Crowd-Feedback in Information Systems Development: A State-of-the-Art Review

Saskia Haug
Karlsruhe Institute of Technology
Karlsruhe, Germany
saskia.haug@kit.edu

Alexander Maedche
Karlsruhe Institute of Technology
Karlsruhe, Germany
alexander.maedche@kit.edu

Abstract

Crowd-feedback is receiving increasing attention in research and practice as a contemporary approach for involving users in information systems development. Thereby, feedback on various aspects of an information system is collected from a non-expert crowd using designated crowd-feedback systems that are able to collect comprehensive and reliable feedback at scale. However, the current body of knowledge on crowd-feedback is scattered and lacks a structured form in which research on crowd-feedback can be classified. To address this gap, this article provides a comprehensive overview of the state-of-the-art of existing crowd-feedback research by: (1) conducting a systematic literature review, (2) developing a morphological box to conceptualize crowd-feedback, and (3) performing a cluster analysis for identifying research streams on crowd-feedback. Analyzing 40 articles, our key contribution resides in the synopsis of the existing crowd-feedback literature. Based on our review, we suggest four research avenues to guide researchers in investigating crowd-feedback in the future.

Keywords: crowdsourcing, IS development, UCD, systematic literature review

Introduction

User involvement is of critical importance in the development of any information system (IS). It has a positive impact on IS project success (Harris and Weistroffer 2009), user satisfaction (McKeen and Guimaraes 1997), and IS acceptance (Ives and Olson 1984). As such, not involving users in IS development may culminate in project failure (Hsu et al. 2013). User-centered design (UCD) is a prominent paradigm that provides methodological guidance for user involvement (Vredenburg et al. 2002). In the iterative UCD process, it is emphasized that users should be involved within all major activities of IS development, namely analysis, specification, design, and evaluation (Brhel et al. 2015; ISO 9241-210 2019). Continuous evaluation of design solutions with potential users is a challenging activity in UCD (Brhel et al. 2015). In particular, designers have to cope with the scalability issues of traditional face-to-face methods when they aim to involve a diverse set of users in the development process. Thus, in recent years, crowdsourcing in user-centered IS evaluation has received growing interest in research (Alyahya 2020; Leicht 2018; Mao et al. 2017; Sarı et al. 2019) and practice (e.g., uTest, UsabilityHub).

Thereby, two main research streams can be identified: crowd-testing and crowd-feedback. While in crowd-testing the crowd interacts with the IS in order to identify errors (Leicht 2018), in crowd-feedback the crowd is asked for explicit, mostly verbal feedback, including opinions and perceptions of the IS design (Xu et al. 2014). Interaction with the IS is thereby often not necessary and a textual description or screenshot of the user interface may be sufficient. However, the transition from crowd-testing to crowd-feedback is fluid and there of course exist systems that include both approaches (e.g. Ayalon and Toch 2019). While huge amounts of user feedback on existing systems are commonly provided in user forums and app stores (Pagano and Bruegge 2013; Tizard et al. 2021; Yen et al. 2016), collecting feedback during the IS development requires dedicated crowd-feedback systems and respective crowds. With these systems, feedback requesters can provide users with structure and guidance for providing dedicated feedback. The outcomes of these systems then comprise quantitative and qualitative data, that contain valuable issues, praises, and ideas for further improvement from the user perspective (Oppenlaender et al. 2021). Therefore, crowd-feedback systems differ from established online forums because the collected feedback is clearly structured and therefore more useful than unstructured comments or aggregated individual preferences collected in online forums (Xu et al. 2014). Overall, existing research on crowd-feedback systems has shown to be able to collect feedback with a quality similar to expert feedback (Wu and Bailey 2016; Yuan et al. 2016). While existing approaches and systems on crowd-testing in IS were recently reviewed (Alyahya 2020; Leicht 2018), there is a lack of a systematic review of existing knowledge on crowd-feedback and corresponding systems.

Although multiple studies already investigated the positive effects of design elements of crowd-feedback systems on outcomes like quality, scalability, and effort (Choi et al. 2020; Greenberg et al. 2015; Oppenlaender et al. 2020), there is a lack of research that synthesizes the current body of knowledge. Specifically, the varying requirements of feedback collection endeavors via crowdsourcing are not well structured and conceptualized. This hinders researchers to identify possible directions for future research efforts in this emerging research stream. In this paper, we seek to focus on these challenges by addressing the following two research questions (RQ):

RQ1: *How to conceptualize crowd-feedback for IS development?*

RQ2: *What is the state-of-the-art of crowd-feedback in IS development and what are future research directions?*

In order to answer the RQs, we perform a systematic literature review (SLR) study based on Webster and Watson (2002) and Kitchenham and Charters (2007) that identifies and investigates 40 articles on crowd-feedback systems. Next, following the approach of Nickerson et al. (2013) and the grounded-theory method proposed by Wolfswinkel et al. (2013) we develop a conceptualization of crowd-feedback in IS. By coding and analyzing the identified 40 articles, we provide a holistic overview of the state-of-the-art of crowd-feedback systems for IS evaluation. Based on a subsequent cluster analysis of the 40 articles, we identify three relevant research streams. Considering the previously provided state-of-the-art overview and the identified research streams, we present four avenues for future research. We contribute to theory by providing a morphological box for crowd-feedback for IS and a state-of-the-art review on this basis. This study additionally contributes to practice by supporting practitioners in applying crowd-feedback systems in IS development by outlining three research streams and related design characteristics. In the remainder of this paper, we first provide an overview of the theoretical foundations and illustrate the methods applied in this study. In section four the results of our study are presented. This is followed by a discussion and outline of future research directions in section five. Lastly, section six is the conclusion of our article.

Related Work and Conceptual Foundations

We first provide conceptual foundations on user-centered evaluation in IS development. Here our focus especially lies on the critical role of feedback in the user-centered evaluation and differentiating it from testing. This is followed by an introduction to crowdsourcing in IS development. Finally, we present existing conceptual frameworks on crowdsourcing systems.

User-Centered Evaluation in IS Development

There exist multiple methods to evaluate design solutions by involving the user, e.g., user interviews, focus groups, or usability testing (Gibbs 1997; Vredenburg et al. 2002). A distinction is usually made between formative and summative evaluation, with formative evaluation being carried out during the process and summative evaluation at the end to evaluate the outcomes (Scriven 1991). A general challenge is that formative and summative evaluation methods are time- and cost-intensive in their application and therefore lack scalability (Gibbs 1997; Scholtz 2001). This issue can be addressed by utilizing online crowds following a crowdsourcing paradigm.

The most fundamental method for evaluation is usability testing with potential users (Brhel et al. 2015; Nielsen 1994, p. 165). In usability testing, participants receive tasks that they must complete by interacting with the IS. During the test, the participant is observed and data, such as the time needed to complete a task and the number of required clicks, is measured. Additional subjective data is collected by recording participants' comments during the usage and optionally via a subsequent questionnaire (Nielsen 1994, pp. 165-206). In comparison, in feedback collection, participants' comments, ratings, votes, and markers are the only data that is collected. For existing systems, users provide feedback directly via pop-ups, app stores, or feedback forums (Almaliki et al. 2014). Feedback forums like Dribble are dedicated to providing feedback on designs and prototypes. The feedback quality on these platforms is often low and feedback requesters are not able to provide any guidance for users (Xu et al. 2014). Therefore, dedicated feedback systems are required to collect valuable feedback that exceeds simplistic statements of "I like it" (Xu et al. 2014). Many of these systems (e.g. Luther et al. 2014; Xu and Bailey 2011) originated from the visual design domain. There, feedback, also known as design critique, is traditionally provided by peers to help designers understand how others perceive their designs (Yuan et al. 2016). Crowd-based feedback systems were initially developed to solve scalability issues of peer feedback (Wauck et al. 2017) and reach a more diverse crowd of feedback providers (Ma et al. 2015). The ongoing challenge, especially in the visual design domain, is to enable the non-expert crowd to provide feedback that is similar to feedback from design experts and investigate the differences between feedback from the crowd and peers (Wauck et al. 2017; Yuan et al. 2016).

Crowdsourcing in IS Development

Crowdsourcing is a method to outsource tasks to a large undefined crowd of people (Howe 2009). For this, the potential of large groups is harnessed. The crowd can have various motivations to contribute to the task, e.g. financial incentives, enjoyment, or social status (Yen et al. 2016). Crowdsourcing is particularly useful for scaling complex work that cannot be handled by computers. As a result, crowdsourcing is being applied in various process areas of the software development process (Sarı et al. 2019). In the last years, the field of crowdsourcing in IS development and software engineering was growing quickly (Mao et al. 2017). The most used platform for crowdsourced software engineering is TopCoder (Sarı et al. 2019). TopCoder uses competitions to find the best solution and rewards the best participants with prize money. However, other crowdsourcing platforms that rely on collaboration instead of competition (e.g. Amazon Mechanical Turk (MTurk)) are also applicable for software development (Sarı et al. 2019). Besides the anonymous crowds that can be reached via dedicated platforms like MTurk or uTest, we consider for our review also crowds like employees, stakeholders, and students that are not necessarily recruited on these platforms. The IS development tasks to which crowdsourcing is mostly applied are requirements analysis, coding, and testing (Ambreen and Ikram 2016; Mao et al. 2017; Sarı et al. 2019).

Applying crowdsourcing in IS design and development yields several advantages. For instance, TopCoder is reported to deliver software artifacts with a lower defect rate and higher quality at lower cost in less time compared to in-house development or outsourcing (Lakhani et al. 2010; Lakhani et al. 2013). Common concerns regarding crowdsourcing in IS development involve the intellectual property, quality, uncertainty, limited interaction, and collaboration overhead (Mao et al. 2017). There are already multiple reviews on crowdsourcing in IS development in general (Mao et al. 2017; Sarı et al. 2019) and on crowdsourcing in IS evaluation in particular (Alyahya 2020; Leicht 2018). However, so far, the focus has been put on crowdsourced software testing. Crowd-testing is an emerging trend in software engineering that enables companies to outsource different software testing activities to a large pool of workers (Alyahya 2020). Dedicated platforms like uTest provide contact to thousands of workers and ensure that the individual testing requirements are included in the tasks (Alyahya 2020). Crowd-feedback has been considered only marginally, if at all, as an aspect of crowd-testing. However, crowd-feedback goes beyond crowd-testing and should therefore be considered in a separate literature review. Additionally, a conceptualization is required to get a holistic understanding of existing approaches for crowdsourcing feedback within IS evaluation.

Conceptualization of Crowdsourcing Systems

Existing conceptualizations of crowdsourcing systems (e.g., Morschheuser et al. 2017; Pedersen et al. 2013; Zuchowski et al. 2016) all follow the same structure: First, a task or problem defined as "a statement of an initial condition and a desired ending condition" (Pedersen et al. 2013) is identified. This is followed by a specification of the crowdsourcing task and system, and finally, specific outcomes are achieved.

Pedersen et al. (2013) did the first effort to conceptualize crowdsourcing research in general. The first element of their conceptual model is the problem which defines the requirements for all other model elements. The main part of the model includes a process (the design of a step-by-step plan to solve the problem), governance (the actions and policies that are applied to manage the crowd), people (including the problem owner and the crowd that consists out of many individuals), and the technology (the technical capabilities that enable the formation of the crowd and the interaction and collaboration between individuals in the crowd). The final element of the conceptualization is the outcome. This refers to the factual and the perceptual outcome of the crowdsourcing process. While the conceptualization of Pedersen et al. (2013) applies to crowdsourcing tasks in general, Zuchowski et al. (2016) and Morschheuser et al. (2017) focused on specific crowdsourcing tasks. The framework of Zuchowski et al. (2016) conceptualizes IT-enabled crowdsourcing with employees in enterprises, also called 'internal crowdsourcing'. This framework includes similar elements as the model of Pedersen et al. (2013). The main contribution of Zuchowski et al. (2016) is the definition of subdimensions that describe internal crowdsourcing tasks. Morschheuser et al. (2017) developed a conceptual framework for gamified crowdsourcing which is based on the previous two conceptualizations. As Morschheuser et al. (2017) put their focus on the crowdsourcing system this element replaces the main component which was initially defined by the governance, IT, process, and people. The gamified crowdsourcing system is mainly defined by the type of crowdsourced work. The design of the crowdsourcing system is not only influenced by the initial problem and tasks as in the other two frameworks but also by the crowds' motivation and the resulting behavior which is, in turn, a result of the gamification affordances and additional incentives. As we appreciate the approach of Morschheuser et al. (2017) of separating the crowdsourcing system and the crowd configuration, our framework will mainly be based on their framework on gamified crowdsourcing.

Research Methodology

To answer RQ1 and RQ2, we followed a multistep research approach. In the first step, we sought to review the current state-of-the-art and to derive a set of relevant papers on crowd-feedback systems. This set of papers provided our foundation to answer RQ1 and RQ2. In the second step, we used these papers to develop a conceptualization of crowd-feedback in the form of a morphological box (RQ1). Finally, we used the set of papers as well as the morphological box and applied a cluster analysis to identify the existing research streams (RQ2). We outline each of these steps in more detail in the following sections.

Systematic Literature Review

To conduct the SLR we followed the guidelines of Kitchenham and Charters (2007) and first developed our search strategy. Therefore, we created the search string in several iterations. We started with an exploratory search using Google Scholar with the search string "*crowd AND feedback AND system*". After reviewing the results, we iterated the search string several times using Google Scholar. The final search string consisted out of four parts: The first part is for ensuring that a crowd is involved in the feedback collection process. In the second part, we added 'critique' and 'comment' as synonyms for feedback and in the third part, we added 'method', 'process', and 'tool' as further means to crowdsource feedback. The fourth part was added to specify about what the feedback is collected. For searching for papers that collect feedback on IS, we additionally included specific types and characteristics of IS like 'website', 'software', 'interactive', 'app', and 'interface'. To provide a holistic overview we also included studies on crowdsourcing feedback on graphic and product design. Due to the similarity to user interface design, we expect the feedback systems and their features to be also applicable to feedback on user interface design. Therefore, we included the term 'design' in the fourth part of our search string. Finally, applying Boolean operators and wildcards led us to the final search string:

crowd AND (feedback OR critique OR comment) AND (system OR process OR method OR tool) AND ("information system" OR website OR software OR interactive OR design OR app OR interface).*

In the next step, we selected ACM Digital Library, IEEE Explore, and AISeL as databases for our SLR. These databases are well-established and were already used by scholars as reliable sources for literature reviews (Bandara et al. 2015). We decided to not limit the results to a specific time period or publication outlet in order to get a holistic overview. To refine the initial set, we then scanned the title, abstract, and keywords followed by reviewing the full text of the remaining papers. For the filtering, we applied six selection criteria: (1) the paper implements a prototype or develops a conceptual framework for collecting feedback, (2) the paper investigates an artifact mainly used to explicitly collect feedback, (3) the feedback is provided by a human crowd and assesses information systems, visual designs or product designs, (4) the article is peer-reviewed, (5) the article has more than three pages, (6) the article is written in English. Next, we conducted a backward and forward search following the same criteria. In our final set, we identified multiple articles that refer to the same crowd-feedback system. For the subsequent analysis of the papers, we inspect these papers jointly. Finally, we coded the resulting set of articles by their main research methodology. The main result of this step is a comprehensive set of papers investigating crowd-feedback systems.

Concept Creation

In the next step, we analyzed the identified set of papers in order to conceptualize crowd-feedback systems (RQ1). Therefore, we developed a morphological box that captures all relevant dimensions of crowd-feedback systems. A morphological box provides a structured overview of all potential solutions to a problem (Zwicky and Wilson 1967) and is commonly used for SLRs in the domain of IS to illustrate the diversity of solutions. To develop a morphological box, we followed the approaches of Wolfswinkel et al. (2013) and Nickerson et al. (2013). Based on their recommendations, we applied a three-step development approach:

In the **first step**, we followed Nickerson et al. (2013) and conducted a conceptual-to-empirical development approach to create an initial conceptual framework as a foundation for the following steps. Therefore, we started with the three frameworks introduced in the conceptual foundations from Morschheuser et al. (2017), Pedersen et al. (2013), and Zuchowski et al. (2016), extracted all of their dimensions and developed an initial conceptualization comprising of four overarching dimensions (i.e., input, crowd configuration, design characteristics, and effects) to guide our next steps.

In the **second step**, we again followed Nickerson et al. (2013) and conducted an empirical-to-conceptual development approach. Therefore, we used an inductive coding approach to create new subcategories for our morphological box and to identify codes for these subcategories based on Wolfwinkel et al. (2013). This step is necessary to develop a morphological box for crowd-feedback accounting for their characteristics which are not captured by the initial coding scheme yet. For that, we iteratively reviewed each of the identified papers and continuously refined the initial coding scheme until the concepts reached an acceptable level of abstraction.

In the **third step**, all studies included in the final set were coded according to the concepts that we defined in the previous steps and a morphological box, as well as a concept matrix as described by Webster and Watson (2002), were created.

Cluster Analysis

Based on the identified set of papers and the derived morphological box, we were seeking to identify the existing research streams on crowd-feedback systems in order to answer RQ2. Thereby we aimed to understand which characteristics of crowd-feedback systems are usually combined and which effects are achieved by doing so. This shall help future researchers and practitioners to select appropriate design combinations when developing new crowd-feedback systems. Due to the relatively low number of papers, we first clustered the papers manually by identifying characteristics that often occur in combination and grouping these papers together. To verify the results, we decided to apply the two-step clustering analysis developed by Chiu et al. (2001). The two-step clustering is an effective approach to identify clusters and is often applied in literature reviews (e.g., Knaeble et al. 2020; Rissler et al. 2017). The advantage of this approach compared to pure hierarchical clustering is that it automatically detects the optimal number of clusters and provides the silhouette measure of cohesion and separation as a quality measure. Additionally, it summarizes the influence of each characteristic on the cluster allocation which helped us to verify if we identified the correct characteristics as the main drivers of the clustering.

To conduct the two-step clustering, we first transformed our concept matrix into a binary form by changing every ‘X’ to a ‘1’ and every empty cell to a ‘0’. Then we applied the two-step clustering to separate our articles into homogenous groups (clusters) using IBM SPSS Statistics 27. The two-step clustering is based on two distinct steps: First, the entire dataset is scanned, and based on sequential clustering preclusters are created. In this step, the log-likelihood distance measure is applied as the similarity criterion which is appropriate as our input data is binary (Mooi and Sarstedt 2011; Theodoridis and Koutroumbas 2009). Second, agglomerative hierarchical clustering is applied to the created preclusters (Chiu et al. 2001). We applied Akaike’s information criterium (AIC) to determine the appropriate number of clusters (Akaike 1998).

Results

In this section, we describe the results of our three-step research approach. First, we outline the results of the SLR and describe the identified set of papers. This serves as the foundation for the following two steps. Second, we present the morphological box of crowd-feedback systems based on the discovered articles and the iterative refinement of the coding dimensions (RQ1). Third, we introduce three research streams that were identified via the cluster analysis (RQ2).

Results of Systematic Literature Review

Applying our search string to the identified databases (i.e., AISeL, ACM Digital Library, IEEE Explorer) resulted in 274 studies. From the initial 274 studies, we excluded 227 by carefully scanning title, abstract, and keywords and thereby applied our inclusion criteria (47 remained). We applied the same criteria when reviewing the full texts and kept 24 studies. Most of the excluded studies either focus on crowd-testing and address feedback only marginally, or collect feedback on something else than information systems, visual designs, or product designs (e.g., university courses). Finally, we conducted a backward and forward search following our criteria and identified 16 additional studies. Consequently, in total, we identified 40 relevant articles. Since some of these studies refer to the same system, only 34 different crowd-feedback systems are included in our set of studies.

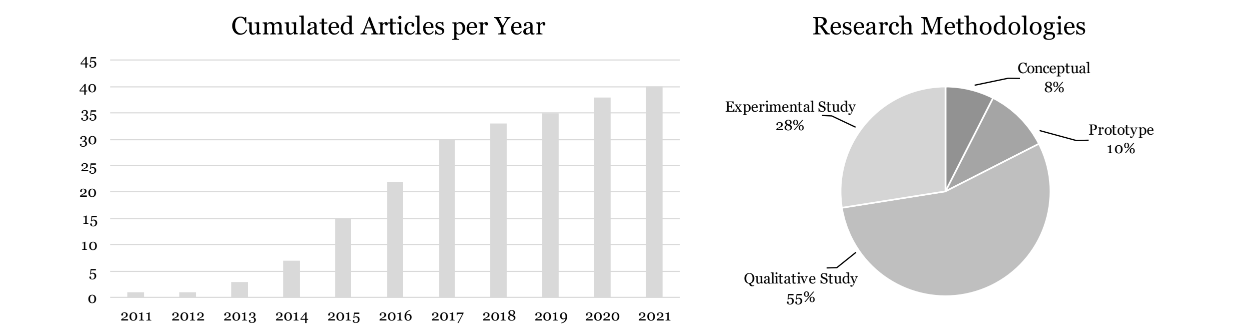

Figure 1. Cumulated Number of Articles reporting Crowd-Feedback Systems (left) and applied Research Methodologies (right)

For the descriptive information on our paper set, we considered all 40 articles without excluding articles on the same crowd-feedback system. This is necessary to account for a holistic overview of all existing studies. A complete list of all identified papers is depicted in Appendix Table 1. The analysis of the publication dates of the articles (see Figure 1, left) shows that the topic of crowdsourcing feedback emerged around ten years ago and had a peek between 2015 and 2017. The most common research methodology applied is the qualitative study including case studies as well as grounded theory (see Figure 1, right). It is followed by the experimental study methodology which is applied in 28% of the studies. Our set also includes articles on prototype development and conceptual models for crowd-feedback systems.

Results of Concept Creation

In the first step, we developed an initial conceptual framework (see Figure 2) based on the existing conceptualizations of crowdsourcing tasks and systems (Morschheuser et al. 2017; Pedersen et al. 2013; Zuchowski et al. 2016) and the identified 40 papers to describe the existing research on crowd-feedback. We retained the established components 'problem' and 'outcomes' and renamed them 'input' and 'effects'. In our case, the input includes characteristics of the IS on which the feedback shall be collected and the specification of the feedback that is requested. The effects describe the effects of applying crowd-feedback systems on the crowd, the feedback, and the resulting IS design. The crowd-feedback system which we put in the center of our conceptualization is mainly described by its design. The crowd-feedback system is usually intertwined with the crowd configuration as the design of the system might restrict the crowd configuration and vice versa.

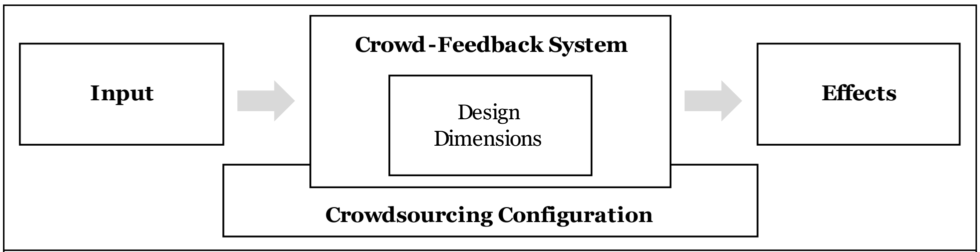

Figure 2. Conceptual Framework of Crowd-Feedback

In the next step, we iteratively developed subdimensions and concepts for each of the four dimensions of our initial conceptual framework to capture the complete landscape of crowd-feedback systems. The development of the subdimensions was completely based on the 40 papers of our SLR. The subsequent coding of all papers resulted in a concept matrix (see Appendix Table 1) and a morphological box (see Figure 3) that includes all (sub)dimensions and characteristics. Studies may include several characteristics of one subdimension (e.g., they collect qualitative and quantitative feedback). For each concept, the figure shows the absolute frequency of the concept in our set of papers (indicated by the number behind the concept). The different shades visualize this frequency and help to illustrate the current focus of research. In the following, we provide an analysis of each subdimension.

Dimension	Subdimension	Characteristic				
Inputs	IS Lifecycle Stage	Development (28)	Operations (10)			
	Feedback Type	Qualitative (29)	Quantitative (17)			
	Scope of Feedback	Non-functional Attributes (25)	Content (17)	Functional Attributes (15)		
Crowd Configuration	Crowd Type	Annoymous (21)	'Proxy' Users (13)	Students (3)	Convenience (2)	
	Incentive	Money (22)	Involvement and Improvement (8)	Interest and Social Compensation (5)	Credits (3)	Gamification (2)
Design Characteristics	Feedback Collection Mechanisms	Questionnaire (10)	Free Text Field (8)	Categories (8)	Selection (7)	Direct Manipulation (2)
	Interactivity Cues	Collaboration (7)	Marker (7)	Context (4)	Recording (4)	
Effects		Outcome Effects (22)	Process Effects (16)	Intermediate Effects (7)		

Applied in ... of studies: < 25% < 50% < 75% ≥ 75%

Figure 3. Morphological Box for Crowd-Feedback

Dimension: Inputs

The input dimension describes characteristics of the initial situation of the feedback collection process. Thereby, we distinguish between characteristics of the IS that needs to be evaluated, like its *lifecycle stage*, and the characteristics of the feedback that is sought. While the *type of feedback* indicates if the collected feedback is of quantitative or qualitative nature, the *feedback scope* represents the attributes of an IS that the feedback is focused on. Some of the crowd-feedback systems are developed for collecting feedback on one specific IS like an 'interactive Energy Saving Account' (Stade et al. 2017) or e-services of public administrations (Pretel et al. 2017). Others focus on an IS class like conversational user interfaces (Choi et al. 2020; Yu et al. 2016), mobile apps (Ayalon and Toch 2018, 2019; Oppenlaender et al. 2021; Seyff et al. 2014), reinforcement learning systems (La Cruz et al. 2015) or adaptive software systems (Muñante et al. 2017). However, most of the presented crowd-feedback systems are not limited and should apply to a wide range of IS.

IS Lifecycle Stage. For the IS lifecycle stage, we distinguish between systems that collect feedback for IS during its *development* process and systems that collect the feedback during *operations* of the IS for further improvement. Most of the articles (28) cover collecting feedback during the development stage, while only ten systems collect feedback during operations. Only four of the feedback systems apply to both lifecycle stages. For instance, Snijders et al. (2015) built a gamified online platform that can be applied to elicit requirements for new IS as well as requirements to further improve existing IS. Wu and Bailey (2016) evaluated their system by collecting feedback on an existing web page. However, their crowd-feedback system is designed in a way that it is also applicable to evaluating IS during development.

Feedback Type. Most papers (29) seek *qualitative* feedback which is usually done via text fields. However, *quantitative* feedback is also often collected and is requested in various ways. A common way is to vote designs 'up' or 'down' as it is applied in the CrowdUI system (Oppenlaender et al. 2020). There, participants can vote on the design suggestions of other participants that were created by manipulating the initial design. The system of Jansson and Bremdal (2018) asks users to vote on web page elements to help a system based on artificial intelligence to learn user preferences. Easterday et al. (2017) implemented a feedback system similar to a forum where users can add qualitative comments, but also vote on the comments of other participants. Besides the voting, the selection of items is a common method to collect quantitative feedback. By asking users to select abstract and emotional images that represent their emotional reaction to a particular design, Robb et al. (2015a, 2015b; 2017) introduced an innovative way to collect quantitative feedback. Paragon (Kang et al. 2018) enables participants to enrich their feedback by selecting exemplary designs. Finally, the most common way to collect quantitative feedback are Likert-scales as used by Schneider et al. (2016) to indicate how severe a usability problem is or by Xu et al. (2014) and Oppenlaender et al. (2021) to rate if design guidelines are considered.

Scope of Feedback. Feedback can be collected on *non-functional attributes,* thus aesthetics and human values, *functional attributes* that summarize feedback about features and functionalities of the IS or *content* of the IS. The code content is meant for systems that collect feedback about the information that is provided by the IS, for example, the content of a website. While non-functional attributes are considered in 25 of our 34 studies, feedback on functional attributes and the content is less often included. As many of the papers in our set focus on design feedback, they usually ask only for feedback on visual design and aesthetical aspects like layout, consistency, balance, readability, and simplicity (Luther et al. 2014) or aim to understand the first notice and impressions of viewers (Xu et al. 2014). Besides these aspects, users' human values are another non-functional attribute. Here, crowd-feedback systems are used to evaluate if the respective values are considered in the IS design. While Ayalon and Toch (2018, 2019) examine the social and institutional privacy of IS, Hosseini et al. (2016) developed a system to collect feedback on the implementation of transparency requirements. Feedback on functional attributes focuses on the capabilities of the IS. Citizenpedia (Pretel et al. 2017) is a platform for citizens to comment on current procedures of e-services. For this purpose, the platform provides a hierarchical overview of the services offered and the corresponding flow of interactions between the citizens and the public administration. There exist also two studies on collecting feedback on the functionality of conversational user interfaces (Choi et al. 2020; Yu et al. 2016). Both tools focus on providing feedback on answers of the conversational user interface. Yu et al. (2016) ask participants to rate the appropriateness of chatbot reactions. Choi et al. (2020) change the conversation flow or provide new suggestions for chatbot reactions. Feedback on content is only used as an addition to feedback on functional or non-functional attributes. None of the articles in our set collected feedback only on the content. The collection of feedback on the content of the IS is especially useful for IS that aim to provide information like a university website (Wu and Bailey 2016), posters and logos (Yen et al. 2016), or weather dashboards (Krause et al. 2017).

Dimension: Crowd Configuration

The crowd configuration describes how the crowdsourcing task is configured by analyzing which *type of crowd* is asked for feedback and how this crowd is *incentivized* to contribute.

Type of Crowd. The type of crowd can be *anonymous*, when dedicated crowdsourcing platforms are used or can consist out of *'proxy' users*, *students,* or friends and family (coded as *convenience)*. Most of the papers in our set use an anonymous crowd to collect feedback. These crowds are recruited on platforms like MTurk, Mobile-Works, or Upwork (Greenberg et al. 2015; Krause et al. 2017). Besides crowdworkers that were recruited on dedicated crowdsourcing platforms, 'proxy' users are also frequently used for feedback collection. This term includes actual and potential users as well as other stakeholders like developers, analysts, clients, and regulatory bodies (Snijders et al. 2015). Students are only used as feedback providers when the IS design is part of an educational class as in the studies of Oppenlaender et al. (2021) and Robb et al. (2015a). In the studies of Wauck et al. (2017) and Yen et al. (2016), designers use social media platforms to crowdsource feedback from social contacts. In both studies, convenience feedback is not used as the main source of feedback but to compare the feedback with the feedback from other sources. Another way to involve contacts from social networks in the feedback process is to promote the feedback system on these platforms as done by Haukipuro et al. (2016).

Incentives. For incentives we distinguish between *money*, *involvement* in the IS development and *improvement* of the IS, *interest and social compensation*, *course credits*, and *gamification*. The incentives are highly related to the type of crowd used. While the anonymous crowd is usually incentivized by a financial reward, 'proxy' users are motivated by the prospect of improvement of the IS and the feeling of being involved in the development process. Students are usually incentivized by course credits and social contacts contribute because of personal interest and social compensation. Gamification is the only incentive that we identified that is not related to one type of crowd. Although Pretel et al. (2017) and Snijders et al. (2015) apply gamification only in the combination with actual user feedback, Morschheuser et al. (2017) show that gamification can also be applied to motivate other types of crowds.

Dimension: Design Characteristics

The design characteristics describe the features of the proposed crowd-feedback systems. During the analysis of our set of articles, we learned that crowd-feedback systems usually consist out of a *feedback collection mechanism* and some additional functionalities that aim to support the crowd during the process of providing feedback, here called *interactivity cues*. In this section, we analyze both parts of crowd-feedback systems separately.

Feedback Collection Mechanisms. For the feedback collection we distinguish between five mechanisms: With *questionnaires*, participants are asked a series of questions. These can be of a qualitative or quantitative nature. *Categories* enable participants to select a suitable category or rubric for their feedback. Compared to questionnaires, this mechanism offers more freedom to feedback providers. Systems that provide only one single text field for feedback, are coded with *free text field*. Another mechanism is the *selection*, where the feedback is provided by selecting items. The most complex mechanism to provide feedback is *direct manipulation* where the crowd can edit a system according to their needs and wishes. The frequency of occurrences of the five feedback collection mechanisms that we identified is evenly distributed. Only direct manipulation is less used. Questionnaires include either closed questions (Haukipuro et al. 2016; Nebeling et al. 2013) or open questions (Greenberg et al. 2015; Xu and Bailey 2011). Using questionnaires can be advantageous compared to categories or more rigid structures, especially when dealing with a large range of themes and possible critiques (Greenberg et al. 2015). On the other hand, categories offer more freedom for the participants, since they can often specify several feedback points under one category and can also choose the order that they want to provide feedback themselves (Luther et al. 2014; Yuan et al. 2016). The study of Yuan et al. (2016) shows that rubrics help feedback providers to contribute feedback that is nearly as valuable as expert feedback. They show that feedback collection via rubrics leads to feedback with higher quality than feedback collected in free text fields. Feedback systems that collect feedback via a free text field do not ask the users specific questions but just provide a field to enter feedback (Krause et al. 2017; Seyff et al. 2014; Wu and Bailey 2021). This mechanism is also often used in user forums and similar communities like the gamified platform to derive user requirements of Snijders et al. (2015). A very different way to collect feedback is selecting items, e.g., design elements (Jansson and Bremdal 2018), concrete improvements (Choi et al. 2020; La Cruz et al. 2015), or labels for chatbot answers (Yu et al. 2016). The selection mechanism enables the collected data to be easily quantifiable and may therefore reduce the effort to analyze the feedback. Finally, direct manipulation is the most direct way for the crowd to provide feedback. Oppenlaender et al. (2020) developed a feedback system that enables the crowd to change the UI design of an existing IS and the ProtoChat system (Choi et al. 2020) allows the user to change the conversation flow of the chatbot.

Interactivity Cues. We identified four frequently used interactivity cues: *collaboration*, *marker*, *context,* and *recording*. There exist more than these four, but we chose to focus on cues that appeared in more than one single article. Collaboration means that the participants can interact with the feedback of other crowd members by rating or commenting on it. Systems that enable the crowd to mark their feedback visually in the system are coded with marker. Context means that the crowd receives a context of use in form of a specific scenario or a persona before providing their feedback. Recording is for systems that include a feature that allows participants to do voice or video recordings of their feedback. Collaboration is often applied in crowdsourcing systems that resemble a user forum or community. There, users can see the feedback of others and react to it by commenting and voting (Haukipuro et al. 2016; Pretel et al. 2017; Snijders et al. 2015) or just use it as inspiration (Xu and Bailey 2011). Another way to include collaboration is to let the crowd rate design suggestions of other crowd members (Oppenlaender et al. 2020). Markers can either be in the form of flags that can be put onto the IS user interface to indicate what specific element is meant by the feedback (Luther et al. 2014; Wu and Bailey 2016; Xu and Bailey 2014) or in the form of screenshots and photos that are taken of the IS when the user encounters a problem and wants to provide feedback (Schneider et al. 2016; Seyff et al. 2014; Stade et al. 2017). Oppenlaender and Hosio (2019) include the marker feature in their system as the actual answer to tasks like 'Touch the areas of the artwork that you like best!'. The context of use aims to help the participant to imagine himself in a real usage scenario while providing feedback. The context can either be created by providing a persona (Ayalon and Toch 2018; Muñante et al. 2017) or a scenario that explains the design and its context (Ayalon and Toch 2019; Wu and Bailey 2021). Recording feedback is an optimal way to better capture the emotions of feedback providers (Ma et al. 2015) and reduce the effort for them to provide feedback (Seyff et al. 2014). While Ma et al. (2015), Easterday et al. (2017), and Dow et al. (2013) let participants record videos of their feedback, Seyff et al. (2014) enable users to provide short audio recordings of their feedback.

Dimension: Effects

Finally, this dimension specifies the consequences of applying the described crowd-feedback system. For the effects, we did not define subdimensions, but three codes: *process effects*, *intermediate effects,* and *outcome effects*. As the studies usually only investigate a subset of possible effects, we can only report the effects that are described in the studies although the feedback system might lead to further effects as well. Process effects address the well-known problems of user-centered evaluation methods like scalability (Easterday et al. 2017; La Cruz et al. 2015; Oppenlaender et al. 2020; Schneider et al. 2016; Yen et al. 2016), effort (Ayalon and Toch 2019; Greenberg et al. 2015; Haukipuro et al. 2016; Hosseini et al. 2016; Jansson and Bremdal 2018; Ma et al. 2015) and costs (Greenberg et al. 2015; La Cruz et al. 2015; Ma et al. 2015; Schneider et al. 2016) of the feedback collection process. Additionally, the diversity of the feedback providers and the resulting feedback is often seen as one big advantage of applying crowdsourcing (Dow et al. 2013; Haukipuro et al. 2016; Ma et al. 2015; Nebeling et al. 2013; Oppenlaender et al. 2021; Schneider et al. 2016; Wauck et al. 2017). These outcomes are in most cases elicited via qualitative interviews with the respective feedback requesters. Intermediate effects are outcomes that might mediate outcome effects like the feedback quality but are not the direct goal of applying crowd-based feedback systems. These effects include user engagement (Hosseini et al. 2016; Oppenlaender and Hosio 2019; Robb et al. 2015b; Robb et al. 2017; Snijders et al. 2015; Yu et al. 2016) and inspiration for the designer (Jansson and Bremdal 2018; Kang et al. 2018; Robb et al. 2015a, 2015b). Finally, the outcome effects are usually measured via a quantitative evaluation, often including experts or the feedback requesters to rate the quality and helpfulness of the feedback or the final IS designs. The most mentioned outcome of crowd-feedback systems is the quality of feedback, including reliability and helpfulness of feedback (e.g., Ayalon and Toch 2018; Choi et al. 2020; Krause et al. 2017; Luther et al. 2014). Additional aspects of outcome effects are the quantity of feedback (Kang et al. 2018), a better final design (Lekschas et al. 2021; Luther et al. 2015; Xu et al. 2015), and increased user satisfaction when using the improved IS (Muñante et al. 2017).

Results of Cluster Analysis

By manually grouping papers with similar characteristics, we identified three clusters. Thereby, we identified the input and crowd configuration, especially the feedback scope and crowd type, as the main drivers for the cluster affiliation. The two-step cluster analysis confirmed our assumption and identified similar research streams while the optimal number of clusters was found to be three. The most important categories are the crowd type, the feedback scope, and the incentive which is highly related to the crowd type. The research streams based on the two-step clustering are displayed in Appendix Table 1. The silhouette measure of cohesion and separation of the analysis is 0.3 which indicates a medium solution (Mooi and Sarstedt 2011). However, since we obtained similar results from the manual analysis, we consider the results reliable and present them in the following. Thereby we highlight the characteristics of each research stream to complement the comprehensive state-of-the-art overview provided in the previous chapter.

Stream 1 – Anonymous Crowd-Feedback: The first stream (11 studies) is dominated by crowd-feedback systems that are designed to ask an anonymous crowd for feedback. These studies mostly originate from the field of visual design. Here, the crowd is usually incentivized by money. Feedback is collected during the development stage on non-functional attributes and often additionally on content. Thereby, all studies collect qualitative feedback. Consequently, these systems are mainly focused on formative feedback to further improve the design. In these studies, the focus lies on achieving outcome effects.

Stream 2 – Real User Crowd-Feedback: The 10 studies in this stream are mainly performed to collect feedback from real and potential users on systems during development as well as systems in operations. The crowd is incentivized by involvement and improvement of the system and is asked for qualitative and quantitative feedback mostly on functional attributes. Most of these systems apply collaboration as an interactivity cue. Therefore, these crowd-feedback systems resemble user forums but provide more guidance. The outcomes of user forums are in most cases outcome effects and sometimes process effects.

Stream 3 – Hybrid Crowd-Feedback: This stream includes 13 studies and is less clearly defined than the other two streams. In this stream, the studies ask all types of crowds for feedback with no limitation on specific attributes. The studies are mainly connected by the goal to achieve process effects. This is consistent with the fact that feedback is mostly collected through questionnaires and selection. Besides process effects, some studies additionally achieve intermediate effects like an increased user engagement.

Discussion

This study synthesizes characteristics of crowd-feedback systems for IS development from articles reporting the results of research projects in this field. Regarding the conceptualization of crowd-feedback (RQ1), we developed a conceptual framework and a morphological box for crowd-feedback. These conceptualizations are not limited to the configuration of crowd-feedback systems but include associated aspects like the crowd configuration as well. Our morphological box provides a comprehensive structure of crowd-feedback systems and visualizes where the focus in recent research was put. The morphological box can be applied to future research projects to consider possible design choices. The subsequent cluster analysis helped us to identify patterns in the existing studies and showed which characteristics of crowd-feedback are usually connected. This might also support crowd-feedback system developers in making the right design choices considering their specific use case. Based on the insights we gained during our study, we now propose four main future research directions (RQ2).

Effects of Design Characteristics

First, we argue that there is a need to better understand the effects of specific design characteristics. Our analysis revealed several feedback collection mechanisms and interactivity cues that can be included in crowd-feedback systems. By clustering our papers, we identified a connection between the inputs, the crowd configuration, and the resulting effects. However, we could not find a pattern for most of the design characteristics. While questionnaires and selection are related to process effects as they are easily quantifiable and therefore scalable, the other feedback collection mechanisms do not seem to be connected to inputs, the crowd configuration, or effects. The same applies to the interactivity cues. Collaboration occurs mostly when real users are asked for feedback, but the remaining interactivity cues seem to not follow any pattern. Although there exists already some research that investigates specific design characteristics, such as the effect of rubrics on the feedback quality (Yuan et al. 2016), there is still a lack of systematic research on the design characteristics of crowd-feedback systems. For the application of crowd-feedback systems in practice, it is important to understand how to achieve specific effects by selecting the appropriate design characteristics, not only considering the feedback quality and quantity, but also the effects on the crowd. At the same time, we consider it important to learn how the design characteristics are related to the input characteristics and crowd configurations to provide recommendations according to the selected inputs.

While we assume most interactivity cues having a positive impact on the feedback, there might also be some drawbacks. For example, we would assume that a collaboration feature positively impacts the provided feedback. However, the crowd might either be inspired by the comments of others or could be influenced by the perceptions of ‘opinion leaders’ (Bodendorf and Kaiser 2009) and consequently not share their own opinions. They could even get the feeling that their feedback is not required anymore. To investigate individual effects as well as the effects of combining characteristics, dedicated experimental studies are required.

Intermediate Effects on the Crowd

Second, our concept matrix shows that intermediate effects are only investigated by a few crowd-feedback studies. In these studies, only user engagement and the inspiration of the designer were investigated as intermediate effects. Additionally, the connection between the interactivity and the intermediate effects, especially concerning the crowd's behavior and perceptions of the crowd-feedback system, needs to be further investigated. Related studies in the field of IS development already identified a positive influence of interactive features on user engagement and the resulting behavior for employee participation (Feine et al. 2020). According to the insights of this study, intermediate effects could also serve as mediators for outcome effects like feedback quantity and quality. Besides the user engagement, we suggest exploring additional intermediate effects such as the perceived interactivity, the effort of using the crowd-feedback system, as well as its usability. Knowledge of the connection between these constructs, design characteristics, and the resulting feedback will help feedback requesters to design better feedback systems.

Crowd-Feedback System Configurators

Third, we propose to research crowd-feedback system configurators to enable novices to build and adapt crowd-feedback systems according to their individual use case. All existing crowd-feedback systems consist out of a fixed set of design characteristics and provide no functionality to adapt them to a specific use case. The configurator should be based on the results of the two previously suggested avenues for future research and consider the three clusters that we identified in this paper. Adaptable crowd-feedback systems might not only make crowd-feedback applicable to a more diverse set of use cases (Luther et al. 2015) but also increase the feedback quality and user satisfaction (Almaliki et al. 2014).

Continuous Feedback Collection

Fourth, we identified the need to research further support for continuous IS evaluation. As we highlighted in the beginning, continuous user involvement is crucial for IS acceptance and success (Harris and Weistroffer 2009; Ives and Olson 1984). During the analysis, we learned that most studies that investigate crowd-feedback focus on feedback during the development process. Only four studies in our set developed a system that applies to the feedback collection during the development and operations of the IS. However, none of these studies explicitly decided to focus on the entire lifecycle, but the respective crowd-feedback systems can just be used to evaluate IS during development or operations. To ensure continuous user involvement a crowd-feedback system that supports the development team during the whole lifecycle is necessary. This system could be combined with the crowd-feedback system configurator. Thereby, this system should guide the researcher in adapting the crowd-feedback systems' features to the specific requirements of the context and the lifecycle stage of the IS.

Limitations

We are aware that our literature review has limitations. Firstly, our results are highly dependent on the search string, the selected databases, and the chosen selection criteria. Our selections may induce a bias in the extracted literature and impact the identified research streams. The high number of studies that we identified via the backward and forward citations shows that our search string had some shortcomings. However, to reduce the probability of bias, we applied established methodological recommendations (i.e., Kitchenham and Charters 2007; Webster and Watson 2002). All decisions during the three stages of our literature review are made explicit. Secondly, for the interactivity cues, we considered only the most common ones. This restricts the holistic overview we aimed to provide of crowd-feedback systems as well as it might influence the clustering. Thirdly, we are aware that the cluster quality of the three research streams that we identified is rather low. Nevertheless, we decided to report the three research streams as they were consistent with the results of the manual clustering. Additionally, we want to guide future researchers who can revise the clusters in further studies.

Conclusion

Besides crowd-testing, crowd-feedback is a promising approach to scale the continuous evaluation of IS. As current research lacks a comprehensive overview of existing crowd-feedback systems, we aimed to structure and analyze existing literature with three main contributions: First, we provided an overview of existing crowd-feedback systems by conducting an SLR and identifying 40 relevant papers. Second, we proposed a morphological box for structuring crowd-feedback in IS development. Third, we identified three main research streams for crowd-feedback systems. Based on the insights gained by these three contributions, we finally highlighted avenues for feature research. We believe that our SLR can serve as a reference in the broader field of crowd-feedback systems and the dimensions that should be considered when researching such systems.

References

Akaike, H. 1998. "Information Theory and an Extension of the Maximum Likelihood Principle," in *Selected Papers of Hirotugu Akaike*, New York, NY, USA: Springer, pp. 199-213.

Almaliki, M., Ncube, C., and Ali, R. 2014. "The Design of Adaptive Acquisition of Users Feedback: An Empirical Study," in *2014 IEEE Eighth International Conference on Research Challenges in Information Science (RCIS)*, IEEE, pp. 1-12.

Alyahya, S. 2020. "Crowdsourced software testing: A systematic literature review," *Information and Software Technology* (127) (doi: 10.1016/j.infsof.2020.106363).

Ambreen, T., and Ikram, N. 2016. "A State-of-the-Art of Empirical Literature of Crowdsourcing in Computing," in *Proceedings - 11th IEEE International Conference on Global Software Engineering, ICGSE 2016*, IEEE, pp. 189-190.

Ayalon, O., and Toch, E. 2018. "Crowdsourcing Privacy Design Critique: An Empirical Evaluation of Framing Effects," in *Proceedings of the 51st Hawaii International Conference on System Sciences*.

Ayalon, O., and Toch, E. 2019. "A/P(rivacy) Testing: Assessing Applications for Social and Institutional Privacy," in *Extended Abstracts of the 2019 CHI Conference*, pp. 1-6.

Bandara, W., Furtmueller, E., Gorbacheva, E., Miskon, S., and Beekhuyzen, J. 2015. "Achieving Rigor in Literature Reviews: Insights from Qualitative Data Analysis and Tool-Support," *Communications of the Association for Information Systems* (37), pp. 154-204 (doi: 10.17705/1cais.03708).

Bodendorf, F., and Kaiser, C. 2009. "Detecting opinion leaders and trends in online social networks," in *Proceedings of the 2nd ACM workshop on Social web search and mining*, New York, NY, USA: ACM, pp. 65-68.

Brhel, M., Meth, H., Maedche, A., and Werder, K. 2015. "Exploring Principles of User-Centered Agile Software Development: A Literature Review," *Information and Software Technology* (61), pp. 163-181.

Chiu, T., Fang, D., Chen, J., Wang, Y., and Jeris, C. 2001. "A robust and scalable clustering algorithm for mixed type attributes in large database environment," in *Proceedings of the Seventh ACM SIGKDD International Conference on Knowledge Discovery and Data Mining / KDD-2001*, pp. 263-268.

Choi, Y., Monserrat, T. J. K., Park, J., Shin, H., Lee, N., and Kim, J. 2020. "ProtoChat: Supporting the Conversation Design Process with Crowd Feedback," *Proc. ACM Hum.-Comput. Interact.* (4:CSCW3) (doi: 10.1145/3432924).

Dow, S., Gerber, E., and Wong, A. 2013. "A Pilot Study of Using Crowds in the Classroom," in *CHI 2013: Changing Perspectives: Conference Proceedings of the 31st Annual CHI Conference on Human Factors in Computing Systems*, New York, NY, USA: ACM, pp. 227-236.

Easterday, M. W., Rees Lewis, D., and Gerber, E. M. 2017. "Designing Crowdcritique Systems for Formative Feedback," *International Journal of Artificial Intelligence in Education* (27:3), pp. 623-663 (doi: 10.1007/s40593-016-0125-9).

Feine, J., Morana, S., and Maedche, A. 2020. "Designing Interactive Chatbot Development Systems," in *International Conference on Information Systems*.

Gibbs, A. 1997. "Focus Groups," *Social Research Update* (19:8), pp. 1-8.

Greenberg, M. D., Easterday, M. W., and Gerber, E. M. 2015. "Critiki: A Scaffolded Approach to Gathering Design Feedback from Paid Crowdworkers," in *C & C - Proceedings of the 2015 ACM SIGCHI Conference on Creativity and Cognition*, New York, NY, USA: ACM, pp. 235-244.

Harris, M. A., and Weistroffer, H. R. 2009. "A New Look at the Relationship between User Involvement in Systems Development and System Success," *Communications of the Association for Information Systems* (24:1), pp. 739-756 (doi: 10.17705/1cais.02442).

Haukipuro, L., Pakanen, M., and Väinämö, S. 2016. "Online User Community for Efficient Citizen Participation," in *Proceedings of the 20th International Academic Mindtrek Conference*, New York, NY, USA: ACM, pp. 78-85.

Hosseini, M., Shahri, A., Phalp, K., and Ali, R. 2016. "Crowdsourcing transparency requirements through structured feedback and social adaptation," in *Proceedings of the International Conference on Research Challenges in Information Science*, IEEE.

Howe, J. 2009. *Crowdsourcing*: *Crowdsourcing: Why the Power of the Crowd is Driving the Future of Business*, New York, N.Y.: Three Rivers Press.

Hsu, J. S., Hung, Y. W., Chen, Y.-H., and Huang, H.-H. 2013. "Antecedents and Consequences of User Coproduction in Information System Development Projects," *Project Management Journal* (44:2), pp. 67-87 (doi: 10.1002/pmj.21330).

ISO 9241-210 2019. *Ergonomics of human-system interaction — Part 210: Human-centred design for interactive systems*, ISO/TC 159/SC 4 Ergonomics of human-system interaction.

Ives, B., and Olson, M. H. 1984. "User Involvement and MIS Success: A Review of Research," *Management Science* (30:5), pp. 586-603 (doi: 10.1287/mnsc.30.5.586).

Jansson, A. D., and Bremdal, B. A. 2018. "Genetic Algorithm for Adaptable Design using Crowdsourced Learning as Fitness Measure," in *2018 International Conference on Smart Systems and Technologies*, pp. 1-6.

Kang, H. B., Amoako, G., Sengupta, N., and Dow, S. P. 2018. "Paragon: An Online Gallery for Enhancing Design Feedback with Visual Examples," in *CHI 2018*: *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA: ACM, pp. 1-13.

Kitchenham, B., and Charters, S. 2007. "Guidelines for performing Systematic Literature Reviews in Software Engineering," in *EBSE Technical Report*.

Knaeble, M., Nadj, M., and Maedche, A. 2020. "Oracle or Teacher? A Systematic Overview of Research on Interactive Labeling for Machine Learning," in *WI2020*, GITO Verlag, pp. 2-16.

Krause, M., Garncarz, T., Song, J., Gerber, E. M., Bailey, B. P., and Dow, S. P. 2017. "Critique Style Guide: Improving Crowdsourced Design Feedback with a Natural Language Model," in *Proceedings of the Conference on Human Factors in Computing Systems*, New York, NY, USA: ACM, pp. 4627-4639.

La Cruz, G. V. de, Peng, B., Lasecki, W. S., and Taylor, M. E. 2015. "Towards Integrating Real-Time Crowd Advice with Reinforcement Learning," in *Proceedings of the 20th International Conference on Intelligent User Interfaces Companion*, New York, NY, USA: ACM, pp. 17-20.

Lakhani, K. R., Boudreau, K. J., Loh, P. R., Backstrom, L., Baldwin, C., Lonstein, E., Lydon, M., MacCormack, A., Arnaout, R. A., and Guinan, E. C. 2013. "Prize-based contests can provide solutions to computational biology problems," *Nature Biotechnology* (31), pp. 108-111 (doi: 10.1038/nbt.2495).

Lakhani, K. R., Garvin, D. A., and Lonstein, E. 2010. "TopCoder (A) Developing Software through Crowdsourcing," *Harvard Business School Case* (610-032), pp. 1-22.

Leicht, N. 2018. "Given Enough Eyeballs, all Bugs are Shallow - A Literature Review for the Use of Crowdsourcing in Software Testing," in *Proceedings of the 51st Hawaii International Conference on System Sciences*.

Lekschas, F., Ampanavos, S., Siangliulue, P., Pfister, H., and Gajos, K. Z. 2021. "Ask Me or Tell Me? Enhancing the Effectiveness of Crowdsourced Design Feedback," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA: ACM, pp. 1-12.

Luther, K., Pavel, A., Wu, W., Tolentino, J. L., Agrawala, M., Hartmann, B., and Dow, S. 2014. "CrowdCrit: Crowdsourcing and Aggregating Visual Design Critique," in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, New York, NY, USA: ACM.

Luther, K., Tolentino, J. L., Wu, W., Pavel, A., Bailey, B. P., Agrawala, M., Hartmann, B., and Dow, S. P. 2015. "Structuring, Aggregating, and Evaluating Crowdsourced Design Critique," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, New York, NY, USA: ACM, pp. 473-485.

Ma, X., Li, Y., Forlizzi, J., and Dow, S. 2015. "Exiting the Design Studio: Leveraging Online Participants for Early-Stage Design Feedback," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, New York, NY, USA: ACM, pp. 676-685.

Mao, K., Capra, L., Harman, M., and Jia, Y. 2017. "A survey of the use of crowdsourcing in software engineering," *Journal of Systems and Software* (126), pp. 57-84 (doi: 10.1016/j.jss.2016.09.015).

McKeen, J. D., and Guimaraes, T. 1997. "Successful Strategies for User Participation in Systems Development," *Journal of Management Information Systems* (14:2), pp. 133-150.
Mooi, E., and Sarstedt, M. 2011. *A Concise Guide to Market Research*: *The Process, Data, and Methods Using IBM SPSS Statistics*, Berlin: Springer.
Morschheuser, B., Hamari, J., Koivisto, J., and Maedche, A. 2017. "Gamified Crowdsourcing: Conceptualization, Literature Review, and Future Agenda," *International Journal of Human Computer Studies* (106), pp. 26-43 (doi: 10.1016/j.ijhcs.2017.04.005).
Muñante, D., Siena, A., Kifetew, F. M., Susi, A., Stade, M., and Seyff, N. 2017. "Gathering Requirements for Software Configuration from the Crowd," in *2017 IEEE 25th International Requirements Engineering Conference Workshops (REW)*, pp. 176-181.
Nebeling, M., Speicher, M., and Norrie, M. C. 2013. "CrowdStudy: General Toolkit for Crowdsourced Evaluation of Web Interfaces," in *EICS '13: Proceedings of the ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, New York, NY, USA: ACM, p. 255.
Nickerson, R. C., Varshney, U., and Muntermann, J. 2013. "A method for taxonomy development and its application in information systems," *European Journal of Information Systems* (22), pp. 336-359 (doi: 10.1057/ejis.2012.26).
Nielsen, J. 1994. *Usability Engineering*, San Francisco, CA, USA: Morgan Kaufmann.
Oppenlaender, J., and Hosio, S. 2019. "Towards Eliciting Feedback for Artworks on Public Displays," in *C&C' 19*: *Proceedings of the 2019 Creativity and Cognition*, New York, NY, USA: ACM.
Oppenlaender, J., Kuosmanen, E., Lucero, A., and Hosio, S. 2021. "Hardhats and Bungaloos: Comparing Crowdsourced Design Feedback with Peer Design Feedback in the Classroom," in *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA: ACM.
Oppenlaender, J., Tiropanis, T., and Hosio, S. 2020. "CrowdUI: Supporting Web Design with the Crowd," *Proceedings of the ACM on Human-Computer Interaction* (4) (doi: 10.1145/3394978).
Pagano, D., and Bruegge, B. 2013. "User involvement in software evolution practice: A case study," in *35th International Conference on Software Engineering (ICSE)*, IEEE, pp. 953-962.
Pedersen, J., Kocsis, D., Tripathi, A., Tarrell, A., Weerakoon, A., Tahmasbi, N., Xiong, J., Deng, W., Oh, O., and Vreede, G. J. de 2013. "Conceptual Foundations of Crowdsourcing: A Review of IS Research," in *Proceedings of the 46th Hawaii International Conference on System Sciences*, IEEE, pp. 579-588.
Pretel, I., Lopez-Novoa, U., Sanz-Yagüe, E., López-de-Ipiña, D., Cartelli, V., Di Modica, G., and Tomarchio, O. 2017. "Citizenpedia: A human computation framework for the e-government domain," in *2017 IEEE SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI*, IEEE, pp. 1-6.
Rissler, R., Nadj, M., Adam, M., and Maedche, A. 2017. "Towards an Integrative Theoretical Framework of IT-Mediated Interruptions," in *Proceedings of the 25th European Conference on Information Systems (ECIS)*.
Robb, D. A., Padilla, S., Kalkreuter, B., and Chantler, M. J. 2015a. "Crowdsourced Feedback With Imagery Rather Than Text: Would Designers Use It?" in *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, New York, NY, USA: ACM, pp. 1355-1364.
Robb, D. A., Padilla, S., Kalkreuter, B., and Chantler, M. J. 2015b. "Moodsource: Enabling Perceptual and Emotional Feedback from Crowds," in *Proceedings of the 18th ACM Conference Companion on Computer Supported Cooperative Work & Social Computing*, New York, NY, USA: ACM, pp. 21-24.
Robb, D. A., Padilla, S., Methven, T. S., Kalkreuter, B., and Chantler, M. J. 2017. "Image-based Emotion Feedback: How Does the Crowd Feel? And Why?" in *DIS 2017 - Proceedings of the 2017 Conference on Designing Interactive Systems*, New York, NY, USA: ACM, pp. 451-463.
Sarı, A., Tosun, A., and Alptekin, G. I. 2019. "A systematic literature review on crowdsourcing in software engineering," *Journal of Systems and Software* (153), pp. 200-219 (doi: 10.1016/j.jss.2019.04.027).
Schneider, H., Frison, K., Wagner, J., and Butz, A. 2016. "CrowdUX: A Case for Using Widespread and Lightweight Tools in the Quest for UX," in *DIS 2016 - Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, New York, NY, USA: ACM, pp. 415-425.
Scholtz, J. 2001. "Adaptation of traditional usability testing methods for remote testing," in *Proceedings of the 34th Hawaii International Conference on System Sciences*, IEEE, p. 134.
Scriven, M. 1991. "Beyond Formative and Summative Evaluation," *Teachers College Record* (92:6), pp. 18-64.
Seyff, N., Ollmann, G., and Bortenschlager, M. 2014. "AppEcho: a user-driven, in situ feedback approach for mobile platforms and applications," in *Proceedings of the 1st International Conference on Mobile Software Engineering and Systems*, New York, NY, USA: ACM, pp. 99-108.

Snijders, R., Dalpiaz, F., Brinkkemper, S., Hosseini, M., Ali, R., and Ozum, A. 2015. "REfine: A Gamified Platform for Participatory Requirements Engineering," in *2015 IEEE 1st International Workshop on Crowd-Based Requirements Engineering (CrowdRE)*, IEEE, pp. 1-6.
Stade, M., Oriol, M., Cabrera, O., Fotrousi, F., Schaniel, R., Seyff, N., and Schmidt, O. 2017. "Providing a User Forum is not enough: First Experiences of a Software Company with CrowdRE," in *2017 IEEE 25th International Requirements Engineering Conference Workshops (REW)*, pp. 164-169.
Theodoridis, S., and Koutroumbas, K. 2009. "Pattern Recognition," *IEEE Transactions on Neural Networks* (2).
Tizard, J., Rietz, T., Liu, X., and Blincoe, K. 2021. "Voice of the users: an extended study of software feedback engagement," *Requirements Engineering*, pp. 1-23 (doi: 10.1007/s00766-021-00357-1).
Vredenburg, K., Mao, J. Y., Smith, P. W., and Carey, T. 2002. "A Survey of User-Centered Design Practice," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, New York, NY, USA: ACM, pp. 471-478.
Wauck, H., Yen, Y.-C., Fu, W.-T., Gerber, E., Dow, S. P., and Bailey, B. P. 2017. "From in the Class or in the Wild?" in *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, New York, NY, USA: ACM, pp. 5580-5591.
Webster, J., and Watson, R. 2002. "Analyzing the Past to Prepare for the Future: Writing a Literature Review," *MISQ* (26:2).
Wolfswinkel, J. F., Furtmueller, E., and Wilderom, C. P. 2013. "Using Grounded Theory as a Method for Rigorously Reviewing Literature," *European Journal of Information Systems* (22) (doi: 10.1057/ejis.2011.51).
Wu, Y. W., and Bailey, B. P. 2016. "Novices Who Focused or Experts Who Didn't?" in *Proceedings of the 2016 Conference on Human Factors in Computing Systems*, New York, NY, USA: ACM, pp. 4086-4097.
Wu, Y. W., and Bailey, B. P. 2021. "Better Feedback from Nicer People: Narrative Empathy and Ingroup Framing Improve Feedback Exchange," *Proc. ACM Hum.-Comput. Interact.* (4:CSCW3), pp. 1-22 (doi: 10.1145/3432935).
Xu, A., and Bailey, B. P. 2011. "A Crowdsourcing Model for Receiving Design Critique," in *CHI '11 Extended Abstracts on Human Factors in Computing Systems*, New York, NY, USA: ACM, pp. 1183-1188.
Xu, A., and Bailey, B. P. 2014. "A System for Receiving Crowd Feedback on Visual Designs Abstract," in *Proceedings of the Companion Publication of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, New York, NY, USA: ACM, pp. 37-40.
Xu, A., Huang, S.-W., and Bailey, B. 2014. "Voyant: Generating Structured Feedback on Visual Designs Using a Crowd of Non-Experts," in *Proceedings of the 17th ACM Conference on Computer Supported Cooperative Work & Social Computing*, New York, NY, USA: ACM, pp. 1433-1444.
Xu, A., Rao, H., Dow, S. P., and Bailey, B. P. 2015. "A Classroom Study of Using Crowd Feedback in the Iterative Design Process," in *Proceedings of the 18th ACM Conference on Computer Supported Cooperative Work & Social Computing*, New York, NY, USA: ACM, pp. 1637-1648.
Yen, Y. C. G., Dow, S. P., Gerber, E., and Bailey, B. P. 2016. "Social Network, Web Forum, or Task Market? Comparing Different Crowd Genres for Design Feedback Exchange," in *DIS 2016 - Proceedings of the 2016 ACM Conference on Designing Interactive Systems*, New York, NY, USA: ACM, pp. 773-784.
Yen, Y.-C. G., Dow, S. P., Gerber, E., and Bailey, B. P. 2017. "Listen to Others, Listen to Yourself: Combining Feedback Review and Reflection to Improve Iterative Design," in *Proceedings of the 2017 ACM SIGCHI Conference on Creativity and Cognition*, New York, NY, USA: ACM, pp. 158-170.
Yu, Z., Xu, Z., Black, A., and Rudnicky, A. I. 2016. "Chatbot Evaluation and Database Expansion via Crowdsourcing," in *RE-WOCHAT: Workshop on Collecting and Generating Resources for Chatbots and Conversational Agents-Development and Evaluation Workshop Programme*.
Yuan, A., Luther, K., Krause, M., Vennix, S. I., Dow, S. P., and Hartmann, B. 2016. "Almost an Expert: The Effects of Rubrics and Expertise on Perceived Value of Crowdsourced Design Critiques," in *Proceedings of the 19th ACM Conference on Computer-Supported Cooperative Work & Social Computing*, New York, NY, USA: ACM, pp. 1005-1017.
Zuchowski, O., Posegga, O., Schlagwein, D., and Fischbach, K. 2016. "Internal crowdsourcing: Conceptual framework, structured review, and research agenda," *Journal of Information Technology* (31:2), pp. 166-184 (doi: 10.1057/jit.2016.14).
Zwicky, F., and Wilson, A. G. 1967. *New Methods of Thought and Procedure*, Berlin: Springer.

Appendix

		Input							Crowd Configuration									Design Characteristics									Effects		
		IS Lifecycle Stage		Type of Feedback		Feedback Scope			Crowd Type				Incentive					Feedback Collection Mechanisms					Interactivity Cues						
Stream	Author	Development	Operations	Qualitative	Quantitative	Non-functional Attributes	Content	Functional Attributes	Anonymous	Proxy Users	Students	Convenience	Money	Improvement/Involvement	Interest/Social Compensation	Course Credit	Gamification	Questionnaire	Free Text Field	Categories	Selection	Direct Manipulation	Collaboration	Marker	Context	Recording	Outcome Effects	Process Effects	Intermediate Effects
1 – Anonymous Crowd-Feedback	Greenberg et al. 2015	X		X		X	X		X				X					X									X	X	
	Kang et al. 2018	X		X		X	X		X				X							X							X		X
	Krause et al. 2017	X		X		X	X		X				X						X								X		
	Lekschas et al. 2021	X		X		X	X		X				X		X					X							X		
	Luther et al. 2014; Luther et al. 2015; Yuan et al. 2016	X		X		X			X				X							X				X			X		
	Wauck et al. 2017	X		X		X	X		X		X	X	X		X	X			X								X	X	
	Wu and Bailey 2016	X		X		X	X		X				X		X				X						X		X		
	Wu and Bailey 2021	X	X	X		X	X		X				X						X					X			X		
	Xu et al. 2014; Xu et al. 2015; Xu and Bailey 2014	X		X	X	X	X		X				X							X				X			X		
	Yen et al. 2016	X		X		X	X		X			X	X		X				X									X	
	Yen et al. 2017	X		X		X			X				X							X							X		
2 – Real User Crowd-Feedback	Easterday et al. 2017	X		X	X		X	X		X										X							X	X	
	Haukipuro et al. 2016	X		X	X			X		X				X				X					X					X	
	Muñante et al. 2017	X		X	X			X		X								X							X		X		
	Nebeling et al. 2013	X	X	X	X	X	X	X	X	X			X	X				X					X				X	X	
	Oppenlaender et al. 2020		X	X	X	X				X				X								X	X				X	X	
	Pretel et al. 2017		X	X				X		X				X			X		X				X						
	Schneider et al. 2016	X		X	X	X	X	X		X			X							X				X			X	X	
	Seyff et al. 2014		X	X				X		X				X					X					X		X	X		
	Snijders et al. 2015	X	X	X	X			X		X				X			X		X				X				X		X
	Stade et al. 2017		X	X	X	X	X	X		X				X						X			X	X					
3 – Hybrid Crowd-Feedback	Ayalon and Toch 2018		X		X	X			X				X					X							X		X		
	Ayalon and Toch 2019	X			X	X			X				X					X							X			X	
	Choi et al. 2020	X		X			X	X	X				X								X	X					X	X	
	La Cruz et al. 2015	X	X		X			X	X				X								X						X	X	
	Dow et al. 2013	X		X		X		X		X			X					X								X		X	
	Hosseini et al. 2016		X		X	X				X				X							X							X	X
	Jansson and Bremdal 2018	X			X	X															X							X	X
	Ma et al. 2015	X		X		X		X	X				X					X								X		X	
	Oppenlaender and Hosio 2019	X		X	X	X				X					X						X			X		X			X
	Oppenlaender et al. 2021	X		X	X	X	X	X	X		X		X			X		X										X	
	Robb et al. 2015a, 2015b; Robb et al. 2017	X		X		X	X		X		X		X			X					X						X		X
	Xu and Bailey 2011	X		X		X			X									X					X						
	Yu et al. 2016	X		X	X		X	X	X				X								X						X		X
	Σ	28	10	29	17	25	17	15	21	13	3	2	22	8	5	3	2	10	8	8	7	2	7	7	4	4	22	16	7
	Percentage	78%	28%	81%	47%	68%	47%	42%	58%	36%	8%	6%	61%	22%	14%	8%	6%	28%	22%	22%	19%	6%	19%	19%	11%	11%	61%	44%	19%

Appendix Table 1. Concept Matrix for Identified Articles Targeting Crowd-Feeedback Systems