, 13(5), 315.
- Howe, J. (2006). The rise of crowdsourcing. *Wired magazine*, 14(6), 1-4.
- Järvinen, P. (2007). Action research is similar to design science. *Quality & Quantity*, 41(1), 37-54.
- Kleemann, F., Voß, G. G., & Rieder, K. (2008). Un (der) paid innovators: The commercial utilization of consumer work through crowdsourcing. *Science, technology & innovation studies*, 4(1), PP. 5-26.
- Malan, R., & Bredemeyer, D. (2001). Defining non-functional requirements.
- Malone, T., & Crowston, K. (1994). The Interdisciplinary Study of Coordination. *ACM Computing Surveys*, 26(1), 87-119.
- March, S. T., & Smith, G. F. (1995). Design and natural science research on information technology. *Decision Support Systems*, 15(4), 251-266.
- Mason, W., & Suri, S. (2012). Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods*, 44(1), 1-23.
- Mumford, E. (1985). Defining system requirements to meet business needs: a case study example. *The Computer Journal*, 28(2), 97-104.
- Nakatsu, R., Grossman, E., & Iacovou, C. (2014). A taxonomy of crowdsourcing based on task complexity. *Journal of Information Science*, 40(6), 823-834.
- Niederman, F., & March, S. (2015). Reflections on Replications. *AIS Transactions on Replication Research*, 1(1), 7.
- Peffers, K., Tuunanen, T., Rothenberger, M. A., & Chatterjee, S. (2007). A design science research methodology for information systems research. *Journal of management information systems*, 24(3), 45-77.
- Penker, M. (2000). *Business Modeling with UML: Business Patterns at Work*: New York, NY: John Wiley & Sons.
- Power, D. J. (2008). Decision support systems: a historical overview. *Handbook on Decision Support Systems 1*, 121-140.
- Sanyal, P. (2016). Characteristics and Economic Consequences of Jump Bids in Combinatorial Auctions. *Information Systems Research*, 27(2), 347-364.
- Schenk, E., & Guittard, C. (2011). Towards a characterization of crowdsourcing practices. *Journal of Innovation Economics & Management*(1), 93-107.
- Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., & Lindgren, R. (2011). Action design research. *MIS quarterly*, 37-56.
- Smith, D., Manesh, M. M. G., & Alshaikh, A. (2013). How can entrepreneurs motivate crowdsourcing participants? *Technology Innovation Management Review*, 3(2).
- Sonnenberg, C., & vom Brocke, J. (2011). *Evaluation patterns for design science research artefacts*. Paper presented at the European Design Science Symposium.
- Von Alan, R. H., March, S. T., Park, J., & Ram, S. (2004). Design science in information systems research. *MIS quarterly*, 28(1), 75-105.
- Vukovic, M., Lopez, M., & Laredo, J. (2010). *People cloud for the globally integrated enterprise*. Paper presented at the Service-Oriented Computing, ICSOC/ServiceWave 2009 Workshops.
- Wexler, M. N. (2011). Reconfiguring the sociology of the crowd: exploring crowdsourcing. *International Journal of Sociology and Social Policy*, 31(1/2), 6-20.