

AUTHENTIC AND CONCURRENT EVALUATION – REFINING AN EVALUATION APPROACH IN DESIGN SCIENCE RESEARCH

Carina Ihlström Eriksson, School of Information Science, Computer and Electrical Engineering, Halmstad University, Sweden, carina.ihlstrom_eriksson@hh.se

Maria Åkesson, School of Information Science, Computer and Electrical Engineering, Halmstad University, Sweden, maria.akesson@hh.se

Karlheinz Kautz, School of Information Science, Computer and Electrical Engineering, Halmstad University, Sweden and Department of Informatics, Copenhagen Business School, Denmark, karl.kautz@cbs.dk

Abstract

This paper addresses the need for more detailed accounts for evaluation in design science research literature. By revisiting a design project regarding the future e-newspaper we give detailed descriptions of its authentic and concurrent evaluation approach by illustrating the what, why and how of all evaluation activities throughout the whole project. The project produced seven different design artifacts that were evaluated. The utility and theoretical outcomes of the evaluation activities clearly influenced design decisions regarding newspaper design, user value and business model design as well as decisions on strategic levels. We emphasize a holistic and concurrent approach to evaluation compared to the general design science research thinking and argue that reflecting on how to seek authenticity is important. By authenticity we refer to the notion of how closely an evaluation captures the context and actual use of an artifact. With the holistic approach we encourage evaluation to be inclusive of different aspects and relationships between stakeholder groups in the evaluation activities. Further we think that concurrency is not narrowed to evaluation but also regards theorizing. While it makes sense for planning to distinguish between phases and stages of evaluation and theorizing, in practice they are intrinsically interlinked and concurrent.

Keywords: Evaluation, Authentic, Concurrent, Design Science.

1 INTRODUCTION

Design science research in Information systems (IS) seeks to make theoretical contributions as well as contributing with relevant design artifacts. In this ambition, design science research literature put a lot of emphasis on evaluation to gather evidence of design theory and the utility of design artifacts. This emphasis on evaluation is reflected in reviewing premier literature in the IS field on design science research (see Venable 2006; Baskerville *et al.* 2007; Pries-Heje *et al.* 2008).

Evaluation is accentuated both as specific phases in design research cycles (Hevner *et al.* 2004; Baskerville *et al.* 2007) and as a concurrent activity following the entire design research process (Sein *et al.* 2011). Yet, there is neither much guidance on strategies to choose evaluation methods, nor on how to perform the actual evaluations to meet the requirements of rigorous and relevant design science research. Rather, the reporting on evaluation in design science literature concerns the purpose of evaluation and the results of evaluation.

We argue that there is a need for more detailed accounts for evaluation in design science research literature. In this paper we revisit a design project and give a detailed description of its evaluation approach. The project explored the design of the future e-newspaper, i.e. a newspaper published on e-paper technology. Inspired by Sein *et al.* (2011) we demonstrate the concurrent evaluation activities throughout the whole project and how authenticity increased as the design progressed. We will describe which evaluation methods were used in different phases of the design process, the rationale behind the choices of evaluation methods as well as how the chosen methods were used and the outcomes of the evaluations.

This paper proceeds as follows. In the next section we present a literature review on how evaluation is treated in design science literature in IS. Next we give an overview of the project setting. Thereafter, we demonstrate the concurrent and authentic evaluation approach applied in design project. The paper is concluded with a discussion of the evaluation approach and its implications for design science research.

2 LITERATURE REVIEW

Evaluation together with building is one of the two defining activities of design science research and its importance is unanimously acknowledged to appraise design theory (Walls *et al.* 1995), assess IT artifacts such as constructs, models, methods, and instantiations (March & Smith 1995), to demonstrate the utility, quality, and efficacy of the IT artifact (Hevner *et al.* 2004) and to provide feedback to the build phase of the IT artifact (Hevner *et al.* 2004; Vaishnavi & Kuechler 2004; Peffers *et al.* 2007; Kuechler & Vaishnavi 2008). The following review of the literature on the topic of evaluation in design science research is based on Baskerville *et al.* (2007) and Pries-Heje *et al.* (2008) who have provided two very useful summaries on the subject matter.

In the conventional literature the design science research process is described as iterative, incremental and/or cyclic (Hevner *et al.* 2004; Vaishnavi & Kuechler 2004; Peffers *et al.* 2007; Hevner 2007; Kuechler & Vaishnavi 2008). The research activities are organized sequentially in stage-gate models with evaluation as a separate stage or activity. In one of these models, as a variation and an explicit link between building and evaluation Peffers *et al.* (2007) in addition distinguish between demonstration and formal evaluation. The emphasis for evaluation is on empirical testing of falsifiable hypotheses (Walls *et al.* 1992; March & Smith 1995; Hevner *et al.* 2004) to provide evidence that the developed theory leads to IT artifacts which ‘work’ and yield the benefits which were anticipated.

Walls *et al.* (1992) who introduced the concept of information systems design theory were the first to demand that testable hypotheses have to be developed and subjected to empirical refutation by some combination of empirical investigation and mathematical proof. Walls *et al.* (1992) assert that a

design theory can only be validated by conducting experiments with the information system constructed according to the developed theory's principles. They further propose a two phase testing procedure where performance of experimental groups using various features of the novel IT artifact is compared with the performance of groups who are not using those features. Walls *et al.* (1992) provide some examples for information systems design theories, but otherwise they provide little to no guidance on how to evaluate and test developed hypothesis.

March and Smith's (1995) emphasis on evaluation regards the development of criteria and the assessment of the artifact's performance in comparison to these criteria. They postulate that beyond establishing that an artifact works or not, evaluation also has to determine how and why it works or not. For this purpose they in general terms propose to use natural science methods.

Hevner *et al.* (2004) also put forward that the design artifact must be rigorously examined via well-executed evaluation methods. They assume that building the artifact is based on a well defined and well structured problem for which the business environment has established clear requirements on which the artifact evaluation should be based. Hevner *et al.* (2004) provide seven guidelines for design science research and summarize in one of them five kinds of evaluation methods: observational including case and field studies, analytical, experimental, testing, and descriptive. They demonstrate their design principles through three cases from the literature where evaluation is performed either as formal proof (Gavish & Gerdes 1998); in a descriptive and analytical manner (Aalst & Kumar 2003) or observational throughout the whole development process of the artifact (Markus *et al.* 2002). The presentations of the evaluation activities are brief and Hevner *et al.* (2004) do not provide much guidance for choosing among the evaluation methods and how to perform the actual evaluations.

On this basis Venable (2006) classified design science research evaluation approaches into two primary forms: artificial and naturalistic evaluation. He argues that artificial evaluation assesses a design solution in a contrived and nonrealistic way and puts forward that artificial evaluation nearly always is reductionist and used to test design hypotheses as most prominently proposed by Walls *et al.* (1992). Artificial evaluation includes laboratory experiments, field experiments, simulations, criteria-based analysis, theoretical arguments, and mathematical proofs. Approaches such as observations, case or field studies may also be used, but generally these only supplement the main goal of proving or disproving the design theory and the utility of the artifact. As they are derived in a quixotic environment and may not deal with realistic requirements and the complexities of organizational life, artificial evaluation results might not be applicable for real use and thus necessitate naturalistic evaluation which embraces the complexities of human practice and explores the performance of an IT artifact in its real environment, i.e. within an organization.

Naturalistic evaluation is always empirical and includes case studies, field studies, surveys, ethnography, phenomenology, hermeneutic methods, and action research. As IT artifacts are placed in human activity settings, there may be multiple, conflicting accounts of the 'truth' regarding any evaluation of the artifact. Naturalistic evaluation may be difficult and costly, because it has to take into account the effects of many interrelated variables in the real world. Pries-Heje *et al.* (2008) also acknowledge that naturalistic evaluation is affected by confounding variables or misinterpretation, and thus its results may not be precise or truthful about an artifact's utility or efficacy in real use. Venable (2006) still considers naturalistic evaluation as the definitive approach to gather conclusive evidence concerning design theory and the utility of a design artifact.

Hence, Baskerville *et al.* (2007) argue that in a naturalistic setting, the evaluation process mirrors information systems development, with the activities analysis, design, building and evaluating, but based on the particular and explicit design theory to be tested and justified. Accordingly they demand a 'soft' design science research process which takes into account the social and organizational issues relating to problem formulation, the conduct of information systems and technology development, the adoption of such information systems, and the evaluation of these IT artifacts in use in natural organizational settings. The approach recognizes that all activities contribute to evaluation, but still has a separate evaluation phase.

Pries-Heje *et al.* (2008) conclude that the discussion of evaluation activities and methods in the literature is limited and typically assumes an ex post perspective, in which evaluation occurs only after the construction of an IT artifact. They put forward that evaluation can also take place prior to construction of the IT artifact based on the design specification alone. This however assumes that a design specification is produced. Their framework encompasses both ex ante and ex post orientations as well as naturalistic and artificial settings for design science research evaluation. The framework proposed offers a strategic view on design science research evaluation and while it makes up for the omission in the literature to provide guidance for the choice of strategies and overall approaches for evaluation in design science research, it gives little directions about how to perform evaluations in more detail.

Hardless *et al.* (2007) provide an example of a soft design science research approach applying naturalistic evaluation. In two organizations they developed three prototypes for IT-mediated learning systems in two action research cycles. In the building phase they used participatory design workshops which included initial prototype reviews and the evaluation of the final prototypes was carried out in the form of surveys, interviews, follow-up meetings as well as documents reviews. While the focus with regard to evaluation in this work is very much on the evaluation results, it provides some detail about the evaluation process and clearly demonstrates the close relationship of evaluation to the building activities.

Sein *et al.* (2011) follow up on this work and an earlier published action research project by Lindgren *et al.* (2004) and argue that building and evaluating an IT artifact are inseparable and inherently interwoven activities which take place concurrently. IT artifacts emerge, are shaped and are continuously refined in the organizational context during development and use. Sein *et al.* (2011) therefore propose a method in which theory building is inherently linked to intervention in organizations. They name this approach action design research (ADR).

Their research method consists of four stages – problem formulation; building, intervention, and evaluation; reflection and learning; formalization of learning - and accompanying 7 principles to provide guidance and rigor. In distinction to action research where evaluation is a separate phase, the second stage of action design research - building, intervention, evaluation - interweaves the building of the IT artifact, the intervention in the organization and its work practices, and evaluation. The accompanying principles comprise the acknowledgment of the inseparable influences which are mutually exerted by the IT artifact and by the organizational context and the mutual influence and learning among the different stakeholder groups, in particular the researchers and practitioners.

To underline the contrast between stage-gate and concurrent evaluation as well as between controlled artificial, and naturalistic authentic evaluation Sein *et al.* (2011) introduce a principle called authentic and concurrent evaluation which emphasizes a key characteristic of action design research: evaluation is not a separate stage of the research process that follows building and takes place in a laboratory environment. As controlled evaluation can be difficult to achieve in an action design research project due to the emergent nature of the artifact, Sein *et al.* (2011) propose that evaluation opportunities are to be sought following natural controls of the project where possible; authenticity is however more important than control. The specific format of evaluation may vary, f. ex. evaluation cycles for an alpha version of an IT prototype can be formative, contributing to the refinement of the artifact and surfacing anticipated as well as unanticipated consequences. Later evaluations of beta versions can be summative, assessing value and utility outcomes. Sein *et al.* (2011) demonstrate their approach by re-conceptualizing the above mentioned action research project which developed design principles for competence management systems (Lindgren *et al.* 2004), but only provide a very short account of the actual evaluation process and activities.

Authenticity in evaluation is primarily addressed in the field of education and medicine. Lincoln and Guba (1986), f. ex. discuss authenticity in naturalistic evaluations and suggest five unique criterions for authenticity; Fairness, Ontological Authentication, Educative Authentication, Catalytic Authentication and Tactical Authentication. Regarding fairness it is essential to expose conflicting constructions and value structures. Ontological authentication concerns applying the conscious experience of the world from an individual or group perspective, and educative authentication implies

the need to increase the knowledge among stakeholders to understand various constructions. The feed-back validity is addressed in the catalytic authentication and empowerment of stakeholders is related to tactical authentication.

In general, only brief reports mostly about the evaluation results and little information about the actual evaluation process are documented in the literature. We ourselves have also described the evaluation of design principles for an IT-artifact in an action research project with a focus on the evaluation outcome as a phase although evaluation took place throughout the whole project (Åkesson *et al.* 2010). Inspired by Sein *et al.* (2011) we revisited our project and are now able to describe a more fine-grained evaluation process. In the following we will therefore demonstrate an authentic and current approach to design science research evaluation and provide a more detailed account of an evaluation process.

3 THE DIGINEWS PROJECT

The DigiNews project was a design science research endeavor initiated by the Swedish Newspaper Publishers' Association and a device producer. The project group consisted of the project leader from Philips Applied Technologies, researchers and newspaper representatives from several countries in Europe, as well as a few tech companies. The evaluation activities presented in this paper was conducted by the research team from Halmstad University.

In the project the research team cooperated with Swedish newspaper publishers including their management, their editor-in-chiefs, their newspaper layout designers and their web designers, as well as newspaper readers and newspaper advertisers. Two different focus groups worked closely with the researchers during the whole project. The first focus group consisted of six readers with different gender and ages, and the second was a design focus group consisting of four newspaper designers. The project is an example of soft design science research (Baskerville *et al.* 2007). As it also intervened in both the newspaper publishers' organizations and their production processes as well as in the newspaper readers' daily reading activities in their homes, it can be understood as action design science, too.

The project developed design visions for the future e-newspaper, i.e. the future newspaper published on e-paper technologies (Ihlström *et al.* 2004; Ihlström *et al.* 2005). E-paper is the common term for several different technologies that can be used to produce screens with a number of specific characteristics, of which many have properties that can be compared with print on paper, making it especially interesting for newspapers. Compared to transmissive displays such as liquid crystal display (LCD) and thin film transistor (TFT) displays, e-paper is a reflective display technology, i.e. it has no back light; it gives the same reading experience as paper such as high contrast and the possibility to read in sunlight.

Current e-newspapers are presented on eReader devices such as the Sony Reader and Amazon Kindle in the format of a book rather than a newspaper. Content is downloadable online and updated in regular intervals as editions. The existing e-newspapers are portable and mobile, contain very few pictures and advertisement presented on a grayscale, provide no interactivity, but some search and audio possibilities, and are not yet bendable. Both bendable and color e-paper exist in the laboratories and are expected to reach the market soon.

The DigiNews project roughly followed a 5 phase design science research approach developed for the design of a mass-market consumer IT artifact (Åkesson *et al.* 2010) where the first phases of exploring the design challenge of the unexplored e-newspaper concept and deriving design principles for the e-newspaper as an IT artifact destined to be used beyond specific organizational boundaries correspond to the problem formulation stage of the original ADR approach (Sein *et al.* 2011). The next two phases of building and evaluation correspond to the building, intervention, and evaluation stage in ADR, and the theorizing phase matches up with the reflection and learning as well as the formalization of learning stages in ADR.

The project produced 7 different (types of) design artifacts (A1-A7), namely the e-newspaper concept itself, mock-ups, scenarios, navigation prototypes, functional prototypes, movie illustrations, and an eReader prototype which were all evaluated with a naturalistic evaluation approach including the following methods and techniques: future workshops, focus groups, interviews, surveys, laboratory and real life tests and assessments. In Figure 1 the seven design artifacts are placed in relation to the first four phases of the design science research approach.

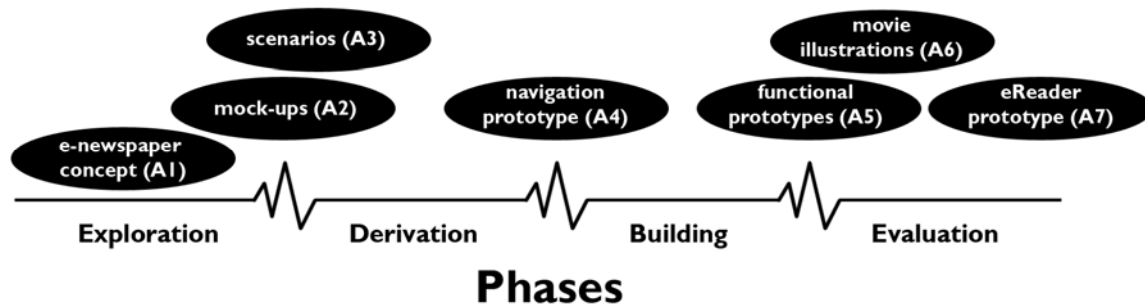


Figure 1. Design artifacts vs. project phases

The assessments were not limited to an evaluation phase or a building, intervention and evaluation stage, but took place from the very beginning of the project in the exploration phase, ex ante already prior to the construction of executable IT artifacts (Pries-Heje *et al.* 2008) and continued in a concurrent manner throughout the whole project.

4 AN AUTHENTIC AND CONCURRENT EVALUATION APPROACH

The concurrent and naturalistic evaluations exhibited an increasing degree of authenticity starting from stakeholder involvement and negotiations in workshops and focus groups of concepts and principles ranging to the construction of prototypes by newspaper publishers and prototype assessments in readers' homes.

In table 1, the evaluation approach is summarized. In the following we present in more detail the what, why and how of all evaluation activities throughout the whole project. The researchers acted as facilitators and conducted all the evaluations.

Design artifact	Designed by	Evaluated by	Evaluation method	Utility outcome	Theoretical outcome
E-newspaper concept (A1)	Readers, newspaper publishers, device producer, advertisers	Readers, newspaper publishers, advertisers	Future workshops, questionnaires, interviews, focus groups	Concept articulation Inspiration to mock-ups, scenarios and business concepts	Design challenge conceptualization Model of a media multi-channel environment
Mock-ups (A2)	Readers, newspaper publishers	Interaction designers, design focus group, advertisers, device producer	Workshops, focus groups	User requirements Input to prototypes and movie illustrations	Design approach for future media
Scenarios (A3)	Readers, newspaper publishers	Interaction designers, DigiNews project group	Workshops	User requirements Refinement of concept Inspiration to prototypes and movie illustrations	
Navigation prototype (A4)	Students	Readers, focus group of readers, design focus group	Focus groups, usability tests	User requirements Input to navigation support in functional prototypes	E-newspaper navigation concept
Functional prototypes (A5)	Design focus group	Readers, advertisers, newspaper publishers	Laboratory tests, interviews, workshops, questionnaires	Refinement of concept Reader preferences Design implications	Design principles for calm e-reading. Initial target group model Business model framework Mobile service value model
Movie illustrations (A6)	Interaction designers	Readers, newspaper publishers, advertisers	Questionnaires, interviews, workshops	Proof of concept Input to business model and diffusion	
eReader prototype (A7)	Newspaper publisher	Readers, newspaper publishers	Two week real life tests, including 2 questionnaires + daily online activities + 60 min interview	Final proof of concept	Key user values of e-newspapers Proof of concept of the design principles

Table 1. Evaluation approach

E-newspaper concept

Several different stakeholders, such as readers, newspaper publishers, device producer and advertisers were engaged to explore and articulate the e-newspaper concept. To evaluate the concept, a broad approach was chosen to get as much input as possible for the forthcoming activities in the project due to the explorative nature of the design problem. As the e-newspaper concept started to take form it was evaluated with the focus group of readers and the design focus group, in future workshops with newspaper publishers and readers, and in interviews with newspaper management and advertisers as well as through questionnaires to readers and newspaper designers.

In these evaluations the e-newspaper concept was appraised by identifying pros and cons related to print and online newspapers and questions like “Which are the advantages with the print newspaper that you find essential to consider when designing the e-newspaper?”, “What aspects do you consider most important regarding the design of the e-newspaper?” and “What is the e-newspapers role in relation to other publication channels such as print and web?” were asked.

These evaluations resulted in a concept articulation of the e-newspaper as a convergence between the printed and the online newspaper, where the layout from the printed newspaper were preferred in combination with functionality from the online newspaper such as updated news and navigation support. Business model related aspects were also evaluated such as the demand for several newspaper brands and distributional aspects. These evaluation outcomes later worked as inspiration to

mock-ups, scenarios and business concepts and the concept which was concurrently evaluated and refined throughout the project. This phase also resulted in theoretical outcomes such as conceptualization of the design challenge, and a model of a multi-channel environment.

Mock-ups

To visualize and make the e-newspaper concept more authentic readers and newspaper publishers were asked to make mock-ups of their idea of a future e-newspaper in future workshops. The mock-ups functioned to communicate design ideas, and the workshop participants were provided with a variety of material such as paper in different sizes and colors, different pens, newspaper clips and printouts from online newspapers, tape in different colors etc. The mock-ups represented ideas of the device design as well as interface, navigation, and newspaper layout and advertising. No limitations were set for size, orientation, navigation support etc. These workshops were concluded with each participant presenting their self assessment of their mock-up and discussed in the group. These presentations were video-recorded.

The mock-ups were evaluated in different ways in order to provide as rich data for later activities as possible. The first evaluation was done with the design focus group together with interaction designers (from an external firm that produced the movie illustrations described below) where the mock-ups and video presentations were studied. The primary goal of this evaluation was to identify new innovative ideas and interesting design solutions. In the second evaluation the mock-ups were evaluated from an advertising point-of-view in a workshop with advertisers. The mock-ups were thereafter evaluated by the device producer at a project meeting, this evaluation focused mainly on the device design. Finally, the mock-ups were evaluated and analyzed by researchers to extract user requirements. This analysis resulted in a compiled list of user requirements which was used as input for prototypes, movie illustrations, and the e-Reader device. Furthermore, as a theoretical output a design approach for future media was suggested.

Scenarios

Scenarios techniques were used to capture ideas of future use cases from different perspectives from readers and newspaper publishers in future workshops. The rationale behind creating and evaluating scenarios was to demonstrate the future visions of the e-newspaper use realistically.

These scenarios were evaluated in two different workshops, one with the DigiNews project group and one with interaction designers. In these workshops the scenarios were analyzed by first categorizing according to different locations such as at home, commuting, at the office, travelling etc, and thereafter to different situations. For example, in one scenario an elderly couple was spending their winters in Spain and their e-newspaper consisted of local news from Sweden in combination with the Spanish TV schedule. Through these scenarios many design related aspects were illuminated e.g. position based services etc which added to the user requirements and refinement of the e-newspaper concept that were input to the prototype design. Finally, three new scenarios were constructed that compiled most of the aspects identified. These three scenarios were later refined and form the base for the movie illustrations. The scenarios were also a part of the design approach for future media mentioned above.

Navigation prototype

As an outcome of the evaluations of the three previously described design artifacts navigation design was identified early in the project as a key critical success factor by all stakeholders. Therefore it was of interest to explore this aspect further with prototypes illustrating different navigational models. As the eReader devices were not yet available in the project the prototypes were implemented in tablet PCs with authentic newspaper content by master students in informatics.

The prototypes and navigation models were evaluated in a usability test with 10 readers and later by the reader focus groups. Each test was divided into the two sessions which lasted about 30 minutes each. In the first session the readers were given some time to get acquainted with the prototypes and then given a few task-based questions concerning navigating to a specified location. In the second session the readers were interacted with an altered version of the prototype were some navigation

support had been removed. The reader focus group that had followed the progression in the project also tested their prototypes in a similar way. This evaluation resulted in different requirements for the device and the screen layout, e.g. hardware buttons and software buttons, indexing, hyperlinks etc. The prototypes were also evaluated by the design focus group. These evaluations contributed to user requirements and also worked as inspiration for the functional prototypes designed by the design focus group. The theoretical output was an e-newspaper navigation concept.

Functional prototypes

Based on the results and requirements compiled from the previous evaluations, three functional prototypes were designed by the newspaper designers in the design focus group and the researchers in collaboration with the aim to design an authentic prototype of the e-newspaper. As eReaders still were not available in the project, tablet PCs were used to present the prototypes. These prototypes were evaluated in a laboratory test with 36 users as well as in a questionnaire to readers which is described below. Thereafter we also evaluated them in workshops with advertisers and newspaper publishers.

In the laboratory test with 36 users each respondent tested two different e-newspaper prototypes with a task oriented think aloud technique. The test was designed to test different design solutions as well as to investigate diffusion and adoption aspects. The evaluation consisted of a 75 minutes test, a 45 minutes structured interview, and was ended with a questionnaire. Questions like “Did you get a feeling of the amount of content, of how much you have read or not read?”, “Did you feel in control over your reading?”, “Which possibilities do you see with the e-newspaper?”, and “What does it take for you to replace your printed newspaper with the e-newspaper?”.

These prototypes were also evaluated in a questionnaire that was available at the web sites of the corresponding newspaper. In conjunction to the questionnaire the e-newspaper prototype for the respective newspaper was presented together with the movie illustrations described below. As the prototypes and the movie illustrations were very realistic and authentic, the purpose of these questionnaires was to get a wider range of user preferences with more focus on diffusion and adoption. In total, 3626 valid responses were retrieved. Examples of questions in the questionnaires are; “At which cost are you willing to change to the e-newspaper?”, “Which additional services to you consider important for the e-newspaper?” and “Would you consider replacing your printed newspaper with an e-newspaper in the future?”.

The outcome of these evaluations directed us towards reader preferences such as preferred design elements which had design implications, input to business model thinking as well as refinement of the e-newspaper concept. As a theoretical outcome a set of three design principles for e-reading were formalized. Furthermore, an initial target group model, a business model framework and a mobile service value model was suggested based on the results from the questionnaires.

Movie illustrations of future use

Three movie illustrations were produced by interaction designers in collaboration with researchers to visualize the e-newspaper in authentic future use situations. These illustrations were based on the outcome of the evaluations of the first four design artifacts. Each illustration builds on a persona, a male student, a female business woman and a female senior citizen and on the three compiled scenarios.

The movie illustrations were evaluated by readers in the questionnaire described above as well as in interviews and workshops with newspaper publishers and advertisers. As the illustrations incorporated many different aspects previously evaluated, they mediated a more complete vision of the e-newspaper concept. They were very helpful in the workshops as the respondents easily could relate to the e-newspaper, and therefore also add additional aspects that refined the e-newspaper concept. Examples of questions asked in the interviews and workshops are; “Do you believe that it is important to initially choose selected target groups?”, “What is your opinion of the e-newspapers possibility to personalization?” and “What new opportunities for advertising do you see with the e-newspaper?”. The outcome of these evaluations was a first proof of concept as well as input to business model and diffusion. The evaluation of the movie illustrations also contributed to the models described above.

eReader prototype

The eReader prototype was implemented in an eReader device made available by the device producer for testing even though not yet launched on the market. Since the technology was still immature, there were limitations to the functionality of this eReader prototype. For example, it was not possible to implement the navigation support and overview as previous work in the project had illuminated as important, it was presented in a 16 level grayscale and did not support columns, resulting in a layout reminding more of a book with few pictures. However, the readability of the screen technology was experienced by the respondents as very good and similar to reading on paper giving a more authentic feeling.

This prototype was evaluated two real-life tests, the first by 12 readers and the second with six newspaper staff. The participants were provided with eReaders for a period of two weeks to use in their homes, commuting, at work etc. Real newspaper content was published every morning with an extra update in the evening. The rationale behind this evaluation approach was to accomplish a test situation as close and authentic as a real e-newspaper reading as possible.

During the two week evaluation period the participants answered two questionnaires (one before and one after the test), performed daily online activities (diary, questions and critical incidents) and a 90 minutes semi-structured interview at the end of the test period. The daily questions were designed to follow the respondents learning process of handling the device and the e-newspaper content as well as to cover the different design aspects. Initially the questions were task oriented, e.g. "Read the editorial. Did you experience any difficulties finding the article?" or technology based, e.g. "How did you find the readability of the screen?". In the later phase of the test the questions concerned their preferences and experiences of the e-newspaper, e.g. "If you were to consider exchanging your printed newspaper to an e-newspaper, would it be enough with one device in your household?" or "Do you perceive the feeling of controlling your reading has changed during these two weeks?".

After two weeks of e-newspaper use, the respondents were visited in their homes for a 90 minute interview about their experiences and preferences of the e-newspaper. In these interviews focus was on daily use and to identify opportunities as well as problems. Furthermore, business model aspects such as the price and possible delivery of the e-newspaper were discussed. A follow-up workshop with the participating readers was conducted where 5 of the 12 participants participated. The purpose of the workshop was to evaluate their experiences of the eReader and the e-newspaper and to introduce them to the three functional prototypes which were much more mature regarding functionality. As these prototypes were more advanced regarding navigation, layout and content, they were used for comparison to the eReader prototypes used in the test and functioned as a basis of discussion. The outcome of this workshop was that almost every problem or negative aspect that was raised during the real-life test due to the limited functionality of the eReaders was considered to be solved with the functional prototypes.

These real-life tests resulted in rich data that taken together with previous results became the final proof of concept of the e-newspaper concept and the design principles which provided input to both device producers and service producers. As theoretical outcome key user values of e-newspapers and proof of concept of the design principles were presented.

5 DISCUSSION

This paper addresses the limited attention to evaluation in design science literature. To this end we have inspired by Sein *et al.* (2011) in detail described an authentic and concurrent evaluation approach. By revisiting a design science research endeavor in the project DigiNews, we have illustrated the what, why and how of all evaluation activities throughout the whole project. As a result of this retro perspective visit to the DigiNews project we can summarize our learning and reflections for an authentic and concurrent evaluation approach. In table 2 we summarize the increasing degree of

authenticity with the different evaluation methods. By authenticity we refer to the notion of how closely an evaluation captures the context and actual use of an artifact.

Design artifact	Evaluation method	Degree of authenticity
E-newspaper concept (A1)	Future workshops, questionnaires, interviews, focus groups	Imagining context and use
Mock-ups (A2)	Workshops, focus groups	Use context imagined and use simulated with mock-ups
Scenarios (A3)	Workshops	Illustrations of context and use
Navigation prototype (A4)	Focus groups, usability tests	Tangible and naturalistic use illustrations
Functional prototypes (A5)	Laboratory tests, interviews, workshops, questionnaires	Tangible and naturalistic use illustrations
Movie illustrations (A6)	Questionnaires, interviews, workshops	Realistic context and use illustrations
eReader prototype (A7)	Two week real life tests, including 2 questionnaires + daily online activities + 60 min interview	In authentic use context and realistic actual use

Table 2. Degree of authenticity

As recognized by Sein *et al.* (2011) ongoing evaluation may take many different forms. The design scope in the DigiNews project was unexplored and therefore the exploration phase included many different explorative approaches. In order to understand the design space evaluation methods were sought that generated data material that could be analyzed into theoretical conceptualizations of the class of problem rather than defined artifact design solutions. Hence we can here see a very close relationship between evaluation and the explorative activities in line with *ex ante* oriented evaluation as put forward by Pries-Heje *et al.* (2008).

Given the explorative nature of the whole project the approach to concurrently evaluate in many different ways was continued through the whole research process. These experiences show that the conceptualizations drawn from these evaluation activities clearly influenced design decisions regarding newspaper design, user value and business model design as well as decisions on strategic levels. Further, they were very important means for intervention in the newspaper organizations even though the IT artifact was very rudimentary. This evaluation approach did not only enable investigating the IT artifact evolution over time and in use but also concurrent intervention in strategic decisions in the participating organizations as underlined by Sein *et al.* (2011).

In the building phase the evaluation activities were designed to inform the e-newspaper prototype design. To inform the design process the evaluation methods were chosen with the aim to emphasize naturalistic (Venable 2006) and authentic (Sein *et al.* 2011) evaluation. This ambition did not only regard evaluation to be naturalistic in the meaning that it draws on empirical material as described by Pries-Heje *et al.* (2008) and authentic in the sense that it is interwoven with organizational practices (Sein *et al.* 2011). This ambition was expanded with a naturalistic and authentic evaluation including the relationships between all stakeholder groups and their interests.

Even though we here put much attention to concurrency there was still a final evaluation phase in accordance with 'soft design science research' approach (Baskerville *et al.* 2007). In the evaluation phase the degree of authenticity was advanced into a real-life testing situation as called for by for example Hevner *et al.* (2004). Although the IT artifact was not fully developed, this provided important input to the confirmation of concepts and design principles. In the real-life situation over a period of two weeks aspect of design and business models were exhibited that were not anticipated in the derivation and building phases. These activities consolidated the utility and theoretical outcomes. However, the evaluation was not separated into this phase but progressed in authenticity throughout the process. The activities in this consolidating evaluation were a continuance of the concurrent evaluation throughout the project. Moreover, the conceptualization and theorizing was not separable to a phase or stage succeeding building and evaluation (see e.g. Sein *et al.* 2011). Theorizing was concurrent throughout the whole process but gained in conclusiveness in the final evaluation.

As demonstrated here, authenticity was more challenging in the earlier phases of the project. Still striving for authenticity contributed to an increased level of authenticity throughout the concurrent evaluation activities in the project. We here emphasize the advice to strive for authentic evaluation not only in the formal evaluation phases commonly conducted in the later phases of research but also to incrementally increase the degree of authenticity along the whole process. The rationale of the effort, we argue, is that the gap between the design solution, the use context and actual use are reduced.

6 CONCLUSION

In summary, we can see a pattern of evaluation strategies and activities changing character as the design process progressed. Even though not planned before the project started, this approach proved to advance the authenticity of evaluation along the design process. The explorative character of the DigiNews project contributes to that this pattern appears very clearly. As described, the early phases of the design research process the artifacts evaluated were abstract and intangible and as the process proceeded the artifacts became more and more tangible and concrete. Accordingly, the evaluations increased in authenticity because the possibilities of illustrating actual use situations and hands-on experiences were created. When executing the final evaluation we further increased the authenticity in choice of evaluation methods. In other words, the evaluation phase was characterized by a very high degree of authenticity compared to the evaluations of the visionary loose concepts that were evaluated in the early phases. Authenticity is challenging early in the design process, but we still argue that reflecting on how to seek authenticity is an important part of concurrent evaluation and that it increases in importance as the design progresses.

In conclusion, we emphasize a holistic and concurrent approach to evaluation compared to the general design science research thinking. With the holistic approach we encourage evaluation to be inclusive of different aspects and relationships between stakeholder groups in the evaluation activities. Further we think that concurrency is not narrowed to evaluation but also regards theorizing. While it makes sense for planning to distinguish between phases and stages of evaluation and theorizing, in practice they are intrinsically interlinked and concurrent.

References

- Aalst, W. and Kumar, A. (2003). XML-Based Schema Definition for Support of Interorganizational Workflow. *Information Systems Research* (14:1), 23-46.
- Baskerville, R., Pries-Heje, J. and Venable, J. (2007). Soft Design Science Research: Extending the Boundaries of Evaluation in Design Science Research. In *Proceedings from the 2nd International Conference on Design Science Research in IT, May 2007, Pasadena, California, USA*.
- Gavish, B. and Gerdes, J. (1998). Anonymous Mechanisms in Group Decision Support Systems Communication. *Decision Support Systems* (23:4), 297-328.
- Hardless, C., Lindgren, R. and Schultze, U. (2007). Technology-Mediated Learning Systems for Project Work - A Design Theory. *Scandinavian Journal of Information Systems*, 19(2), 3-36.
- Hevner, A.R., March, S.T., Park, J. and Ram, S. (2004). Design Science in Information Systems Research. *MIS Quarterly*, (28:1), 75-105.
- Hevner, A. (2007). A Three Cycle View of Design Science Research. *Scandinavian Journal of Information Systems* (19:2), 87-92.
- Ihlström, C., Svensson, J. and Åkesson, M. (2005). How would you like your e-newspaper? - converging the best of two worlds. In *Proceedings of HCI International 2005, Las Vegas*.
- Ihlström, C., Åkesson, M. and Nordqvist, S. (2004). From Print to Web to e-paper - the challenge of designing the e-newspaper. In *Proceedings of ICC 8th International Conference on Electronic Publishing, ELPUB 2004, Brasilia*, 249-260.
- Kuechler, B. and Vaishnavi, V. (2008). Theory Development in Design Science Research: Anatomy of a Research Project, In *Proceedings of the Third International Conference on Design Science Research in Information Systems and Technology, Westin, Buckhead, Atlanta, Georgia, USA*.

- Lincoln, Y.S. and Guba, E.G. (1986). But Is It Rigorous? Trustworthiness and Authenticity in Naturalistic Evaluation. *New directions for program evaluation* (39), 73-84.
- Lindgren, R., Henfridsson, O. and Schultze, U. (2004), Design Principles for Competence Management Systems: A Synthesis of an Action Research Study. *MIS Quarterly*, (28:3), 435-472.
- March, S. T. and Smith, G. (1995). Design and Natural Science Research on Information Technologies. *Decision Support Systems* (15:4), 251-266.
- Markus, M. L., Majchrzak, A. and Gasser, L. (2002) .A Design Theory for Systems that Support Emergent Knowledge Processes. *MIS Quarterly* (26:3), 179-212.
- Peffer, K., Tuure Tuunanen, T., Rothenberger, M. and Chatterjee, S. (2007). The Design Science Research Methodology For Producing And Presenting Is Research. *Journal of Management Information Systems*, 24(3), 45-87.
- Pries-Heje, J., Venable, J. and Baskerville, R. (2008). Strategies for Design Science Research Evaluation. In *Proceedings of the 16th European Conference on Information Systems (ECIS 2008)*, Galway, Ireland.
- Sein, M. K., Henfridsson, O., Puro, S., Rossi, M and, Lindgren, R. (2011). Action Design Research. *MIS Quarterly*, (35:1), 37-56.
- Vaishnavi, V. and Kuechler, W. (2004). Design Research in Information Systems. January 20, 2004, last updated August 16, 2009. URL: <http://www.isworld.org/Researchdesign/drisISworld.htm>
- Venable, J. (2006). The Role of Theory and Theorising in Design Science Research. In *Proceedings of the First International Conference on Design Science Research in IT (DESRIST)*, Claremont, CA.
- Walls, J., Widmeyer, G. and El Sawy, O. (1992). Building an Information System Design Theory for Vigilant EIS. *Information Systems Research* (3:1), 36 -59.
- Åkesson, M., Kautz, K. and Ihlström Eriksson, C. (2010). Engaged Design Science: Developing Visions for the future e-Newspaper. Accepted to ICIS 2010, St. Louis, Missouri, December 12-15.