In-app Reflection Guidance: Lessons Learned across Four Field Trials at the Workplace

Angela Fessl,Gudrun Wesiak,Verónica Rivera-Pelayo,Sandra Feyertag,and Viktoria Pammer

Abstract—This paper presents a concept for in-app reflection guidance and its evaluation in four work-related field trials. By synthesizing across four field trials, we can show that computer-based reflection guidance can function in the workplace, in the sense of being accepted as technology, being perceived as useful and leading to reflective learning. This is encouraging for all endeavours aiming to transfer existing knowledge on reflection supportive technology from educational settings to the workplace. However, reflective learning in our studies was mostly visible to limited depth in textual entries made in the applications themselves; and proactive reflection guidance technology like prompts were often found to be disruptive. We offer these two issues as highly relevant questions for future research.

Index Terms—Workplace learning, reflective learning, reflection guidance, reflective note analysis.

◆

1 INTRODUCTION

REFLECTIVE learning is the conscious re-evaluation of past experiences with the goal to learn from them. This is in line with the definition of Boud et al. [1], who define reflective learning as *"those intellectual and affective activities in which individuals engage to explore their experiences in order to lead to new understandings and appreciations"*. In the area of workplace learning, reflective learning has been identified as a crucial method for informal learning, as it does not rely on explicitly available learning material, curricula, or teachers [2], [3], [4]. *"Like many other cognitive activities, reflection is often spontaneous, and, at times, outside an individuals awareness"* as stated by Daudelin [5]. Reflective learning is a cognitive process based on the individual's intrinsic motivation and cannot be directly enforced. However, external impulses can be given to stimulate learners' motivation.

Techniques like prompts or reflective diaries can foster reflection by reminding, helping and motivating learners to reflect [6], [7], [8]. Such techniques can help to make reflective learning happen more systematically and bring reflection to learners' awareness. Computer support for reflective learning, and in particular for triggering reflective learning are well investigated in formal learning environments, but are less well investigated in workplace settings. Our research goal is to address this gap by exploring reflection guidance technology in workplace settings. Transferring successful reflection guidance technologies, like prompts or diaries, from formal learning environments to workplaces is not trivial. While in learning settings, the learning activity is the operative activity, in workplace settings, learning is a meta-activity. Thus, it needs to be carefully considered when and how reflection can and should happen within daily working routines. Such considerations

impact on the design of reflection guidance technology.

In this work, we present an in-app reflection guidance concept that designs reflection guidance as adaptive software components (e.g. different type of prompts, diaries, contextualisation). The concept was instantiated and evaluated in four workplace settings. The contribution of this paper is twofold: First, in synthesizing results across four field trials to highlight the underlying common concept for reflection guidance technology. Second, in providing empirical data and insights about reflection guidance technology in the workplace to learning technologies literature.

Three of the four trials have been reported in previous conference and journal publications [9], [10], [11], [12], [13], [14], but without emphasis on reflection guidance technology. A shorter version of this paper was published at the European Conference on Technology Enhanced Learning 2015 [15].

2 RELATED WORK

In this paper, we are interested in technology that can be used to guide reflective learning. There are two essentially distinct types of reflection guidance technologies. First, diaries, journals and e-portfolios are technologies that structure (some more, some less) learners' description of experiences and aim to engage learners in reflecting on those or reflecting while describing those. Second, prompts are proactive invitations or challenges that motivate learners to reflect. Below we describe existing literature on both types (Sections 2.1 and 2.2); and we take them up as interventions in our reflection guidance concept. In our concept, we conceptualise diaries as "reflection-on-action" components and prompts as "reflection-in-action" components. The terms "reflection-on-action" and "reflection-in-action" lead back to Schön [16]. By framing diaries and prompts in relation to Schön, we emphasize the different timing of diaries and prompts with regard to the activity into which reflective learning is embedded: Reflection-in-action components aim to engage users in reflecting close to the operative activity, while reflection-on-action components aim to engage users

- *A. Fessl, G. Wesiak. and S. Feyertag are from the Know-Center, Inffeldgasse 13, 8010 Graz, Austria.*
  E-mail: afessl@know-center.at
- *Verónica Rivera-Pelayo is from the Forschungszentrum Informatik, Karlsruhe, Germany.*
- *Viktoria Pammer is from the Knowledge Technologies Institute, TU Graz A-8010 Graz, Austria.*

in reflecting later in time with respect to the activity.

We also relate to open learner modelling literature (Section 2.3) in the sense that reflection guidance technology that we analyse in this paper is embedded in applications that collect and present information about the activity in which reflective learning is embedded.

We conclude the related work section with a re-statement of this paper's contribution in the context of such related work, and a statement of research questions (Section 2.4.

2.1 Diaries, Journals and E-Portfolios

Diaries are personal notes not meant to be shared, hence most studies focus on the effect diary writing has on learning e.g. [17]. In contrast, the term "journal" is mostly used to describe diaries that are intended for sharing. They are used in different training settings (e.g. athletes) [18] and sometimes also in work-related settings especially in medical education of nurses [19], [20], [21]. E-Portfolios are mostly defined as a collection of information and (physical) artefacts, gathered for specific purposes by a user over time [22], in educational [23] as well as work-related settings [24].

The joint goal of diaries, journals and e-portfolios is twofold. First, they support describing experiences by providing a technology for content creation (e.g., allowing text). Second, they need to provide easy accessibility and sophisticated visualisations to engage learners in re-visiting these experiences. Creating diaries, journals or e-portfolios is very time consuming, however. This might be the main reason why they are very widely used in educational settings (all above cited references except [24]), and very little reported in professional settings.

2.2 Prompts

By reflection prompts, we understand interventions such as small text messages that try to motivate a user to reflect. Core challenges of good prompts are right timing in the sense of not interrupting a user where (s)he should not be interrupted, and context-awareness in the sense of adapting (referring to) context where possible. Thillmann et al. [25] showed that the timing of a prompt affects the learning outcome. To reduce the stress of interruptions through prompts on mobile devices, Ho and Intille [26] presented prompts directly after the completion of actions, while Pejovic and Musolesi [27] identify opportune moments for interruption by classifying relevant sensor data for context recognition.

Often, prompts consist of a question, sentence starter (e.g. "What I am thinking about now is...") or a direct instruction. The goal of prompts is to focus user attention on relevant aspects or experiences of their learning and working activities. In learning management systems, prompts emerged to be a viable and appropriate approach for guiding and initiating reflective learning [6], [8]. Here, prompts are used very often to organise, retrieve, monitor or evaluate knowledge as well as to reflect on students' learning [8], [28], [29], [30].

Davis [6] distinguishes between *self-monitoring prompts* and *activity prompts*. Self-monitoring prompts encourage students to reflect on their own learning by asking "Thinking ahead" or "Checking our understanding" questions. Activity prompts motivate students to reflect on their progress in the activity and specifically about whether they have devoted attention to each aspect of their project. In both cases, different types of sentence starters are used. Davis [7] differentiates between *generic* and *directed* prompts. While the generic prompts only ask learners to stop and reflect about their current problem solving activities, the directed prompts also provide them with an expert model of reflective thinking in the problem solving process. Furthermore, Ifenthaler et al. [8] follow the categorisation of prompts from Davis [7] and refer to *generic* and *directed* prompts. They investigated the efficiency of the prompts in form of effective instructional interventions for self-regulated learning in problem solving processes. They stated that generic prompts give learners more guidance to use different problem solving strategies but at the same time retain their autonomy to a certain extent. In contrast, directed prompts give more step-by-step instructions and restrict autonomous work. For our work, we follow Verpoorten et al. [31], who created the term *reflection amplifier*, which is a *"deliberate and well-considered prompting approach, which offers learners a structured opportunity to examine and evaluate their own learning"*.

2.3 Open Learner Modelling

User models are traditionally understood as models that computer systems have of their users [32]. In learning systems, user models often contain information about users' "knowledge, interests, goals, background and individual traits" [33]. Such models are then used by computers to adapt their behaviour or information representation, but for such system adaptation the user models need not be accessible to users. However, one strand of research is to investigate the usefulness of making user models accessible to learners. Such user models in learning systems are then called "open learner models". The core idea for making learner models open is that users can see what the learning system knows about their knowledge, and use this as basis for reflecting on their learning status and progress, and to plan further learning activities [34], [35], [36]. Relating such open learner modelling to Boud et al. [1], the learner models' content represents the learning experience. In learning settings, the learning experience is the object of reflection, i.e. what the learners reflect on. In this sense, open learner models are similar to what the applications do, when reflection guidance technology is embedded: they collect information about a user activity, and represent it to users as basis for reflection.

Typical challenges addressed in the field of open learner modelling are system architecture and technical challenges such as privacy and security, or the use of ontologies and standards to describe user characteristics [36]. A second large research interest is that of easy to understand visualisations of user model content (see e.g., [37], [38]).

While research on open learner models is mostly focussed on educational settings however, our research directly targets learning in self-regulated workplace settings. Kay & Kummerfield [36] are one of the few papers from open learner modelling research who describe, albeit visionary, lifelong learning settings that go beyond formal education (the example they give is on physical exercise).

2.4 Contribution and Research Questions

As described above, most related research has been carried out in formal learning settings. A transfer of technologies and support strategies from formal learning environments to workplaces is not straightforward: Investigating and evaluating prompts in learning environments can benefit from the fact that the learning activities of the learners are known beforehand and prompts can easily be adapted according to these activities. In contrast, in a work-related setting tasks are not always known beforehand or only vaguely known. Secondly, because a learner uses the learning environment actively while learning, prompts can be displayed according to the current learning task. In a working environment, learning is a secondary activity in relation to working. This means that prompts risk to be more disruptive with respect to the users primary activity than in learning settings. Third, most workplaces are stressful and work-intensive. Thus, it has to be very well considered, if and how time consuming approaches like diaries are applicable in work related settings.

The contribution of this paper lies in providing empirical data and insights about reflection guidance technology in the workplace to learning technologies literature. We have conceptually aligned reflection guidance technology with Schön's distinction between reflection-in-action and reflection-on-action. Parts of the overall concept can also be related to activity and self-monitoring prompts by Davis [6] and to reflection amplifiers as described by Verpoorten [31]. This overall concept has been instantiated in four field trials. Our research questions, evaluation plan, and tools follow Kirkpatrick's model [39] for assessing training effectiveness in organisations. Kirkpatrick argues that learning should be evaluated at four levels that build up on each other in the sense of one level needing to be evaluated "positively" before success can be achieved at the next level. The levels are reaction (how users react to technology), learning (whether and what learning occurs), behaviour (how behaviour of people is changed) and results (work performance).

We have evaluated the implemented reflection guidance technology at the first two levels "reaction" and "learning", and instantiated them with the following research questions:

- **RQ1: Usage** (How) Do participants use the different reflection guidance components?
- **RQ2: Learning - Support by Intervention** Which reflective learning activities are supported by interventions?
- **RQ3: Learning - Depth** To what depth does reflective learning occur?
- **RQ4: Design challenges for interventions** What are the design challenges for reflection guidance technology in the workplace?

3 METHODOLOGY

We carried out four field trials in which we place interventions in terms of applications that include reflection guidance technology. In each trial we explore the impact of reflection guidance technology on learning along the four research questions. The apps and reflection guidance technology was developed in an iterative design process individually for each field trial prior to the actual trials. The field trials are bound together by the same research questions, evaluation procedure and concept for designing reflection guidance technology. Below we first describe the common evaluation procedure (Section 3.1) and the common reflection guidance concept (Section 3.2). Then we describe the specific instantiations of reflection guidance in the apps (Section 3.3) and the specifics of each trial (Section 3.4).

3.1 Evaluation Procedure and Tools

All four field trials followed the same procedure and used the following evaluation tools: At the beginning of each field trial, the corresponding application was introduced to the participants. During this phase, all participants were asked to fill in the pre-questionnaire. Afterwards the participants were asked to integrate and use the presented applications in their daily working routines. At the end of the evaluation period, all participants were asked to fill in a post-questionnaire. Several of them were also asked to participate in an interview. In the clinic, an on-site workshop was conducted additionally. Table 1 gives an overview of the tools and their relation to the research question.

TABLE 1
Evaluation tools and relation to the research question

Research Questions	Tool	Field trial	Stage	Content
-	pre-questionnaire	all	at the beginning	consent form, demographic data
RQ1	log data	all	during usage	interactions with the apps, quiz results
RQ2	post-questionnaire	all	at the end	questions about the apps and reflection guidance components
RQ3	coding schema	Mood-map, Know-Self	after the evaluations	analysis of the inserted text snippets
RQ4	post-questionnaire	all	at the end	questions about the role of the reflection guidance components
RQ4	interviews/ workshop	all	at the end	additional feedback about the apps and the guidance components

A pre-questionnaire was used to gather demographic data about the participants.

To answer research question RQ1 on usage, the objective usage rates were captured via users' logdata and written text entries were collected within the applications.

To answer research question RQ2 on reflective learning supported by apps, the post-questionnaire contained questions on how supportive the apps were with respect to reflective learning, so-called app-specific reflection questions. For these questions we used an existing item pool of 43 questions (see [40], p. 38), which cover a wide variety of features reflection apps may provide. Out of this, an adequate subset of items was selected for each evaluation (see Table 4). All items were presented as 5 pt. rating scales ranging from 1 - "I strongly disagree" to 5 - "I strongly agree". The questions ask in what way the overall intervention (not just the particular reflection guidance technology in the app)

supported reflective learning. They also refer to activities that are relevant in reflective learning, i.e. reconstructing work experience. Note that this means, that we measured the support for learning via self-assessment of users.

To answer research question RQ3 on learning in terms of depth of reflection, we used a qualitative analysis schema for written statements of reflection [41]. We applied this schema to the notes that participants inserted in the apps.

To answer research question RQ4 on the design challenges for interventions of the implemented reflection guidance components, the post-questionnaires, interviews and workshops cover questions adapted to the implemented intervention per app and field trial.

3.2 Reflection Guidance Concept

The reflection guidance concept describes three app-independent component types. The three types of component take into account the challenges stated above and address in particular the consideration of timing (reflection-in-action and reflection-on-action components) and context (contextualisation components).

3.2.1 Reflection-in-Action Components

Schön [16] defines *reflection-in-action* as reflection which takes place while doing own work. "Reflection-in-action" components present triggers while doing work to initiate reflective learning. These triggers in form of prompts make people aware of topics worth being reflected on including but not limited to unusual behaviour or changes, monitor progress, comparing and encouraging oneself or others. We define two types of reflection-in-action components:

Reflection interventions: Reflection interventions are prompts that motivate users to do something, typically to use the app into which reflection guidance is implemented, such as adding content. Reflection interventions are similar to the *activity prompts* by Davis [6], which have the goal to facilitate the completion of specific aspects of an activity.

Reflection amplifiers: We understand reflection amplifiers as prompts that do some sort of content or behaviour analysis, and adapt to the results of this analysis. As they relate to actual statements made by or activities carried out by the user, they are ideally more on target (in the sense of personalisation) than reflection interventions. This understanding of reflection amplifiers is in line with Verpoorten et al. [31] who describe reflection amplifiers as providing *"a structured opportunity to examine and evaluate their own learning"*. Reflection amplifiers are also similar to Davis [6] *self-monitoring prompts*, which provide questions or sentence starters to initiate reflection with the goal of knowledge integration when answering these questions.

3.2.2 Reflection-on-Action Components

Schön [16] defines *reflection-on-action* as analysing reactions to a situation and exploring the reasons and consequences well after the situation. "Reflection-on-action" components are features that aim to motivate users to reflect about the gathered data after having used the application into which reflection guidance technology is embedded, e.g. at the end of a working day. These components consist of reports as

well as reflection diaries or reflection journals. "Reflection-on-action" components encompass:

Reports: By reports we mean automatically created summaries of past experiences or events captured during work. The aim of reports is to provide an overview data basis for reflection in relation to different time periods (e.g. hourly, daily, weekly reports). Re-evaluating own captured experiences or comparing them with others' experiences can lead to the detection of behavioural patterns, insights about one's own working behaviour and result in learning outcomes.

Reflection Diary: A reflection diary stores insights, thoughts, artefacts, or learning outcomes in a structured way. It is typically suggested to using such a diary regularly (e.g. at the end of each shift). Often, a diary makes visible learning outcomes. Re-evaluating these diary entries again can trigger reflective learning.

3.2.3 Contextualisation Components

In order to be able to return to experiences, which is a relevant step for reflection as suggested by Boud et al. [1], it is important to be able to recall sufficiently the context of the experience. All mechanisms for recall are by nature faulty - human memory is, automatic activity tracking methods are as well - however they still provide enough hints to be able to re-evaluate past experiences. In software tools, we distinguish contextualisation components that track context automatically and those that provide support for manual descriptions of context. In either case, contextualisation components aim to provide users with context to the main data captured within an application / the main activity supported by an application.

3.3 Instantiating the Reflection Guidance Concept
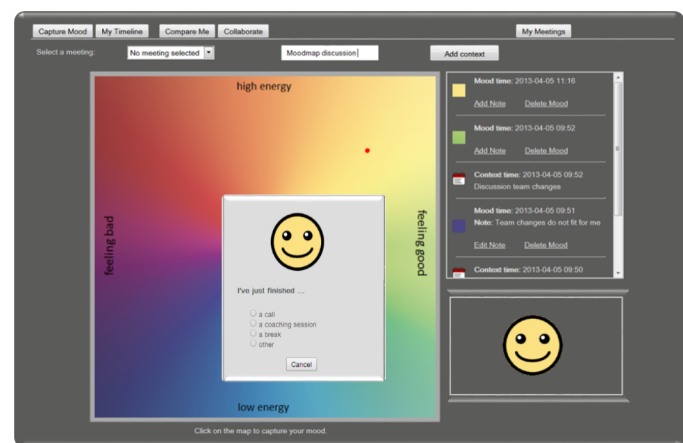
3.3.1 Reflection Guidance in the MoodMap App

Fig. 1. The MoodMap App: Mood can be stated by clicking on the colourful map. Afterwards the user has to select a mandatory context.

The MoodMap App was used in two field trials, and therefore developed in two versions, Version A and B. The baseline functionality is the same in both versions, but they contain different reflection guidance components.

The *MoodMap App* [9], [14] is a web application that maps mood on a coloured map along the two dimensions of Russell's Circumplex Model of Affect: valence (feeling good

- feeling bad) and arousal (high energy - low energy) [42]. Mood is captured by clicking on a bi-dimensional mood map coloured according to Ittens system [43] (see Figure 1). Various implemented visualisations (e.g. own mood development over time, own mood in comparison to average team mood) present the captured moods on an individual as well as collaborative level to provide different perspectives about the data and to trigger reflective learning about the mood development. Furthermore, two reports are implemented summarizing the mood development during a day from different perspectives.

Reflection Guidance Components:

Version A - Contextualisation: The contextualisation component is implemented in a mandatory way. Making the contextualisation mandatory came up in user-centered software design iterations conducted together with the target groups. Thus, for each inserted mood, a pre-defined context has to be chosen out of the following four options especially tailored to a call centre setting: "after a call", "after a coaching session", "after a break" or "other" (shown in Figure 1, top). Afterwards, a note has to be added, which is motivated by a reflection amplifier question like *"What are you currently thinking?"* or *"What influences your current mood?"*. The purpose of this mandatory contextualisation is to 'force' the user to think about the current inserted mood and add the user's thoughts into the application.

Version B - Reflection-in-action components: The reflection interventions occur automatically when a user enters the MoodMap App for the first time of a day, when the user enters a meeting for the first time, or when the application was not used for a longer time period, e.g. *"Welcome to the MoodMap App. Express yourself for the first time today."*. Furthermore reflection amplifiers automatically pop up during the application usage to motivate users to reflect about the current mood. This is done by making the user aware of a significant mood change of the own mood but also in relation to the collaborative mood and by directly asking for input. The amplifiers encompass questions like *"You are more energized now, but you are not feeling very good. What happened?"* or *"Your mood changed to a more positive mood than the team mood. What was the reason for this change?"*.

Version A and B - Reflection-on-action components: Two types of reports are implemented. One report summarizes the individual mood development of a day. The other report presents the individual mood development with regard to the team mood development in order to stimulate reflection. In both reports, a reflection diary is available to store insights gained.

3.3.2 Reflection Guidance in the Medical Quiz

The *Medical Quiz* [10], [11] is a web-based quiz especially developed for nurses working at a Stroke Unit in German hospitals (see Figure 2). Four different quiz types were implemented: A Quiz-against-time (answer as many questions as possible in 5 minutes), the Quiz-of-20 (answer twenty questions), the Quiz-of-10 (answer 10 questions) and the Quiz-of-5 (answer 5 questions). The content-based questions, randomly chosen out of a database, are multiple-choice or single-choice questions, which provide users the possibility to check out their current knowledge status. All content-based questions were developed by nurses and

Fig. 2. The Medical Quiz: On top, a content-based question is shown, below a reflective question is presented to motivate to reflect.

physicians working at the German Stroke Unit.

Reflection Guidance Components:

In the Medical Quiz only reflection-in-action components are implemented, neither contextualisation nor reflection-on-action components are available. These reflective questions - an integrated form of reflection amplifiers - are presented at the beginning, during and at the end of a quiz. At the beginning of the quiz, the presented reflective question depends on the current knowledge status (tracked in previous quiz attempts) and the user's play frequency (how often the user already played the quiz). Additionally, a motivating introduction statement is presented. The goal of the reflective question is to make the player aware of the individual knowledge status and play frequency and to reflect about it. For example: *"You are very motivated and you play the quiz at least once per week - your results are rather good, but could be further improved. What is your recipe for success?"*. The in-between questions (shown in Figure 2), only available in the Quiz-of-20, put focus on the content-based questions and how they refer to past working situations and working experiences, e.g. *"If you consider the question above and think back to the qualification program or a situation during work, which knowledge can you gain out of it?"*. The questions at the end of the quiz ask explicitly for gained insights or new knowledge with regard to the currently played quiz or try to draw a relation to the player's work, e.g. *"Reflect on the currently played quiz. In what regards are the quiz questions related to your work?"*. This type of question is used in all quiz types but the Quiz-against-time.

3.3.3 Reflection Guidance in KnowSelf

KnowSelf [12], [13] is designed to support individual reflective learning regarding time management and self-organisation of knowledge workers during work (see Figure 3). The application automatically records working activities on a personal computer by capturing resources and applications used during work. Manual project and task
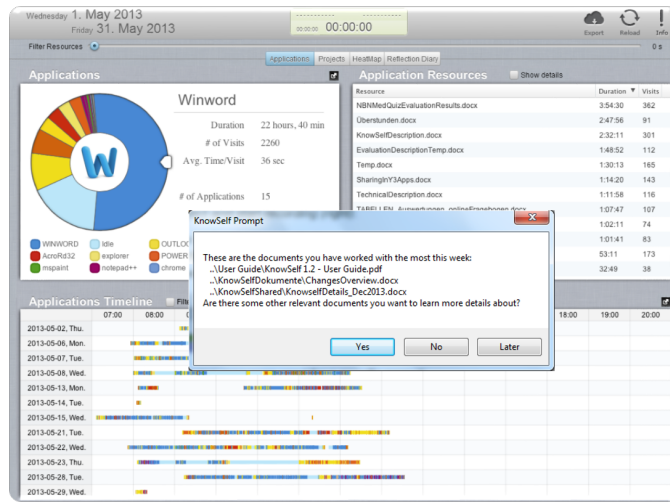
Fig. 3. KnowSelf: Used resources during work are shown. The reflection amplifier makes aware of the most used resources of the week.

recording, as well as manually inserted notes and comments complete the data captured by KnowSelf. The application visualizations expose, in particular, the frequency of application and resource switching, the time spent in numerous applications and the time spent on different activities.

Reflection Guidance Components:
In KnowSelf, contextualisation, reflection-in-action as well as reflection-on-action components are implemented.
Contextualisation: As KnowSelf is an activity tracking tool it automatically captures the current context of the user including currently used applications and resources.
Reflection-in-action Components: In KnowSelf, time-triggered reflection interventions and event-triggered reflection amplifiers are implemented. The time-triggered reflection interventions (see in Figure 3) are shown on a specific day at a specific time which can be adjusted by the user and remind the user to reflect by offering feedback on various topics regarding the user's behaviour. One example for a time-triggered prompt is "On _DATE_ you reached the highest level of switches between _FROMTIME_ and _TOTIME_. Do you have an idea what is the reason for it? Would you like to see your recordings in the App which could help you to reflect on it?". This prompt might show up every Friday. The event-triggered reflection amplifiers are generated when a significant change has been detected, e.g. when the number of switches is higher than usual or after long periods of idle time. An example for an event-triggered prompt is "On _DATE_ between _FROMTIME_ and _TOTIME_ you had higher switches than you usually have in that time. Do you have any idea why? Would you like to take a look at your captured activities in the App and reflect on it?". It appears when the number of switches during a 15 min interval gets unusual high compared to the user's average switching frequency.
Reflection-on-action components: A reflection diary is implemented, where all notes, thoughts or reflection insights can be stored. Different visualisations presenting the captured applications and resources used serve as report of the user's working activities.

3.4 Field Trial Settings

Table 2 gives an overview of the four conducted field trials, the corresponding work-settings, the duration and the number of participants. Below, we describe each setting in terms of organisational setting, study participants, and specifics of the evaluation procedure. The latter describes per field trial who introduced the reflection intervention (app including instantiation of reflection guidance concept), and what were concrete instructions to participants.

TABLE 2
Evaluation Overview

Field trial setting	App	Intervention	Duration	Participants
Call Centre (GB)	MoodMap App (Version A)	Contextualisation, Reflection-on-action	4 weeks	34 participants (30 call takers, 2 coaches, 2 managers)
IT Company (IT)	MoodMap App (Version B)	Reflection-in-action, Reflection-on-action	6 weeks	32 employees from five departments
Stroke Unit (DE)	Medical Quiz	Reflection-in-action	4 months	21 nurses
IT Company (DE)	KnowSelf	Contextualisation Reflection-in-action, Reflection-on-action	6 weeks	10 consultants

3.4.1 MoodMap App in a Call Centre

Setting: Four teams out of two call centres of a large telecommunications company in Great Britain participated in this trial (see [44], p. 52). The two call centres are situated in Scotland and together they have about 450 employees. They are part of a large telecommunications company, serving customers in more than 170 countries. The first call centre employs more than 300 people (divided in 19 teams) and 157 people (divided in 7 teams) work in the second centre. Such call centres can handle an average of 27.000 calls every day, working in a range of functions from directory enquiries to residential and business broadband. Each team consists of call takers, coaches and managers. The call takers are responsible for taking the calls and solving any issues directly with the customers. The coaches support and coach the call takers for their work, while the managers ensure that the call takers are performing against their targets, review their performance and supervise the training sessions.

Participants: From 34 participants (2 managers, 2 coaches, 30 call takers) who used the application and filled out the post-questionnaire, 2 gave interviews and 24 shared their demographics. Of these, 14 are male, 10 female, 16 were aged between 20 and 29, 6 between 30 and 39, 2 between 40 and 49. On average they were 3 years in their position.

Learning Domain: The learning domain is to reflect on own mood in relation to calls, and on the overall mood in the call taking team. The participants are professionals in their work, and consideration of emotions and emotion regulation are core parts of work in call centres [45].

Procedure specifics: The MoodMap App with contextualisation and reflection-on-action components (version A) was introduced by the project manager of the telecommunications company. All participants (call takers, coaches and managers) were asked to integrate the MoodMap App in their daily working routines. Their task was to state their moods, choose a mandatory context, add a note during all calls, and reflect about them within the evaluation period of 4 weeks. Managers were asked to use the team visualisations and take actions if necessary, while coaches were asked to use the app during the coaching session.

3.4.2 MoodMap App in an IT Company

Setting: The Italian company is leading in the development of software-technology and cloud services in the emergency domain. This encompasses the development of information and communication technology systems for emergency centres and volunteering associations. The company consists of altogether 38 employees distributed over 5 departments. In this trial, the MoodMap App with the implemented reflection interventions and reflection amplifiers was evaluated.

Participants: 32 employees, 24 male and 8 female, have participated in this field trial. They were aged between 20 and 59, with 3 of them aged between 20 and 29, 27 between 30 and 39, 4 between 40 and 49 and 1 between 50 and 59. They had worked on average 7 years in their position.

Learning Domain: The learning domain is to reflect on own and others' mood in relation to work procedure, and thus on overall wellbeing and collaboration at work. The participants are professionals in their work.

Procedure specifics: The MoodMap App with reflection-in-action and reflection-on-action components (version B) was introduced by a responsible project manager. The participants were asked to insert their moods during their working shifts and to reflect about their inserted moods and notes individually. The managers were additionally instructed to use the team visualisations in order to reflect about the mood development of their departments and take actions if necessary with regard to each team member.

3.4.3 Medical Quiz in a Stroke Unit

Setting: The Medical Quiz was introduced to the nurses participating in a qualification program for stroke nurses, which takes place once a year in a German hospital. (see [44], p.137). All participants were nurses working in different German hospitals and were in education to become a nurse for a Stroke Unit. The qualification program lasted for 4 months, starting in October 2013 lasting until January 2014. In one week of each month the participants came together for the training at the organising hospital.

Participants: 18 nurses (16 female, 2 male) participated in this evaluation and 3 gave interviews; 12 were aged from 20 − 29 and 6 between 30 and 59. The average time in their current position was 4 years.

Learning Domain: The learning domain is medical knowledge for stroke nurses. The quiz aims to help nurses prepare for a formal exam; the reflective questions in the quiz aim to help nurses reflect on their learning state and strengthen the connection between theory and practice (see [10]). Nurses are already experienced in the learning domain.

Procedure specifics: At the beginning of the qualification program, the Medical Quiz was introduced to the participants and they were asked to play the quizzes consistently. They should use the content questions for memorising and strengthening their newly gained knowledge and the integrated reflective questions for reflecting about the quiz content to put theoretical knowledge into practice. At the end of the evaluation, a half-day workshop and interviews were conducted to collect deeper insights from the participants with regard to the quiz and its reflection support.

3.4.4 KnowSelf in an IT Company

Setting: The evaluation took place in a German consulting company that consults, sells, and personalises customer relationship management software to help analyse and optimise the marketing, sales and service processes of their customer companies (small and medium enterprises). Employees mainly work in small teams of two to three people. Altogether, the company has about 60 employees, most of them based in the headquarters. The employees have many meetings with customers at the customers' site and their daily work is heavily focused on customers' needs. This requires a high level of flexibility and the development of individual best practices. Consulting and sales involve a high degree of reflection regarding interactions with customers.

Participants: 10 employees (3 females, 7 males) participated in the evaluation and 7 gave interviews; 1 was younger than 19 years, 4 each were aged from 20 − 29 and from 30 − 39 and 1 were aged from 40 − 49. All of them were employed full-time with three-quarter of them on the management level. On average, the participants have been working in their current positions for more than one year with total work experience of about 4 years.

Learning Domain: The learning domain is time management, which is an essential part of consultants' work.

Procedure specifics: In a kick-off meeting at the company, the application was presented and explained to the participants by the evaluation responsible of the company (see [44], p. 25). They were asked to install KnowSelf on their business PCs and to use the application to daily reflect on the collected data.

4 RESULTS

The presentation of the results is structured along the research questions and the respective evaluation tools. First, we present the objective usage rates captured via users' log-data. Second, we present the results of the post-questionnaire with regard to the support participants felt was given to them for reflective learning activities by the interventions. Third, we describe the results of applying the coding schema for reflection on textual entries made in the apps. Fourth, we describe results of the design challenges for reflection guidance technology based on the post-questionnaire in 3 trials, on interviews from one trial, and on usage of reflection guidance components from 4 trials.

4.1 RQ1: Usage

The MoodMap trial (Version A) at the call centre had 34 participants; the MoodMap trial (Version B) at the IT

company had 32 participants; the Medical Quiz trial had 18 participants and the KnowSelf trial 10 participants. The average number of interactions (e.g. visiting different visualisations) of the participants with the respective intervention is summarized in Table 3, including the mean, standard deviation, minima and maxima. In the MoodMap trial at the call centre participants interacted with the MoodMap on average $M = 42.67$ times ($SD = 72.7$). The MoodMap trial at the IT company showed on average 49.78 interactions with the application ($SD = 82.75$). KnowSelf was used on average for $M = 38.5$ minutes (SD = 23.8) by each participant. The participants of the Medical Quiz answered on average $M = 461.9$ ($SD = 341.0$) questions. The high standard deviation shows that the usage of the apps was very polarized, having very active user, whereas others almost did not use the app.

TABLE 3
Usage data per trial

Setting	Usage type	Mean (Std)	Min	Max
MoodMap trial in a call centre	Interactions with app	$M = 42.67$ ($SD = 72.7$)	1	338
MoodMap trial in IT company	Interactions with app	$M = 49.78$ ($SD = 82.75$)	1	391
KnowSelf trial in IT company	Minutes of app usage	$M = 38.5$ ($SD = 23.8$)	7.76	53.72
Medical Quiz trail at Stroke Unit	Number of answered questions	$M = 461.9$ ($SD = 341.0$)	25	1358

4.2 RQ2: Learning - Support by Interventions

The post-questionnaires contained 7 app-specific reflection questions for both MoodMap trials and the Medical Quiz trial, and 5 questions for the KnowSelf trial (see Table 4).

TABLE 4
Application Specific Reflection Questions

Questions	Apps
[The app] helped me to collect information relevant to reconstructing experiences from work.	MoodMap, KnowSelf
[The app] helped me to reflect on experiences from work.	MoodMap
[The app] helped me to collect information that could help me decide when to reflect.	MoodMap
[The app] helped me to reconstruct a work experience.	MoodMap, Quiz
[The app] helped me by capturing my reflection outcomes.	MoodMap, KnowSelf
[The app] helped me by making reflection outcomes available for later use.	MoodMap, Quiz
[The app] helped me by capturing information for evaluation of learning/reflection.	Quiz
[The app] helped me by reminding me to reflect.	Quiz, KnowSelf
[The app] helped me by providing information relevant for the decision to reflect.	Quiz
[The app] helped me by providing accurate information about my work.	Quiz, KnowSelf
[The app] guided me in capturing information about my work experiences.	MoodMap
[The app] guided me in deciding whether/when to reflect.	Quiz
[The app] provided relevant content for reflection.	KnowSelf

Table 5 presents the mean ratings per trial as well as the

Pearson coefficents of correlating the extent of app usage and how users perceive the support for reflective learning. The correlations for the Medical Quiz ($r = .612$, $p = .015$) and KnowSelf ($r = .850$, $p = .015$) are significant while the correlations for both MoodMap trials are not significant.

TABLE 5
Pearson correlation of app usage and perceive app support for reflective learning (** - significant for $p < 0.01$)

Setting	Mean (SD)	Pearson coefficients
MoodMap trial in a call centre	$M = 3.25$ ($SD = .92$)	$r = .226$, $p = .247$
MoodMap trial in IT company	$M = 2.91$ ($SD = .75$)	$r = .24$, $p = .177$
KnowSelf trial in IT company	$M = 3.28$ ($SD = .66$)	$r = .850$, $p = .015$**
Medical Quiz trail at Stroke Unit	$M = 3.51$ ($SD = .42$)	$r = .612$, $p = .015$**

4.3 RQ3: Learning - Depth

We analysed all textual entries in the apps used in the field trials according to the coding schema [41]. This coding schema describes three levels of depth of reflection (low, medium, high), each of which are characterised by the content of the reflective expression. For instance, expressions that describe experiences and own emotions related to experiences count as low depth of reflection; expressions that contain interpretations or justifications of actions count as medium depth of reflection; expressions covering learning from reflection and drawing conclusions count as high depth of reflection.

Two raters rated all notes from the MoodMap trial (Version B) in the IT company, and three raters rated all notes from the MoodMap trial (Version A) in the call centre and from the KnowSelf trial. Below we describe results only on those notes on which we achieved full agreement between all raters. Agreement was achieved either in a first step of categorising notes independently or in a second step after discussing deviating assignments. For both MoodMap trials, the discussions resulted in 100% agreement.

Some of the notes belong to more than one category, however in those cases all raters agreed on all categories. Table 6 presents the number of notes per category and field trial, distributed according the three levels of reflection. Table 6 also shows for each trial how many notes we could not agree on as well as the number of notes that could not be applied to any categories of the coding schema. Textual entries in the Medical Quiz were too short to be analysed according to the coding schema; we provide a more descriptive analysis towards the end of this section. Table 7 shows examples of notes for each category of the coding schema.

In the MoodMap trial at the call centre, participants attached 475 non empty notes to 548 inserted moods. Out of these, 283 could be identified as individual reflective items after applying the coding schema. 95% of the reflective notes were assigned to the level of low reflective depth, which includes descriptions of experiences or emotions. The remaining 5% belong to the level of medium reflective depth, which refers to interpretations and justifications of actions and working on solutions.

TABLE 6
Analysis schema: number of notes per category and evaluation

Categories of reflection elements	Mood-Map call centre	Mood-Map IT company	Know-Self
Low reflective depth			
1. Description of an experience	141	150	54
2a. Own emotions	185	48	0
2ap. Emotions with physical conditions (e.g. pain, illness)	0	18	0
2b. Other emotions	50	0	0
Medium reflective depth			
3. Interpretation or justification of actions	17	14	33
4. Linking an experience explicitly to other experiences	3	0	1
5. Linking an experience to different pieces of knowledge, rules, values, organisational documents	0	0	0
6a. Responding to interpretation of the action (inquiry/different/alternate perspectives)	0	0	1
6b. Responding to interpretation of the action (challenging or supporting assumptions / opinions / attributions)	0	0	0
7a. Working on a solution based on assumptions, insights (explanation of reasons)	0	0	7
7b. Working on a solution (giving suggestions without proposing to set them in practice/referring to an experience)	1	0	4
High reflective depth			
8a. Insights / learning from reflection (different / better understanding of experience)	0	0	5
8b. Insights/learning from reflection (generalising from experiences, finding patterns across experiences)	0	0	6
9. Drawing conclusions and implications from reflection	0	0	8
Non-reflective notes			
	182	15	33
No agreement			
	0	0	11

TABLE 7
Analysis schema: examples

App & Organisation	Category	Example of notes
MoodMap - Call Centre	1: experience or issue	Coach is now dealing with the horrible case and its Friday! :)
MoodMap - Call centre	2a: own emotions	Talk with manager, feeling a bit more positive
MoodMap - IT company	2a, 2ap: own emotions and physical condition	I have a little bit of stomach ache, but I am feeling positive :-)
MoodMap - Call Centre	1, 3: interpretation of actions	Back and forth we go, another day of getting nowhere with our control desks. Honestly not sure why the customer wants to stay with [company name] at this stage, no one.
MoodMap - IT company	1,3: interpretation of actions	Emergencies that have distorted the work plan, overload of tasks and little fluidity (lack of feedback etc.).
MoodMap - Call Centre	1, 7b: solution suggestion	Process for important job and customer did not have much of a clue and had unrealistic expectations. Will have to refer to sales to move
KnowSelf - IT company	1: experience	Today I finished the references for the homepage
KnowSelf - IT company	1, 8b: experience and insights	On the subjective experience, less fragmentation during the day happened. Tasks can be better performed in a blocked way.
KnowSelf - IT company	7b, 9: solution suggestion and conclusion	Store documents solely in sharepoint, easier for versioning and without multiplication of documents

The coding schema could not be applied to the text entries inserted in response to the reflective questions in the Medical Quiz, because the text entries were too short, often consisting of one word only. Thus, for the quiz, we analysed the content of 603 answers regarding the most frequent words to get a general impression of participants' thoughts. For the questions at the beginning of the quiz the following keywords were used the most: repetition (40) and learning (27), concrete answers given were *"strengthen my knowledge through repetition"* and *"check my knowledge status"*. Very short answers were given to the in-between questions as for example: yes (145), no (38), and some longer answers stated *"I can apply theoretical knowledge in practice"* and *"answer questions of patients"*. Questions posed at the end of the quiz, were mostly answered with yes (55), practice (13), or more concrete: *"I can recognize my state of knowledge by answering the questions several times and enhance my knowledge accordingly."* or *"partly better understand medical orders"*.

4.4　RQ4: Design Challenges for Interventions

To identify design challenges for interventions, we used the post-questionnaires, the usage of the reflection guidance technologies and the conducted interviews.

4.4.1　Reflection-in-action Components

Reflection interventions in form of time-triggered prompts were implemented in KnowSelf and evaluated with two explicit questions in the post-questionnaire. Users rated the

In the MoodMap trial at the IT company 203 non empty notes were captured. 188 notes could be identified as reflective items. Beside the categories defined in the reflection coding schema, the category 2p was added to the schema. It represents the physical condition (e.g. pain, illness) related to own emotions and is thus defined as subcategory of 2a (own emotions). 94% of the reflective notes were assigned to the level of low reflective depth and 6% to the level of medium reflective depth.

From the 103 statements inserted in KnowSelf, 33 statements were classified as non-reflective and for 11 entries there was no inter-rater agreement. The remaining 59 statements were classified as individual reflection items. 45% of them were assigned to the level of low reflective depth, 39% to the level of medium depth of reflection, and 16% to the level of high reflective depth. For the latter, the diary entries included insights, learning outcomes, or conclusions drawn from reflecting.

reflection intervention reminding them to use KnowSelf in general slightly positive with $M = 3.22$ ($SD = 1.09$) whereas the reminder of project recording was perceived slightly negative with $M = 2.75$ ($SD = 1.28$) as shown in Table 8. Reflection interventions in the MoodMap trial at the IT company were not explicitly evaluated.

Reflection amplifiers were implemented in all four apps. For the MoodMap trial (Version A) at the call centre, the amplifier in form of a prompt to add a note to the mood resulted in 475 inserted notes. In this trial, the perceived usefulness of the prompts was not explicitly evaluated.

In the MoodMap trial (Version B) at the IT company two questions about the reflection amplifiers were inserted in the questionnaire. On average participants slightly disagreed that the questions about their own mood changes ($M = 2.84$, $SD = 1.05$) and their mood changes with regard to the team's mood ($M = 2.54$, $SD = 1.36$) motivated them to reflect. Altogether 2067 moods and 203 notes were inserted. Furthermore 1106 prompts (reflection interventions and reflection amplifiers) were automatically presented and 136 of them were answered. To get more detailed information about how the reflection amplifiers were perceived, seven participants, including two managers, four staff members and one evaluation responsible took part in an interview. All of the participants mentioned that the evaluation was conducted in a very stressful and work-intensive time, where they nearly had no time to use the MoodMap App properly. The managers mentioned that they have seen the reflective prompts, but used them very rarely. Also the coordinator admitted that he had never used the reflection amplifiers, nevertheless he stated that *"...it should be a good thing to do that"*. The staff members - except one - did use the amplifiers and one stated that *"Sometimes I feel the pop-up was very useful to remember me to write something about my mood. In other case I felt it was very annoying"* and another found it useful for giving context. These feedback showed that the opinions of participants who used the reflection-in-action components were very different ranging from very useful to annoying.

In KnowSelf, the usefulness of the event-triggered reflection amplifiers was explicitly investigated (see Table 8). Participants rated the notifications about most used resources as rather helpful ($M = 3.30$, $SD = 1.25$), whereas the notifications about unusual amounts of idle time and notifications about specific amounts of switches were not perceived as very helpful. The general rating regarding KnowSelf prompts and their potential to motivate participants to reflect was rated rather low $M = 2.00$ ($SD = 1.06$).

In the Medical Quiz, we integrated reflection amplifiers by presenting reflective questions at the beginning, during and after a quiz play. Figure 4 shows the results of the answered (bottom bars) and not answered (top bars) reflection amplifiers per quiz type.

For the Quiz-of-20, altogether 205 reflection questions at the beginning were shown and more than 50% were answered (Figure 4, "Q20 B"). For the Quiz-of-10 only 18% of the 53 posed reflective questions were answered (Figure 4, "Q10 B"), for the Quiz-of-5 38% out of 47 questions (Figure 4, "Q5 B") and for the Quiz-against-time 13% out of 51 questions were answered (Figure 4, "QaT B"). The two in-between reflective questions were only added to the Quiz-of-20, which all had the task to motivate the player
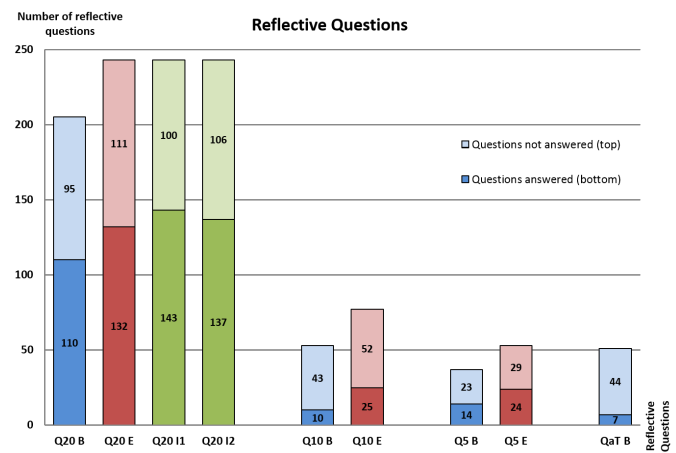

Fig. 4. Number of presented (full scale bars) and answered (bottom part) reflection questions: split into questions shown at the beginning (B), during (I1, I2) and at the end (E) of the quizzes.

to relate the presented content-based question to possible situations during work. For both in-between questions, the participants had also answered more than 50% of the presented questions (Figure 4, "Q20 I1" and "Q20 I2").

At the end of each quiz, except the Quiz-against-time, a reflective question was presented with the goal to motivate the users to reflect about the currently completed quiz and to evaluate whether they could gain any benefits or insights out of it. These reflective questions were chosen randomly out of altogether 8 questions. In the Quiz-of-20, 243 of these reflective questions were presented to the players and 54% of them were answered (Figure 4, "Q20 E"). For the Quiz-of-10, 77 questions were presented and 32% were answered (Figure 4, "Q10 E"), and for the Quiz-of-5, 45% of the 53 posed questions were answered (Figure 4, "Q5 E").

Out of 1205 reflective questions, 603 (52%) were answered with a short sentence or a single word only (see Section 4.3). This indicates that the quiz players thought about the questions, i.e. that reflective learning could be triggered. The 48%, which indicated only a letter or nothing at all, must be regarded as missing values; but only as missing values in terms of measurement, not in terms of reflective learning. Looking at the time in the game when the questions were asked, the results indicate that users preferred questions at the beginning of the quiz (59.3% answered) over questions at the end (51.5% answered) and were most reluctant to deal with questions while playing the quiz (42.4% answered).

4.4.2 Reflection-on-action components

In KnowSelf (see Table 8), the participants documented their experiences, insights and comments regarding the application in the Reflection Diary. Six participants were willing to share their explicit data and altogether 103 statements were inserted in the diary (the content is discussed in the next section). In contrast, the MoodMap App provided reports including a reflection diary but these were not used in any trial. In the quiz, no diary was implemented.

TABLE 8
Assessment of KnowSelf Prompts

Questions	KnowSelf	
	Mean	SD
Please indicate your agreement with the following statement (scale: 1 - strongly disagree, 5 - strongly agree):		
The KnowSelf Prompts motivated me to reflect.	2.00	1.06
To what extent were the following categories of KnowSelf Prompts helpful to you? (scale: 1: not helpful at all to 5: very helpful):		
Reflection Intervention: Reminder for using KnowSelf generally (e.g. look at heatmap, write diary entries)	3.22	1.09
Reflection Intervention: Reminder of project recording	2.75	1.28
Reflection Amplifier: Notification about specific amount of switches	2.30	1.34
Reflection Amplifier: Notification about unusual amount of idle time	2.57	0.79
Reflection Amplifier: Notification about most used resources	3.30	1.25

4.4.3   Contextualisation

Mandatory contextualisation in the MoodMap trial (Version A) at the call centre was evaluated by means of three questions. The participants rated the questions as follows on a Likert scale with 1 being the lowest and 5 the highest value. Participants were slightly positive on how helpful the contextualisation was for associating the inserted mood with the respective situation ($M = 3.39; SD = 0.86$). Participants were also slightly positive on how helpful contextualisation was for recalling the past experience ($M = 3.21; SD = 0.74$). Participants were also slightly positive on whether pre-defined context definitions were adequately chosen (Q3: $M = 3.2; SD = 0.94$). 31% of the mood entries were inserted "after a call", 6% "after a break", 2% "after a coaching session" and 61% in "other", not pre-defined contexts. In the category "other", participants could give a free-text description of context (the pre-defined context descriptors were for convenience sake). The following situations were most common free-text entries in the category "other": start or end of shift, before break or lunch, back from lunch or a certain event, problem or issue (e.g. crash, waiting for other departments), feeling better after dealing with a problem, successful events, or having finished a task.

5   DISCUSSION

5.1   RQ1: Usage

Usage as described in the results section is the prerequisite for discussing the subsequent research questions RQ2, RQ3 and RQ4. Usage indicates a basic level of success of applications including reflection guidance technology, as all participants were volunteers, and not using the apps would have been a viable way for them. We saw large differences in usage (high standard deviations). This indicates that some participants did indeed see benefit and had interest in using technologies, while others did not. In the present study setup we were rather interested in quality and quantity of learning that did occur rather than on cases of non-successful technology take-up and learning. In future studies, given that we now know reflection guidance technology can induce learning in workplace settings, we would look at factors that influence whether or not users take up technology and learning.

5.2   RQ2: Learning - Support by Interventions

In all four field trials, participants tended to agree that the respective intervention gave support to activities that are part of reflective learning, such as reconstructing experiences, collecting relevant information for reflecting on experiences, reflecting on experiences, etc.

Across all field trials, participants frequently asked in the introductory phase "What is in for me?". This relevance of perceived usefulness was the reason why all four field trials had been preceded by an iterative design process in which interventions were adapted to the specific use case. Despite this, we found it challenging to bring across the benefit of reflective learning in the introductory phase prior to each field trial. In the case of the Medical Quiz, perceived usefulness of reflection was made difficult by the fact that participants were learning for an exam: So, a quiz without reflective questions would have been more time efficient for the immediate goal of passing the exam, while reflective questions had a longer-term goal of supporting the take-up of theory into practice. In the MoodMap App trial at the IT company, we were successful in helping users perceive the usefulness of the intervention by presenting a success story of reflective learning with the app prior to the trial.

The lesson learned is to assign significant resources prior to a field trial for an introduction to reflective learning and the concrete technology used.

Results from two out of four evaluations (Medical Quiz and KnowSelf) show a positive correlation between the extent of app usage and how users perceive the apps' support for reflective learning. We are faced here with two items that need to be perceived as useful: First, reflective learning in the context of work and second, the concrete intervention. While we had ensured the first (see prior publications [5], [13], [46], [47]), we suspect that there is a circularity in that those users who engage more with an app, perceive its benefits more clearly, or are at least more able to use the app in a way that supports them in reflective learning.

5.3   RQ3: Learning - Depth

In all trials, notes with reflective content were identified (see Table 6). The reflection depth that was visible in the textual entries was limited: In both MoodMap trials, reflection took place to a large extent on the lowest depth of reflection (94% and 95%), with only a few notes in the medium depth, and none of high depth. These notes deal with work experiences, own emotions, emotions of others and some notes regarding interpretation and justification of actions. In the MoodMap trial at the IT company, the own physical well-being was also reported. This shows that participants become aware of their own and others' mood and reflected about them.

In the KnowSelf trial, again most notes were on the lowest depth of reflection, but the picture is much more balanced (low: 45%; medium: 39%; high: 16%), and a few notes achieved a high depth of reflection. Most of the notes were describing experiences (low), followed by suggesting solutions to observed problems (medium), and documenting

some insights and conclusions drawn from personal experiences (high). These findings confirm that the participants gained new insights and a better understanding of their work experiences, which could then be used as a basis for changes in behaviour and sustainable improve related work. We were not able to apply the analysis schema to the answers given to the reflective questions within the quiz. The descriptive analysis we carried out showed anecdotal evidence that reflective learning had taken place. However, short textual notes from which reflection cannot be judged does not necessarily mean that the reflective learning that did take place was similarly limited. In a qualitative exploration [10] we identified two reasons for the very short answers: First, from a lack of computer skills (typing takes time!); and second because participants used the Medical Quiz to prepare themselves for the exam at the end of the workshop. Third, participants stated in interviews that they had discussed Medical Quiz content informally.

Overall, we conclude that reflective learning was observable in all field trials, but of limited depth. We cannot conclusively determine across all trials whether deeper reflection had happened outside what was observable; for the Medical Quiz at least we have indications that deeper reflection happened in face-2-face discussions. The fact that reflection of high depth was much less frequently observed can plausibly be assumed to stem from an inherent increasing difficulty in levels of increasing reflection depth, i.e. it is more difficult to derive insights from reflection (high reflective depth) than to merely describe an experience (low reflective depth). Additionally, we would like to put forward a second explanation, which complements the previous one: In professional work one also needs to ask: What insights could actually be gained from a particular experience? We argue that in professional practice, a majority of reflection by necessity cannot lead to deep insights if the actor is indeed a professional; rather, the professional should be able to derive insights from the "odd experience out of the majority of experiences". Therefore, it might be a very natural phenomenon for the reflective practitioner to mostly reflect on lower depths of reflection; and the trick will be not to miss insights, patterns, and implications from those experiences that offer an opportunity for deeper learning.

5.4   RQ4: Design Challenges for Interventions

The relevance of context (context-awareness) has been postulated as central to improve human-computer interaction in general (see e.g., [48]). In technology that supports reflective learning, context-awareness typically means that technology has some way of observing those experiences that users want to reflect on, and presenting those activities in ways that are conducive to learning - *"context becomes content"*, as Müller et al. [49] put it. This is in line with our work, as we see that contextualisation plays two different roles in reflection technology:

First, for promoting reflection, data about experiences need to be analysed in relationship to other relevant information about the past experience, that is *in context*. Note that the context that was used in the MoodMap App, namely a short textual description of the setting in which the mood entry was taken, represents only a very simple "shorthand"

for describing the complex conceptual entity of context. A number of contexts were predefined, such as "after a call". Alternatively, the category "other" could be chosen, and a free text description of context could be given. We see this combination of predefined context descriptions with the possibility to alternatively define context more freely as a good design practice, as i) contextualisation was perceived as helpful by participants and ii) predefined contexts were used in 39% of all mood entries. The second means that the predefined contexts were useful enough to serve as convenience functionality, but that despite prior definition of contexts together with target stakeholders, there was still a need for specification. During longer usage, it might be possible to converge on a longer list of predefined contexts that would capture a higher percentage of relevant contexts [50].

Second, proactive reflection guidance technology like prompts can benefit from information about the users' context such that prompts do not interrupt users in their primary activity (=working). The correct timing of presenting interventions is an open research question, shared amongst all research on proactive technology [25]. In our studies, we have systematically noted, that the less interaction with the reflection intervention is required, the more the intervention is "usable" for users even at stressful working times and embedded into work practice: Users in the MoodMap trial at the IT company captured many moods during the day, but did not use the prompts to reflect. The reason given was that the trial took place in a very stressful and work-intensive time. Reflective questions in the quiz were used more when placed in the beginning and at the end of the quiz, as playing the quiz itself was perceived as an activity that should not be disrupted. In KnowSelf, notifications were often perceived as a further source of work fragmentation. Particularly in settings where the reflection intervention is not technically integrated with the environment in which the primary activity takes place, it is difficult for the reflection intervention software to "know" when a user has time to reflect and when not. This is the reason why it was easier for us as designers to time the reflection questions in the quiz well, than to time the reflection interventions in KnowSelf well.

5.5   Methodology

By nature, field trials are messy; in this section we discuss what we see as most serious issues in the methodology.

Each field trial was run in different organisations in different industry sectors. The interventions we tested were, in every case, adapted to the field trial settings. Such adaptations were necessary in order to convince stakeholders in the settings to participate in the trials at all. Even though the same reflection guidance concept underlies each intervention, the concrete interventions were different. On the one hand, this means that a statistical comparison would be meaningless. On the other hand, as results are overall positive across these very heterogeneous settings, chances are high that the overall results are transferable to other workplace settings.

The durations of trials was different. The study participants were volunteers who had to manage the extra time needed for the field trials (questionnaires, interviews, reflection

activities) either within their working time or in their spare time. This means that we could not order participants to participate for a given length but had to convince, which ultimately lead to different durations of field trials. Additionally, the medical quiz trial was set in a discontinuous setting (training was spaced across four months).

Finally, we measured the impact of interventions on learning and the design challenges for interventions only very indirectly, namely by asking participants for self-assessments. This means that we did not have any objective means to measure whether and what learning actually occurred. However, the learning domains in all field trials, except the medical quiz, are very fuzzy, to the extent that it would not be possible to devise a test for assessing knowledge before and after the interventions. After all, what is learned by reflection is highly individual in the case of workplace learning in professionals.

6 CONCLUSION

Across all four field trials at the workplace, we found that reflection guidance technology was used and accepted. Participants found interventions to be supportive for reflective learning. There are very few examples of successful reflection guidance technology available in the literature. Therefore, our interventions in the applications as well as the insights gained through the trials constitute a novel contribution in the field.

We found the following two issues to be shared amongst all four field trials and relevant for future research: Firstly, only limited depth of reflection was observable in textual entries. We offer two possible interpretations for this: One the one hand, deep reflection might be very difficult to achieve, which would point to professionals being more educated in reflective learning. Or, on the other hand, professionals do achieve deep reflective learning, but do not articulate it in the form of textual entries and thus the challenge is in detecting it. Secondly, we see the development of a good timing for reflection guidance technology as a major challenge. This issue is particularly relevant in workplace settings, where the employees face disruptions on many fronts.

ACKNOWLEDGMENTS

REFERENCES

[1] D. Boud, R. Keogh, and D. Walker, *Reflection: Turning Experience into Learning*. New York: Routledge Falmer, 1985, ch. Promoting Reflection in Learning: a Model., pp. 18–40.

[2] M. Eraut, "Informal learning in the workplace," *Studies in Continuing Education*, vol. 26, no. 2, pp. 247–273, 2004.

[3] S. Hoyrup, "Reflection as a core process in organisational learning," *Journal of Workplace Learning*, vol. 16, no. 8, pp. 442–454, 2004.

[4] K. Knipfer, B. Kump, D. Wessel, and U. Cress, "Reflection as catalyst for organisational learning," *Studies in Continuing Education*, vol. 35, no. 1, pp. 30–48, 2013.

[5] M. W. Daudelin, "Learning from experience through reflection," *Organizational Dynamics*, vol. 24, no. 3, pp. 36–48, 1996.

[6] E. A. Davis, "Scaffolding students' knowledge integration: prompts for reflection in KIE," *Int. J. of Science Education*, vol. 22, no. 8, pp. 819–837, 2000.

[7] E. Davis, "Prompting middle school science students for productive reflection: Generic and directed prompts." *J. of the Learning Sciences*, vol. 12, no. 1, pp. 91–142, 2003.

[8] D. Ifenthaler, "Determining the effectiveness of prompts for self-regulated learning in problem-solving scenarios," *Educational Technology & Society*, vol. 15, no. 1, pp. 38–52, 2012.

[9] A. Fessl, V. Rivera-Pelayo, V. Pammer, and S. Braun, "Mood tracking in virtual meetings," in *Proceedings of the 7th European conference on Technology Enhanced Learning*, ser. EC-TEL'12. Berlin, Heidelberg: Springer-Verlag, 2012, pp. 377–382.

[10] A. Fessl, M. Bratic, and V. Pammer, "Continuous learning with a quiz for stroke nurses," *Int. J. Technology Enhanced Learning*, vol. 6, no. 3, pp. 265–275, 2014.

[11] A. Fessl, G. Wesiak, and V. Pammer, "Training reflective practice: Improving stroke nurses' work with a medical quiz," in *Proceedings of the EC-TEL 2016 (accepted as poster)*, 2016.

[12] V. Pammer and M. Bratic, "Surprise, surprise: Activity log based time analytics for time management." in *CHI'2013 , 27 April - 2 May 2013, Paris, France*, 2013.

[13] V. Pammer, M. Bratic, S. Feyertag, and N. Faltin, "The value of self-tracking and the added value of coaching in the case of improving time management," in *Proceedings of the 10th European Conference on Technology Enhanced Learning (ECTEL '15)*, G. Conole, T. Klobucar, C. Rensing, J. Konert, and E. Lavouee, Eds. Springer Int. Publishing, 2015, pp. 467–472.

[14] V. Rivera-Pelayo, *Design and Application of Quantified Self Approaches for Reflective Learning in the Workplace*. KIT Scientific Publishing, Karlsruhe, 2015.

[15] A. Fessl, G. Wesiak, V. Rivera-Pelayo, S. Feyertag, and V. Pammer, "In-app reflection guidance for workplace learning," in *Proceedings of the 10th European Conference on Technology Enhanced Learning (ECTEL '15)*, G. Conole, T. Klobucar, C. Rensing, J. Konert, and E. Lavouee, Eds., vol. 9307. Springer Int. Publ., 2015, pp. 85–99.

[16] D. A. Schön, *The Reflective Practitioner: How Professionals Think In Action*, 1st ed. Basic Books, 1984.

[17] C. Tang, "Reflective diaries as a means of facilitating and assessing reflection," *Pacific Rim Conference on Higher Education Planning and Assessment, Hilo, Hawaii,*, vol. 2, 3-7 June 2000.

[18] S. E. Walker, "Journal writing as a teaching technique to promote reflection." *J. of athletic training*, vol. 41, no. 2, pp. 216–221, 2006.

[19] K. Thorpe, "Reflective learning journals: from concept to practice." *Reflective Practice*, vol. 5, no. 3, pp. 327–343, 2004.

[20] K. D. Chirema, "The use of reflective journals in the promotion of reflection and learning in post-registration nursing students." *In Nurse Education Today*, vol. 27, p. 192202, 2007.

[21] M. Harris, "Scaffolding reflective journal writing negotiating, power, play and position." *Nurse Education Today*, vol. 28, p. 314326, 2008.

[22] A. Alexiou and F. Paraskeva, "Enhancing self-regulated learning skills through the implementation of an e-portfolio tool." in *Procedia - Social and Behavioral Sciences*, vol. 2, no. 2, 2010, pp. 3048–3054.

[23] S. Elango, R. C. Jutti, and L. K. Lee, "Portfolio as a learning tool: students' perspective." *Annals of the Academy of Medicine, Singapore*, vol. 34, no. 8, pp. 511–4, 2005.

[24] D. J. Greaves and S. K. Gupta, "Portfolios can assist reflective practice and guide learning." *Current Anaesthesia & Critical Care*, vol. 14, pp. 173–177, 2003.

[25] H. Thillmann, J. Kuensting, J. Wirth, and D. Leutner, "Is it merely a question of "what" to prompt or also "when" to prompt?" *Zeitschrift für Pädagog. Psychologie.*, vol. 23, no. 2, pp. 105–115, 2009.

[26] J. Ho and S. S. Intille, "Using context-aware computing to reduce the perceived burden of interruptions from mobile devices," in *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, ser. CHI '05. New York,, USA: ACM, 2005, pp. 909–918.

[27] V. Pejovic and M. Musolesi, "InterruptMe: Designing Intelligent Prompting Mechanisms for Pervasive Applications," in *Proceed. of*

the 2014 ACM Int. Joint Conf. on Pervasive and Ubiquitous Computing, ser. UbiComp '14. New York, NY, USA: ACM, 2014, pp. 897–908.

[28] M. Bannert and P. Reimann, "Supporting self-regulated hypermedia learning through prompts," *Instructional Science*, vol. 40, no. 1, pp. 193–211, 2012.

[29] A. King, "Facilitating elaborative learning through guided student-generated questioning," *Educational Psychologist*, vol. 27, no. 1, pp. 111–126, 1992.

[30] G. Xun and S. Land, "A conceptual framework for scaffolding iii-structured problem-solving processes using question prompts and peer interactions," *Educational Technology Research and Development*, vol. 52, no. 2, pp. 5–22, 2004.

[31] D. Verpoorten, W. Westera, and M. Specht, "Reflection amplifiers in online courses: a classification framework." *Journal of Interactive Learning Research*, vol. 21(4), no. 56, 2010.

[32] G. Fischer, "User modeling in human-computer interaction." *User Modeling and User-Adapted Interaction*, vol. 11, pp. 65–86, 2001.

[33] P. Brusilovsky and E. Millán, *User Models for Adaptive Hypermedia andAdaptiveEducational Systems*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2007, pp. 3–53.

[34] S. Bull, V. Dimitrova, and G. McCalla, "Open learner models: Research questions," *International Journal of Artificial Intelligence in Education (IJAIED Special Issue)*, vol. 17 (2), pp. 83–87, 2007.

[35] S. Bull and J. Kay, "Student models that invite the learner in: The smili:() open learner modelling framework," *Int. J. Artif. Intell. Ed.*, vol. 17, no. 2, pp. 89–120, Apr. 2007.

[36] J. Kay and B. Kummerfield, *Adaptive Technologies for Training and Education*. Cambridge University Press, 2011, ch. Lifelong learner modelling, pp. 140–165.

[37] J. Kay, K. Yacef, and P. Reimann, "Visualisations for team learning: Small teams working on long-term projects," in *Proceedings of the 8th Iternational Conference on Computer Supported Collaborative Learning*, ser. CSCL'07. International Society of the Learning Sciences, 2007, pp. 354–356.

[38] J. Uther, "On the visualisation of large user models in web-based systems," Ph.D. dissertation, University of Sydney, 2002.

[39] D. L. Kirkpatrick and J. D. Kirkpatrick, *Evaluating trainings programs: The four levels. 3rd Ed.* Berrett-Koehler Publ., Inc., 2006.

[40] K. Knipfer, D. Wessel, and K. DeLeeuw, *D1.5 Specification of Evaluation Methodology and Research Tooling*, June 2012.

[41] M. Prilla and B. Renner, "Supporting collaborative reflection at work: A comparative case analysis," in *Proceedings of ACM Conference on Group Work (GROUP '14)*. ACM, 2014.

[42] J. A. Russell, "A Circumplex Model of Affect," *Journal of Personality and Social Psychology*, vol. 39, pp. 1161–1178, 1980.

[43] J. Itten, *Kunst der Farbe*, 1st ed., D. Boud, R. Keogh, and D. Walker, Eds. Ravensburg, Germany: Otto Maier Verlag, 1971.

[44] B. Renner and G. Wesiak, *D10.3 Summative Evaluation of MIRROR Appsphere usage and effectiveness at test beds*, 2014.

[45] D. V. Jaarsveld and W. R. Poster, "Call centers : emotional labor over the phone," *Emotional labor in the 21st century : diverse perspectives on the psychology of emotion regulation at work.*, 2013.

[46] V. Pammer and A. Fessl, *D4.1 - Results of the user studies and requirements on Individual Reflection at Work*, 2011.

[47] A. Fessl, O. Blunk, M. Prilla, and V. Pammer, "The known universe of reflection guidance: a literature review," *Int. J. of Technology Enhanced Learning.*, 2016 (in press).

[48] G. Fischer, "Context-aware systems: The 'right' information, at the 'right' time, in the 'right' place, in the 'right' way, to the 'right' person," in *Proceedings of the International Working Conference on Advanced Visual Interfaces*, ser. AVI '12. New York, NY, USA: ACM, 2012, pp. 287–294.

[49] L. Müller, M. Divitini, S. Mora, V. Rivera-Pelayo, and W. Stork, "Context becomes Content: Sensor Data for Computer Supported Reflective Learning," *IEEE Transactions on Learning Technologies*, 2014.

[50] V. Rivera-Pelayo, A. Fessl, L. Müller, and V. Pammer, "Introducing mood self-tracking at work: Empirical insights from call centers," *ACM Trans. Comput.-Hum. Interact.*, vol. 24, no. 1, pp. 3:1–3:28, Feb. 2017.

Angela Fessl is a post-doc researcher in the area of Technology Enhanced Learning with focus on reflective learning at the work place and has a proven track of research and development Document/Content Management and eLearning Systems. Angela received her Master for Computer Science (MSc) in the field of Telematik and her PhD (with distinction) in Informatics at Graz University of Technology, Austria

Gudrun Wesiak was senior researcher at the Know-Center and is lecturer at Graz University. She holds a MSc and PhD (with distinction) from the University of Graz. Her research focuses on psychological aspects in technology enhanced learning and assessment, as well as on the evaluation of knowledge technologies in learning and workplace settings.

Verónica Rivera-Pelayo, currently at IBM, was a scientist at the FZI Research Center for Information Technology in Karlsruhe, Germany during this work. She holds a PhD (Dr.-Ing.) from the Karlsruhe Institute of Technology and a Degree in Informatics Engineering from the Universitat Politècnica de Catalunya BarcelonaTech (UPC). Her research is focused on the area of Technology Enhanced Learning and the use of self-tracking tools to support reflective learning on the job.

Sandra Feyertag is a software developer and researcher at the Know-Center. She holds a BSc for Computer Science from Graz University of Technology and is currently doing her Master in the same field of study. Sandra is engaged in various aspects of the area Ubiquitous Personal Computing, lately focusing on requirements analysis and user studies.

Viktoria Pammer is assistant professor at the Knowledge Technologies Institute, Graz University of Technology and area manager at the Know-Center. She is responsible for teaching, research and innovation in the field of computer-supported working, learning and creativity, with an emphasis on knowledge work. Viktoria holds a PhD and an MSc (both with distinction) from the Graz University of technology.