Article:
Thalmann, S, Ley, T, Maier, R et al. (3 more authors) (2018) Evaluation at scale: An approach to evaluate technology for informal workplace learning across contexts. International Journal of Technology Enhanced Learning, 10 (4). pp. 289-308. ISSN 1753-5255

https://doi.org/10.1504/IJTEL.2018.095124

Evaluation at Scale: An Approach to Evaluate Technology for Informal Workplace Learning Across Contexts

Stefan Thalmann

Graz University of Technology & Know-Center GmbH., Graz, Austria, stefan.thalmann@tugraz.at

Tobias Ley

Tallinn University, Estonia, tley@tlu.ee

Ronald Maier

University of Innsbruck, Austria, ronald.maier@uibk.ac.at

Tamsin Treasure-Jones

University of Leeds, United Kingdom, T.Treasure-Jones@leeds.ac.uk

Christina Sarigianni

University of Innsbruck, Austria, christina.sarigianni@uibk.ac.at

Markus Manhart

Apracor GmbH, Munich, Germany, markus@apracor.com

Abstract:

Technology for informal learning at the workplace is designed to support work processes and the learning that can occur within these processes anytime and anywhere. The implicit, spontaneous and hidden nature of informal learning in addition to the large and less predictable number of application scenarios challenge the evaluation of learning technology. Therefore, we require cross-case comparisons to draw conclusions beyond the immediate context of informal learning in a single case. A further challenge for evaluation is added if a user-centered design method had been employed that already had involved users in large numbers and has led to a high level of expectations regarding the potential of the tool to fulfil their needs. In this article, we suggest a participatory approach for the evaluation of technology for informal learning at the workplace across sectoral and country borders in order to address these evaluation

challenges. The proposed evaluation approach is based on the ideas of active involvement of stakeholders, an iterative planning process, a continuous, open and collaborative interpretation of the data collected on the basis of representations shared across cases plus a pluralism of methods for evaluating learning technology, and is particularly intended for large scale evaluations in workplace settings. We present a case, in which we applied the proposed evaluation approach in the context of a large scale European research project on technology supported informal learning across professional, sectoral and country borders. Finally, we reflect on our lessons learned and provide recommendations.

Keywords: Participatory evaluation, informal learning, workplace learning, participatory design

1. Introduction

Shorter innovation cycles increase the need to continuously and quickly learn at work (Malecki, 2010). In workplace settings, informal learning is seen as the most important way to acquire and develop skills and competencies (Boud and Middleton, 2003, Eraut and Hirsch, 2007). Informal learning is situated, closely connected to practices at the workplace, it is driven by tasks and interests of the learner connected to work (Marsick and Watkins, 2015). As a result, any technological support for learning at the workplace needs to blend into existing practices (Lindstaedt et al., 2010). Moreover, introducing technology often changes these practices to a significant degree (Waizenegger et al., 2016). For example, when technologies allow collaboration across time and space, new learning practices will result, such as when teams are set up quickly to solve problems; or when mobile technologies allow connecting learning to a physical place of work, then learning practices become much more contextualized and are moved outside the classroom (Ley et al. 2014).

It is now widely accepted that in these cases a participatory, design-based approach that puts a focus on understanding situated practices and changing them is key (Ehret and Hollett, 2016). Such an approach involves a diverse set of stakeholders early on in the research who collaboratively design artefacts, technologies and new practices together with designers and developers (Leinonen and Durall-Gazulla, 2014). When it comes to the evaluation of the effects

of these approaches, however, this type of participatory research often has its limitations because of the highly contextual nature and a tight connection to a particular case and setting. As a consequence, evaluation often switch back to a traditional waterfall approach (Ley et al. 2014).

In this paper, we address this challenge by introducing Eval@Scale, a scalable evaluation method that is in line with participatory, design-oriented and stakeholder-driven research and development approaches. Scalable in our context means to conduct the evaluation over a long time, over several contexts and involving diverse people. Our proposed evaluation method is responsive to stakeholder concerns and takes into account the less predictable learning situations, limited awareness of learning and the situated character of learning that results from a participatory design approach. At the same time, our proposed evaluation method allows researchers to conduct evaluation activities across a number of pilots in order to generalize findings across organizational, professional, industry-sector and national boundaries.

After introducing the related work, we first describe our proposed evaluation method on a conceptual level in chapter 3. In chapter 4, we describe a complex case in which our evaluation approach was applied. We present how we involved a diverse set of stakeholders early in the evaluation design, how we ran distributed evaluation studies across contexts and how we collaboratively interpreted the data collected from diverse cases for generalizations beyond the immediate context of a single case. Finally, we reflect on the proposed evaluation method in the light of our case and give recommendations for its application.

2. Background

Informal learning is no new phenomenon, it has always had an important role in workplace learning, for example as apprentices learned their craft from their masters (Marsick et al., 1999). Informal learning has been increasingly promoted over the recent years because of the changes in work organizations (Garrick, 1998) and becomes even more important in contexts in new

work paradigms such as industry 4.0 or smart factory (Schuh et al., 2015). Informal learning has also received more attention from TEL research as mobile devices and social technologies allow supporting this important but somehow hidden type of learning (Ley et al., 2014).

Informal learning provides a contrast to formal learning and has a stronger focus on the social significance of learning from others, as it takes place in a much wider variety of settings compared to formal learning (Eraut, 2004). The important interplay of both the informal and the social character further emerges in research on learning at the workplace (Hart, 2011). Therefore, informal learning is defined as unintentional, non-institutional, less structured, experiential learning which is primarily under the control of the learner (Steurer et al., 2015). Most informal learning occurs within social contexts (Eraut, 2004, Marsick, 2009) and its outcomes are difficult to predict (Clarke, 2004, Marsick, 2009). People lack awareness of their own learning and subsequently might not be able to report about their own learning (Eraut, 2004).

Informal learning is often an unintended consequence of a task conducted in the daily routine (Marsick and Watkins, 2015). Hence, informal learning is deeply rooted in the work situation and creating supportive technology is a significant challenge (Littlejohn and Margaryan, 2014).

User satisfaction with technology can be assessed using questionnaires, but assessing how learning technology changed actual learning practices is very challenging. In that case, an intensive interaction with end-users and other stakeholders is needed, which makes large-scale cross-contextual evaluation studies very time and resource demanding as many different users need to participate. In addition, the complex interrelation of work and learning processes requires an in-depth domain knowledge when measuring and interpreting the effects of technology designed for informal learning at the workplace. Hence, an intensive involvement of domain experts is needed as well as time to understand the work processes to be able to frame the evaluation in a meaningful way.

Further, informal learning at the workplace is very much influenced by the organization of work itself. As work requires more and more inter-organizational collaboration, also a huge variety of collaboration tools and especially social media tools are used for informal learning at the workplace (Schäper and Thalmann, 2014). Hence, introducing technology for informal learning at the workplace should also take the need for inter-organizational collaboration into account. To even go a step further, to successfully introduce such technology a critical mass of people needs to be engaged to realize the network effects that offset start up effort invested to embrace the new technology. Thus, scaling of learning technology is key to successfully introduce learning technology for collaborative informal learning at the workplace that crosses the boundaries of the immediate workplace setting. As informal learning is directly related to the work situations and work situations more and more cross borders of organizations, industry sectors and countries, evaluation cannot longer just stick to one single case. This means for TEL research that more and more multifaceted evidence collected across borders is needed.

In contrast to formal learning, informal learning situations cannot be well defined before the learning situation actually occurs and also assessment criteria are difficult to define because informal learning is often spontaneous, serendipitous and interest-driven rather than planned in advance. Informal learning not only takes place in classroom settings, but in unscheduled and more unpredictable ways in the working context without explicit learning goals. Hence, we need a richer set of data sources and data collection activities suited to explore the often hidden nature of informal learning which can be applied in workplace settings (Gopalakrishnan et al., 2013). Based on the idea of an active involvement and a pluralism of methods for evaluating learning technology (Gopalakrishnan et al., 2013), we propose a participatory evaluation approach that includes stakeholders in both, iterative planning and continuous, open and collaborative interpretation of the diverse data collected to evaluate learning technology for informal learning across different contexts.

Participatory approaches can be used in any evaluation design – they are not exclusively bound to specific evaluation methods or restricted to quantitative or qualitative data collection and analysis. Participation by stakeholders can occur at any stage of the evaluation process: in its design, in data collection, in analysis, in reporting and in managing the study (Guijt, 2014). The purpose of participatory evaluation is to focus on informal learning at the workplace and to draw lessons learnt, which could guide future decisions (Campilan, 2000). The aim is here to involve large numbers of stakeholders throughout the evaluation, ranging from strong supporters of the introduced learning technology to sceptical stakeholders. The goal of such extensive and prolonged stakeholder involvement is to gain insights into the changes in work and learning processes that are associated with the learning technology used in workplace settings (including the knowledge exchange networks of the participants) and to draw conclusions beyond the immediate context for theory as well as tool design and sustainability.

3. The performed approach to evaluate learning technology across contexts

The participatory evaluation approach is a joint, collaborative stakeholder approach with domain experts, project staff, and stakeholder groups (Guijt, 2014, Campilan, 2000). The basic assumption is that informal learning is multifaceted and takes places in various situations involving several stakeholders. Hence, to ensure a preferably large coverage of observation and reflections about the effects of learning technology, the goal is to include not just the end-users, but also developers, educational institutions, networks, associations and other organizations in the evaluation process. Therefore, we selected a representative type of participation to give the stakeholders a voice i.e., stakeholder involvement in planning the evaluation, to ask them to comment on findings, to identify the lessons learned, to determine the appropriate steps, and to finally create a joint learning experience (Guijt, 2014).

Generalizability is important but mostly not achievable in such evaluation studies. In this regard, a cross-case comparison seems important and desirable to stabilize findings. With cross-

case comparison we refer to a comparison of the findings across cases in different sectors and countries. Following these thoughts, our approach aims to create a shared understanding not only in a single case, but also across case settings and therefore enhances the overall validity of the evaluation. Figure 1 provides an overview of our proposed evaluation method on a conceptual level. The goal is to scale evaluation of learning technology by conducting the evaluation over a long time, over several contexts and by involving diverse people

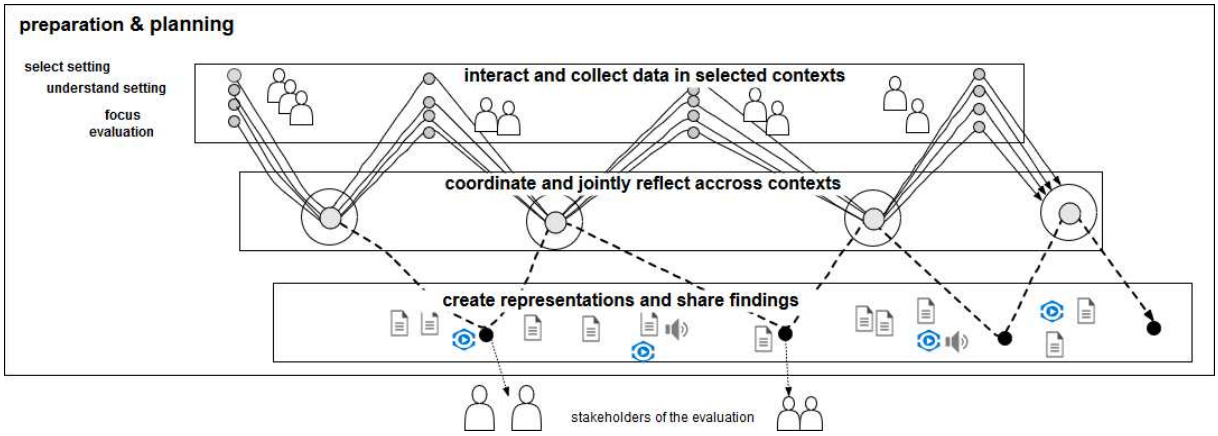

Figure 1: Eval@Scale activities and interactions

Our proposed approach consists of four overlapping and interacting activities: a preparation & planning activity, an interaction and data collection activity, a coordination and reflection activity and a creation and sharing activity. Thereby, joint reflection starts just after the first data are collected and directly informs the next round of interaction and data collection. Further, the first round of joint reflection also triggers the creation of representations and the sharing of findings. Thereby, the findings are shared with the evaluation team, with participants as well as with external stakeholders. This procedure requires massive coordination between all involved actors until enough data has been collected and interpreted, so that the findings are saturated, trustworthy and valid across case settings. The four activities of our approach include well defined communication and exchange procedures between the cases and will now be described on a conceptual level in following and by using an in-depth example in chapter 4.

Preparation & planning

To prepare the evaluation, we first brought the gathered experience from the (co-) design and the past end user interactions on board. Thereby, the preparation had three major activities: (1) the selection of the evaluation setting, (2) the understanding of the context and (3) focussing the evaluation.

Select the setting

According to our definition of scaling, the first goal was to find settings with diverse contexts, diverse people and in which we can conduct the evaluation study over a longer period of time. The different cases we achieved by selecting different networks in two different industry sectors in two different countries. The diverse set of people was ensured by conducting a stakeholder analysis, focussing on individuals, groups, organisations or institutions that are considered to have a stake in the evaluation (Vos and Achterkamp, 2006). Stake in this regard means, how stakeholders exert power, as well as, how urgent and legitimate their claims are (Mitchell et al., 1997). Therefore, we first analysed past documents and involved members of the co-design team. Based on this investigation, we held a workshop involving members from all initial stakeholder groups such as designers, developers, end user interactors or researchers for each case. In this workshop we identified the relevant stakeholder groups and their stakes. Based on this local workshop for each case, an overarching workshop involving the domain experts from other cases was needed to share experiences across cases. In addition to the identification of the stakeholders also their potential interests in the evaluation needed to be clarified. Based on these interests, the major focus areas of the evaluation were defined for each case.

Understanding of the setting

After the setting was selected an in-depth inquiry and understanding was needed. Therefore, we conducted sessions with the end-users and we engaged domain experts in the evaluation team. As part of this understanding, we aimed at finding evaluation criteria, as evaluation involves “the identification, clarification, and application of defensible criteria to determine an

evaluation object‘s value (worth or merit) in relation to those criteria” (Fitzpatrick et al., 2004). In this respect, evaluation criteria are objectives to be achieved relating to concerns or issues of the stakeholders (Mortimer et al., 2008). The criteria need to be converted into one or more measurable indicators (Mortimer et al., 2008) that permit the assessment of how criteria are met. Hence, we considered such “measures” as more detailed specifications that help us to assess the fulfilment of the evaluation criteria.

Taking the points from above into account, planning the evaluation required the definition of the evaluation object’s value, relating dimensions of “measures”, i.e. the evaluation criteria, and the measures itself. In a participatory evaluation approach, all stakeholder groups should be incorporated in this central task. Hence, our preparation started with the identification of relevant evaluation criteria for each stakeholder in a workshop for each case. Based on this draft, additional focus groups with each group of stakeholders were performed in which we identified the stakeholders rank the (initial) evaluation criteria from their perspective or added new criteria if necessary. This way, the evaluators obtained insights about the importance of different evaluation criteria for the stakeholders. Subsequently, the stakeholders from the focus groups defined one or several measures for each evaluation criterion and proposed one or several data collection methods for each measure which they deem relevant in these workshops. The basic idea behind this approach is, that the end users know their workplace and their knowledge sharing networks best and within a joint development and refinement their knowledge should be incorporated. However, the understanding of the context was ongoing, and we expected to return to, reflect on and refine these initial evaluation criteria as the evaluation progressed.

Focussing the evaluation

After understanding what is important in every setting and what is important for the evaluation as well as knowing the organisational restrictions, an overall focussing of the evaluation

activities was required. For this purpose, a joint meeting in which the evaluators of the different settings bring their work together proved to be valuable. The meeting and also the document created was the outcome of a collaborative journey and represented the collective understanding of the agreed procedure. Thereby, our goal was not to create one unique and static plan (which would contradict the participatory idea plus the ideas of iterative planning and continuous collaborative interpretation), rather to learn from each other and to relate the foci of each setting. By doing so, relationships between the cases can be detected and kept in mind during the data collection and analysis.

Interact and collect data in selected cases:

In contrast to traditional evaluations, we did not just collect data from but we interacted with the evaluation participants. In addition to formal interaction formats such as interviews, informal approaches such as regular accompanying talks or diaries were especially important. The mixture of interaction and data collection methods needs to be specified in close cooperation with the stakeholders in advance. Suitable collection methods should be chosen for collecting evidence for the suggested evaluation criteria as well as being applicable by the evaluation team in the workplaces of the end users. Applicable here means that they are also practical and can be integrated into the daily work practices of the end users without raising major interruptions.

Overall, we interacted and collected data in (1) face to face interactions in workshops, (2) accompanying conversations and (3) log data of involved systems. Further the interactions were strongly intertwined with the ongoing analysis and joint reflections conducted in parallel.

Workshops

We structured our formal evaluation interactions around three evaluation workshops. First, a kick-off workshop to introduce the learning technology and to assess the status quo. Second, an intermediate workshop to discuss reported changes in practice and possible required adaptations

of the learning technology or the evaluation process. Third, a final workshop to discuss perceived changes in practice and anticipated changes and effects caused by a future and long term usage.

Accompanying conversations

The aim of our proposed evaluation method is to continuously incorporate stakeholder feedback and to jointly reflect about the implications. For this reason, accompanying conversations on a regular (i.e. weekly basis) and a continuous analysis of log data were performed. The conversations aimed at establishing and keeping constant communication between the evaluation team and participants during the entire evaluation period. This approach allows the evaluation to be undertaken in a reactive manner and to investigate (during the evaluation period itself) for example why certain features are heavily used or why new practices of informal learning emerged. Based on these insights questions and areas of interest were iteratively adapted to incorporate the ongoing reflections and put emphasis on promising phenomena accordingly. Furthermore, the continuous interaction with the evaluation participants strives to help solving technical issues as quickly as possible as well as ensuring the awareness and engagement of the participants. This approach can be combined with other forms of data collections, i.e. questionnaires.

Log data

In addition to the direct data collection with the end users, we also used indirect ways of collecting evaluation data. Here log data proved to be a valuable source of evidence. On the one hand we could collect quantitative evidence and on the other hand we could identify valuable triggers for the accompanying conversations.

Coordinate and jointly reflect across contexts:

Similar to the joint interaction and data collection, also the data analysis was performed in a collaborative manner for each setting and also across the cases. Thereby, (1) the joint interpretation and reflection, and (2) the coordination meetings were important activities.

Joint interpretation and reflection

The goal of our evaluation was of an explanatory nature and thus we wanted to adjust our data collection quickly to emerging insights and assumptions. To fulfil this need, we analysed our data very soon after the collection in small teams. Thereby it turned out, that the domain knowledge is crucial. Hence, the interpretation of log data for each tool component was coordinated by the respective developers and the ad-hoc interpretation of accompanying conversations and other informal interaction coordinated by those team members conducting the end user interactions. However, the joint interpretation discussions that took place on a regular basis helped to ensure that richer understandings were formed, including multiple perspectives on the same sets of data. Finally, in larger (cross-context) meetings a joint discussion, comparisons and reflection was triggered and resulted in a continuous readjustment of the interaction and data collection.

Coordination meetings

For the cross-case dimension of the study, overarching reflection workshops were conducted. Particularly, to reflect on the success and failure stories in each setting and to jointly reflect upon the similar perspectives. By doing so the data collection is more aligned and it is ensured that (to a certain degree) comparable evaluation evidence is collected.

The meetings were not only intended to produce evaluation artefacts, but also to continuously adjust the focus of all evaluation activities. Presenting and discussing the intermediate findings with the end users is considered as an additional feedback loop to validate the intermediate findings.

For the cross-case analysis, the focus in the post-data collection shifted from the detection of effects and changes to a more in-depth and theory rooted interpretation of the collected data. Thereby, the previously identified similarities and differences across cases helped to sharpen the focus. All collected data, e.g. interview transcripts, field notes or diaries were rigorously analysed using a qualitative content analysis (Patton, 2002, Mayring, 2014). Also the post-data collection interpretation involved a broad range of project members to jointly discuss findings and to collaboratively develop suitable evaluation artefacts. To successfully accomplish this task the role of a study coordinator (Maier and Thalmann, 2012) proved to be valuable.

Create representations and share findings

The participatory and distributed character of the evaluation also created the need to agree on common representations of data and findings that could be more widely discussed and shared. We used different forms of online collaboration tools to share and discuss outcomes (for example themes identified, reflections on theory, preliminary frameworks/models) regularly. For the short feedback cycles during the evaluation, for example, we created stories showing the usage of tools in different settings and shared them via an online collaboration tool. More detailed reports (such as thematic analysis or hierarchies of barriers identified) were created towards the end for a more widespread dissemination. In addition to traditional reports, also regular feedback in social media channels accessed by the end users, such as Twitter or Facebook was used. This regular feedback triggers a reflection about the intermediate findings and can be seen as a validation activity. But the feedback is also an evaluation activity, which can influence the ongoing evaluation, e.g. by setting the focus to a certain feature, by motivating or making changes to deal with any demotivation encountered.

4. Application example of Eval@Scale: A large scale informal learning evaluation

4.1 Overview of the evaluation pilots and cross-case analysis

We tested our evaluation method in a large scale EU funded TEL research and development project in which learning technology for informal learning at the workplace was developed and rolled out. Table 1 gives an overview of the six cases, called pilot studies in the context of this application example, performed in the scope of the project.

Table 1: Overview of evaluation pilots

Pilot	No. of participants	Sector	Pilotgroup	Country
A	18	Healthcare	Practice Manager Network	UK
B	6	Healthcare	TEL Strategy Team	UK
C	5	Healthcare	Quality Improvement Training Team	UK
D	9	Construction	A class of carpentry apprentices	GER
E	17	Construction	A class of well-builder apprentices	GER
F	18	Construction	A class of machine operators	GER

Pilots A, B and C focussed on teams (working across organisations and/or geographically distributed) who used the tools to support their work and learning on joint projects and the development of new services. This involved collaborative problem-solving and knowledge building. Their main concerns were to keep track of their work on their projects and to keep the work focused and flowing. Pilots D, E, and F focussed on informal learning and training in construction in the context of a big construction training centre. The tools focussed on supporting the apprentices during their time outside the training centre in their companies.

Further, the tools focussed on experience sharing in the distributed groups and joint reflection in the training centre.

Table 2 gives an overview of the four different roles of participants (evaluation specialists, learning science researchers, technology researchers and developers and end users and other stakeholders) and the levels of participation in the evaluation.

Table 2: levels of participation in the evaluation

Level of participation / Role	**Evaluation Specialists**	**Learning Science Researcher**	**Technology Researchers and Developers**	**End-users and other stakeholders**
Select setting	Informed the selection process	Supported the identification & recruitment of the pilot groups	Identified first use cases	Led on the identification & recruitment of the pilot groups
Understand context	Led the process and consolidated lists	Provided evaluation criteria	Specified use cases for each case	Provided evaluation criteria
Focus evaluation	Proposed the plan and asked input	scoped according to learning theories	Scoped according to technologies	Checked for organisational constraints
Creation of interaction and data collection instruments	Recommended instruments and guided the selection	Instantiated instruments in cases	Technical integration	Organisational integration
Collecting the data	Co-ordinate the collection of data	Interact and collect qualitative data	Collect log-data	Support activities
Analysing and joint reflection of the data	Moderate discussions	Prepare domain input and guide learning discussion	Prepare log-data and solve technical issues	Interaction and feedback during evaluation
Create representations	Coordinate creation processes and focus on evaluation criteria	Create learning-oriented feedback	Create technical feedback	Give feedback
Share findings	Coordination	Share to end	Share technical	Give feedback

	and formal evaluation report	users	feedback	

The overall coordination (cross-country and cross-sectoral) took place via physical meetings. First, we had a series of preparation meetings in which we jointly discussed the evaluation criteria from all pilots. As a result, we grouped these criteria into four major research areas. In doing so we ensured a common understanding of the overall project idea and created awareness for these topics in all pilot teams. Further, the project agreed on the overall evaluation plan. A few weeks before the first evaluation activities started, we had an additional meeting to synchronize the detailed evaluation plans for each pilot. Thereby, the focus and scope of the workshops and the accompanying conversations were synchronized. We had an additional meeting after the conduction of the intermediate workshops, to jointly reflect about the progress and first insights. This workshop was particularly important to exchange ideas and best practices and to have first reflections about joint patterns at this stage, as these insights could still be used for scoping the data collection for the rest of the evaluation. Finally, we had a meeting with the focus on overarching (theory-guided) reflection. Therefore, first findings of each pilot were discussed and suitable artefacts produced.

In the following section we will present one pilot study in detail, to show how we applied the proposed evaluation approach in practice.

4.2 The implementation of the participatory evaluation approach within Pilot A

Background: The evaluation reported here followed 3 years of empirical and co-design work with healthcare professionals. Over those 3 years we had partnered with 3 General Practices (medical centres providing Primary Care - family doctors) to explore (through observations, focus groups and interviews) how informal learning was currently supported within their

workplaces and networks, and to work with small groups from these General Practices to co-design technology tools (co-design workshops and field-tests) to better support their learning.

Selecting the setting

Nearing the end of the 3 years, representatives from the key healthcare organisations with whom we had worked were invited to take part in an evaluation planning meeting, at which the pilot groups who would take part in the evaluation were identified. The Practice Managers Network was proposed by several stakeholders, who then worked with us to recruit the network to the pilot.

The network consisted of 18 Practice Managers (PMs), one from each of the General Practices within their federation. The PMs have monthly meetings, during which they discuss the progress of their projects and prioritise their activities. The tools that the pilot group used had been co-designed to support collaborative work leading to the production of a document (e.g. developing and defining a new protocol or a new shared service).

The location, timing and length of the key evaluation workshops was decided jointly with the lead of the Practice Managers Network and reviewed at each stage of the pilot. Therefore, the main evaluation events (namely kick-off, intermediate and final workshops) were organized as part of the meetings that the Practice Managers Network had already arranged for themselves. Such an approach ensured that the extra effort required from the PMs (in terms of time, organisation and travel) was minimised and therefore gave us the opportunity to involve as many people as possible. It also meant that the tools use could be discussed and explored within the context of the real ongoing project work that the Practice Managers were bringing to their own meeting.

Understanding the setting and focusing the evaluation

This had already been started within the co-design work prior to the evaluation, but was continued within the kick-off workshop. The goal of the kick-off workshop was to introduce

the collaborative tools and discuss with the PMs their current ways of working, their needs and challenges, as well as their expectations from the tools. The PMs also identified in this workshop the specific project (the development of a new shared locum service) for which they would use the tools.

Interact and Collect Data

Workshops

Kick-off workshop

The kick-off workshop was divided in two parts 1) focus group and 2) training. As we had large numbers, the focus group was run in the form of a knowledge cafe. This approach allowed the PMs to work in smaller groups and reflect on predefined topics, describe concrete cases from their network, express their expectations of the tools and set concrete goals. After the focus group, the PMs received training and got familiar with the tools by performing various exercises. At the end of the workshop the PMs themselves chose the key projects/work for which they would use the tools during the evaluation period. Therefore, the planned use of the tools was determined by the PMs themselves and was tied to their real work.

Intermediate workshop

Midway through the evaluation timeframe we returned to the network for an intermediate workshop. The focus of the intermediate workshop was to gain a wider view (from the whole group) of their current and planned use of the tools and any benefits or challenges encountered, as well as to support them in making further progress. The intermediate workshop was conducted in the form of a focus group. In addition, the PMs had networked laptops and access to the tools in the workshop itself, meaning that they could pull up onscreen examples of their current work in the tools. This allowed them to provide the researchers with clear illustrations of what they had done and their future plans. It also enabled the researchers to support the group in using the tools to move their specific work forward.

Final workshop

The evaluation was completed with a final workshop. The purpose of the final workshop was to facilitate a profound reflection on the entire period of summative evaluation and discuss in detail with the PMs about their experiences. The aims were to understand how the tools had been used, what barriers or attitudes had prevented greater use of the tools, what the effect had been of the tools use and in particular whether the tools had led to any changes in the working/learning practices of the pilot group. The PMs shared their experiences, expressed "wow" moments and challenges and made recommendations on further improvements. The final workshop was followed by some individual interviews with members of the group. The individual interviews gave individuals (particularly those who had made most use of the tools) the opportunity to expand on their descriptions of the changed learning and working practices that occurred during the evaluation. During all three workshops, the discussions and relevant activities were audio-recorded and transcribed for analysis.

Accompanying conversations and self-descriptions

Between the three main workshops, we conducted accompanying conversations with members of the network, in order to acquire real-time feedback and reflections on their current and planned use of the tools and any benefits or challenges encountered. A short interview guideline was used and the accompanying conversations were performed initially via telephone once a week. However, some of the PMs wished to send the answers via email instead. Such an approach seemed convenient for both the PMs and the researchers, since it removed the difficulties around scheduling call times and still allowed for follow-up questions to be sent by email or addressed in the next call/email.

Apart from the accompanying conversations, all PMs received regular nudge emails, which had the purpose to facilitate and stimulate the activity. However, several PMs reported (in the accompanying conversations) that the group felt overwhelmed by this level of email. Therefore,

responding to their request, the number of nudge emails sent was reduced. PMs were all also provided with a support email address and phone number which they could use to send in questions, ask for help or send in further (unprompted) feedback.

The accompanying conversations and the question and answers from the support emails were all included in the ongoing data analysis.

Log data

Across the evaluation period, the PMs' use of the tools was observed through the collection and analysis of log data. The log data were collected and reported in a bi-weekly basis. The continuous observation of the log data provided a detailed overview of the PMs' activities in the tools and allowed a better preparation for the next steps. Thus, during the workshops and given the time limits, the discussion was optimized by focusing on the performed activities and the encountered challenges. This process resulted in receiving valuable feedback for improving the tools' performance and providing necessary support.

Coordination and Joint reflection

Coordination meetings

For better facilitation of the evaluation, we organized weekly video conferences with the group, who were involved in supporting the pilot, collecting and analysing the data (e.g. the evaluation co-ordinator, the learning science researchers and the technology researchers and developers). The call covered all three healthcare pilots. This bundling already allowed a first cross tool and cross case reflection within one country and one industry sector.

Joint interpretation and reflection

The regular topics of the weekly meetings involved updates from the pilot groups (news from accompanying conversations, latest insights from the analysis of log data, ongoing collaborative interpretation of qualitative data) and discussion over future planning (required interventions, questions for the accompanying conversations or self-descriptions). Data analyses were shared using collaboration tools and then discussed and refined during these meetings. Therefore, the

agreed findings emerged from this iterative process and were written up (collaboratively) by the evaluation co-ordinator and researchers into an evaluation report.

Create representations and share findings

During the evaluation period we continuously created and refined representations to both guide and report on the analyses and findings. This included theory-driven representations such as an initial framework of the anticipated changes in practice that was created at the start of the evaluation period and filled in, discussed and reshaped during the evaluation. It also included emergent, data driven representations, such as the grouped themes induced from the qualitative data and actual barriers reported to changes in practice. These representations were used to focus our discussions during the co-ordination meetings. The findings and reflections were shared not only within the involved research group, but also with the end users. Such activities included discussions during the workshops (which also acted as an opportunity to validate or change our understanding) and short emails with interesting insights that emerged during the process. After the evaluation period we also reported back to the leaders of each pilot with a summary of our findings and offered them the chance to provide further clarification and feedback. We also shared our findings with a wider set of healthcare stakeholders to obtain their feedback as part of our regular participation in relevant conferences and workshops and specific presentations and discussions of interim findings on such events. Within the healthcare setting we did not make use of additional channels such as social media or blogs to share preliminary findings with the evaluation participants since they had made it clear to us that they had limited time and would appreciate having the reporting/engagement kept to the scheduled sessions. In contrast, the participants in the construction setting welcomed regular updates and engagement via blogs and social media as well within and between the scheduled sessions.

5. Discussion

The characteristics of informal learning in workplace settings, that is learning that can happen unplanned, intertwined with work processes and with outcomes that are difficult to assess (Steurer et al., 2015), turn evaluation of learning technology into a challenging activity. As a consequence, we need an evaluation approach being able collecting evaluation data in all these situations to provide high quality evaluation outcomes. For this purpose, we presented a participatory evaluation approach which allows us to better align tool development, stakeholder requirements and evaluation activities across different workplace contexts. Our approach is a

participative approach involving all relevant stakeholders in the evaluation activities. In contrast to known participative approaches we propose to extend the focus across the organisational boundaries as this reflects the nature of informal learning. Further, our approach allows the coordination and reflection about several evaluation settings. This is a particular benefit as the intensive coordination allows a joint reflection and theorizing across contexts. Therefore, various methods for data collection are recommended to have a rich and multifaceted picture. But this requires also intensive interaction with representatives of all relevant stakeholder groups as well as careful coordination of all evaluation activities.

The **data collection activities** are multifaceted to address the specific characteristics of informal learning across contexts. This means combining quantitative data collection such as log data and questionnaires with qualitative data collection, such as interviews and focus groups to collect the needed evaluation data. Thereby, it turned out that the continuous analysis of the quantitative data creates relevant focal points for the qualitative inquiry. Further, also the cross-case discussion about data collection provided many insights on how the framing and spotting could be improved. Further, it was valuable to apply the knowledge cafe approach during the kick-off workshop by bringing the pilot groups together since participants took advantage of such settings to openly discuss their problems. Such an approach could also improve tool trainings since professionals recognize that some of their previously discussed issues could be addressed by the deployed tools. We also learned to flexibly react to the needs of the participants and to switch from accompanying conversation phone calls to self-description forms. The essence is really to co-develop the data collection with the participants and all stakeholders to fit to their preferences and conditions of work settings. Our lessons learned are, that if they are involved in the decision process, they tend to be more open and engaged throughout the evaluation process. Additionally, some of the end-users highly appreciated incentives to encourage participation, for example a joint BBQ event. Participatory evaluation also means end user engagement over a long period of time. Depending on the cultural setting, organizing

social events with the end users around the workshops can stimulate engagement. This can offer evaluators additional opportunities to take field notes and allows for informal chats beyond the rather formal workshop settings.

The **data analysis activities** started early on to continuously inform and adjust the data collection and to give feedback to the participants. The weekly update rounds in the form of a video conference were very valuable and allowed us to gain an overview of the ongoing activities, share preliminary insights and discuss our next steps. It was important to have facilitators close to the pilot groups that have the chance to provide constant updates and to reflect the preliminary insights in the light of the pilot group. Also, sharing ideas and thoughts on the preliminary insights was very useful for further data analysis and reflection on the findings. In the beginning, the data analysis was mainly driven by the need to adjust the data collection process and to create appropriate artefacts to inform all stakeholders. Over time, abstraction played a more important role by cross-case analysis as we became more confident in our observations of the relevant phenomena that were associated with usage of the tools and developed a profound understanding.

The **coordination process** benefitted much from taking the feedback from the participants into consideration and helped to adapt our strategies, for example to reduce the amount of emails sent to participants of the evaluation when they signalled that they felt overwhelmed and introduce new ways of communication. Thereby, we learned to treat each group according to their own needs and expectations, as not all pilots reacted in the same way. Hence, project-wide guidelines are important to coordinate all activities, but flexibility is required to adapt all activities to the specific contexts. Taking the specifics of each sector into account also helped in the cross-case analysis (see below). Overall, the comparison between industry sectors is very valuable to reflect on a “fresh perspective” from outside in order to adjust the focus in the data collection as well as in the data analysis. Additionally, the cross-case analysis is very important

to avoid running into the trap of a single context tool. Further, if the tools can be easily customized for different contexts, engagement of users may be easier with new groups of users who first encounter the tools when mature.

The success of the proposed evaluation approach highly depends on the team's ability to involve many stakeholders and individuals, so that the advantages of a multi perspectival collection and analysis of evaluation evidence can address the peculiarities of the context of informal learning in the context of diverse workplace settings. This is, however, the particular strength of our proposed evaluation approach, as it allows an in-depth investigation of how learning technology affects practices of informal learning in workplace settings. Keeping all stakeholders on board even though it requires substantial efforts, especially when it comes to communicating the intermediate evaluation results to the end users, enhances the quality of the findings and reduces the frustration of stakeholders as more (ideally all) perspectives are considered.

The **cross-case analysis** resulted in several insights that were common across the different contexts, and some that differed. First, the analysis led to a model of knowledge appropriation (Ley et al. 2017) that explains how new and innovative practices in a domain are adopted through different social and organisational learning processes. In the healthcare domain, this was observed in the context of introducing new guidelines in a GP practice. In the building and construction domain, this was connected to trainers collaboratively working on introducing new teaching techniques. Particularly, we found evidence in both domains (healthcare and construction) of how scaffolding at the workplace was conducted (through seeking help from colleagues and guiding less experienced co-workers or newcomers), and how knowledge developed and matured in the organisation or network (through knowledge sharing, co-creation and processes of formalization). Underlying these two are various practices of appropriating knowledge for a particular context (creating awareness and building shared understanding of the new practice, and adapting it and validating it in a new context).

The cross-case analysis also revealed factors for the adoption of workplace learning technology in sectors that have been more hesitant to embrace learning technology in the past. Especially insightful was the need to bridge different learning contexts (some more formal, such as workplace trainings, some more informal, such as workgroup meetings). We found evidence for these practices in all studied contexts. Generally, technological solutions were more successfully adopted when a strong relation to the formal learning contexts was established.

6. Conclusion and outlook

This paper proposes the Eval@Scale approach for the evaluation of informal workplace learning. We discussed how traditional evaluation approaches reach their limits when they are applied to evaluate technology for informal learning in diverse work settings, especially in evaluation scenarios that cross sectoral and country borders. We presented a participatory evaluation approach which allows better alignment between tool development, stakeholder requirements and evaluation activities across different contexts. This approach allows a more holistic and in-depth investigation of technology for informal learning in context, but requires also an intensive interaction with all stakeholders as well as a careful coordination of all activities. We applied the proposed approach in an EU funded research project to evaluate the developed technology for informal learning at the workplace in different domains and countries. We presented our procedure and reflected on our lessons learned. In future research, our approach should be applied in different settings to validate the approach and to be able to give recommendations for different settings. Further, the evaluation of the participatory evaluation approach itself should be performed. Such a meta-analysis containing the definition of criteria and measures to systematically investigate and demonstrate how stakeholders perceived an evaluation would be beneficial to support further projects that aim to apply this approach.

Acknowledgement

The research leading to the presented results was partially funded by the European Commission under the 7th Framework Programme, Integrating Project LEARNING LAYERS (Contract no. 318209). The Know-Center is

funded within the Austrian COMET Program - Competence Centers for Excellent Technologies - managed by the Austrian Research Promotion Agency FFG.

7. Literature

BOUD, D. & MIDDLETON, H. 2003. Learning from others at work: communities of practice and informal learning. Journal of workplace learning, 15**,** 194-202.

CAMPILAN, D. 2000. Participatory Evaluation of Participatory Research. Forum on Evaluation of International Cooperation Projects. . Nagoya University, Japan.

CLARKE, N. 2004. HRD and the challenges of assessing learning in the workplace. International Journal of Training and Development, 8**,** 140-156.

EHRET, C. & HOLLETT, T. 2016. Affective Dimensions of Participatory Design Research in Informal Learning Environments: Placemaking, Belonging, and Correspondence. Cognition and Instruction, 34**,** 250-258.

ERAUT, M. 2004. Informal learning in the workplace. Studies in continuing education, 26**,** 247-273.

ERAUT, M. & HIRSCH, W. 2007. The Significance of Workplace Learning for Individuals, Groups and Organisations Oxford & Cardiff, UK: SKOPE Research Centre.

FITZPATRICK, J. L., SANDERS, J. R. & WORTHEN, B. R. 2004. Program evaluation: Alternative approaches and practical guidelines.

GARRICK, J. 1998. Informal learning in the workplace: Unmasking human resource development, Psychology Press.

GOPALAKRISHNAN, S., PRESKILL, H. & LU, S. 2013. Next Generation Evaluation: Embracing Complexity, Connectivity, and Change: A Learning Brief.: FSG.

GUIJT, I. 2014. Participatory Approaches: Methodological Briefs-Impact Evaluation No. 5.

HART, J. 2011. Social learning handbook, Centre for Learning & Performance Technologies.

KHOZA, L. B. 2015. Nursing Students' Perception of Clinical Learning Experiences. Journal of Human Ecology, 51**,** 103-110.

LEY, T., COOK, J., DENNERLEIN, S., KRAVCIK, M., KUNZMANN, C., PATA, K., PURMA, J., SANDARS, J., SANTOS, P. & SCHMIDT, A. 2014. Scaling informal learning at the workplace: A model and four designs from a large-scale design-based research effort. British Journal of Educational Technology, 45**,** 1036-1048.

LEY, T., MAIER, R., WAIZENEGGER, L., MANHART, M., PATA, K., TREASURE-JONES, T., SARGIANNI, C., THALMANN. 2017. Knowledge Appropriation in Informal Workplace Learning, Project Report, http://results.learning-layers.eu/scenarios/knowledge-appropriation/

LEINONEN, T. and DURALL-GAZULLA, E., 2014. Design thinking and collaborative learning. Comunicar, 21(42).

LINDSTAEDT, S., KUMP, B., BEHAM, G., PAMMER, V., LEY, T., DOTAN, A. & HOOG, R. D. Providing Varying Degrees of Guidance for Work-Integrated Learning. In: M. WOLPERS, P. A. K., M. SCHEFFEL, S. LINDSTAEDT, & V. DIMITROVA, ed. 5th European Conference on Technology Enhanced Learning, 2010 Barcelona, Spain. Springer LNCS, 213-228

LITTLEJOHN, A. & MARGARYAN, A. 2014. Technology-Enhanced Professional Learning. International Handbook of Research in Professional and Practice-based Learning. Springer.

MAIER, R. & SCHMIDT, A. 2014. Explaining organizational knowledge creation with a knowledge maturing model. Knowledge Management Research & Practice.

MAIER, R. & THALMANN, S. 2012. Collaborative Ethnography for Information Systems Research Studying Knowledge Work Practices and Designing Supportive Information Systems. Australasian Journal of Information Systems, 17**,** 137-160.

MALECKI, E. J. 2010. Global knowledge and creativity: new challenges for firms and regions. Regional studies, 44**,** 1033-1052.

MANUTI, A., PASTORE, S., SCARDIGNO, A. F., GIANCASPRO, M. L. & MORCIANO, D. 2015. Formal and informal learning in the workplace: a research review. International Journal of Training and Development, 19**,** 1-17.

MARSICK, V. J. 2009. Toward a unifying framework to support informal learning theory, research and practice. Journal of Workplace Learning, 21**,** 265-275.

MARSICK, V. J., VOLPE, M. & WATKINS, K. E. 1999. Theory and practice of informal learning in the knowledge era. Advances in Developing Human Resources, 1**,** 80-95.

MARSICK, V. J. & WATKINS, K. 2015. Informal and Incidental Learning in the Workplace (Routledge Revivals), Routledge.

MAYRING, P. 2014. Qualitative content analysis-theoretical foundation, basic procedures and software solution, Klagenfurt, Austria, Social Science Open Access Repository.

MITCHELL, R. K., AGLE, B. R. & WOOD, D. J. 1997. Toward a theory of stakeholder identification and salience: Defining the principle of who and what really counts. Academy of management review, 22**,** 853-886.

MORTIMER, S., PARK, J., MAUCHLINE, A., HAYSOM, K., WESTBURY, D., PURVIS, G., LOUWAGIE, G., NORTHEY, G., FINN, J. & KNICKEL, K. 2008. The Agri-Environmental Footprint Index: User's Manual.

MUUKKONEN, H., HAKKARAINEN, K. & LAKKALA, M. Collaborative technology for facilitating progressive inquiry: Future learning environment tools. Proceedings of the 1999 conference on Computer support for collaborative learning, 1999. International Society of the Learning Sciences, 51.

PATTON, M. Q. 2002. Qualitative Research & Evaluation Methods, Thousand Oaks, Sage.

POULOUDI, A. & WHITLEY, E. A. 1997. Stakeholder identification in inter-organizational systems: gaining insights for drug use management systems. European journal of information systems, 6**,** 1-14.

SCHÄPER, S. & THALMANN, S. 2014. Technology Support for Informal Learning in Mobile Work Situations. In: KUNDISCH, D., SUHL, L. & BECKMANN, L. (eds.) MKWI 2014. Paderborn, Germany.: Universität Paderborn.

SCHUH, G., GARTZEN, T., RODENHAUSER, T. & MARKS, A. 2015. Promoting Work-based Learning through INDUSTRY 4.0. Procedia CIRP, 32**,** 82-87.

STEURER, M., THALMANN, S., MAIER, R., TREASURE-JONES, T., BIBBY, J. & KERR, M. Assessing Informal Social Learning at the Workplace–A Revalidation Case from Healthcare. In: KÖHLER, T., KAHNWALD, N. & SCHOOP, E., eds. Professional Knowledge Management Conference (ProWM 2015), 2015 Dresden, Germany.: TUDpress Verlag des Wissenschaften Dresden, 251 - 266.

VOS, J. F. & ACHTERKAMP, M. C. 2006. Stakeholder identification in innovation projects: Going beyond classification. European Journal of Innovation Management, 9**,** 161-178.

WAIZENEGGER, L., THALMANN, S., SARIGIANNI, C., ECKHARDT, A., KOLB, D., MAIER, R. & REMUS, U. 2016. FROM ISOLATION TO COLLABORATION-HOW THE INCREASING DIFFUSION OF MOBILE DEVICES HAS CHANGED PRACTICES OF KNOWLEDGE SHARING IN NON-OFFICE SETTINGS. 24th European Conference on Information Systems. Istanbul, Turkey: AIS.