Design and Development of a DSS Supporting the Integration of Crowdsourcing in Theory Testing: A Design Science Perspective

Completed Research Paper

Ijeoma Enwereuzo
Victoria University of Wellington
Wellington, New Zealand
Ijeoma.enwereuzo@vuw.ac.nz

Pedro Antunes
Victoria University of Wellington
Wellington, New Zealand
Pedro.antunes@vuw.ac.nz

David Johnstone
Victoria University of Wellington
Wellington, New Zealand
David.johnstone@vuw.ac.nz

Gert-Jan De Vreede
University of South Florida
Florida, USA
gdevreede@usf.edu

Abstract

The integration of crowdsourcing in behavioral research in the IS field offers several advantages and opportunities. This paper builds on prior study, employing a design science research (DSR) paradigm to design, develop and evaluate a tool that assists researchers in adopting crowdsourcing when testing theory about behavioral phenomena. The proposed tool is based on an extensive review of literature on how theory has been tested, and a pattern model that standardizes extracted concepts, activities, processes and relationships into patterns. In particular, we discuss the architecture of the proposed tool and present two prototypes, one used for knowledge articulation by representing, extracting, organizing and acting on relevant information and the other on decision making and recommendation for the tool users. Evaluation results show the applicability and utility of the tool.

Keywords: Decision support system, crowdsourcing, theory testing, design science

Introduction

Developing and testing theory about behavioural phenomena is common in the information systems (IS) field (Gregor 2006). However, it faces many conceptual and practical challenges. Researchers must consider a diversity of variables, settings and data necessary to form robust evidence about systems and human behaviour. Researchers also have to reflect on the diversity of methodologies, methods, processes, instruments, and tools available in the IS methodological toolkit (Bhattacherjee 2012). Furthermore, the operationalisation of data collection procedures, which may involve a large number of people, systems and processes, often require long periods of trial and error (Peer et al. 2017; Witschey et al. 2013). Considering all these difficulties, a crowdsourcing strategy may seem attractive to help researchers to accomplish their theory testing objectives.

Crowdsourcing is a managerial model that relies on information technology to outsource tasks to a large number of participants using different types of incentives such as remuneration, prizes and peer-esteem (Kietzmann 2017). Tasks can be massively distributed, done in parallel and fulfilled in short time periods. Furthermore, the task and the crowd can be easily managed through crowdsourcing platforms like Wikipedia, MTurk, Upwork, and InnoCentive.

Several researchers have already noted several advantages of adopting crowdsourcing when testing theory about behavioural phenomena. Some areas in social studies are using crowdsourcing as a de facto standard for delivering questionnaires (Bates and Lanza 2013; Shank 2016). Lowry et al. (2016 pointed out that crowdsourcing could increase the quality of data collection, e.g. by crowdsourcing quality control, diversifying methods and instruments, and ultimately collecting more data. Steelman et al. (2014) also highlighted several advantages of crowdsourcing in dealing with psychometrics, demographics and structural properties of data samples.

Enwereuzo et al., 2017 considered the feasibility of applying the crowdsourcing model to theory testing. In particular, the study developed a systematic procedure for checking if theory testing activities, either individually or collectively, could be crowdsourced. This study builds on prior work with the goal to *design, develop and evaluate a tool that assists researchers adopting crowdsourcing when testing theory about behavioural phenomena*. This research provides two complementary contributions: 1) an innovative decision support tool addressing a problem relevant to behavioural researchers in the IS field; and 2) results from the tool evaluation, which demonstrate the utility of the tool in resolving the research problem.

The remainder of this paper is structured as follows. Section two describes the research context of the study. Section three describes the tool development method. In Section four, we describe the design, development and evaluation of the tool. Finally, Section five highlights the research contributions and conclusions.

Research Context

Behavioural researchers in the IS field work with various types of empirical data, which they have to gather and analyse in order to determine whether the propositions articulated by theory are supported or not by the data. A variety of methods can be used by researchers. A common perspective splits methods into quantitative, qualitative and mixed categories (Creswell and Creswell 2017). Some of these methods require a considerable effort to apply. For instance, observation often requires shadowing the activities of multiple workers in their natural work settings over long periods of time. Then, qualitative analysis of video footage will require many hours going through video footage, indexing, transcribing, coding, condensing, and synthesizing data (Derry et al. 2010). To increase consistency and reliability, quality checks and duplication of activities, such as parallel coding, are often necessary (Thomas 2006). And in the end, the whole data gathering/analysis process may have to be repeated until reaching acceptable quality.

Investigating how crowdsourcing can alleviate the work done by behavioural researchers could have a significant impact in the IS field. In particular, expediting the theory building/testing cycle could foster new IS theory; and could also contribute to improving existing, or developing new, IS research methods. 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 to study participants (Bates and Lanza 2013; Behrend et al. 2011; Jarmolowicz et al. 2012; Peer et al. 2017). It has also started to be used as an instrument to collect system usage logs (Kittur et al. 2008; Stewart et al. 2017). And it is already an important component of citizen science, supporting distributed collection of field data (Bonney et al. 2009; Gura 2013). However, the relationship between crowdsourcing and theory building/testing has not yet been systematically explored and established.

Research Background and Goals

The whole design science research project focuses on the development of the DSS, some intermediate steps to achieve this goal has been developed in past studies. Early steps in establishing this connection were reported by Enwereuzo et al., 2017, 2019. These studies addressed three fundamental goals: 1) develop a conceptual framework of theory building/testing; 2) develop a model characterising how researchers have been testing theory; and 3) develop a procedure for checking which theory testing activities could be crowdsourced. These previous studies form the foundation for the development of this last phase of the project (see Figure 1). The following briefly outlines each of these three prior outcomes.

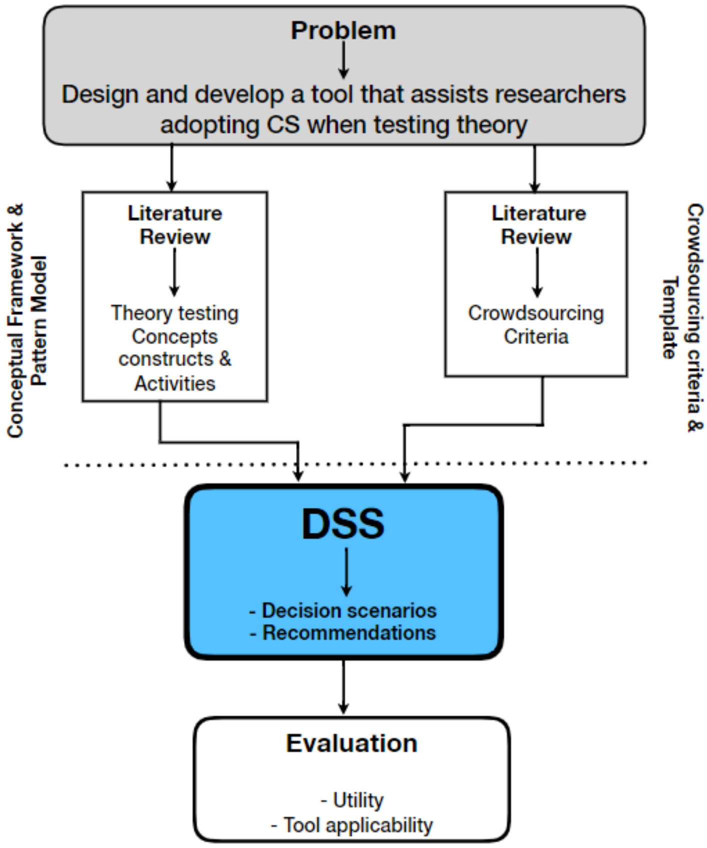

Figure 1. Research framework

Conceptual framework. The conceptual framework characterises the main concepts and constructs involved in theory development. The framework is significant because it characterises theory testing as a distinct component of theory building. Earlier frameworks blend theory testing with theory building. By separating theory building from theory testing, we make the research process more amenable for crowdsourcing, since theory testing activities are the ones that could benefit most from mass distribution to the crowd.

Model. The model is a key element in understanding how crowdsourcing can actually be applied to theory testing. The model synthesises an extensive literature review on the different ways in which researchers have been testing theory within the IS discipline. The model characterises theory testing as a system consisting of various research activities interconnected by a set of patterns, which realise several research goals. The model is significant because it codifies and integrates knowledge on how theory has been tested in practice. The model is also instrumental in our pursuit of bringing crowdsourcing into theory testing because it supports researchers identifying regularities in theory testing activities, which may then be individually or collectively checked if they can be crowdsourced or not.

Procedure. The question asked was "how can a researcher decide if a theory testing activity can be crowdsourced or not?" The developed procedure systematically uses a set of crowdsourcing criteria that can be used by researchers to decide if a theory testing activity, or a set of activities, can be crowdsourced or not. The main contribution of this procedure is to give clarity and systematicity to the decision-making process.

This study builds on these outcomes in order to:

- Assist researchers establish if Crowdsourcing (CS) is a good strategy to adopt for any of the theory testing steps. This assistance could be given as a guideline or recommendation.

- Build a comprehensive and integrated view of the theory testing activities and its associated crowdsourcing attributes.
- Provide a means for effective processing and presentation of knowledge as it relates to theory testing activities and crowdsourcing.

To achieve the outlined goals, we briefly consider what type of decision support system is needed.

Decision Support Systems View

Decision Support Systems (DSS) represent a research area in Information Systems (IS) with a long history that can be traced back to Simon's intelligence-design-choice model developed in 1960 (Hosack et al. 2012; Simon 1960). The main focus in this research area is centred on support and improving decision-making (Arnott and Pervan 2005). In many cases, these systems are not meant to replace the decision makers (in our case, the researchers), but to help them extend their capabilities and to make better and more informed decisions, highlighting why the term 'support' is important (Hosack et al. 2012).

A large number of DSS have been studied and developed in IS and its related fields, and this led to several taxonomies being proposed. Five types of DSS were suggested by Power (2008, which are: data driven, model driven, knowledge driven, document driven and communication driven DSS. While Arnott and Pervan (2005 developed a seven-type taxonomy, which are: 1) personal DSS for individual managers; 2) group DSS for a group of decision makers; 3) negotiation support systems for group support systems but involves negotiation functions; 4) intelligent DSS, uses artificial intelligence; 5) knowledge DSS, provides knowledge for storage, retrieval, transfer, and application; 6) data warehousing for processing large-scale (big) data; and 7) enterprise reporting and analysis systems. Based on these taxonomies, we note that our tool is meant to support individual researchers, and as such it is a personal, model-driven DSS. It is a model driven DSS because it uses limited data and parameters, and it does not require large data bases (Power 2008).

DSS are most often software based, and assist with decisions mostly by illustrating possible outcomes, which can be presented visually, numerically or by leading users through logical decision steps (Dicks et al. 2014). Some of these tools are complex models, mainly operated by their developers, while others have simple interfaces that can be used by non-experts. We adopt the latter description because the type of users that are targeted are relatively inexperienced users such as PhD students. We intend to develop our tool in such a way that it is user-friendly and easy to use.

Research Method

This study follows the Design Science Research (DSR) paradigm (Hevner and Chatterjee 2010b; Hevner et al. 2004). This paradigm was adopted because DSR concerns problem-solving (Gregor and Hevner 2013): 1) it seeks to create and evaluate innovative, first-of-a-kind IS artefacts; 2) which solve relevant organizational problems; while 3) founding the artefact design on a rigorous knowledge basis, which simultaneously informs the artefact design with existing theory, methods and artefacts, and contributes to knowledge by delivering new theory, methods and artefacts.

In our case, the artefact design is anchored on a conceptual framework of theory testing, a pattern model of theory testing, and a procedure for checking if theory testing activities can be crowdsourced. These three artefacts can be considered *secondary artefacts*, which inform the design of a decision support tool, which is the *primary artefact* of the study. The design of an artefact based on secondary artefacts provides support and guidance (Hevner and Chatterjee 2010a; Hevner et al. 2004) and establishes a solid knowledge foundation on which to design the primary artefact.

The DSR paradigm also distinguishes two primary research activities: *build* and *evaluate* (March and Smith 1995; Peffers et al. 2007). The build activity involves generating a preliminary problem frame: a set of requirements and design rules which then guide the artefact design. The build activity is followed by the evaluation activity, which confronts the design against the research goals. The evaluation results may suggest new problem frames, which may then lead towards new build activities, until a satisficing solution is obtained (Sein et al. 2011). Our build activity involves the design of decision support tool using a set of crowdsourcing attributes, feasibility checks as to activities that can be crowdsourced, while the evaluate activity assesses the *utility* brought by the decision support tool.

When choosing a DSR evaluation method, it is important to balance the interests of practitioners and researchers (Sonnenberg and vom Brocke 2011). While practitioners are interested in aspects such as the applicability and usefulness of the artefact, researchers are more interested in the validity of the research process and outputs. Furthermore, the evaluation may be done ex ante, where artefacts are evaluated prior to their implementation or actual construction, and ex post, where artefacts are evaluated after they have been designed and constructed (Pries-Heje et al. 2008).

Evaluation criteria for this study would focus on applicability, understandability, utility and impact of artefact on users (Prat et al. 2015). This evaluation would involve real users using real systems to accomplish real tasks in real settings (Pries-Heje et al. 2008). The evaluation method chosen are card sorting, observation and interviews. This method is used in this study due to the success of subsequent ones used in previous study and the amount of useful feedback received.

DSS Artefact

The DSS is based on the three secondary artefacts, which are the conceptual framework, the pattern model and the CS template. The pattern model and the CS template feeds the DSS and serves as its knowledge base. The DSS basically allows users to play with the model using a set of restrictions and requirements.

Tool's Conceptual Design

The conceptual design of the tool has three main components as shown in Figure 2, based on a framework proposed by Holsapple (2008): GUI (Graphic User Interface), Problem processing component (PPC) and the knowledge component (KC). We adopt the framework proposed by Holsapple (2008) because it has the essential components required of a decision support syste. The GUI component is responsible for the interaction between the users (researchers) and the tool. Inputs and parameters are accepted by this component. These inputs are then processed by the PPC. Input flows are also controlled by the PPC by choosing and adapting what elements the GUI presents. The PPC acquires the required knowledge from the KC, processes the knowledge, and returns the output to the user.

The KC consists of procedural knowledge extracted from the conceptual model (Holsapple 2008), which characterizes steps for theory testing. It is a step-wise specification of what to do to accomplish theory testing. We also have the descriptive knowledge obtained from the pattern model – this provides the definitions and descriptions of concepts that must be considered in the process of decision-making. It also includes relationships among concepts and activities. Reasoning knowledge provides the CS requirements and criteria constraining these elements. Using the KC, researchers can make decisions as to what pathway to select and its related activities and recommendations as to what activity can be CS is provided.

Tool Development

The tool development consists of two phases: template-based DSS and the user interface-based DSS. Lee and Bui (2000 suggests that templates are useful tools for representing, extracting, organizing and acting on relevant information. Therefore, the domain knowledge articulation is demonstrated through the template, of which knowledge from the pattern model is transferred into computer-based formulation.

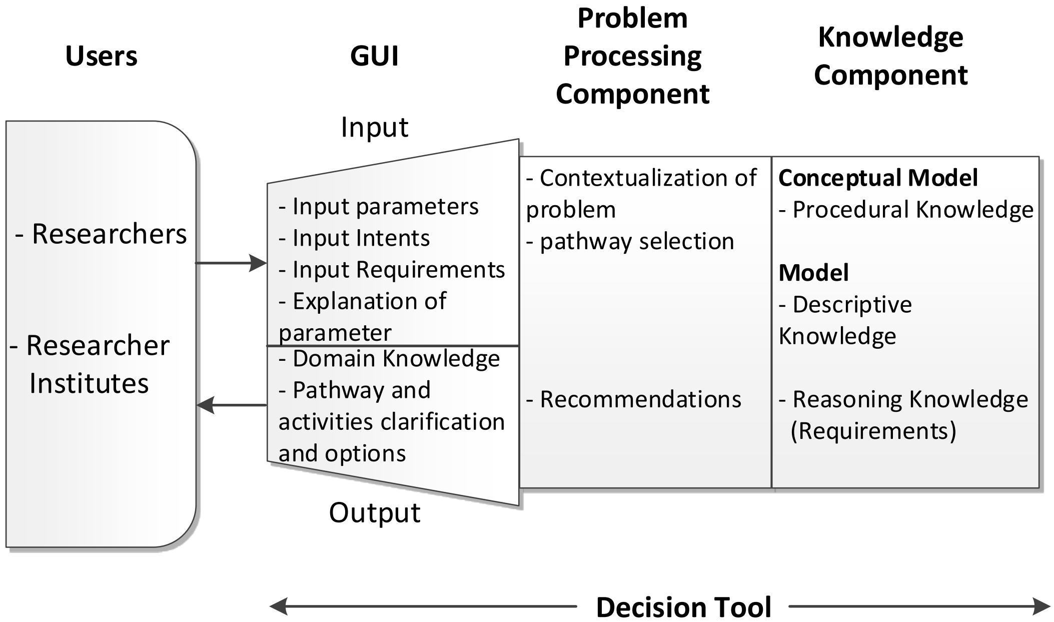

Figure 2. Tool's Conceptual Design (adapted from (Holsapple 2008))

Template-based Prototype

Templates were used because of the numerous advantages it provides as a preliminary stage in design and development of a tool. By providing a checklist of relevant information to look for, templates support proactive processing of information. They make it easier to compare multiple cases and infer generalizations from them. They support the design of a system, by providing information needed for enacting or automating the process (Lee and Bui 2000).

We develop a set of templates (See Figure 3) to systematically analyse how to adopt crowdsourcing for the various theory testing patterns previously defined. The templates allow for systematic checking of the properties of the crowdsourcing process (defined in the next section) while at the same time addressing the specific theory testing activities defined for a particular pattern. A template may suggest three types of results: cannot crowdsource the activity; difficult to crowdsource the activity; and easy to crowdsource the activity.

Crowdsourcing Criteria

The crowdsourcing process can generally be described as a sequence of four steps. The first step considers the adoption of crowdsourcing at the strategic level. Then, we consider the specification and execution of tasks. Finally, the outcomes from the different tasks must be consolidated. In order to be able to complete this process with success, several properties must be fulfilled.

Linked to the adoption of crowdsourcing at the strategic level, we find the important topic of ethics. First, we should consider if it may be *unethical* to crowdsource. A variety of situations may suggest crowdsourcing to be unethical, e.g., gathering data from inside a hospital without consent. If the strategy is accepted from an ethical perspective, it would typically involve acquiring ethical approval when necessary, or sticking to ethical guidelines especially as relates to research. We need to further consider if the ethical mandate, such as ethical approval obtained by the researcher is *transferable* or *non-transferable* to the crowd. We suggest that both unethical goals and non-transferable ethical mandates should not be crowdsourced.

After considering the strategy, we then have to examine the specification of one or more crowdsourcing tasks. One critical property to consider at this level is task complexity (Ghezzi et al. 2018; Jiang et al. 2015; Vondrick et al. 2013). A task can be classified as (Antunes et al. 2012; Reason 2008): *skills-*

based, if it can be accomplished as a routine by the crowd, in an almost unconscious way; *rules-based*, if it requires following a plan, with some latitude of decision-making by the crowd; and *knowledge-based*, if it is novel, requiring considerable problem analysis and solution-based decision-making by the crowd. We suggest that, as we move from skills-based to knowledge-based tasks, crowdsourcing becomes more difficult to achieve. This recommendation is based on research showing that the crowd prefers shorter tasks (Gadiraju et al. 2015); and also that micro tasks - which are tasks that have been broken down into smaller tasks, enable the crowd to generate higher quality work (Cheng et al. 2015).

Another facet to analyse are the inputs and outputs required by tasks. Input is the problem the crowd is asked to resolve, while output is the solution developed by the crowd, both of which can be classified as either *well-defined* (e.g. classifying photos using a predefined set of tags) or *ill-defined* (e.g. generating new research ideas) (Nakatsu et al. 2014). We contend that tasks with well-defined inputs and outputs can be more easily crowdsourced. Ill-defined tasks may be more difficult to crowdsource, since the crowd is more involved in defining what the task is about, which may lead to unwanted outcomes. This recommendation is based on research showing that the crowd prefers tasks with low associated risks (McInnis et al. 2016), and should be considered here especially when such tasks are related to theory testing.

Regarding execution, two important properties must be considered: context and support. Some tasks are completely independent of the context in which they are executed, e.g. classifying photos according to 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 in public libraries, driving a car, or gathering data 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 in private places such as military installations, coding data sets using proprietary tools, and selecting records from company databases. In these cases, we designate the tasks as *context-restricted*.

We suggest that context-free tasks are easy to crowdsource, while context-dependent tasks are difficult to crowdsource. The rationale for considering a context-dependent task difficult to crowdsource is that certain constraints have to be enforced, which make the task more difficult if not impossible to specify. In principle, we consider a context-restricted task not crowdsourceable. This is because, if it would be possible to circumvent the restrictions, then they would be dependencies, not restrictions. Consider, as an example, that you plan to crowdsource data collection in a hospital. Sending the crowd to the hospital without permission is not advisable, and therefore the task is context-restricted. However, if you would get permission from the hospital, then the task would be context-dependent.

Support concerns the methods, tools and training required to perform the task. Regarding this criterion, we consider the following values: support is *not needed* when the task can be carried out by the crowd without any support (e.g. counting birds in a forest); support *exists* when training, methods and tools exist and can be used by the crowd to execute the task (e.g. a photo sharing tool is available to upload bird photos); support *must be developed*, when it is needed but must be developed by the researcher (e.g. a website must be created to classify birds in a certain way); and support is *not unavailable*, when it is needed but the researcher is either unwilling or unable to provide it. We consider that crowdsourcing can be easily done when support is not needed or exists, crowdsourcing is difficult to accomplish when support must be developed, and crowdsourcing cannot be accomplished when support is necessary but not available.

			Activity 1	Activity 2	Activity n	Intent	Cannot crowdsource	Difficult to crowdsource	Easy to crowdsource
Strategy	Ethics	Transferable		Y		Proceed			X
		Non-transferable		Y		Stop	X		
		Unethical		Y		Stop	X		
Specification	Complexity	Skills-based	N	Y	N	Proceed with activities with "Y"			X
		Rules-based	Y	N	Y	Proceed with activities with "N"		X (1)	X (1)
		Knowledge-based	Y	N	Y	Proceed with activities with "N"	X (1)	X (1)	
	Input	Well-defined		Y		Proceed			X (2)
		Ill-defined		Y		Modify else stop		X	
	Output	Well-defined		Y		Proceed			X (2)
		Ill-defined		Y		Modify else stop		X	
Execution	Context	Free		Y		Proceed			X
		Dependent		Y		Modify if possible		X	
		Restricted		Y		Consider alternative else stop	X		
	Support	Exist		Y		Proceed			X
		Not needed		Y		Proceed			X
		Must be developed		Y		Proceed if possible		X	
		Unavailable		Y		Consider alternative else stop	X		
Consolidation	Coordination	Pooled		Y		Proceed			X
		Coordinated		Y		Proceed			X
		Shared		Y		Stop if support is unavailable, else proceed	X (3)	X (3)	
	Quality control	Peer-reviewed		Y		Proceed			X
		Averaging-effect		Y		Proceed			X
		Review		Y		Modify if possible			
		Notes:	1. Decision depends on required rules or knowledge. 2. Both input and output must be well-defined. 3. Depends if support is unavailable or not.						

Figure 3. Template-based tool on theory testing and CS

After executing the tasks, their outputs must be consolidated. Here, we have to consider what type of coordination is required to consolidate the various task outputs. We consider the following values (Crowston 2012; Malone and Crowston 1994): the task outputs may be *pooled* together (e.g. individual ideas can be merged), the outputs from different tasks may be *coordinated* (e.g. the outputs from idea generation may be fed to a selection task), or the outputs from different tasks may be *shared* (e.g. collaborative editing). We consider that crowdsourcing can be easily done when the consolidation is pooled or coordinated. Crowdsourcing will be difficult to accomplish in cases where consolidation requires a shared approach, and appropriate support must be developed. Crowdsourcing cannot be accomplished when a shared approach is needed, and support is not available.

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

User Interface-based Prototype

The user interface-based prototype was developed as an improved and revised version of the template-based prototype. This prototype was implemented using Visual Basic for Applications (VBA) Excel, providing wider access to the knowledge base. VBA is the programming language used on an excel interface. It enables building of user-defined functions (UDFs), automating processes and assessing windows API through dynamic-link libraries (DLLs).

The prototype consists of two decision functions (Figure 4), the left-hand side provides the list of theory testing activities, which the user has an option of selecting the appropriate ones relating to the user's

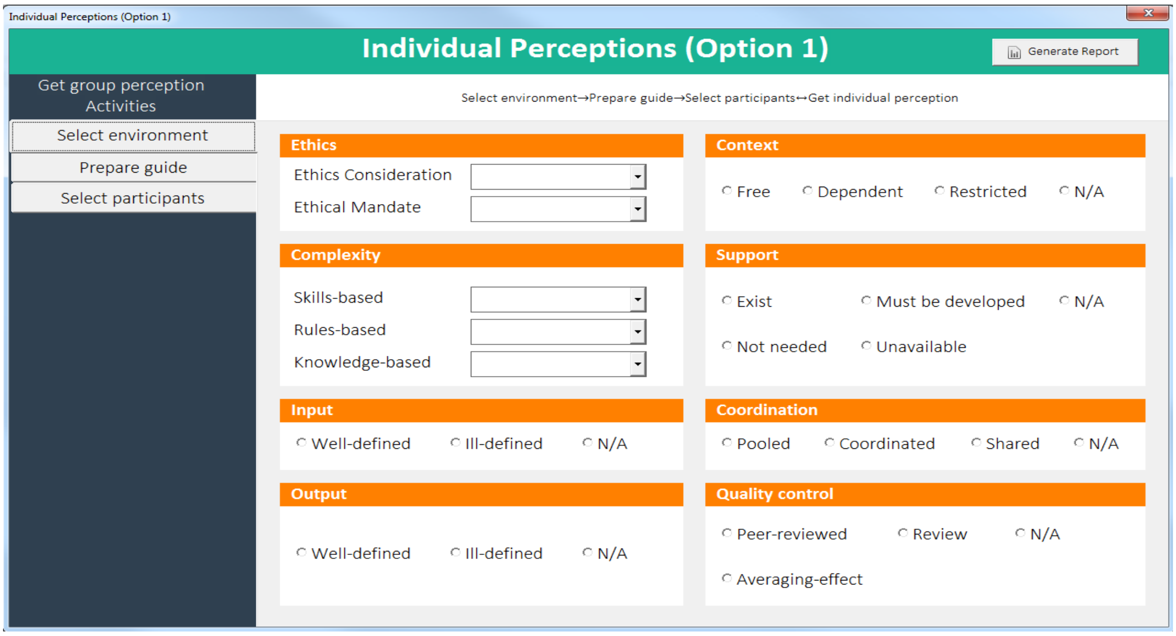

Figure 4. User Interface-Based Prototype

needs, of which when an activity is selected, it creates or generates options on the right-hand side. The right-hand side is option-based generated based on the crowdsourcing criteria.

This gives different options to the user, to assist in decision making as to if the selected activity could be crowdsourced or not, and the level of difficulty if it can be crowdsourced. There is a generate report button, that when clicked generates a visual report to the user, which the user can then decide if the chosen activity is worth crowdsourcing or not. There is also an option that allows the user to receive the report through email as an attachment.

Evaluation

The evaluation strategy adopted for this study includes card sorting, interviews and observation. Card sorting is a qualitative evaluation method that has been widely used in various fields such as psychology, knowledge engineering, and software engineering (Barrett and Edwards 1995), as well as DSR (Prat et al. 2015). Insights on how users view the problem and the solution addressed by an artefact can be gotten through card sorting. This is extremely important in the DSR context, since DSR addresses wicked problems for which there are no single best problem definitions and solutions (Rittel and Webber 1973). We use card sorting to evaluate how our developed artefact relates to what exists in the researchers' minds. Our focus was evaluating the idea, what the tool is meant to achieve, and the process and not on the efficiency or effectiveness of the software itself. We structured the card sorting approach as a combination of card sorting exercises, interviews and observation. They were conceived to acquire expository knowledge regarding the utility, understandability and applicability of the artefact by:

E1: Establishing how useful the participants find the DSS in making a decision based on their research project.
E2: Testing user's understanding of the how the DSS process works and what it is meant to achieve;
E3: Testing the applicability of the DSS to their study, and research in general;

Setting and participants. The card sorting exercises were conducted in a meeting room, which had a large table where to lay out the cards and several computers for the exercise. PhD students conducting research projects in the IS field were the participants used. PhD students were selected because they represent the main target audience for the DSS. 11 were selected by convenience.

Materials. The card sorting materials involved sets of 95 cards with words on them and a unique identifier number for recording purposes. Blank cards were also provided to write on if needed.

Procedure. The exercises were done in one-on-one sessions moderated by the first author (Figure 5). At the beginning of each session, the purpose of the exercises, the artefacts under evaluation, and the card sorting technique was expounded to the participant. The participants were then encouraged to express their thoughts and opinions during the exercises, to provide detailed feedback about what they were thinking and doing. The exercises were then operationalised according to the following steps.

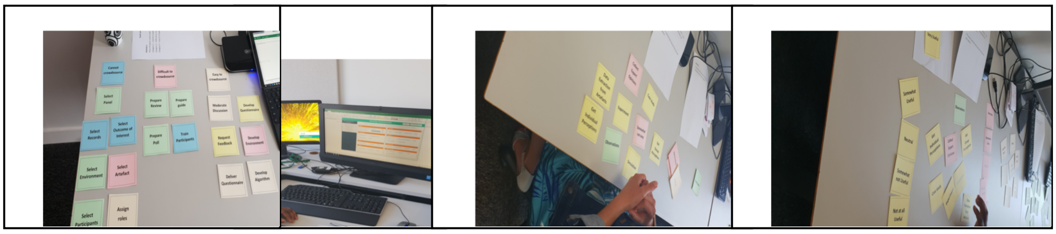

Figure 5. Card sorting exercise

The purpose of each exercise was explained before handing over a deck of cards to the participant. The participant was then given some time to read through the cards for familiarization of contents. The first exercise contained a deck of cards with theory testing activities, which the participant should group in three categories: Cannot CS, difficult to CS and easy to CS. The participant is then given sometime to use the developed tool based on their research project. The second exercise and subsequent exercises, interview and observations were aimed at getting participants reactions, and feedback regarding how beneficial the DSS was in making decisions if a chosen theory testing activity can be crowdsourced.

Along with information about the card sorts, we also gathered the participant's reactions and comments made throughout the exercises, focusing on the positive and negative reactions. The exercises were audio recorded and captured to ease later analysis. The moderator also observed the impact of the tool on the users, focusing on their reactions, and facial expressions, notes were taken. All participants completed the entire procedure and each session took an average of 40 minutes.

Usefulness of the DSS in Decision Making (E1)

This exercise was aimed at determining if the participants found the DSS useful in deciding to CS the selected theory testing activity based on their research project Participants were told to sort the cards into three categories: Cannot CS, difficult to CS and easy to CS. The cards contained different theory testing activities. The cards had unique numbers, making it easy for analysis. The participants were then asked to use the tool to determine if such activities were placed in the right category based on the recommendation of the tool. The result suggests that 20% had little mismatch, because they had prior knowledge of CS, while 80% had a lot of mismatch because they concept of CS was new to them. Using the tool helped in categorizing the mismatch activity (Pries-Heje and Baskerville) in its right category, especially based on their research project. From the second exercise conducted, 40% of the participants said the tool was very useful, while 50% said it was somewhat useful, the last 10% said it was somewhat not useful.

Interviews were done to understand the reason behind their choices, especially for those that considered it somewhat not useful. Reason behind their response was due to the fact of being at the tail-end of their study, and have used other methods to acquire data, so the tool is not beneficial to them now but would be beneficial for other studies. Generally, the participants felt it was useful and advised that the tool be used at the beginning of one's study, to help with decision making.
Observation showed that the participants were very keen to see the result of their selection, and what the tool will recommend at the end. The "wow" expression could be seen when the report page of the tool was generated.

Ease of use

We evaluated the ease of use, focusing more on the process it took for the tool to actual help in making a decision. We added this evaluation because ease of use could affect how users see the usefulness of the tool, it serves as a moderating factor (Wei 2009).

The result suggests that, 30% of the sample said that it was very easy to use the tool and understand the procedures, while 60% said it was somewhat easy, another 10% responded that it was not easy at all to use and understand. We interviewed the participants that said it wasn't easy to use and understand to know the reason behind their response, and most of them said it was easy to use and understand because the moderator was there to clarify somethings they don't understand, but if the moderator wasn't there, that would make it less easy. This feedback was well noted and crucial for further development of the software, since this was just a prototype. Generally, most of the participants found it easy to use and understand.

Understanding of the Intent of the DSS (E2)

This step evaluated the degree to which the artefact can be comprehended, if the participants had a clear understanding of the intent of the tool, that is, what it is meant to do. Participants were given cards with various words that implied understandability and were told to pick out those that they feel expressed their level of understanding. Based on content analysis, we found that all the participants fully understood the intent of the tool, some words like "I get the idea, I get the point, I comprehend etc., were used to express their opinions. From the interviews, participants gave reasons why they gave such responses, most of which were they being able to use the tool, and the tool helping them with making decisions, which they would naturally not have been able to make. Also, some of the CS criteria, which the tool highlighted, they would not have thought about them, but were really important to consider before deciding to CS.

Applicability of the DSS (E3)

The last exercise performed by the participants was aimed at determining how applicable this tool is to research in general, if the participants found the tool applicable to use in their research project, next project or for other new researchers. Using content analysis, we found that all the participants found it applicable, not necessarily to their immediate project, as some of them were rounding up, but to research in general. Words like "relevant, I can relate to it, suitable, etc.", were used to describe their opinion.

Some of the participants were thrilled about the tool, as such wanted to discuss their project more, and how they could use CS in some of their testing activity to reduce the time it would cost them if they were to do that alone. The exercise encouraged the participants, experienced or not, to think carefully about testing activities that they though could not be CS, how that could possibly be done, and to also consider the CS criteria, and how that affects the outcome of their decisions. Finally, the participants expressed their general feelings and opinion of the tool and exercise, which were mainly that, it was simple, great, good, interesting, etc. There was also considerable positive feedback about the card sorting exercises, as it provided an effective opportunity to compare their thoughts with the use of the tool.

Discussion and Conclusion

Gathering and analysing data to test a theory can be a daunting endeavour. 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 empirical data pertaining to complex environments, considering a multitude of factors, contextual elements and stakeholders, which require new, innovative approaches.

We developed the tool consisting of three main components: GUI, information processing component, and knowledge component. This was utilized in two prototype implementations: the template-based prototype and the user-interface based prototype. While the template-based prototype was used for knowledge articulation by representing, extracting, organizing and acting on relevant information, the

DSS was targeted for decision making and recommendation for researchers using the tool. Therefore, the two prototypes make complementary contributions to research and practice.

The developed template and DSS help researchers to systematically check if crowdsourcing can be applied to a variety of theory testing patterns. So, our proposition is not just to check if data collection per se can be crowdsourced or not. Our proposition is to check if the patterned activities related to theory testing can be crowdsourced or not. The developed artefact provides constructs and methods, and design principles and rules that may be adopted by researchers in a variety of situations. It also contributes with operational knowledge highlighting how researchers may benefit from the crowdsourcing strategy.

Implementing these prototypes from a Design Science perspective represents a form of evaluation for the pattern model, which demonstrates its applicability (Peffers et al. 2012). Since the prototype helps with decision making, and can be used as a research tool, it contributes to crowdsourcing research.

Despite 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, researchers may wish to take into consideration other contextual elements influencing the research, such as institutional polices, culture etc. That introduces a contingency factor in the decision to crowdsource that complicates the decision-making process.

Furthermore, our procedure takes into consideration a set of properties that cannot be considered complete. Many other properties could eventually be added, some of them addressing operational issues such as the characteristics of specific platforms used to crowdsource. Other properties could consider characteristics of the crowd and a more comprehensive taxonomy of crowdsourcing tasks. Another issue to consider, which is illustrated in our examples, is that the development of task support may be considered along with the automation of some theory testing activities. However, automation was not considered in our research.

Finally, we should also consider that researchers may also find creative ways to overcome some of the constraints addressed by the crowdsourcing process, e.g. adopting games, role playing, etc. Therefore, we suggest that both the templates and the prototype described in this paper are just initial approaches to the endeavour of bringing design science into theory building and theory testing, and also bringing crowdsourcing into theory testing. These limitations can serve as directions for future research.

References

Antunes, P., Herskovic, V., Ochoa, S., and Pino, J. 2012. "Structuring Dimensions for Collaborative Systems Evaluation," *ACM Computing Surveys* (44:2).

Arnott, D., and Pervan, G. 2005. "A Critical Analysis of Decision Support Systems Research," *Journal of information technology* (20:2), pp. 67-87.

Barrett, A., and Edwards, J. 1995. "Knowledge Elicitation and Knowledge Representation in a Large Domain with Multiple Experts," *Expert Systems with Applications* (8:1), pp. 169-176.

Bates, J., and Lanza, B. 2013. "Conducting Psychology Student Research Via the Mechanical Turk Crowdsourcing Service," *North American Journal of Psychology* (15:2), p. 385.

Behrend, T., Sharek, D., Meade, A., and Wiebe, E. 2011. "The Viability of Crowdsourcing for Survey Research," *Behavior research methods* (43:3), p. 800.

Bhattacherjee, A. 2012. "Social Science Research: Principles, Methods, and Practices,").

Bonney, R., Cooper, C., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K., and Shirk, J. 2009. "Citizen Science: A Developing Tool for Expanding Science Knowledge and Scientific Literacy," *BioScience* (59:11), pp. 977-984.

Brunt, L., and Meidell, E. 2018. "When Are Crowdsourced Data Truthful, Accurate, and Representative," *The Journal of Business Inquiry* (17:1), pp. 55-71.

Cheng, J., Teevan, J., Iqbal, S., and Bernstein, M. 2015. "Break It Down: A Comparison of Macro-and Microtasks," in *Proceedings of the 33rd Annual Acm Conference on Human Factors in Computing Systems*. ACM, pp. 4061-4064.

Creswell, J., and Creswell, J. 2017. *Research Design: Qualitative, Quantitative, and Mixed Methods Approaches*. Sage.

Crowston, K. 2012. "Amazon Mechanical Turk: A Research Tool for Organizations and Information Systems Scholars," in *Shaping the Future of Ict Research. Methods and Approaches*. Springer, pp. 210-221.

Derry, S., Pea, R., Barron, B., Engle, R., Erickson, F., Goldman, R., Hall, R., Koschmann, T., Lemke, J., Sherin, M., and Sherin, B. 2010. "Conducting Video Research in the Learning Sciences: Guidance on Selection, Analysis, Technology, and Ethics," *The Journal of the Learning Sciences* (19:1), pp. 3-53.

Dicks, L. V., Walsh, J. C., and Sutherland, W. J. 2014. "Organising Evidence for Environmental Management Decisions: A '4s'hierarchy," *Trends in ecology & evolution* (29:11), pp. 607-613.

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.

Enwereuzo, I., Antunes, P., & Johnstone, D. (2019). Patterns of Tesing Theory with Human Subjects: A Design Science Perspective. Paper accepted at AMCIS.

Gadiraju, U., Fetahu, B., and Kawase, R. 2015. "Training Workers for Improving Performance in Crowdsourcing Microtasks," in *Design for Teaching and Learning in a Networked World*. Cham: Springer, pp. 100-114.

Ghezzi, A., Gabelloni, D., Martini, A., and Natalicchio, A. 2018. "Crowdsourcing: A Review and Suggestions for Future Research," *International Journal of Management Reviews* (20:2), pp. 343-363.

Gregor, S. 2006. "The Nature of Theory in Information Systems," *MIS quarterly*), pp. 611-642.

Gregor, S., and Hevner, A. R. 2013. "Positioning and Presenting Design Science Research for Maximum Impact," *MIS quarterly* (37:2), pp. 337-355.

Gura, T. 2013. "Citizen Science: Amateur Experts," *Nature* (496:7444), pp. 259-261.

Hevner, A., and Chatterjee, S. 2010a. *Design Research in Information Systems: Theory and Practice*. Springer Science & Business Media.

Hevner, A., and Chatterjee, S. 2010b. "Design Science Research in Information Systems," in *Design Research in Information Systems*. Springer, pp. 9-22.

Hevner, A., March, S. T., Park, J., and Ram, S. J. M. q. 2004. "Design Science in Information Systems Research," (28:1), pp. 75-105.

Holsapple, C. W. 2008. "Dss Architecture and Types," in *Handbook on Decision Support Systems 1*. Springer, pp. 163-189.

Hosack, B., Hall, D., Paradice, D., and 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), p. 315.

Jarmolowicz, D., Bickel, W., Carter, A., Franck, C., and Mueller, E. 2012. "Using Crowdsourcing to Examine Relations between Delay and Probability Discounting," *Behavioural processes* (91:3), pp. 308-312.

Jiang, L., Wagner, C., and Nardi, B. 2015. "Not Just in It for the Money: A Qualitative Investigation of Workers' Perceived Benefits of Micro-Task Crowdsourcing," *48th Hawaii International Conference on System Sciences*: IEEE, pp. 773-782.

Kietzmann, J. 2017. "Crowdsourcing: A Revised Definition and Introduction to New Research," *Business Horizons* (60:2), pp. 151-153.

Kittur, A., Chi, E., and Suh, B. 2008. "Crowdsourcing User Studies with Mechanical Turk," in *Proceedings of the Sigchi Conference on Human Factors in Computing Systems*. ACM, pp. 453-456.

Lee, J., and Bui, T. 2000. "A Template-Based Methodology for Disaster Management Information Systems," *System Sciences, 2000. Proceedings of the 33rd Annual Hawaii International Conference on*: IEEE, p. 7 pp. vol. 2.

Lowry, P. B., D'Arcy, J., Hammer, B., and Moody, G. D. 2016. "“Cargo Cult” Science in Traditional Organization and Information Systems Survey Research: A Case for Using Nontraditional Methods of Data Collection, Including Mechanical Turk and Online Panels," *The Journal of Strategic Information Systems* (25:3), pp. 232-240.

Malone, T., and Crowston, K. 1994. "The Interdisciplinary Study of Coordination," *ACM Computing Surveys* (26:1), pp. 87-119.

March, S. T., and Smith, G. F. 1995. "Design and Natural Science Research on Information Technology," *Decision support systems* (15:4), pp. 251-266.

McInnis, B., Cosley, D., Nam, C., and Leshed, G. 2016. "Taking a Hit: Designing around Rejection, Mistrust, Risk, and Workers' Experiences in Amazon Mechanical Turk," in *Proceedings of the 2016 Chi Conference on Human Factors in Computing Systems*. ACM, pp. 2271-2282.

Nakatsu, R., Grossman, E., and Iacovou, C. 2014. "A Taxonomy of Crowdsourcing Based on Task Complexity," *Journal of Information Science* (40:6), pp. 823-834.

Peer, E., Brandimarte, L., Samat, S., and Acquisti, A. 2017. "Beyond the Turk: Alternative Platforms for Crowdsourcing Behavioral Research," *Journal of Experimental Social Psychology* (70), pp. 153-163.

Peffers, K., Rothenberger, M., Tuunanen, T., and Vaezi, R. 2012. "Design Science Research Evaluation," *International Conference on Design Science Research in Information Systems*: Springer, pp. 398-410.

Peffers, K., Tuunanen, T., Rothenberger, M. A., and Chatterjee, S. 2007. "A Design Science Research Methodology for Information Systems Research," *Journal of management information systems* (24:3), pp. 45-77.

Power, D. J. 2008. "Decision Support Systems: A Historical Overview," *Handbook on Decision Support Systems 1*), pp. 121-140.

Prat, N., Comyn-Wattiau, I., and Akoka, J. 2015. "A Taxonomy of Evaluation Methods for Information Systems Artifacts," *Journal of Management Information Systems* (32:3), pp. 229-267.

Pries-Heje, J., and Baskerville, R. 2008. "The Design Theory Nexus," *MIS quarterly*), pp. 731-755.

Pries-Heje, J., Baskerville, R., and Venable, J. R. 2008. "Strategies for Design Science Research Evaluation," *ECIS*, pp. 255-266.

Reason, J. 2008. *The Human Contribution: Unsafe Acts, Accidents and Heroic Recoveries*. Surrey, England: Ashgate.

Rittel, H., and Webber, M. 1973. "Dilemmas in a General Theory of Planning," *Policy sciences* (4:2), pp. 155-169.

Sein, M. K., Henfridsson, O., Purao, S., Rossi, M., and Lindgren, R. 2011. "Action Design Research," *MIS quarterly*), pp. 37-56.

Shank, D. B. 2016. "Using Crowdsourcing Websites for Sociological Research: The Case of Amazon Mechanical Turk," *The American Sociologist* (47:1), pp. 47-55.

Simon, H. A. 1960. "The New Science of Management Decision,").

Sonnenberg, C., and vom Brocke, J. 2011. "Evaluation Patterns for Design Science Research Artefacts," *European Design Science Symposium*: Springer, pp. 71-83.

Steelman, Z., Hammer, B., and Limayem, M. 2014. "Data Collection in the Digital Age: Innovative Alterantives to Student Samples," *Mis Quarterly* (38:2), pp. 355-378.

Stewart, N., Chandler, J., and Paolacci, G. 2017. "Crowdsourcing Samples in Cognitive Science," *Trends in cognitive sciences* (21:10), pp. 736-748.

Thomas, D. 2006. "A General Inductive Approach for Analyzing Qualitative Evaluation Data," *American journal of evaluation* (27:2), pp. 237-246.

Vondrick, C., Patterson, D., and Ramanan, D. 2013. "Efficiently Scaling up Crowdsourced Video Annotation," *International Journal of Computer Vision* (101:1), pp. 184-204.

Wei, W.-C. 2009. "A Technology Acceptance Model: Mediate and Moderate Effect," *Asia Pacific Management Review* (14:4), pp. 461-476.

Witschey, J., Murphy-Hill, E., and Xiao, S. 2013. "Conducting Interview Studies: Challenges, Lessons Learned, and Open Questions," in *1st International Workshop on Conducting Empirical Studies in Industry*. IEEE, pp. 51-54.