

# ON THE ADOPTION OF CROWDSOURCING FOR THEORY TESTING

*Research paper*

Enwereuzo Ijeoma, Victoria University of Wellington, New Zealand,  
[ijeoma.enwereuzo@vuw.ac.nz](mailto:ijeoma.enwereuzo@vuw.ac.nz)

Pedro Antunes, Victoria University of Wellington, New Zealand, [pedro.antunes@vuw.ac.nz](mailto:pedro.antunes@vuw.ac.nz)

David Johnstone, Victoria University of Wellington, New Zealand, [david.johnstone@vuw.ac.nz](mailto:david.johnstone@vuw.ac.nz)

## Abstract

*This paper examines the possibilities of using the crowdsourcing strategy for theory testing. We first analyse the relationships between theory building and theory testing activities. Then, based on a systematic review of 248 papers published in MISQ, we characterise the intents and pattern systems of activities that have been used for theory testing. Finally, we ascertain which activities can be crowdsourced or not and pinpoint a set of pathways supporting partial and total crowdsourcing. The obtained results show that a large number of activities related to data gathering can be crowdsourced, and that a number of intents have viable pathways supporting partial crowdsourcing.*

*Keywords: Theory, Theory testing, Crowdsourcing.*

## 1 Introduction

Theory concerns “abstract entities that aim to describe, explain, and enhance understanding of the world and, in some cases, to provide predictions of what will happen in the future and to give a basis for intervention and action” (Gregor, 2006). Theory also provides a linguistic tool for organizing scientific knowledge and scientific research (Colquitt and Zapata-Phelan, 2007). Finally, theory can be used as an educational device to raise consciousness about a set of concepts. Though only validated theory can be considered useful and helpful.

There are two general forms of validating theory: formative and summative (Lee and Hubona, 2009). The former is part of the theory building process while the latter seeks closure with a statement whether there is support or not to a theory. Theory building is an on-going process of observing, confirming, applying, and adapting theory (Lynham, 2002). It also involves the development of new ideas, concepts, conceptual frameworks, methods, and models (Nunamaker Jr et al., 1990). Theory testing is a natural complement to theory building; just as Bacharach (1989) said, “if it is not testable, no matter how profound or aesthetically pleasing it may be, it is not theory”.

Theory testing always involves collecting data either confirming or disconfirming the set of propositions articulated by a theory (Lokke and Sorensen, 2014). We are particularly interested in theory testing that involves human participants producing a wide range of data such as opinions, thoughts, expertise, attitudes, communications, observations, actions, and task logs. Theory testing involving human participation is ethically, conceptually and practically challenging. Especially in social sciences, one has to respect best practices regarding the participants' involvement, the nature of data being collected,

diversity of variables influencing human behaviour, and adequacy of data collection processes, methods and tools (Bhattacharjee, 2012). A more practical but nevertheless important challenge is recruiting and engaging study participants (Witschey et al., 2013).

Although many robust methods have already been developed to test theories with human participation, which may include interviews, surveys, and many others, we are still curious to know to what extent the crowdsourcing (CS) strategy, can be used to support the theory testing process, especially in the Information Systems (IS) field, and to understand the advantages and constraints this method brings to theory testing. CS is a strategy in which a single task maybe fragmented into multiple tasks delivered to “a large group of people in the form of an open call” (Howe, 2006). The relationship between the requestor and performers is usually loose, with little or no restrictions on who may perform the task (Yuen et al., 2011, Whitla, 2009, Schenk and Guittard, 2011, Zhao and Zhu, 2012). Besides, the whole process is supported by information technology, which helps reaching the crowd and managing the task execution. CS has been recognised to take advantage of parallelisation and to harness the collective intelligence of the crowd (Parvanta et al., 2013).

From the outset, the potential advantages of adopting CS for theory testing could be significant. CS may help controlling the timing and the scale of data collection in an elastic way, from very short periods but huge scale to long periods and smaller scale. This elasticity also extends to spatial coverage, regarding countries, organisations, contexts, etc. (Smith et al., 2013, Ford et al., 2015, Brabham, 2008a). Further, CS may also reduce the overhead associated to the testing process. This includes removing administrative overhead related to procuring participants, supervision, scheduling, training, managing experimental treatments, selecting samples, distributing research instruments, and collecting results.

Even though the adoption of CS to theory testing seems a compelling endeavour, we must address an important challenge: **how to align theory testing with crowdsourcing**. Research communities have developed theory-testing patterns, which may have to be adjusted to the CS strategy and re-checked again for robustness. And in some cases we may find that CS may not be feasible.

In this paper, we investigate this alignment problem in the IS field. We analyse various ways in which theories have been tested. Then, we seek to identify **patterns** that could help us discover how theory testing can be aligned with the CS method. To achieve this goal, we undertake a descriptive literature review in the IS field. Since CS may have a particular impact on how theory testing is conducted, we focus our review on the operational dimension of theory testing, rather than philosophical, epistemological or methodological dimensions.

The remainder of this paper is organized as follows. Section two discusses related work on theory types, theory building, theory testing, and how they may be related to CS. Section three describes the method and procedure used for the literature review, as well as the obtained results. Section four contains the discussion of the results and Section five contains some concluding remarks and future work.

## 2 Related Work

An academic discipline will not have much leverage if it does not have useful theory to contribute (Niederman and March, 2015). Theory helps scientists to logically describe and explain a phenomenon, process or sequence of events. Gregor (2006) suggests a classification into five theory types. This study will focus mainly on theories with testable proposition, therefore, we will consider type III and IV theory.

Type III theory aims to predict “what will be”. These theories are able to predict the outcomes, results or consequences using some explanatory factors but without necessarily explaining the connections among factors, or explaining the independent and dependent variables. Some research approaches falling into this category use statistical techniques such as correlation or regression analysis (Gregor, 2006) to predict but not to explain phenomena.

Type IV theory is concerned with “what is”, “how”, “why” and “what will be”. A theory in this category is capable to provide both the prediction and the scaffolding of the underlying causes of a phenomenon using theoretical constructs. To many, this is the ideal theory as it seeks to define the concepts and propositions underlying a phenomenon, while specifying a set of fundamental relations among different variables that help predicting the phenomenon (Gregor, 2002).

## 2.1 Focus on theory with testable propositions

Type III and type IV theories will be our primary targets for identifying theory-testing patterns, since researchers develop testable propositions with the explicit purpose to test them. The other types of theory do not have to pass the testing challenge, since propositions only need to be either identified (types I and II) or applied (type V).

Goode and Hatt (1952) define testable proposition as an imaginative idea, a guess, a statement of a solvable problem or any thinking that can be put to test to determine its validity. A testable proposition can also be an hypothesis: a tentative generalization which has to be tested to determine its validity (Khan, 2011). A testable proposition should have the following characteristics (Khan, 2011): 1) refer to observable and measurable events, which are pre-conditions for the formulation of testable propositions; 2) be conceptually comprehensible and provide a solution to a defined problem; 3) be formulated in a way that can be tested and verified directly; 4) be related to the existing body of knowledge; and 5) have logical unity and comprehensiveness (Khan, 2011).

The test of a theory is concerned with establishing or refuting the validity of the theory’s core propositions. A theory is tested by determining to what degree it provides or not a close fit to empirical data (Colquitt and Zapata-Phelan, 2007). These binary outcomes are the basis for codifying knowledge and creating the opportunity for further research. Testing and re-testing theory are considered key foundations to progress scientific knowledge (Niederman and March, 2015).

Researchers spend time finding and then explaining fundamental relationships in empirical data (Viswesvaran and Ones, 1995). Researchers often also try to explore the mediators and moderators that define the boundary conditions of a theory. De Vaus (2013) and Bitektine (2008) suggested a stage-gate framework describing both theory building and testing. We adapted that framework to focus primarily on theory testing, noting however a set of pre and post conditions necessary to realise theory testing (Figure 1):

### Pre-conditions:

1. Have a statement with a brief explanation of the theory and what it intends to achieve. Have a set of conceptual propositions. A conceptual proposition is a statement that shows the relationship between two factors. The more conceptual propositions to be tested, the stronger the test of a theory (Stinchcombe, 1987).

### Theory testing steps:

1. If necessary, restate the conceptual propositions as testable propositions. This involves translating abstract concepts into concrete, observable and measurable variables. This step is often called operationalization (De Vaus, 2013).
2. Decide what data are relevant or appropriate to test the testable propositions.
3. Collect relevant data.
4. Analyse the data. Data are analysed to see: a) how much support there is for the testable propositions; b) how much support there is for the conceptual propositions; and c) how much support there is for the whole theory.

### Post-conditions:

1. Have a statement assessing the whole theory. The theory may be or may not be supported completely.

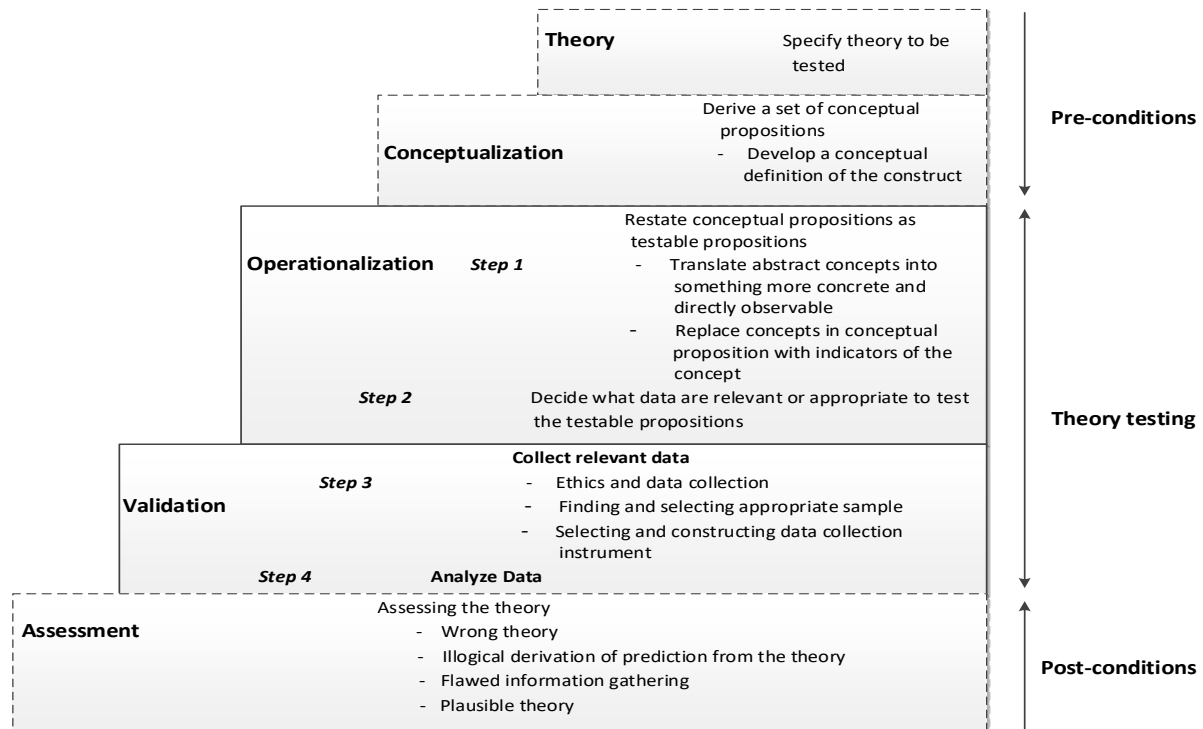


Figure 1. Steps to theory testing, adapted from (Bitektine, 2008, De Vaus, 2013, Mackenzie et al., 2011, Niederman and March, 2015)

Most theories in social sciences are tested quantitatively (Bitektine, 2008, Hyde, 2000, Piper, 2006), while others are qualitatively tested, especially using the case study method. Bitektine (2008) identified some important steps to consider for theory testing when adopting the case study method that significantly align to the one mentioned above, with the following amendments:

**Pre-conditions:**

3. Identify the case(s), unit of analysis and subjects.

**Theory testing steps:**

4. Analyse the case(s) and identify patterns using the foundations provided by the theory.
5. Analyse internal and/or external validity. Realise that single-case studies tend to generate type-1 errors (accept false hypotheses) while cross-case studies tend to generate type-2 errors (reject true hypotheses) (Gerring, 2006).

## 2.2 Considering crowdsourcing in theory testing

To be able to relate CS to theory testing, it is important to first look at the main concepts pertaining to CS, which we summarise in Table 1. The task which includes data collection, could assume multiple forms: questionnaires, social marketing, fact checking, tagging, activity logging, voting, playing, simulation, ideation, commentary, expert opinions, quantified self, etc. This concept of relying on the crowd for data collection is not completely new, since it is also found in citizen's science (Bonney et al., 2009) and living labs (Følstad, 2008). Furthermore, tasks may also involve related activities such as data analysis and quality checking, which in essence collect meta-data (data about data).

Vukovic et al. (2010a) discussed the kind of contract the crowd is involved in when doing the task, which could be internal, external or mixed. In this study, we assume most theory testing utilizes an external contract.

Concept	Description
Crowdsourcer	The entity that seeks to carry out a specific task by harnessing the crowd. In our context, the crowdsourcer is the researcher (Parvanta et al., 2013, De Vaus, 2013).
Task, Crowd task	The task is what fulfils the crowdsourcer's goals, while the crowd task is what is actually assigned to the crowd. Usually, the crowd task is a decomposition of the task.
Crowd	The people assigned to the task. They tend to be unknown to each other and often unknown to the crowdsourcer.
Incentive	The remuneration, motivation or compensation the crowd receives as a result of fulfilling the task.
Open call	The general invitation to become a member of the crowd, which will fulfil the designated task.
Platform	The technological platform that mediates the crowdsourcer and the crowd.

*Table 1. Crowdsourcing concepts*

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 in 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 has not yet been systematically explored and established.

By adopting CS, we see that the researcher has to fulfil certain expectations and obligations, which can be expressed as a set of requirements. To align theory testing and CS, it is therefore important to be explicit about such requirements, which serve as a checklist for judging the viability of using CS. Furthermore, the list of requirements also ensures uniformity when judging viability across many different studies. In Table 2 we elaborate a list of requirements we consider relevant for theory testing.

We have mandatory requirements – these are requirements that if one of the requirements is not met, then crowdsourcing will not be a viable option, they reflect conceptual and practical limitations that cannot be circumvented without breaking the whole notion of crowdsourcing (e.g., crowdsourcing without the Internet is not crowdsourcing). While the desirable requirements are requirement that even if one is not met, CS is possible, but to get a good result, it is better to meet such requirements. The mandatory requirements are mandatory because some of those requirements are what CS is made up of, and also platform limitations and conditions while the desirable requirements are also platform conditions, but are flexible depending on what the researcher wants.

Our selection of mandatory requirements reflects a relatively consensual view in the related literature, but the list of desirable requirements can be seen as more controversial. For instance, many crowdsourcing tasks have been implemented using project websites and social media. However, we suggest that the costs of not using a crowdsourcing platform are just too high for theory testing purposes. Even though many crowdsourcing tasks have been implemented without any particular skills requirements (e.g. photo classification), in our particular context we see matching skills as very important to avoid uncontrolled setting and biased results, which could reduce data quality. In our particular context, training can be a problem. In many research setting, especially in experimental research, some training is necessary, e.g. training participants to gather environmental data using specific instruments. However, it may be difficult to assess if training was or not successful, especially because tasks are one-off (so, no repeated training sessions). For that reason, we suggest that simple or no training should be the target.

The crowd can also be classified according to the required domain knowledge. By domain knowledge we mean specialization in an area of knowledge necessary to complete a task (De Boer et al., 2012, Ross et al., 2009). The selection of the crowd members may depend on the task requiring domain knowledge or not, and how CS platforms support this feature. In this study, we do not consider the specific characteristics of CS platforms.

This study contributes to previous research on theory generation and testing by investigating how crowdsourcing can alleviate the challenges of data collection and analysis in theory testing. Using crowdsourcing as a research instrument has been around for some time. In particular, crowdsourcing is becoming a common way to deliver questionnaires in behavioural research (Bates & Lanza, 2013; Behrend et al., 2011; Jarmolowicz et al., 2012). In user studies, it

	Requirements	Description	Justification	Examples
Mandatory	Horizontal decomposability	The task must be decomposable into a number of simple, independent tasks that can be executed in parallel by the crowd	Task decomposition leads to faster turn-around time, which is one of the main advantages of CS (Vukovic et al., 2010b, Heer and Bostock, 2010)	(Turner et al., 2012), (Filatova, 2012)
	One-off	The crowd task must be executed only once for each element of the crowd.	Repeated tasks are usually not allowed (platform limitation) (Yang et al., 2008, Mason and Suri, 2012)	(Filatova, 2012), (Jarmolowicz et al., 2012)
	Limited interaction	The crowd task must be executed by the crowd using a minimal communication channel with the crowdsourcer	Communication loops are usually not possible (platform limitation) (Wexler, 2011, Wu et al., 2013)	(Turner et al., 2012)
	Computer mediation	The crowd task must be remotely executed, using the Internet for communication and coordination	Most CS platforms are web-based (Kleemann et al., 2008, Andriole, 2010)	(Turner et al., 2012), (Jarmolowicz et al., 2012)
	Simplicity	The crowd task must be 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)	(Filatova, 2012), (Biemann, 2013)
Desirable	Time-boundness	It should be possible to complete the crowd task in a bounded period of time	Timing is crucial to any CS task (Kittur et al., 2008, Bernstein et al., 2011)	(Turner et al., 2012), (Jarmolowicz et al., 2012)
	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 (Schenk and Guittard, 2011, Geiger et al., 2011)	(Turner et al., 2012), (Lin et al., 2012)
	Training	Training, if required, should not be complex or prolonged	In current CS platforms, training is still considered challenging (Park et al., 2014)	(Filatova, 2012), (Biemann, 2013)

Table 2. Crowdsourcing requirements

### 3. Literature Review

A descriptive literature review is adopted as the research method for analysing different ways in which theory has been tested in the IS field. This method of literature review is used because it helps to determine the extent to which a body of empirical studies supports or reveals any interpretable trends or patterns with respect to pre-existing findings (Paré et al., 2015). This is done by collecting, codifying and analysing data that reflects the frequency of the topics found in the literature (Paré et al., 2015). The use of literature review is important because it grounds the research in the existing knowledge base. A similar approach was used by Thuan (2016) to identify factors influencing the decision to CS. A systematic approach is used instead of the narrative because it increases rigor and transparency and it is a well-structured and defined process (Kitchenham et al., 2009, Okoli and Schabram, 2010).

Our method of systematic literature review is based on (Okoli and Schabram, 2010, Kitchenham et al., 2009), who recommend searching the literature using the following stages: 1) select articles; 2) filter articles; 3) data extraction; and 4) data synthesis. These stages are described below.

### *Select articles*

We are interested in articles that test theory to show the plausibility of such theory. This led to the selection of MIS Quarterly as the focus of the review. We selected this journal because it is known to have a strong theoretical basis and has a strong reputation. Furthermore, this journal follows a positivist tradition of research emphasising theory testing (Mingers, 2003). We selected a 10-year period for the review because this frame seems sufficient to develop a comprehensive view over the topic. The total number of articles published in the period was 307.

Articles were searched using a set of keywords, a procedure that other published reviews have adopted (Estellés-Arolas and González-Ladrón-De-Guevara, 2012). We searched for papers using a combination of keywords such as ‘theory’, ‘test’, ‘theoretical model’, ‘theoretical framework’, ‘conceptual model’, and ‘conceptual framework’. The search was done on the full body. Articles with any of these keywords were selected. The search identified 298 papers, which underlines the strong theoretical contents of MISQ and supports our selection.

### *Filter articles*

This stage eliminated articles unrelated to this study. As mentioned earlier, we were interested in articles whose theory was tested one way or the other, and which theories belonged to type III and IV categories. Two intermediate steps were followed in this stage. Firstly, after keyword search, an article was taken from the pool and a check was done to determine if it contained a theory with testable propositions (Khan, 2011, Muntermann et al., 2015). We read through the methodology section to determine this. Secondly, we checked if the article was tested using human participants, since our research is restricted to theory testing using human participants. After this filtering stage, 248 articles were selected for coding. In Table 3 we summarise the results from the filtering process.

Stages	Excluded	Selected
Articles published in MISQ between 2007 and 2017	-	307
Applying list of keywords	9	298
Checking for testable propositions and human participants	50	248
Selected articles	-	248

Table 3. Summary of literature review

### *Data extraction*

This step involved a detailed reading of specific sections of each article and coding of relevant data for posterior analysis. According to Paré et al. (2016), there are three main aspects to consider during coding. The first one is deciding what parts of an article provide most relevant data. We focused primarily on the methodology, discussion

and concluding sections. Coding was centred on the activities done by the researchers and related to the testing process.

We were not so much interested in the conceptual parts of theory testing, but more on the activities and patterns involved in the theory testing process. We were concerned with questions like what types of participants were used, what type of resources or instruments were used, what types of research methods and techniques were used, data collection methods, procedures followed during testing, and what were the outcomes of theory testing.

The second aspect considered by Paré et al. (2016) involved operationalizing the coding procedure. We adopted a combination of open coding with predefined data extraction forms. Open coding was used to ensure that no relevant information would be left out, and to ensure the search was not confined to a small set of questions and assumptions. Data extraction forms were also used to ensure that

pre-defined relevant elements could be extracted efficiently. Figure 2 shows a fragment of the data extraction form.

Detail of Coding on Theory testing using Crowdsourcing																	
General information				Research Methodology				Theory Type				Theory testing context					
No	Reference	Year	Date	Main findings (according to the data in the abstract)  Hypothesis was tested using humans, using the Microsoft usability guideline model, to test how participants rate 5 specific categories: content, ease of use, promotion, made-for-medium and emotion for	Method of Data collection	Method of Data Analysis	Others	Type III and IV	Theory Tested	Others	Application			Task			Note
											Inputs	Agents	Function	Outputs	Participants	Complexity	
1	Web and wires	2006	16-Jun		Online and Physical	Statistical analysis	Yes	TAM/MUG (Microsoft Usability guideline)	Instrument students	Guideline: Data gathering	Internet and cell	Simple	yes				

Figure 2. Sample of Coding Form

The third aspect considered by Paré et al. (2016) concerned organizing the extracted data in preparation for analysis. NVivo was adopted to manage the extracted data. Coding was done by the first author. To increase the reliability of the analysis (Krefting, 1991), an independent researcher coded a sample selection of articles in parallel and the codes were compared and adjusted for clearer understanding.

### Data synthesis

Data synthesis focussed on identifying patterned activities in theory testing. A pattern is a regular and intelligible form or sequence in which something is done. It can also be seen as a generalized solution to recurrent problems (Penker, 2000).

Some vital elements associated with patterns are *intent*, *structure* and *implementation* (Kafura, 1998, Penker, 2000). As implementation is usually very specific to a problem context, in this study we only consider the first two elements. Intent summarizes the general purpose of a study taking a theory testing perspective, which does not concern the specific problem under investigation (Penker, 2000). In our synthesis, we identify five different intents (Figure 3).

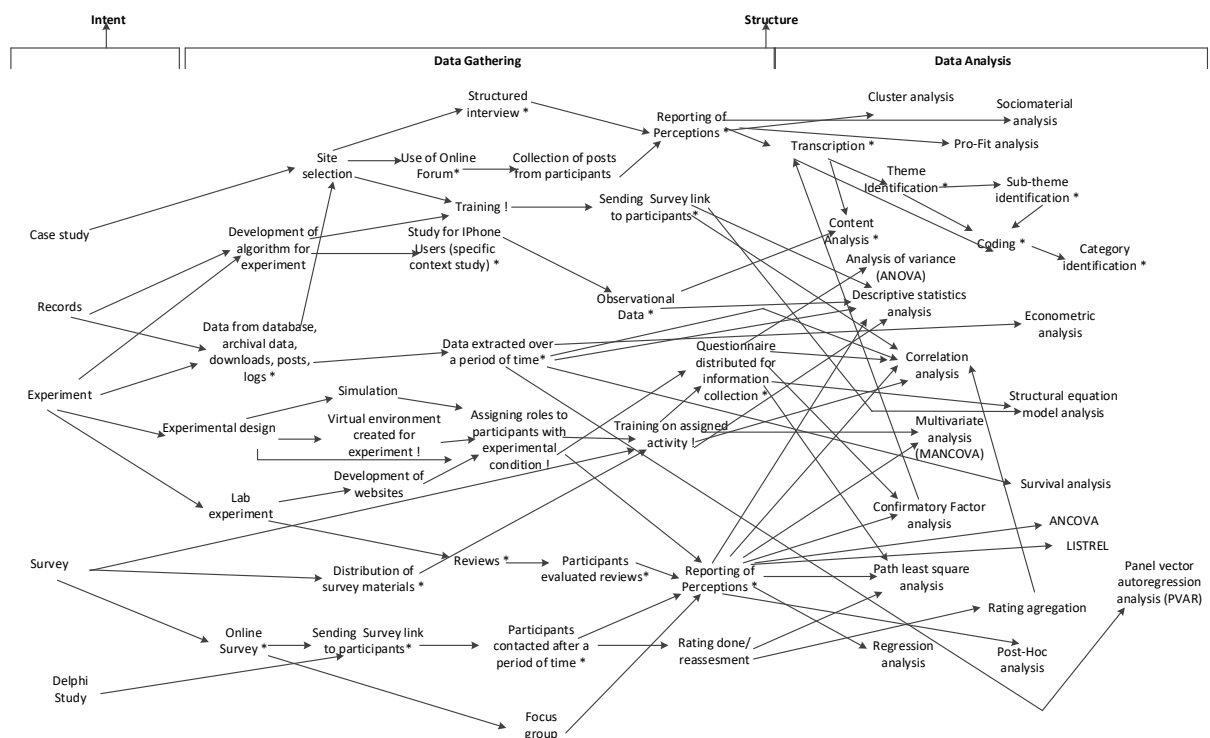


Figure 3. Activities and pattern systems for theory testing (\* indicates the activity can be crowdsourced)



Structure regards the organisation of activities, which builds up pattern systems with linked activities. In our synthesis, we identified a large number of pattern systems involving 52 different activities (Figure 3). To designate the activities, we used generic names commonly recognized by scholars, e.g. interviews, site selection, cluster analysis, etc. The activities were related in patterned systems using directed arrows.

To emphasise the structural features of the collected data, we organised the pattern systems in two categories: data gathering and data analysis. These categories reflect the steps in theory testing discussed in Section 2.1.

### 3.1 Assessment of crowdsourcing requirements

We used the requirements shown in Table 2 to make a systematic decision if an activity could be CS or not. An activity is selected then a requirement check is made, if any of the mandatory requirement is not met and there is no way to modify or get an alternative, then that activity cannot be CS, but if all requirement is met, then the desirable requirements are considered. Such activity can be CS if some desirable requirement is not met, but it is advisable to consider those requirements for good output, this can be seen in figure 4.

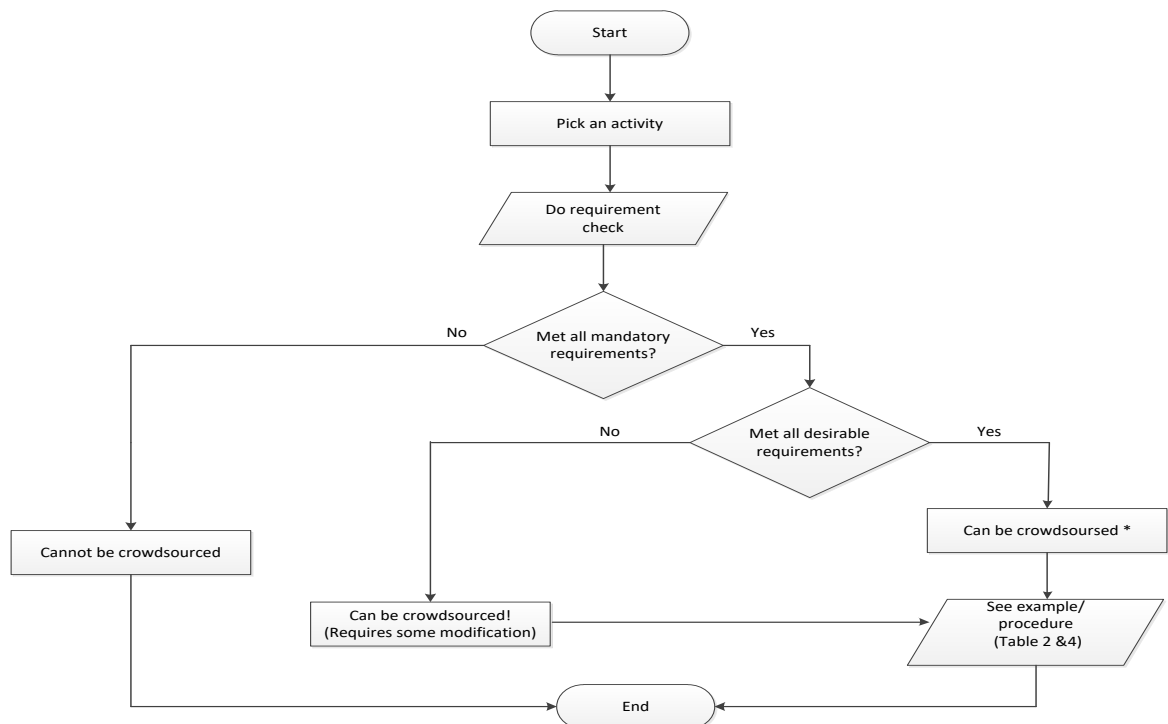


Figure 4. Flowchart for Requirement Check

Some examples have been cited in table 2 and table 4 as a guideline, some examples could not be sourced for some activities, because CS has not been considered in such aspects, which is one of the main contribution of this study, to explore such areas as long as the CS requirements can be met. With the help of the flowchart and the requirement check, decisions were made for each activity to determine if it can be CS or not.

Take for instance the “case study” intent. We have to make a decision on each activity linked to that intent. Consider then “site selection”. We determine if this activity can be CS or not by going through the list of requirements. In this case, the activity is not decomposable and cannot be remotely executed using the Internet. Therefore we can conclude this activity cannot be CS. We then pick another

Activity	Description	Example/Procedure
Structured Interview	Presentation of interview with exactly the same questions in the same order	(Brabham, 2010, De Vaus, 2013)
Perceptions recorded	Recording of perceptions from participants, e.g. collection of user's feedback, which can serve as a valuable input to further improvement or as the needed data	(Snoek et al., 2010)
Online forum	Use of any form of online forum for data collection. Involves online discussion site where people can hold conversations in the form of posted messages, e.g. Ubuntu forums, Nairaland, etc.	(Lampe et al., 2014, Brabham, 2009)
Transcription	Process of producing a written copy of something	(Holley, 2010, Eveleigh, 2014)
Data from posts, database, downloads etc	Collection of data from posts, databases, downloads, logs etc.	(Lampe et al., 2014, Eveleigh, 2014, Chernova et al., 2011)
Virtual environment created for experiment	Use of virtual environment for experiment, e.g. human-robot interaction demonstrated using games	(Chernova et al., 2011)
Assigning roles to participants	Roles are assigned to participants based on experimental condition	
reviews	Participants conducting reviews assigned to them	(Su et al., 2013)
Participants evaluates reviews		
Online survey	Use of online survey for data collection	(Brabham, 2008b, Zheng et al., 2011)
Survey link sent to participants	Sending of survey link to participants, through which data is collected	
Participants contacted after a period of time		
Training	Teaching, demonstrating or undertaking a course of exercise for skill needed for a particular task.	(Filatova, 2012), (Biemann, 2013)

*Table 4. Theory testing activities that could be crowdsourced and its examples*

activity, which is “interview”. This activity can be decomposed in independent tasks, has clear inputs and outputs, can be remotely executed and communicated via the Internet, can be done within a defined time frame, and the domain knowledge of the crowd may be assessed in the interview. We can therefore conclude that this activity can be CS, which is denoted with an asterisk in Figure 3.

The results from this assessment are summarised in Figure 3. The symbol “\*” close to an activity indicates that we found the activity as being able to be CS, as every requirement in Table 2 can be fulfilled. The symbol “!” close to an activity indicates that the activity can be CS if one or more desirable requirements are violated but the mandatory requirements can still be fulfilled. Although the desirable requirements may be violated and CS still possible, it is advisable to consider these requirements before deciding to CS as some of them are also critical if the best possible results are to be expected.

Throughout this process, we found some cases that could not meet all requirements but with some additional constraints which, if applied by the researcher, would allow for CS to be applied. Consider for instance the training requirement. The requirement states that training should not be complex. If a study requires complex training, or training that cannot be recorded and uploaded but necessitates the researcher to do some sort of demonstration for the participants physically, then the activity cannot be

CS. Though such a decision depends more on the researcher than on specific the characteristics of CS and theory testing. We denote these activities with (!) in the model shown in Figure 3.

## 4 Further Analysis and Discussion

Several researchers in IS and other disciplines have been adopting CS for delivering surveys (Ghose et al., 2010, Kim, 2012, Steelman et al., 2012, Ward and Broniarczyk, 2011, Goodman and Malkoc, 2012, Crowston, 2012). Our analysis suggests that CS may be used beyond that activity. As summarised in Table 5, we identified 5 intents involving 52 different activities. Table 5 shows in more detail how the activities are split between data gathering and data analysis, and within each category between CS and non-CS types. We observe that only 10 activities related to data gathering were classified as not able to be CS, and only 7 activities related to data analysis were classified as being able to be CS.

In Table 5 we provide a more detailed analysis of the theory testing systems of patterns using the notion of pathway. A pathway is a set of linked activities that instantiate an intent from beginning to end. Once again, we can split the exercise between data gathering and data analysis. Furthermore, we can divide pathways between fully CS, partially CS, and non-CS. For instance, a fully CS data gathering

Intents	# Path	# DG	# DA
Case study	17=0(F) + 17(P)	0(F) + 3(P)	5(F) + 6(N)
Records	10=0(F) + 10(P)	0(F) + 3(P)	1(F) + 8(N)
Experiment	98=0(F) + 98(P)	1(F) + 19(P)	6(F) + 47(N)
Survey	36=0(F) + 36(P)	3(F) + 2(P)	0(F) + 45(N)
Delphi	11=0(F) + 11(P)	1(F) + 1(P)	0(F) + 11(N)

pathway allows researchers to CS every data gathering activity, while a partial CS pathway only allows them to CS some of the required data gathering activities.

$$Path(F) = DG(F) \cup DA(F) \quad Path(P) = DG(P) \cup DA(P)$$

Where Path = pathway, F = Fully Crowdsourced, P = partially crowdsourced, N = Non-crowdsourcable, DA = Data Analysis, DG = Data Gathering.

Table 5 Crowdsourcing Pathways

gathering and data analysis). However, ‘experiments’, ‘surveys’ and ‘Delphi’ can fully CS the data gathering stage, while ‘case studies’, ‘records’ and ‘experiments’ can fully CS the data analysis stage. The detailed breakdown for the number of pathways that can be fully, partially or non-CS can be seen in Table 6.

Intents	Data Gathering		Data Analysis	
	# CS Activities	# Non-CS activities	# CS Activities	# Non-CS activities
Case study	4	2	6	6
Records	5	1	1	8
Experiment	8	4	-	14
Survey	8	2	-	13
Delphi	3	1	-	11

Table 6. Crowdsourcing activities

We therefore suggest that the developed model, with its various pattern systems, gives researchers different pathways for testing their theories, which allow for circumventing activities that cannot be CS. Take for instance a ‘case study’. This particular pattern system considers all activities starting with site

Through this analytic process, we show that none of the 5 intents can be fully CS (both data

selection and then having training, online forum or structured interviews. We observe that parts of this pattern system may not be viable for CS, for example where training might be too demanding. However, we can still CS by circumventing ‘training’, which can be done by either adopting an online forum activity or sending survey links to participants.

Another interesting finding is the fact that some aspects of data analysis can actually be CS, even though most of the activities relate to qualitative methods, since quantitative methods for analysis do not seem amenable for horizontal decomposition. This finding can be helpful for those engaging in data analysis using any of these activities. Using CS for these activities may actually increase the reliability of the research process, and could also help conducting validity checks by getting different people to code the same data in order to reduce potential researcher bias and increase inter-coder reliability.

This study contributes to previous research by identifying conceptual elements of theory testing. Often theory testing is embedded in theory building, which makes it difficult to focus on it as an autonomous process. In order to avoid this problem we clearly identify the core elements of theory testing, independently from theory building. We also explore how theory testing has been conducted by researchers. An understanding of the types of theory testing tasks lays a foundation for a discussion on how to crowdsource them. We also conceptually characterised crowdsourcing in order to evaluate theory testing tasks for their crowdsourcability. We identified the fundamental requirements for a theory testing activity to be a candidate for crowdsourcing support. We then used these requirements to systematically assess what types of theory testing tasks have already been crowdsourced or can be crowdsourced. Finally, we provide recommendations for researchers who may wish to use crowdsourcing to enhance the effectiveness of their research method. We particularly focus on researchers developing theory who are considering different strategies on how to test their theory. For this purpose, we discussed various theory testing patterns and identified what patterns can be fully crowdsourced, partially crowdsourced, or not crowdsourced at all.

## 5 Concluding Remarks

Theory testing, which involves the validation or falsification of a theory, can be difficult to accomplish, because a good theory needs good data for validation and good data is usually hard to obtain. CS can be used for theory testing in cases where humans are involved in the testing process. However, one has to consider how to align the theory testing goals with the constraints imposed by CS. This alignment problem has been the main focus of our study.

To better understand the problem, we analysed how theories have been tested in the past. MISQ was selected as a target for a systematic literature review, because of its high reputation and editorial focus on theory. 248 articles published in the last ten years in MISQ were analysed. In our analysis, we focussed on the operational dimension of theory testing and not on the particulars of each theory. We then identified the intents and pattern systems of theory-testing activities adopted by each study. The activities were then organised in a model that underlines three groupings: intent, data gathering and data analysis. We also analysed how the activities were actioned in their respective studies, we identified a set of links and pathways for theory testing. The combination of activities and links identifies a set of pattern systems. We then analysed which activities and pattern systems could be CS or not.

This study provides two important insights. One is highlighting how CS can be used across a multitude of activities related to theory testing, using data from prior research. The other one is suggesting that most theory testing intents can be fully CS. Even though some intents may involve pathways that cannot be CS (fully or partially), our analysis shows that in most cases there are alternative pathways that can be CS. Researchers can therefore make the decision to choose pathways that are suitable for their study. This study does not cover the aspect of decision-making, i.e. motives for crowdsourcing and when it’s best to crowdsource. For future study, developing an algorithm and decision tree for pathway selection is an aspect to look into.

## References

- Andriole, S. J. 2010. Business impact of Web 2.0 technologies. *Communications of the ACM*, 53, 67-79.
- Bacharach, S. B. 1989. Organizational theories: Some criteria for evaluation. *Academy of management review*, 14, 496-515.
- Bates, J., & Lanza, B. (2013). Conducting psychology student research via the Mechanical Turk crowdsourcing service. *North American Journal of Psychology*, 15(2), 385.
- Behrend, T. S., Sharek, D. J., Meade, A. W., & Wiebe, E. N. (2011). The viability of crowdsourcing for survey research. *Behavior research methods*, 43(3), 800.
- Bernstein, M. S., Brandt, J., Miller, R. C. & Karger, D. R. Crowds in two seconds: Enabling realtime crowd-powered interfaces. Proceedings of the 24th annual ACM symposium on User interface software and technology, 2011. ACM, 33-42.
- Bhattacharjee, A. 2012. Social science research: principles, methods, and practices. *Textbooks Collection*, Book 3.
- Biemann, C. 2013. Creating a system for lexical substitutions from scratch using crowdsourcing. *Language Resources and Evaluation*, 47, 97-122.
- Bitektine, A. 2008. Prospective case study design: qualitative method for deductive theory testing. *Organizational Research Methods*, 11, 160-180.
- Bonney, R., Cooper, C. B., Dickinson, J., Kelling, S., Phillips, T., Rosenberg, K. V. & Shirk, J. 2009. Citizen science: a developing tool for expanding science knowledge and scientific literacy. *BioScience*, 59, 977-984.
- Brabham, D. C. 2008a. Crowdsourcing as a model for problem solving an introduction and cases. *Convergence: the international journal of research into new media technologies*, 14, 75-90.
- Brabham, D. C. 2008b. Moving the crowd at iStockphoto: The composition of the crowd and motivations for participation in a crowdsourcing application. *First monday*, 13.
- Brabham, D. C. 2009. Crowdsourcing the public participation process for planning projects. *Planning Theory*, 8, 242-262.
- Brabham, D. C. 2010. Moving the crowd at Threadless: Motivations for participation in a crowdsourcing application. *Information, Communication & Society*, 13, 1122-1145.
- Chernova, S., Depalma, N., Morant, E. & Breazeal, C. Crowdsourcing human-robot interaction: Application from virtual to physical worlds. RO-MAN, 2011 IEEE, 2011. IEEE, 21-26.
- 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, 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*. Springer.
- De Boer, V., Hildebrand, M., Aroyo, L., De Leenheer, P., Dijkshoorn, C., Tesfa, B. & Schreiber, G. Nicesourcing: Harnessing the power of crowds of experts. International Conference on Knowledge Engineering and Knowledge Management, 2012. Springer, 16-20.
- De Vaus, D. 2013. *Surveys in social research*, Routledge.
- Estellés-Arolas, E. & González-Ladrón-De-Guevara, F. 2012. Towards an integrated crowdsourcing definition. *Journal of Information science*, 38, 189-200.
- Eveleigh, A. 2014. Crowding out the archivist? Locating crowdsourcing within the broader landscape of participatory archives. *Crowdsourcing our Cultural Heritage*, 211-212.
- Filatova, E. Irony and Sarcasm: Corpus Generation and Analysis Using Crowdsourcing. LREC, 2012. 392-398.
- Følstad, A. 2008. Living labs for innovation and development of information and communication technology: A literature review. *eJOV: The Electronic Journal for Virtual Organization & Networks*, 10.
- Ford, R. C., Richard, B. & Ciuchta, M. P. 2015. Crowdsourcing: A new way of employing non-employees? *Business Horizons*, 58, 377-388.

- Geerts, S. 2009. Discovering crowdsourcing: theory, classification and directions for use. *unpublished Master of Science in Innovation Management thesis, Eindhoven University of Technology*.
- Geiger, D., Seedorf, S., Schulze, T., Nickerson, R. C. & Schader, M. Managing the Crowd: Towards a Taxonomy of Crowdsourcing Processes. AMCIS, 2011.
- Gerring, J. 2006. *Case study research: Principles and practices*, Cambridge University Press.
- Ghose, A., Ipeirotis, P. G. & Li, B. Designing Ranking Systems for Hotels on Travel Search Engines to Enhance User Experience. ICIS, 2010. Citeseer.
- Goode, W. J. & Hatt, P. K. 1952. *Methods in social research*.
- Goodman, J. K. & Malkoc, S. A. 2012. Choosing here and now versus there and later: The moderating role of psychological distance on assortment size preferences. *Journal of Consumer Research*, 39, 751-768.
- Gregor, S. 2002. A theory of theories in information systems. *Information Systems Foundations: building the theoretical base*, 1-20.
- Gregor, S. 2006. The nature of theory in information systems. *MIS quarterly*, 611-642.
- Gura, T. (2013). Citizen science: Amateur experts. *Nature*, 496(7444), 259-261.
- Heer, J. & Bostock, M. Crowdsourcing graphical perception: using mechanical turk to assess visualization design. Proceedings of the SIGCHI Conference on Human Factors in Computing Systems, 2010. ACM, 203-212.
- Holley, R. 2010. Crowdsourcing: How and why should libraries do it? *D-Lib Magazine*, 16.
- Howe, J. 2006. The rise of crowdsourcing. *Wired magazine*, 14, 1-4.
- Hyde, K. F. 2000. Recognising deductive processes in qualitative research. *Qualitative market research: An international journal*, 3, 82-90.
- Jarmolowicz, D. P., Bickel, W. K., Carter, A. E., Franck, C. T. & Mueller, E. T. 2012. Using crowdsourcing to examine relations between delay and probability discounting. *Behavioural processes*, 91, 308-312.
- Kafura, D. 1998. *Object-oriented software design and construction with C++*, Prentice-Hall, Inc.
- Khan, J. A. 2011. *Research methodology*, APH Publishing.
- Kim, J. 2012. The effect of design characteristics of mobile applications on user retention: an environmental psychology perspective.
- Kitchenham, B., Brereton, O. P., Budgen, D., Turner, M., Bailey, J. & Linkman, S. 2009. Systematic literature reviews in software engineering—a systematic literature review. *Information and software technology*, 51, 7-15.
- Kittur, A., Chi, E. H. & Suh, B. Crowdsourcing user studies with Mechanical Turk. Proceedings of the SIGCHI conference on human factors in computing systems, 2008. ACM, 453-456.
- 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, PP. 5-26.
- Krefting, L. 1991. Rigor in qualitative research: The assessment of trustworthiness. *American journal of occupational therapy*, 45, 214-222.
- Lampe, C., Zube, P., Lee, J., Park, C. H. & Johnston, E. 2014. Crowdsourcing civility: A natural experiment examining the effects of distributed moderation in online forums. *Government Information Quarterly*, 31, 317-326.
- Lee, A. S. & Hubona, G. S. 2009. A scientific basis for rigor in information systems research. *MIS Quarterly*, 237-262.
- Lin, J., Amini, S., Hong, J. I., Sadeh, N., Lindqvist, J. & Zhang, J. Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. Proceedings of the 2012 ACM Conference on Ubiquitous Computing, 2012. ACM, 501-510.
- Lokke, A. & Sorensen, P. 2014. Theory testing using case studies. *Electronic Journal of Business Research Methods*, 12, 66-74.
- Lynham, S. A. 2002. The general method of theory-building research in applied disciplines. *Advances in developing human resources*, 4, 221-241.

- Mackenzie, S. B., Podsakoff, P. M. & Podsakoff, N. P. 2011. Construct measurement and validation procedures in MIS and behavioral research: Integrating new and existing techniques. *MIS quarterly*, 35, 293-334.
- Mason, W. & Suri, S. 2012. Conducting behavioral research on Amazon's Mechanical Turk. *Behavior research methods*, 44, 1-23.
- Mingers, J. 2003. The paucity of multimethod research: a review of the information systems literature. *Information Systems Journal*, 13, 233-249.
- Muntermann, J., Nickerson, R. & Varshney, U. 2015. Towards the Development of a Taxonomic Theory.
- Niederman, F. & March, S. 2015. Reflections on Replications. *AIS Transactions on Replication*.
- Nunamaker Jr, J. F., Chen, M. & Purdin, T. D. 1990. Systems development in information systems research. *Journal of management information systems*, 7, 89-106.
- Okoli, C. & Schabram, K. 2010. A guide to conducting a systematic literature review of information systems research. *Sprouts Work. Pap. Inf. Syst*, 10, 26.
- Paré, G., Tate, M., Johnstone, D. & Kitsiou, S. 2016. Contextualizing the twin concepts of systematicity and transparency in information systems literature reviews. *European Journal of Information Systems*, 1-16.
- Paré, G., Trudel, M.-C., Jaana, M. & Kitsiou, S. 2015. Synthesizing information systems knowledge: A typology of literature reviews. *Information & Management*, 52, 183-199.
- Park, S., Shoemark, P. & Morency, L.-P. Toward crowdsourcing micro-level behavior annotations: the challenges of interface, training, and generalization. Proceedings of the 19th international conference on Intelligent User Interfaces, 2014. ACM, 37-46.
- Parvanta, C., Roth, Y. & Keller, H. 2013. Crowdsourcing 101 a few basics to make you the leader of the pack. *Health promotion practice*, 14, 163-167.
- Penker, M. 2000. Business Modeling with UML: Business Patterns at Work. New York, NY: John Wiley & Sons.
- Piper, S. 2006. Qualitative theory testing as mixed-method research. *Journal of Research in Nursing*, 11, 183-193.
- Ross, J., Zaldivar, A., Irani, L. & Tomlinson, B. 2009. Who are the turkers? worker demographics in amazon mechanical turk. *Department of Informatics, University of California, Irvine, USA, Tech. Rep*.
- Schenk, E. & Guittard, C. 2011. Towards a characterization of crowdsourcing practices. *Journal of Innovation Economics & Management*, 93-107.
- Smith, D., Manesh, M. M. G. & Alshaikh, A. 2013. How can entrepreneurs motivate crowdsourcing participants? *Technology Innovation Management Review*, 3.
- Snoek, C. G., Freiburg, B., Oomen, J. & Ordelman, R. Crowdsourcing rock n'roll multimedia retrieval. Proceedings of the 18th ACM international conference on Multimedia, 2010. ACM, 1535-1538.
- Steelman, Z., Soror, A., Limayem, M. & Worrell, D. 2012. Obsessive compulsive tendencies as predictors of dangerous mobile phone usage. AMCIS 2012 Proceedings. 9
- Stewart, N., Chandler, J., & Paolacci, G. (2017). Crowdsourcing Samples in Cognitive Science. *Trends in cognitive sciences*, 21(10), 736-748.
- Stinchcombe, A. L. 1987. *Constructing social theories*, University of Chicago Press.
- Su, A. I., Good, B. M. & Van Wijnen, A. J. 2013. Gene Wiki Reviews: marrying crowdsourcing with traditional peer review. *Gene*, 531, 125.
- Thuan, N. H., Antunes, P. & Johnstone, D. 2016. Factors influencing the decision to crowdsource: A systematic literature review. *Information Systems Frontiers*, 18, 47-68.
- Turner, A. M., Kirchhoff, K. & Capurro, D. 2012. Using crowdsourcing technology for testing multilingual public health promotion materials. *Journal of medical Internet research*, 14.
- Viswesvaran, C. & Ones, D. S. 1995. Theory testing: Combining psychometric meta-analysis and structural equations modeling. *Personnel Psychology*, 48, 865-885.

- Vukovic, M., Laredo, J. & Rajagopal, S. Challenges and experiences in deploying enterprise crowdsourcing service. *International Conference on Web Engineering*, 2010a. Springer, 460-467.
- Vukovic, M., Lopez, M. & Laredo, J. Peoplecloud for the globally integrated enterprise. *Service-Oriented Computing. ICSOC/ServiceWave 2009 Workshops*, 2010b. Springer, 109-114.
- Ward, M. K. & Broniarczyk, S. M. 2011. It's not me, it's you: How gift giving creates giver identity threat as a function of social closeness. *Journal of Consumer Research*, 38, 164-181.
- Wexler, M. N. 2011. Reconfiguring the sociology of the crowd: exploring crowdsourcing. *International Journal of Sociology and Social Policy*, 31, 6-20.
- Whitla, P. 2009. Crowdsourcing and its application in marketing activities. *Contemporary Management Research*, 5.
- Witschey, J., Murphy-Hill, E. & Xiao, S. Conducting interview studies: Challenges, lessons learned, and open questions. *Conducting Empirical Studies in Industry (CESI)*, 2013 1st International Workshop on, 2013. IEEE, 51-54.
- Wu, W., Tsai, W.-T. & Li, W. 2013. Creative software crowdsourcing: from components and algorithm development to project concept formations. *International Journal of Creative Computing*, 1, 57-91.
- Yang, J., Adamic, L. A. & Ackerman, M. S. Crowdsourcing and knowledge sharing: strategic user behavior on taskcn. *Proceedings of the 9th ACM conference on Electronic commerce*, 2008. ACM, 246-255.
- Yuen, M.-C., King, I. & Leung, K.-S. A survey of crowdsourcing systems. *Privacy, Security, Risk and Trust (PASSAT) and 2011 IEEE Third International Conference on Social Computing (SocialCom)*, 2011 IEEE Third International Conference on, 2011. IEEE, 766-773.
- Zhao, Y. & Zhu, Q. 2012. Exploring the motivation of participants in crowdsourcing contest.
- Zheng, H., Li, D. & Hou, W. 2011. Task design, motivation, and participation in crowdsourcing contests. *International Journal of Electronic Commerce*, 15, 57-88.