

Integrated Project Reflective Learning at Work

European Commission Seventh Framework Project (IST-257617)

Deliverable D 1.5

Specification of Evaluation Methodology and Research Tooling

Editor Kristin Knipfer, Daniel Wessel, Krista DeLeeuw

Work Package WP 1

Dissemination Level Public

Status Final

Date 30 June 2012



The MIRROR Consortium

Beneficiary Number	Beneficiary name	Beneficiary short name	Country
1	imc information multimedia communication AG	IMC	Germany
2	Know-Center (Kompetenzzentrum für wissensbasierte Anwendungen und Systeme Forschungs. Und Entwicklungs GmbH) Graz	KNOW	Austria
3	Imaginary srl	IMA	Italy
4	Deutsches Forschungszentrum für Künstliche Intelligenz GmbH Saarbrücken	DFKI	Germany
5	Ruhr-Universität Bochum	RUB	Germany
6	The City University	CITY	UK
7	Forschungszentrum Informatik an der Universität Karlsruhe	FZI	Germany
8	Norges Teknisk-Naturvitenskapelige Universitet	NTNU	Norway
9	British Telecommunications Public Limited Company	ВТ	UK
10	Tracoin Quality BV	TQ	Netherlands
11	Infoman AG	INFOM	Germany
12	Regola srl	REG	Italy
13	Registered Nursing Home Association Limited	RNHA	UK
14	Neurologische Klinik GmbH Bad Neustadt	NBN	Germany
15	Medien in der Bildung Stiftung	KMRC	Germany



Amendment History

Version	Date	Author/Editor	Description/Comments
Preliminary	Preliminary Version of D1.5 Evaluation Framework, due December 2011 per reviewers reques		
V 0.1	14.11.2011	K. Knipfer	First Draft
V 0.2	01.12.2011	K. Knipfer	Draft of Chapter 1 and 2
V 0.3	06.12.2011	K. Knipfer	Draft of Chapter 3
V 0.4	07.12.2011	K. Knipfer	Draft of Chapter 4 based on Input from IMC; Integrated Input from TQ and BT on Evaluation Levels; Integrated Input from NTNU (Chapter 3)
V 0.5	08.12.2011	D. Wessel	Minor Revisions on Chapter 1 and2
V 0.6	08.12.2011	D. Wessel	Consolidation of Chapter 3
V 1.0	09.12.2011	K. Knipfer, D. Wessel	Draft for Internal Review
V 1.0 R	16.12.2011	D. Wessel, K. Knipfer	Revision according to Reviewers' Comments; Final Consolidation of Chapter 3; Major Restructuring of Chapter 3 and 4
V 1.0 RF	19.12.2011	K. Knipfer, D. Wessel	Final Version for Submission to the EC
D 1.5 Evalu	ation Framev	vork	
V 1.0	17.04.2012	D. Wessel, K. DeLeeuw	Added structure for the final deliverable
V 1.1	23.05.2012	K. DeLeeuw	Revision of the Summative Evaluation Criteria
V 1.2	23.05.2012	D. Wessel	Draft version of Evaluation Toolbox (Chapter 4) and Evaluation Procedure (Chapter 5)
V 1.3	25.05.2012	D. Wessel	Added further questions and Chapter 7
V 1.4	25.05.2012	K. DeLeeuw	Edited Chapters 1-6 for consistency
V 1.5	28.05.2012	D. Wessel	Updated Chapters 1-8 and Appendix.
V 1.6	29.05.2012	K. DeLeeuw	Final Version for Internal Review
V1.7	25.06.2012	K. DeLeeuw	Integrated internal review comments
V1.8	26.06.2012	D. Wessel	Added Appendix C
V2.0	30.06.2012	K. DeLeeuw, D. Wessel	Final Version for Submission to the EC



Contributors

Name	Institution
Nils Faltin	IMC
Viktoria Pammer, Angela Fessl	KNOW
Lucia Pannese, Dalia Morosini	IMA
Silke Balzert, Thomas Kleinert	DFKI
Michael Prilla	RUB
Neil Maiden, James Lockerbie	CITY
Veronica Rivera	FZI
Birgit Krogstie	NTNU
Hans Dirkzwager, Ellen Lennarts	ВТ
Roy Ackema	TQ
Andrea Herrmann, Anne König	INFOM
Michele Biolé, Marco Parigi	REG
Kevin Pudney, Malcolm Rose	RNHA
Dominik Cavael	NBN
Nina Feldmann	KMRC

Reviewer

Name	Institution
Roy Ackema	TQ
Stefanie Lindstaedt	KNOW
Samia Drissi	Infoman
Neil Maiden (for preliminary version)	City

Page 5

Legal Notices

The information in this document is subject to change without notice.

The Members of the MIRROR Consortium make no warranty of any kind with regard to this document, including, but not limited to, the implied warranties of merchantability and fitness for a particular purpose. The Members of the MIRROR Consortium shall not be held liable for errors contained herein or direct, indirect, special, incidental or consequential damages in connection with the furnishing, performance, or use of this material.



Executive Summary

MIRROR's vision is to empower employees to learn by reflection on their work practice and on their very personal learning experiences. MIRROR aims at assisting employees in capturing experiences and in developing creative solutions for problems that need to be solved. This will be achieved by complementing personal and organisational learning environments with personal MIRROR apps for individual, collaborative, creativity-based, game-based, as well as organisational learning through reflection. MIRROR will provide new learning technologies for "learning on the job", "learning by doing", "learning from peers" and "experiential learning". With MIRROR applications, the effectiveness of learning will be increased significantly in situations where no teachers, no formal content, and no explicit knowledge are available.

Summative evaluation will ensure high quality and effectiveness of our MIRROR apps. Impact on both an individual level and a business level will be examined on the basis of general and testbed-specific evaluation criteria. Formative evaluation is not the topic of this deliverable, although our work will inform the formative evaluation within the project.

We have modelled the relevant aspects of learning by reflection at work in great detail with the i^* approach (Chapter 2) and derived summative evaluation criteria to construct our summative evaluation framework (Chapter 3). While the i^* provided a bottom-up view, our modification of the Kirkpatrick model allowed us to keep the bigger picture in mind.

Based on these relevant criteria that determine the success of the MIRROR apps, we developed tools to assess these measures. The Toolbox (the Core Questions in Chapter 4 and additional questions in Appendix A) allows us to assess relevant indicators for reflective learning and their impact for individuals and team learners, as well as the organisation as a whole (business impact). It also ensures comparability between the different apps and different testbeds. The evaluation procedure is described and agreed upon within the project (Chapter 5), and the documentation process allows us to conduct the overall integration of the results (Chapter 6).

To assess the impact of MIRROR apps and to ensure high quality of the evaluation process and results, we will apply this evaluation framework for all apps and for all of our five testbeds.



Table of Contents

E	xecuti	ve Summary	6
T	able o	f Contents	7
F	igures		10
T	ables		11
1	Intr	oduction	12
•	1.1	MIRROR Vision	12
	1.2	MIRROR Research Objectives	14
	1.3	Users' Needs and Expectations for Reflective Learning Support	14
2	ldei	ntifying MIRROR Stakeholders and Their Goals	16
	2.1	The <i>i</i> * Methodology	17
	2.2	The MIRROR i* Model of Reflective Learning at Work	19
	2.2.	1 Worker	20
	2.2.	2 Individual reflector	22
	2.2.	3 Individual team reflector	23
	2.2.	4 Collaborative team reflector	24
	2.2.	5 Organisational reflector	25
	2.3	Validating the <i>i*</i> Model	26
3	Spe	ecifying Summative Evaluation Criteria	27
	3.1	Evaluation Criteria Based on the <i>i*</i> Model of Reflective Learning at Work	27
	3.2	Overview of MIRROR Summative Evaluation Criteria	29
	3.3	Methodological Considerations	32
4	Eva	aluation Toolbox	34
	4.1	Demographic Information	34
	4.2	Level 1: Reaction (Usage)	36
	4.3 and L	Level 2: Learning (App-Specific Reflection Questions, Short Reflection Scanearning Outcomes)	ale, 37
	4.3.	·	37
	4.3.	2 Short Reflection Scale	41
	4.3.	3 Learning outcomes	42
	4.4	Level 3: Behaviour	42



Page 8



	4.4	1	Subjective improvement in work	42
	4.5		el 4: Results	42 42
	4.5		BT	43
	4.5		Infoman	43
	4.5		NBN	44
	4.5		Regola	45
	4.5		RNHA	45
	4.6		ticipant Characteristics	46
	4.7		litional Questions	47
5			on Procedure	48
	5.1		onale	48
	5.2		luation Procedure Steps	49
	5.3	Doo	euments	50
	5.3	.1	App Evaluation Sheet	50
	5.3	.2	Evaluation App Implementation List	50
	5.3	.3	Evaluation Data File Template	51
6	Str	ategy	for Overall Integration	52
	6.1	Doc	umentation in D1.7 and D10.3	52
	6.1	.1	Intra- and Inter-App Comparisons	52
	6.1	.2	App-Specific Reflection Support	53
7	Ou	tlook		54
8	Re	feren	ces	55
9	Ар	pend	ix A: Evaluation Toolbox	56
	9.1	Cor	e Questions	56
	9.1	.1	About Yourself	56
	9.1	.2	Usage of the App	56
	9.1	.3	App-Specific Questions	57
	9.1	.4	Reflection Scale	59
	9.1	.5	Learning outcomes	59
	9.1	.6	Work Behaviour	60
	9.1	.7	Results Level	60
	9.2	Pari	ticipant Characteristics	62



Specification of Evaluation Methodology and Research Tooling

Page 9

	9.2.1	Need for Cognition	62
	9.2.2	Ambiguity Tolerance	63
	9.2.3	Mindfulness	64
	9.2.4	Self-Efficacy	66
	9.2.5	IT Attitudes	66
9	.3 Fur	ther Questions beyond the Core Questions	68
	9.3.1	Reaction Level	68
	9.3.2	Learning Level	69
	9.3.3	Behaviour Level	76
	9.3.4	Results Level	77
10	Append	lix B: The Kirkpatrick Evaluation Model	79
11	Append	lix C: i* Model Graphics	80



Figures

Figure 1: Four Types of Process Elements of i*: Actors, Goals, Tasks, Resources, and Soft Goals_	18
Figure 2: i* Model of Reflective Learning at Work	20
Figure 3: "Worker" in i* Model of Reflective Learning at Work	_ 21
Figure 4: "Individual Reflector" in i* model of Reflective Learning at Work	23
Figure 5: "Individual Team Reflector" in i* model of Reflective Learning at Work	_ 24
Figure 6: "Collaborative Team Reflector" in i* model of Reflective Learning at Work	25
Figure 7: "Organisational Reflector" in i* model of Reflective Learning at Work	26
Figure 8: MIRROR Summative Evaluation Criteria based on the i* Model	29
Figure 9: Overview of Levels of Evaluation and Evaluation Criteria	31
Figure 10: Computer Supported Reflective Learning Model (CSRL)	37

Specification of Evaluation Methodology and Research Tooling

Page 11

Tables

Table 1: Emergent Evaluation criteria from the i* Model of Reflective Learning at Work	27
Table 2: Four Levels of Evaluation, i* Criteria and Rationale	32
Table 3: Summative Evaluation Criteria: Description and Evaluation Tools and Methods	32
Table 4: Example of the Core Question Table Allowing Comparison of Core Questions	53



1 Introduction

In this Deliverable (D1.5), we describe the development of the summative evaluation framework for MIRROR. We will first review important aspects of the project, then we will explain the conceptual work of modelling reflective learning at work with the i^* methodology to identify stakeholders and their goals within MIRROR (Chapter 2). This conceptual work formed the basis for our evaluation framework, which specifies summative evaluation criteria (Chapter 3). Chapter 4 describes the framework and lists our Core Questions, which form the centrepiece of our evaluation. Chapter 5 focuses on the procedure for carrying out the evaluation, and in Chapter 6 we describe how the results will be reported. Chapter 7 gives a short outlook of the next steps within the MIRROR project regarding summative evaluation. Finally, in Appendix A we present our Evaluation Toolbox which contains all relevant evaluation items.

1.1 MIRROR Vision

A major part of learning at work happens outside formal training. However, in contrast to formal training, there are no established means to assist employees in learning from their daily work experiences. Reflection on past experiences is an effective mechanism for both employees' individual and collaborative learning and for organisational learning (Argyris & Schön 1996; Dewey 1933; Kolb & Fry 1975). Daudelin (1996) defines reflection as "the process of stepping back from an experience to ponder, carefully and persistently, its meaning to the self through the development of inferences; learning is the creation of meaning from past or current events that serves as a guide for future behaviour". Reflection has the potential to lead to a better understanding of one's own work practice and can guide future behaviour (Järvinen & Poikela, 2001; Moon, 1999).

Boud et al. (1985, p.19) gives a comprehensive definition of reflective learning that builds a major part of the theoretical basis for our work in MIRROR: "Reflection in the context of learning is a generic term for those intellectual and affective activities in which individuals engage to explore their experiences in order to lead to new understandings and appreciations." Reflective learning has the potential to lead to more flexible work routines and higher performance in a rapidly changing work context. Accordingly, reflection has the potential to lead to change and development as it leads to insights into work practices and identifies where work routines need to be modified.

Our vision is to empower employees to learn by reflection on their work practice and on their personal learning experiences. MIRROR aims at assisting employees in capturing experiences and in developing creative solutions for problems that need to be solved immediately. This will be achieved by complementing personal and organisational learning environments with personal MIRROR applications for individual, collaborative, creativity-based, game-based as well as organisational learning through reflection.

MIRROR will provide new learning technologies for "learning on the job", "learning by doing", "learning from peers" and "experiential learning". With MIRROR applications, the effectiveness of learning will be increased significantly in situations where no teachers, no formal content, and no explicit knowledge are available. MIRROR enables employees to identify their individual learning needs through observation and reflection of their own work practice. In several MIRROR use cases, capturing technologies will go beyond purely



amassing recordings and sensor data to assist users in learning from these data. By taking into account easy to obtain physiological data, we extend the understanding of "learning experiences" from a purely cognitive point of view to a holistic, "whole body" point of view.

Captured and recorded work data facilitate learning from others and sharing of tacit knowledge. Teams will be empowered to reflect on shared experiences and work practices collaboratively. MIRROR empowers employees to maintain and increase their competencies independently from static training content or formal training programs that only slowly adapt to personal and emerging learning needs. Whereas previous training investments focus on formal learning (e.g., off-the-job training, courses) the MIRROR project will empower organisations and their employees to leverage and strengthen on-the-job learning through reflective learning.

Furthermore, employees' daily work experiences are rarely tapped for innovations and organisational learning. While organizational top-down processes have their merits in terms of standardization and quality assurance, it diminishes employees' motivation to excel in terms of quality and customer orientation. MIRROR addresses these deficiencies by facilitating the bottom-up integration of individual experiences into continuous process improvement, especially regarding innovation processes. Thus, organisations will be able to learn from the input and reflection outcomes of their employees. From a top-down view, by implementing MIRROR applications (apps), service organisations will create an environment of learning and motivation that makes learning progress observable to the individual, while leaving the individual in control over her learning data. This leads to efficient innovation processes as it leverages the creativity of the individuals and fosters their contribution to corporate goals.

Given that we deal with informal learning, evaluating the success of MIRROR is challenging, as there are no predefined learning goals that can be used as evaluation standards. However, a summative evaluation is needed to determine the quality and effectiveness of our MIRROR apps, on both an individual level and a business level, on the basis of general and testbed-specific evaluation criteria.

We developed a methodology for the evaluation of the impact and effectiveness of MIRROR apps and determined indicators of reflection and its effects at the individual, inter-individual and organizational levels. Integrating the theoretical understanding of reflection (see D1.4 and the i^* work described in Chapter 2.1) we address both the process and outcomes of reflection supported by the MIRROR environment. We also developed a set of tools and instruments for measuring the effectiveness of MIRROR apps in the test beds with qualitative and quantitative data. Given that the apps are still under development in close interaction with the testbed partners, we also need to allow for individual extensions of the research methodology. As the developers of the MIRROR apps will conduct the data gathering themselves (in close cooperation with the testbed partners), we provide them with templates and guidelines for the evaluation process (see the Evaluation Toolbox in Appendix A), and familiarise them with all relevant materials and standard procedures for data gathering, data compiling and results reporting.

Thus, in this deliverable, we give an overview of the evaluation approach and specify evaluation criteria on various levels. We also specify research methods and instruments for thorough assessment of relevant criteria.



Although we are concerned only with the summative evaluation and will describe the summative evaluation results in deliverables D1.7 and D10.3 in month 48 (at the end of the project), we begin evaluating the apps as soon as they are implemented in the testbeds. This provides us with a broader basis of comparison of their effects and allows us to feed back the evaluation results immediately. In contrast with formative evaluation, which is not the topic of this deliverable, we are not concerned with usability but with the proposed effects of the apps. Thus we inform and are informed by formative evaluation within the project.

1.2 MIRROR Research Objectives

Before specifying the evaluation criteria for summative evaluation, we would like to point to the MIRROR research objectives outlined in the description of work:

<u>Objective A:</u> Refine the conceptual understanding of learning by reflection in the workplace and its contribution to a holistic treatment of knowledge in the organisation.

Objective B: Develop a variety of methods and apps for learning by reflection based on historical or captured user data and experiences

- 1. Effortlessly capture and represent tacit work practices and learning experiences
- 2. Support learning by reflection on work practices and learning experiences
- 3. Extend creative problem solving during work tasks with reflective learning processes
- 4. Support collaborative knowledge construction by reflecting on common work practices and learning experiences
- 5. Utilize simulation and games in order to experience new work practices
- 6. Enable organizational learning through collective reflection and generalisation

Objective C: Apply the developed reflection methods and apps in diverse test beds and evaluate learning effectiveness "in the wild"

These research objectives provided the basis for definition of relevant evaluation criteria for summative evaluation as depicted in the description of work. These success criteria outlined in the description of work formed the basis for considerations of how to evaluate MIRROR's effectiveness. Furthermore, the insights that we have gained through the extensive user and design studies during the first year of MIRROR, helped us in specifying users' needs and expectations for reflective learning support. We will thus shortly summarize the results of the user and design studies relevant for the definition of evaluation criteria for MIRROR apps.

1.3 Users' Needs and Expectations for Reflective Learning Support

The user and design studies during the first year of the project MIRROR showed that the project proposal addressed relevant aspects of reflection at work. In all testbeds, employees reflect – but we have found that they did not use the available information to the fullest and the reflection processes were not yet supported by technology. We identified unsolved issues that the technology developed in MIRROR can address to support reflection.

The following statements summarize the key findings of the user studies:

1. We encountered both spontaneous, informal, and incidental reflection and prescheduled, intended, and structured reflection. We also found differences among the employees with regard to the inclination to reflect.



- 2. Reflective learning does not occur automatically during the course of daily working routines. We have identified triggers for reflective learning that, despite the different work contexts, seem to be quite similar in all testbeds.
- 3. The content of reflection can be one's own individual experience, experiences of comparable others, and shared team experience.
- 4. Reflection involves examination of past or current experiences for the purpose of guiding future behaviour. Reflection can involve comparison with other people for the purpose of evaluating one's own performance and re-evaluation of one's own experience.
- 5. The process of reflection is often accomplished collaboratively by a team/working unit or a (loose) group of individuals.
- 6. Organizational learning often evolves by accumulating learning through reflection of individuals and groups/teams within the organization (bottom-up learning). The freedom to change individual work practice seems to be inversely related to the need to affect changes on an organisational level.
- 7. Reflection often led to a new/better understanding of the experience and enabled implications, conclusions, or 'lessons learned' to be derived. However, we also found many reflective incidents that lacked specification of a learning outcome.
- 8. Task performance is reported to be a major reason to reflect. As the MIRROR apps are supposed to support learning by reflection and thereby improve task performance, they fit the requirements.
- 9. Support for reflection was virtually non-existent in the testbeds. Finding means to document and share outcomes of individual and collaborative reflection flexibly and unobtrusively is considered to be a major goal in all testbeds.
- 10. The user studies highlight that participants experience a need to reflect more often or more thoroughly on several issues. These issues include:
 - time management and stress management
 - training of challenging dialogues and dealing with critical incidents
 - professional educational and personal development
 - organisational routines and business processes

Given that reflective learning is not explicitly supported by tools in the testbeds, but that employees are highly motivated to reflect, huge gains for learning and organisational development might be achieved if reflection is supported efficiently.



2 Identifying MIRROR Stakeholders and Their Goals

To assess the impact of MIRROR apps during the summative evaluation, we need a fine-grained view of learning by reflection in the workplace. To achieve this, implemented a theory-based approach; that is, the evaluation of MIRROR will be guided by the conceptual model of reflective learning at work (CSRL model, see Deliverable 1.4). We complement this model with a goal-based approach, i.e., our evaluation is designed to assess whether reflective learning goals and objectives of relevant stakeholders within and beyond the project consortium are met. We will shortly point to some advantages and risks of both approaches:

A **theory-based approach** encourages a greater understanding of fundamental mechanisms of reflective learning (what works when, where, why, for whom, etc.), increases dialogue among all stakeholders, and clarifies underlying assumptions. However, it requires time upfront to specify the conceptual model, links between stakeholders, reflective learning processes, activities, and learning outcomes. Furthermore, there is a risk that our conceptual model may not capture all important aspects of reflective learning at work.

A **goal-based approach** is one means to determine whether results align with goals of stakeholders. Stakeholders include project team members (especially the testbed representatives), the target audience of MIRROR (i.e., staff and top management of our testbeds), and the wider target audience beyond the MIRROR project. This approach helps identify the specific, goal-related evaluation criteria. However, there is the risk that a goal-based approach does not examine unexpected results if not defined properly in advance. It requires clearly articulated goals and objectives. Furthermore, it may not capture some aspects of the reflective learning process, which are of particular importance to MIRROR.

Thus, we faced several challenges with regard to summative evaluation. Our evaluation approach...

- ... requires time upfront to specify the conceptual model, links between stakeholders, processes, activities, and outcomes,
- ... requires clearly articulated goals and objectives, and
- ... must capture all relevant processes.

Furthermore, our evaluation approach must avoid the disadvantages of theory- and goal-based evaluation approaches and ...

- ... must capture all important aspects of reflective learning at work, and
- ... must be able to deal with unexpected results.

To address these challenges, we combined a theory-based and a goal-based approach. We use the goal-based *i*approach* to model relevant aspects of reflective learning at work. This approach models the roles of individuals and teams at work, their goals, and their means of achieving those goals. We specified these criteria based on the Computer-Supported Reflective Learning (CSRL) theory (see Deliverable 1.4).

The *i** approach has several advantages: It was used successfully in APOSDLE, we have experts on the *i** model in our project (CITY), and it also allows for a detailed specification of relevant actors, their resources, tasks, and goals. This model allows for a detailed structural



specification of reflection at work on a deeper and more detailed level than the reflective learning model (see Deliverable 1.4). This specification increases the likelihood that we will address all important aspects of reflective learning at work. However, because this approach leads to a very detailed model, it is necessary to aggregate the results to achieve workable evaluation criteria.

We developed our *i** model in three workshops. The first two were one-day face to face meetings in London at City University, to capitalize on City's expertise and resources. The third workshop was divided into two days of 3-4 hours each, using Skype and Google Docs.

During the first workshop at London City University in October 2011, members of NTNU (reflection model), CITY (i^* and evaluation expertise), and KMRC (psychological processes of reflection and evaluation expertise) developed an i^* model of individual reflection. During the second workshop at London City University in November 2011, members of RUB (collaborative reflection expertise) joined NTNU, CITY, and KMRC to develop an i^* model of collaborative reflection. During the third workshop in November 2011, members of DFKI (organisational reflection expertise) and KMRC (psychological processes of reflection, evaluation expertise, and after two intensive workshops now familiar with the i^* methodology) developed an i^* model of organisational reflection.

These three i^* models were integrated by CITY to form a complex and comprehensive i^* model of reflective learning at work, relating the different actors. As this model is intended to be a means to derive the evaluation criteria, we refrain from describing the whole model here. However, it contributed to our understanding of reflection, and NTNU has collapsed the i^* model into a version for the reflection model (see D1.4). Larger versions of the model graphics are available in Appendix C.

We next describe the syntax of *i** in Chapter 2.1 before explaining the MIRROR *i** model of reflective learning at work in Chapter 2.2, and its validation with the testbed partners in Chapter 2.3.

2.1 The *i** Methodology

The established *i** approach (Yu & Mylopoulos, 1994) has been adopted and extended by City University to model complex technical and social systems. *i** was originally developed to model information systems composed of heterogeneous actors with different, oftencompeting goals that depend on each other to undertake their tasks and achieve the goals. For our purposes, "actors" are always people in a workplace who can take on different roles.

*i** consists of two main modelling components:

- The Strategic Dependency (SD) model describes the dependency relationships among various actors in an organisational context.
- The Strategic Rationale (SR) model describes stakeholder interests and concerns, and how they might be addressed by various configurations of systems and environments (Yu, 1997).

The first type of i^* model is the SD model, which describes a network of dependency relationships among actors or roles identified in the socio-technical system. The opportunities available to these actors can be explored by matching the depender (the actor who "wants")



and the dependee (who has the "ability"). The dependee's abilities can match the depender's requests; thus, the system-wide strategic model is developed. The SD model enables us to model future solutions at a high level in terms of actors and the goals that these actors want to achieve.

One advantage of SD models is that they are relatively easy to read, which allows its joint development and discussion with stakeholders, as shown during the third General Assembly of the MIRROR project where we discussed the *t** models with the whole project. Each SD model has only five different types of concepts, each of which are described and depicted in graphical form in Figure 1.

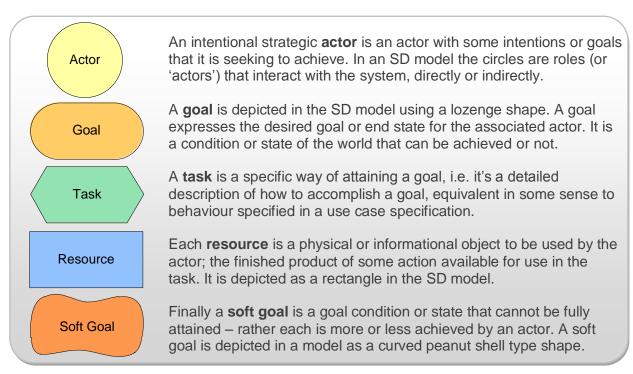


Figure 1: Four Types of Process Elements of i*: Actors, Goals, Tasks, Resources, and Soft Goals

Please note: The i* definition of "goals", that is, a dichotomous differentiation of whether they were accomplished or not, were not used in our approach, as we aim to be more detailed in evaluating the MIRROR apps. Instead we focus on soft goals, which capture the more complex nature of the processes we are interested in. Soft goals also have the advantage that they result in greater variability, which allows for improved data analysis.

These five types of process elements are linked using dependency links between actors in the SD model, which indicate that one actor depends on another for something that is essential to the former actor for attaining a goal. Dependency links are indicated with lines with the letter D inscribed on them, with the curved part of the D facing the depender.

The second type of *i** model is the SR model, which provides an intentional description of how each actor achieves its goals and soft goals. An element is included in the SR model only if it is considered important enough to affect the achievement of some goal. The SR model includes the SD model, so it describes which actors may be able to accomplish something by themselves, or by depending on other actors. It specifies goals, tasks, resources and soft goals (as described earlier) linked via task decomposition links, meansend links, contributes-to soft goal links, and by dependency links from the SD model. There are 4 types of links that can be used in SR modelling:



1) The dependency link works on the same principle as the dependency link in the SD model. In this case, it is a link between a process element within one actor boundary and a process element within another. Where there is a non-decomposed actor, the process element links directly to the actor shape.
Dependee (has)DDepender (wants)
2) The means-end link indicates a relationship between an end and a means for attaining the end. Each means-end link provides a different mean for attaining the end. Whenever there are different ways to achieve a goal, the means-end link should be used to represent this relationship, which acts as a logical 'OR' between the different ways (means) of accomplishing the goal. The means-end link is graphically represented with an arrow, where the arrowhead points from the means to the end. Note: the means cannot be a soft goal or a resource, as they do not describe how to attain an end.
Means (goal or task) ————————————————————————————————————
3) The contributes-to soft goal link is a special means-end link with a soft goal as the end. Since the achievement of a soft goal cannot be clearly defined, no means can be clearly specified for its achievement. Therefore, the contributes-to soft goal link represents a slightly different kind of relationship. This link represents that a goal, a task, a resource, or a soft goal can positively or negatively contribute to the attainment of a soft goal, without ensuring the attainment. The graphical notation of this link is an arrow, where the arrowhead points from the means to the end, which is always a soft goal, with an indication of the direction of the contribution.
Means (any type) — +> End (soft goal)
Means (any type) — — — End (soft goal)
4) The task decomposition link is used to decompose a task into subcomponents. A

4) The **task decomposition link** is used to decompose a task into subcomponents. A task specifies how to achieve a desired state, and when there is enough decomposable knowledge of how to perform a task, this knowledge can be decomposed into subcomponents (sub-processes) by the task decomposition link. Because all subcomponents need to be completed for the task to be performed, the relationship between the subcomponents can be interpreted as a logical 'AND'. A task can be decomposed into lower process elements of any type (resource, soft goal, goal or task). The task decomposition link is graphically represented in i* as a straight line marked with an orthogonal dash on the side of the task to be decomposed.

Sub-component (any type) — Task

2.2 The MIRROR *i** Model of Reflective Learning at Work

As the involved tasks and goals differ according to the different levels of reflection (individual, collaborative, and organisational reflection) we needed to specify roles ("actors" in the *i** syntax) and look at reflection in the context of each role. Thus, the overall reflection model has been arranged into roles, as depicted using the green actor boundaries in the *i** models (see Figure 2). The figures serve only to highlight the complexity of the tasks involved in a specific role. Larger versions of the figures are available in Appendix C.

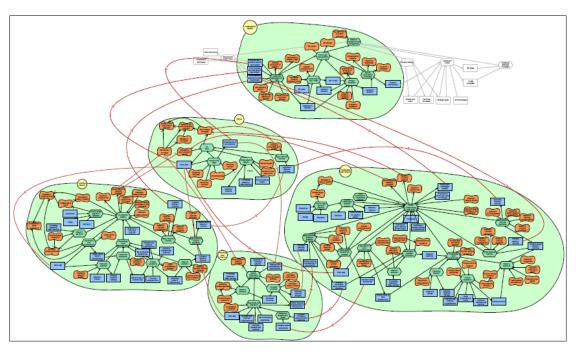


Figure 2: i* Model of Reflective Learning at Work

The different actors are related to each other using dependency links and contributes-to soft goal links. Means-end links were used mostly in relating goals to each other. Dependency links were used to identify dependencies in the achievement of soft goals (to be evaluated in the project). Contributes-to links across actor boundaries highlight relationships between the soft goals of different roles – in our model, we can see how the achievement of a soft goal in one role may benefit another. There is also the possibility for modelling trade-offs by using negative contribution links where appropriate.

The roles modelled in our i^* diagrams are:

- Worker
- Individual Reflector
- Individual Team Reflector
- Collaborative Team Reflector
- Organisational Reflector

Not all roles will be relevant at all testbeds, and individuals can switch between roles.

Although the level of detail of i^* is beyond the CSRL model described in D1.4, we include it in order to determine the evaluation criteria. It might be helpful for the reader to have a look at the reflection model in D1.4 to keep the "big picture" in mind. Note that this model here has a specific function to perform – it assists in identifying MIRROR summative evaluation criteria.

2.2.1 Worker

The "worker" role focuses on the individual employee engaged in his or her work— and not specifically on reflection (main task: "Work", see Figure 3). The worker can move into the "individual reflector" role (see 3.2.2) if a reflection session is initiated. Thus, while we differentiate between the two roles for the purpose of finding evaluation criteria, this "worker"



actor was developed together with the "individual reflector" actor during the first workshop. The main sub-tasks are:

- Undertake work tasks
- Monitor work tasks
- Initiate reflection session

Furthermore, we added a task related to the implementation of any reflection outcome:

Apply reflection outcome

The main soft goals related to this actor concern the qualities of monitoring work and the following high-level soft goals:

- · Work mastery increased
- Self-efficacy increased
- · Satisfaction increased
- · Quality of own work experience improved
- · Quality of work performance improved

Soft goals of the individual reflector, individual team reflector, and collaborative team reflector contribute positively to these high-level goals. Namely:

- Reflection outcome achieved
- · General reflection capability increased

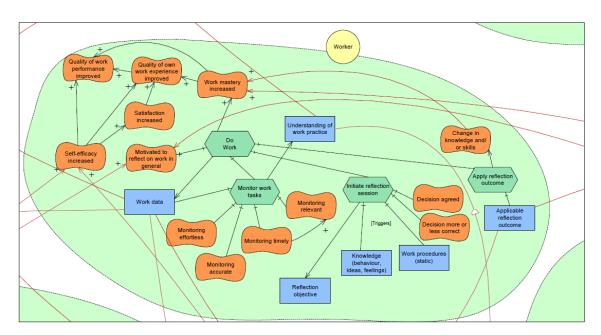


Figure 3: "Worker" in i* Model of Reflective Learning at Work



2.2.2 Individual reflector

The "individual reflector" actor was also developed during the first workshop and focuses specifically on the individual reflection process. The model centres on the main task "Do reflection session" (see Figure 4). The main sub-tasks are:

- Find & create space for reflection
- Reconstruct work experience(s)
- Re-evaluate work experience(s)
- Make reflection outcome applicable

The main soft goals related to the main task "Do reflection session" state that the reflection session shall be:

- Reflection session more successful (in the future)
- Achievable in work constraints (e.g., within the available time/data)
- Efficient
- Effective

Another desired quality of the reflection session is for the individual to be motivated to reflect. Soft goals that are sought to be achieved include:

- General reflection capability increased
- Reflection outcome achieved (it is not necessarily clear what the outcome will be but there is an implicit success criteria)
- Reflection objective met

These soft goals contribute positively to individual worker soft goals such as "Self-efficacy increased". For example, if reflection is done successfully, individuals should have more confidence that they can find solution to work problems, deal competently with them; that is, their self-efficacy should increase.

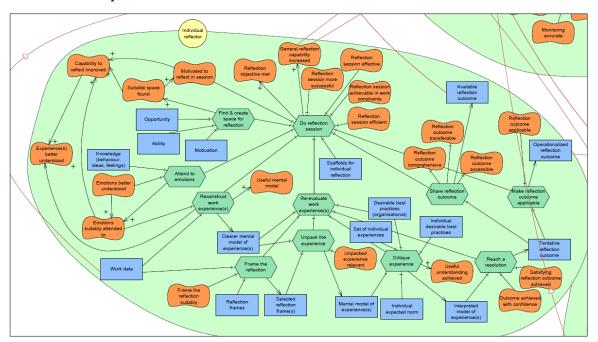




Figure 4: "Individual Reflector" in i* model of Reflective Learning at Work

2.2.3 Individual team reflector

The "individual team reflector" is an individual reflecting individually as part of a group. Even if a person is reflecting with a group, some processes are individual and they are dealt with in this role. The collaborative part is addressed in section 3.2.4 Collaborative Team Reflector Role. An individual switches between both roles frequently while reflecting in a team. This actor was developed during the second workshop to develop an i^* model of collaborative reflection. The main task of the individual team reflector is to "reflect individually within team" (see Figure 5). The main sub-tasks are:

- Attend to emotions
- Re-evaluate own experience
- Articulate reflection outcomes

Also, when the individual shares their experiences with other team members, he/she undertakes the task "Relate own experience to others" as part of re-assessing their own experience.

The main soft goals related to this actor concern the qualities of: attending to emotions, articulating outcomes, finding sufficient time and space to reflect, and increasing their general reflection capability:

- General reflection capability increased
- Emotions suitably attended to
- Emotions better understood
- Re-assessed confidently
- Time and space sufficient
- Articulated clearly to oneself
- Articulated clearly to others

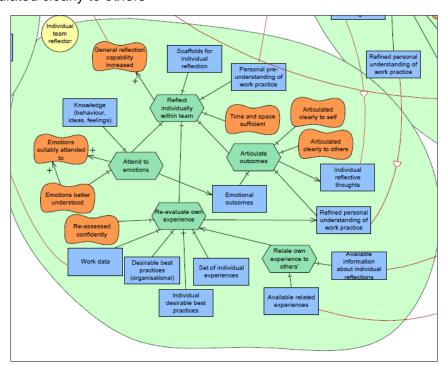




Figure 5: "Individual Team Reflector" in i* model of Reflective Learning at Work

2.2.4 Collaborative team reflector

The "collaborative team reflector" actor was developed during the second workshop to develop an i^* model of collaborative reflection. The collaborative team reflector is a person reflecting with other people, and has the main task "Reflect collaboratively within team" (see Figure 6). The main sub-tasks are:

- Initiate spontaneous collaborative reflection or organise collaborative reflection session
- Make individual reflections available
- Make related experiences available
- Reconstruct work experience(s)
- Make sense of available information
- Re-evaluate work experience(s)
- Make reflection outcome applicable
- Share reflection outcome

There are a number of further subtasks, one of which, "Critique experience", is a key part of the reflection process.

The main soft goals related to this actor concern:

- making reflections and experiences available to the team
- making sense of the available information
- critiquing experiences
- reaching a resolution
- · producing a reflection outcome

The high-level goals include those of the individual reflector and also "Team work practice improved".

The collaborative team reflector actor depends on the individual reflector for "Individual reflective thoughts" and "Refined personal understanding of work practice". There is also a dependency on "Work data" from the individual worker.

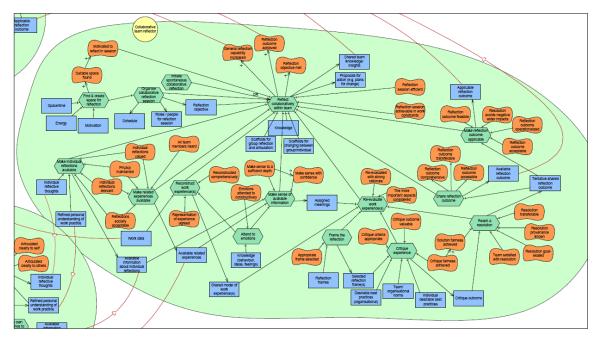


Figure 6: "Collaborative Team Reflector" in i* model of Reflective Learning at Work

2.2.5 Organisational reflector

The "organisational reflector" role is assumed by a person who reflects not on one's own processes (or those of one's own group), but on behalf of the organisation. This actor was developed during the third workshop. The organisational reflector has the main task "Reflection in Business Process Management (BPM)" (see Figure 7). We originally included the whole BPM process in the model but decided to focus on the task "control and improve BPM" as this task is most relevant for reflection. The main subtasks are:

- Control and improve business processes
- Collect and aggregate work data
- Compare AS IS and TO BE
- Analyse deviations
- Specify improvement potential(s)

Soft goals that are sought to be achieved include:

- Data gathering efficient
- Relevant data gathered for reflection
- Key issues/problems identified quickly
- Cause for deviations identified
- Deviations addressed
- Implementation feasible
- Reactions appropriate

The organisational reflector has several dependencies with the worker, individual reflector, and collaborative team reflector, as information gathered for organisational reflection includes data about AS IS processes and individual and team work practice. Implementing improvement ideas in turn affects staff's work practice again.

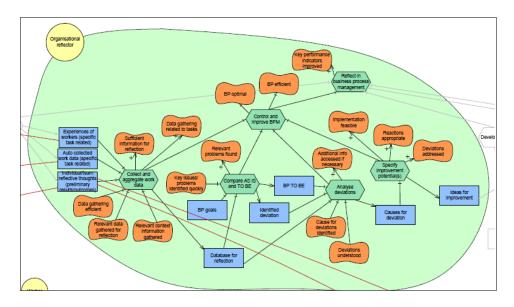


Figure 7: "Organisational Reflector" in i* model of Reflective Learning at Work

2.3 Validating the *i** Model

After constructing the model, we used the third MIRROR General Assembly (November 2011) to review the *i** model with the testbed partners. Five groups, each led by members involved in developing the model (from NTNU, CITY, RUB, and KMRC) and containing members of one of the five testbed partners (Infoman, RNHA, NBN, BT and Regola) discussed the model (involving all of the actors described in section 2.2) together.

The objectives were to identify possible missing tasks or goals and to evaluate the relevance of the task and goals for the testbeds. To make it concrete, we used the storyboards as practical examples. We also began collecting testbed-specific process indicators.

In these discussions, we found that our i^* models already covered almost all relevant task and goals. Minor changes were made to the model, which have been included in Chapter 2.2. We ran a successful trial of using the i^* model in combination with specific app ideas in order to determine relevant evaluation app- and testbed-specific evaluation criteria.

Thus, our *i** approach turned out to be very successful in determining the relevant aspects of learning by reflection and provided a useful basis for our summative evaluation framework.



3 Specifying Summative Evaluation Criteria

Reflection is part of informal learning, and thus, knowledge gains are very context- and person-specific. Employees reflect on their own work experience and the outcomes are highly dependent on the unique nature of these experiences. Thus we used the i^* approach to capture the unique affordances of supporting learning by reflection (see Chapter 3.1).

Even given the informal nature of the content of reflection, it is still possible to evaluate how general work-related criteria are affected by enhancement or facilitation of reflective learning. We therefore refer to the well-established evaluation approach of Kirkpatrick (Kirkpatrick & Kirkpatrick, 2006). However, as it was developed for formal training, we adapted it to our purposes (see Chapter 3.2).

By combining i^* and the Kirkpatrick approach (see Chapter 3.3), we cover both the specific processes and effects of learning by reflection, and the more general effects based on Kirkpatrick's evaluation approach, to construct an evaluation framework with in-depth summative evaluation criteria.

3.1 Evaluation Criteria Based on the *i** Model of Reflective Learning at Work

The i^* model helped us to specify the conceptual model, the links between the stakeholders, processes, activities, and outcomes and clearly articulate goals and objectives on a very detailed level. To facilitate working with the evaluation criteria, we extracted a list of tasks and related subtasks in connection with the associated soft goals from the i^* model (see Table 1). The i^* model contains a goal hierarchy in which achieving lower lever goals will contribute to the achievement of higher level goals.

Table 1: Emergent Evaluation criteria from the i* Model of Reflective Learning at Work

Actor/Role	Task/Subtask	Soft goal
Worker	Perform task	Motivated to reflect on work in general
		Quality of own work experience improved
		Quality of work performance improved
		Satisfaction increased
		Self-efficacy increased
		Work mastery increased
	Monitor work tasks	Monitoring accurate
		Monitoring effortless
		Monitoring relevant
		Monitoring timely
	Initiate reflection session	Decision agreed
		Decision correct
	Apply reflection outcome	Work practice improved
Individual Reflector	Do reflection session	Capability to reflect improved
		Reflection session effective
		Reflection session efficient
		Reflection session successful
		Reflection session achievable in work constraints
	Do reflection session	Reflection objective met
	Find & create space for reflection	Suitable space found
	Attend to emotions	Emotions better understood
		Emotions suitably attended to
	Reconstruct work experience(s)	Useful mental model
	Frame the reflection	Frame the reflection suitably
	Unpack the experience	Unpacked experience relevant
	Critique experience	Useful understanding achieved

	Reach a resolution	Outcome achieved with confidence
	Chara reflection outs	Satisfying reflection outcome achieved
	Share reflection outcome	Reflection outcome accessible
		Reflection outcome comprehensive
	Make reflection outcome applicable	Reflection outcome transferable Reflection outcome applicable
	Make reflection outcome applicable	Tenection outcome applicable
dividual Team Reflector	Reflect individually within team	General reflection capability increased
	•	Time and space sufficient
	Attend to emotions	Emotions better understood
		Emotions suitably attended to
	Re-evaluate own experience	Re-assessed confidently
	Articulate outcomes	Articulated clearly to self
		Articulated clearly to others
ollaborative Team Reflector	Reflect collaboratively within team	Reflection session efficient
		General reflection capability increased
		Motivated to reflect in session
		Reflection objective met
		Reflection outcome achieved
		Reflection session achievable in work constraints
	Find & create space for reflection	Suitable space found
	Make individual reflections available	Privacy maintained
	Make Individual reliections available	Individual reflections valued
		Individual reflections valued
	Make related experiences available	All team members heard
	make related experiences available	Reflections socially acceptable
	Reconstruct work experience(s)	Reconstructed comprehensively
	Reconstruct work experience(s)	Representation of experience agreed
	Attend to emotions	Emotions attended to constructively
	Make sense of available information	Make sense with confidence
	Make sense of available information	Make sense to a sufficient depth
	Re-evaluate work experience(s)	Re-evaluated with strong rationale
	re-evaluate work experience(s)	The more important aspects considered
	Frame the reflection	Appropriate frame selected
	Critique experience	Critique outcome valuable
	Ontique experience	Critique criteria appropriate
		Critique fairness achieved
	Reach a resolution	Solution fairness achieved
	Reach a resolution	Resolution goal-related
		Resolution provenance known
		Resolution transferable
		Team satisfied with resolution
	Share reflection outcome	Reflection outcome accessible
	Share reflection outcome	
		Reflection outcome comprehensive
	Make reflection outcome applicable	Reflection outcome transferable
	wake reliection outcome applicable	Reflection outcome feasible Reflection outcome operationalised
		Resolution avoids negative wider impacts Reflection outcome acceptable
		Removing acceptable
rganisational Reflector	Reflect in business process management	Key performance indicators improved
-	Control and improve BPM	BP efficient
	<u> </u>	BP optimal
	Collect and aggregate work data	Data gathering related to tasks
		Data gathering efficient
		Relevant context information gathered
		Relevant data gathered for reflection
		Sufficient information for reflection
	Compare AS IS and TO BE	Key issues/problems identified quickly
		Relevant problems found
	Analyse deviations	Cause for deviations identified
	•	Deviations understood
		Additional info accessed if necessary
	Specify improvement potential(s)	Deviations addressed
		Implementation feasible
		Reactions appropriate

Note that this table shows the all task-soft goal relationships of the model. Not all task-soft goal relationships are relevant for each testbed and for each App. Each app has a specific



focus and is limited to the specific aspects of reflection it supports; this issue must be addressed by summative evaluation. We will refer to this issue again in Chapter 4.

3.2 Overview of MIRROR Summative Evaluation Criteria

Based on Table 1, the MIRROR stakeholders' soft goals were then generalized and aggregated to derive relevant evaluation criteria (see Figure 8). These criteria are associated with work performance, the reflection process, and the reflection outcome. This clustering is in line with Boud's model of reflective learning (see Boud et al., 1985), who differentiates experience (i.e., work), reflection as a process, and learning as the outcome of reflection. Furthermore, we included criteria that are specific for organisational learning by reflection, and refer to business process management effectiveness.

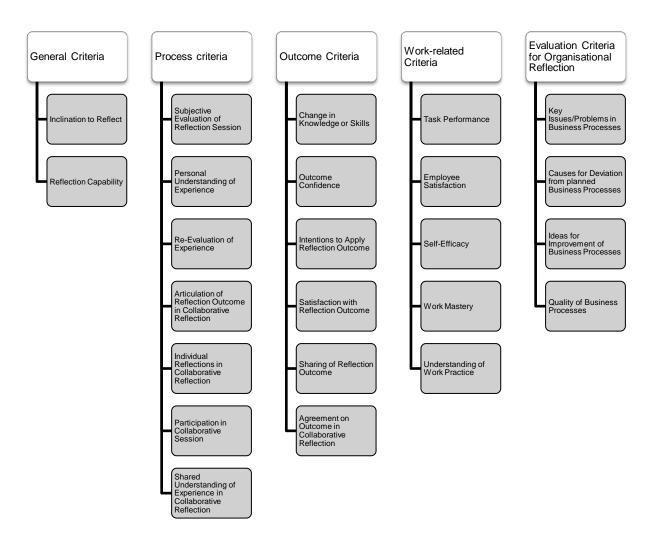


Figure 8: MIRROR Summative Evaluation Criteria based on the i* Model

In line with the **four levels of evaluation developed by Kirkpatrick**, we similarly consider four levels of summative evaluation. Because MIRROR supports informal learning (rather than Kirkpatrick's formal learning) and because we are interested in the reflection process itself, **we have modified these levels slightly**, as described below (see a full description of the original model in Appendix B). The four levels are inspected from a summative evaluation perspective, with the assumption that lower levels are prerequisites for higher levels. That is,



positive results of MIRROR app usage on the business level can only be seen if behaviour has changed as a consequence of reflective learning that occurs when the app is appropriately used.

Level 1: Reaction

To what degree do participants react favourably to our MIRROR apps?

In the case of MIRROR, reactions to instances of the app usage in the testbeds are measured (ranging from a simple one-time test to weeks-long implementations). Because MIRROR is concerned with informal learning, in our evaluation framework Level 1 additionally includes whether participants are motivated to use the app in the first place. This can be measured by recording when and how the apps are used (i.e., with log files). In addition, we can evaluate the user experience through subjective evaluations. There is much overlap with formative evaluation regarding this level, but in the summative evaluation we focus more on usage as a precondition for the next levels.

Level 2: Learning

To what degree do participants acquire knowledge, skills, attitudes, confidence, and commitment?

Level 2 is concerned with what app users learn, and how that learning is aligned with what we intend for them to learn. MIRROR's intention is that users should reflect while using the app and the outcome of that process is learning. For that reason, in Level 2 we evaluate the process of reflection as well as learning.

The reflection sessions can result in either:

- a. Change in knowledge/skills by understanding how specific competencies or processes can be improved
- b. Behavioural intentions indicating how the reflection outcome may be applied in daily work

Learning can be assessed by specific evaluation forms that ask for documentation of learning outcomes or subjective ratings. Quantity and quality of these learning outcomes can then be examined.

Level 3: Behaviour

To what degree do participants apply what they learn?

Assuming learning has occurred, and that the work environment is conducive, it is now possible to measure the conversion rate of behavioural intentions into actions. Level 3 is concerned with whether the new knowledge and skills are put to use. We are interested specifically in behavioural changes' positive impact on work performance. We can measure this in terms of employee-relevant testbed KPIs, via manager ratings of work behaviour, and with subjective ratings of improvement.

Level 4: Results

To what degree do targeted outcomes occur as a result of MIRROR?

In Level 4, the effect of app users' work changes on the business is evaluated. The potential impact of changes in action on the business should be considered during definition of the action. Theoretically, change of organization-level KPIs (sales revenue,



profit, and customer satisfaction) can be impacted by participants' actions, but in some cases it can be difficult to isolate the effects of the specific action because many parameters impact business KPIs. However, by isolating business measures to include only the relevant employees (i.e., MIRROR participants), it becomes possible to see these effects.

Furthermore, we can measure the success of the apps in supporting reflection by measuring the uptake by the community using a loyalty metric like the *Net Promoter* metric ("The Net Promoter Score and System", 2012). We can also measure how often the MIRROR approach is referred to other potential users and we can ask users to rate the apps (as is now common practice in some app stores).

Figure 9 shows these four levels of evaluation, *based on* the Kirkpatrick model and modified to apply to MIRROR's informal, reflective learning situation. The grey boxes on the left specify the specific processes or outcomes that are evaluated at each level; the blue boxes in the middle specify how those are instantiated in MIRROR; and the orange boxes on the right specify relevant types of evaluation criteria.

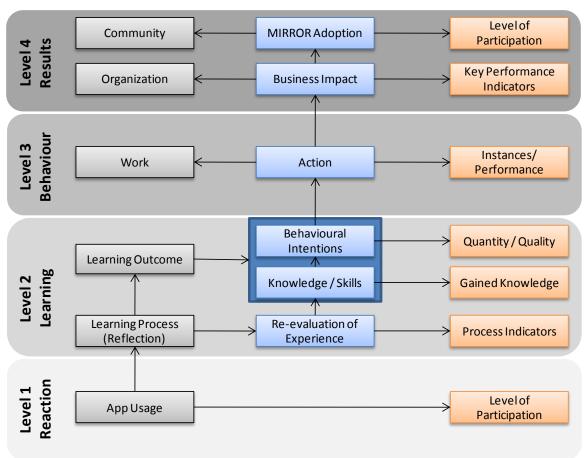


Figure 9: Overview of Levels of Evaluation and Evaluation Criteria

These evaluation criteria are further specified by combining the modified Kirkpatrick model (as shown in Figure 9) with the criteria from the *i** model (as shown in Figure 8). The *i** model criteria add detail and concreteness to the four level approach of Kirkpatrick. This provides a useful evaluation framework that specifies what types of changes we expect at each level. The mapping of how these two models can be combined is explained in Table 2.



Table 2: Four Levels of Evaluation, i* Criteria and Rationale

Level	i* Criteria	Rationale			
1 Reaction	General Criteria	The i^* general criteria are concerned with the motivation and opportunity to			
		reflect; in other words, whether participants are motivated to use the app.			
		This inclination is affected by how well participants like the app.			
2 Learning	Process Criteria	The i^* process criteria shed light onto whether and how participants engage			
		in the process necessary for learning: in our project this is the reflection			
		process.			
	Outcome Criteria	Outcome criteria of i^* are mainly related to learning and include change in			
		knowledge and behavioural intentions.			
3 Behaviour	Work-Related	Work-related i* criteria concern concrete action or they are related to			
	Criteria	behaviour (e.g., self-efficacy, work mastery, and employee satisfaction			
		concern subjective evaluation of performance at work).			
4 Results	Business Impact	Evaluation criteria for organisational learning are related to the high-level			
	•	(business) impact of MIRROR.			

3.3 Methodological Considerations

An evaluation toolbox has been created and is similar to the User Studies Toolbox (cp. D1.1). It specifies research instruments and measures that can be used before and after implementation of MIRROR. Table 3 outlines general methods to assess the evaluation criteria, and the Toolbox is presented in section 4.

Table 3: Summative Evaluation Criteria: Description and Evaluation Tools and Methods

Evaluation leve ⁾ /i*Criteria	Description and characteristics	Possible tools and methods
Reaction/General Criteria	How high is the level of participation? How easy and comfortable is the user experience? How practical and useful is the long-term usage perceived?	Questionnaires Log file data (e.g., user actions).
2. Learning / Process and Outcome Criteria	Did the learner gain a deeper understanding of his / her work experience? What is the extent of advancement or change in the direction or area that was intended? Is the learner aware of their change in behaviour, knowledge, skill level?	Interviews after app usage. Post-usage surveys or questionnaires. Thinking aloud protocols and audio / video recordings of collaborative reflection sessions.
3. Behaviour / Work-Related Criteria	Did the learner put his/her learning into effect when back on the job? Was there noticeable and measurable change in the activity and performances? Was the change in behaviour and new level of knowledge sustained?	Observation and interview over time to assess change, relevance of change, and sustainability of change. Assessments of relevant performance scenarios, and specific key performance indicators or criteria. Self-assessment using carefully designed criteria and measurements.
4. Results / Business Impact	Business or organisational key performance indicators, such as: Volumes, values, percentages, timescales, return on investment, and other quantifiable aspects of organisational performance, for instance; numbers of complaints, staff turnover, attrition, failures, wastage, non-compliance, quality	Usually, KPIs are already in place via normal management systems and reporting.



Specification of Evaluation Methodology and Research Tooling

Page 33

ratings, achievement of standards and accreditations, growth, retention, etc.



4 Evaluation Toolbox

Based on the specification of the summative evaluation criteria, we created a toolbox similar to the "Specification of Research Methodology and Research Tooling" (Deliverable 1.1). We present the toolbox here mostly as a collection of questions suitable for written questionnaires. However, these questions can easily be modified to be used in interviews and focus groups. The toolbox is also complemented by other data sources, e.g., log files or log file equivalents (e.g., learning diaries used by WP7).

Not every question of the toolbox can – or should – be used for every evaluation of an application. The toolbox was developed like a literal toolbox – as a repertoire from which tools are selected for specific purposes. Given that the apps have a specific focus and are used by voluntary participants in the testbed (whose time is valuable), only a limited number of questions are mandatory. These core questions, which make up the centrepiece of our summative evaluation framework, are presented here. We have included the whole Toolbox (including these core questions) similar to D1.1 in Appendix A of this document.

The **Core Questions** presented here are assessed each time a significant app test is done in the testbeds, i.e., there is contact with the actual target audience of the apps, the apps are actually used, and data is produced. Whereas the questions for Levels 1, 2, and 3 of the modified Kirkpatrick model will be answered by participants in a questionnaire, Level 4 can also be addressed by organizational level measures. For the answer scale of the questions see Appendix A.

We have organized the questions (the core questions here and the whole toolbox in Appendix A) according to our modification of the Kirkpatrick levels which was informed by the i^* modelling (see Chapter 3). This provides us with a **hierarchical order of evaluation levels, where each lower-order level is necessary for the next level**. That is, usage of the app is required for it to support reflection and thereby learning; only when learning outcomes are produced can the behaviour be changed in an informed way; and only if employee behaviour changes can the organization change. If support of learning by reflection is hindered, we will be able to see at which level the failure occurred.

We are confident that we cover the most relevant aspects of summative evaluation, but additional questions can still be added to this framework (the apps will continue to be developed over the next two years, see also section 4.7). Note that the procedure for the summative evaluation is similar to the user studies done earlier in the project, following the same code of conduct and ethical guidelines. This approach proved to be feasible. Developers, in collaboration with the testbeds, are responsible for complying with ethical standards (e.g., disclaimers, confidentiality agreements, etc.) as specified in D1.1, Chapters 8 and 9.

4.1 Demographic Information

Some data about the participants is necessary to connect the participant data across different implementations of an app (or different apps). Each question in this and the



following sections is labelled with question identifiers to allow for data integration. Here, **CD** stands for **C**ore Question **D**emographic items.

CD1 Participant ID

The Participant ID consists of the first letter of the participant's place of birth, the first letter of the participant's father's first name, the first letter of the participant's mother's first name, and the participant's own day of birth (two digits). If any of these elements are unknown, the participant uses a placeholder (X for the first unknown, Y for the second, Z for the third).

The following text should be used on every questionnaire administered:

Please write down your Participant Code. Your code consists of:

- 1. The first letter of your place of birth
- 2. Your own day of birth (two digits)
- 3. The first letter of your father's first name
- 4. The first letter of your mother's first name

Example: A person born in London on the 7th of July, with parents named Jake and Sue and born would enter: "L" (for London) in the 1st blank, and "07" (for a birthday on the 7th) in the 2nd blank, "J" (for Jack) in the 3rd blank, and "S" (for Sue) in the 4th blank. So, this person's code would be: <u>L</u> <u>07</u> <u>J</u> <u>S</u>.

If you don't know any of these, use the letter below the correct box shown below.

	1 st letter of your place of birth	Your day of birth	1 st letter of your father's first name	1 st letter of your mother's first name
Your Code				
If unknown, use:	X	00	Y	Z

CD2 Team-ID

Only if the app is used in teams: Assign a common ID if the app is used within teams in testbeds (e.g., "team A", "team B").

- CD3 Current Date (dd-mm-yyyy)
- **CD4 Gender** (1 = Male, 2 = Female), might not be allowed in all testbeds
- **CD5** Age (range: $1 = \le 19$, 2 = 20-29, 3 = 30-39, 4 = 40-49, 5 = 50-59, $6 = \ge 60$)
- **CD6 Job Scope** (1 = Full-time, 2 = Part-time)



CD7 Department

CD8 Position

CD9 Years in current position

CD10 Years in current team (if applicable)

CD11 Years in similar positions (e.g., at another company)

4.2 Level 1: Reaction (Usage)

Although usability is firmly an issue of formative evaluation (see Deliverable 10.2), app usage is a precondition for any higher level in the evaluation model. Thus, we are concerned about the actual usage numbers and qualities.

If possible, **log files** should be used that provide simplified logging of the desired data (e.g. start and stop date/time, allowing quick calculation of usage times and numbers). If this is not possible, we have provided questions that can be asked regarding app usage.

The data can later be aggregated to indicate whether the app was used as intended. Developers should determine a list of features of the app that are necessary to be used in combination to successfully support reflection and to be considered "intended usage". This information could then be rated for quality; depending on the number and order of functions used, developers can use a scale indicating how close to intended use the usage was.

CL stands for Core Question Log File, while CU stands for Core Question Usage.

CF1 Log File Data: Number of times used

CF2 Log File Data: Total time (minutes) used

CF3 Log File Data: Average time (minutes) used

CF4 Log File Data: Number of times each key function of app is used

For example with games: How often was the game completed? For other apps: How often were all necessary functions to support reflection used? What "completely" means is determined by the app developers in collaboration with the test beds and stated in the App Evaluation Sheet.

Note: The following self-report questions should only be used if the log file data for the app is not available.

CU1 Self-Report: Number of times used

How many times have you used [the app]?

CU2 Self-Report: Total time (minutes) used

How many minutes did you spend using [the app] in total?

CU3 Self-Report: Average time (minutes) used

How many minutes did you spend using [the app] on average?

CU4 Self-Report: Number of times app is used completely

How many times did you use [specific function] of the app? (repeat for each key function)



4.3 Level 2: Learning (App-Specific Reflection Questions, Short Reflection Scale, and Learning Outcomes)

4.3.1 App-specific Reflection Questions

A central aspect of our summative evaluation framework is the actual support of reflection that the apps provide. Thus, we refer here to the theoretical model of the project (CSRL model, see Deliverable 1.4). Each app has been designed to support certain processes during reflective learning. We provide questions to evaluate how they addressed these areas of support.

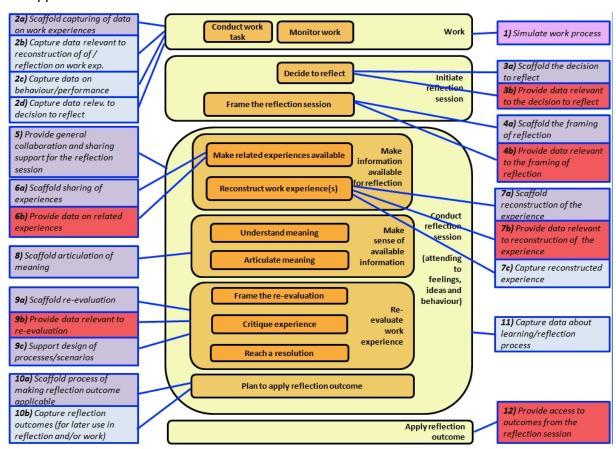


Figure 10: Computer Supported Reflective Learning Model (CSRL)

The amount of questions chosen from this section depends on the breadth of functions the app in question provides to support reflection. The relevant questions (functions the app provides) will be chosen by the developers for each app. Additionally, one to three questions are added where no changes are expected. This will help to identify whether the participants only report positive reactions because they want to please the designers or evaluators (social desirability/demand characteristics). If positive effects are seen only for the supported functions, this gives confidence that it actually is a true effect. The questions are reformulated for the specific app name, i.e., "[The app]" is replaced with the actual name of the app.

Note: The descriptions in italics and the ordering are taken from Deliverable 1.4. **CA** stands for **C**ore Question **A**pp-Specific Reflection Question.



Capture data relevant to reconstruction of / reflection on work experience (2b in CSRL)

This could be any type of data that can be used to aid reconstruction and reflection. The capturing could be done automatically and/or manually.

CA1 [The app] helped me to collect information relevant to reconstructing experiences from work.

CA2 [The app] helped me to reflect on experiences from work.

Capture data on behaviour/performance (2c in CSRL)

This would for instance be with the purpose of evaluating whether performance is improved and/or learning has taken place.

- **CA3** [The app] helped me to collect data on behaviour before the reflection session.
- **CA4** [The app] helped me to collect data on behaviour after the reflection session.

Capture data relevant to decision to reflect (2d in CSRL)

This could be any data that may be used to infer that it might be time to reflect.

CA5 [The app] helped me to collect information that could help me decide when to reflect about my work.

Capture reconstructed experience (7c in CSRL)

For instance, when the learner has been using a tool to aid the reconstruction of experience, the result could be captured in the tool.

CA6 [The app] helped me to reconstruct a work experience.

Capture reflection outcomes (for later use in reflection and/or work) (10b in CSRL)

This could be temporary/partial or final outcomes.

- **CA7** [The app] helped me by capturing my reflection outcomes.
- **CA8** [The app] helped me by making reflection outcomes available for later use.

Capture data about learning/reflection process (11 in CSRL)

This could be any data from the reflection session that can be used to evaluate the process.

CA9 [The app] helped me by capturing information for evaluation of learning/reflection.

Provide data relevant to the decision to reflect (3b in CSRL)

This could be any data that may be used by a learner to infer that it is time to reflect.

- **CA10** [The app] helped me by reminding me to reflect.
- **CA11** [The app] helped me by providing information relevant for the decision to reflect.
- **CA12** [The app] helped me by providing accurate information about my work.

Provide data relevant to the framing of reflection (4b in CSRL)

This includes data that can help set the scene for reflection e.g. by showing the availability of resources (e.g. people, information, rooms...) and helping structure/allocate them.

CA13 [The app] helped me by providing information relevant for the framing of reflection.



CA14 [The app] helped me by showing the availability of resources needed for reflecting.

CA15 [The app] helped me to allocate or structure the resources needed for reflection.

Provide data on related experiences (6b in CSRL)

This is about providing access to relevant data/information on related experiences, for instance others' experience of the same or similar situation.

CA16 [The app] helped me by providing information about related experiences.

Provide data relevant to the reconstruction of experience (7b in CSRL)

This could be data that are somehow seen as elements of the experience itself or that gives context and/or aids recall.

CA17 [The app] helped me to remember and reconstruct the experience/situation.

Provide data relevant to re-evaluation (9b in CSRL)

Provide data that help in evaluating the experience and consider alternatives. Includes descriptions of particular cases/episodes, more general problem-solution patterns (in the same work domain or a different one), and simulations.

CA18 [The app] helped me by providing access to data (e.g., simulations) relevant to the re-evaluation of experience.

CA19 [The app] helped me by providing access to data relevant to the experience.

Provide access to outcomes of the reflection session (12 in CSRL)

This is about making results from the reflection session accessible in the work situation (where considerations include making them available at the right time and in a useful format).

CA20 [The app] helped me by providing access to resources resulting from reflection sessions.

Scaffold capturing of data on work experiences (2a in CSRL)

Provide process guidance and/or templates/structuring to aid the learner(s) in capturing data that can later be used to reconstruct and reflect on work experiences.

CA21 [The app] guided me in capturing information about my work experiences.

Scaffold the decision to reflect (3a in CSRL)

Provide process guidance and/or templates/structuring to aid the learner(s) in deciding whether to initiate a reflection session.

CA22 [The app] guided me in deciding whether/when to reflect.

Scaffold the framing of reflection (4a in CSRL)

Provide process guidance and/or templates/structuring to aid the learner(s) in setting the scene for reflection e.g. by showing the availability of resources (such as people, data, room...) and helping structure/allocate them.

CA23 [The app] guided me in finding the resources needed for reflection.

CA24 [The app] guided me in allocating/structuring the resources needed for reflection.

Scaffold sharing of experiences (6a in CSRL)



Provide process guidance and/or templates/structuring to aid the learner(s) in sharing experiences

CA25 [The app] helped me by supporting sharing of experiences.

CA26 [The app] guided me in sharing experiences with others.

Scaffold reconstruction of experience (7a in CSRL)

Provide process guidance and/or templates/structuring to aid the learner(s) in reconstructing experience.

CA27 [The app] guided me in reconstructing and remembering the experience/situation.

Scaffold articulation of meaning (8 in CSRL)

Provide process guidance and/or templates/structuring to help the learner(s) express what the experience means (to self or others) (in light of their objectives, values etc.). This could also include guidance for negotiation of meaning and documentation of viewpoints.

CA28 [The app] guided me in articulating the meaning of an experience.

CA29 [The app] guided us in negotiating the meaning of an experience.

CA30 [The app] guided us in documenting different viewpoints on the experience.

Scaffold re-evaluation (9a in CSRL)

Provide process guidance and/or templates/structuring to help the learner(s) reevaluate experience and reach a resolution.

CA31 [The app] guided me in re-evaluating an experience.

CA32 [The app] guided me in reaching a resolution.

Scaffold process of making reflection outcome applicable (10a in CSRL)

Provide process guidance and/or templates/structuring to help the learner(s) consider how to make the outcome applicable to the work (and further reflection). This could include considering constraints and the option of not applying the resolution.

CA33 [The app] guided me in making the reflection outcome applicable to my work.

CA34 [The app] guided me in making the reflection outcome applicable to further reflection.

CA35 [The app] guided me in considering constraints of the reflection outcome.

CA36 [The app] guided me in considering the option of not applying the reflection outcome.

The following two categories can also be considered as scaffolding in a broad sense:

Support design of processes/scenarios (9c in CSRL)

Aid learners in describing processes/scenarios, typically those with a potential to lead to desired work experiences/results (but could also include less desirable processes/scenarios, e.g. to cover and compare 'good practice' and 'bad practice')

CA37 [The app] guided me in describing work scenarios that could lead to desired results.

CA38 [The app] guided me in describing both "good practice" and "bad practice" work scenarios.

Provide general collaboration and sharing support for the reflection session (5 in CSRL)



A category designed to cover general process support for the reflection session applying across all its steps, i.e., infrastructure for reflection.

CA39 [The app] provided help with collaboration.

CA40 [The app] provided relevant content for reflection.

CA41 [The app] guided me through the reflection process.

Use of tools for simulating the work process

Among the categories of tool use for reflection we include tools simulating the work process, since experience gained through the simulation can serve as a source of reflective learning about work. This provides the learner with an opportunity to get virtual experience from the work domain.

CA42 [The app] helped me by simulating the work process.

CA43 [The app] helped me by providing me with virtual experience in my work domain.

4.3.2 Short Reflection Scale

The Short Reflection Scale (developed in collaboration with WP4 and WP6) assesses participants' general tendency to reflect and the importance they place on reflection. This scale is different from the App-Specific Reflection Questions in that one does not need to use an App to answer these questions. Rather, this scale will allow us to see whether using the apps prime people to reflect more; that is, whether our apps increase participants' tendency or inclination to reflect both individually and in teams. This scale is intended to be used both pre- and post-implementation in order to assess this expected increase. The phrases in square brackets should be replaced with the testbed's relevant work task(s). Note that organisational reflection is not included in this scale, as it is intended to assess factors directly related to the individual employee. **CR** stands for **C**ore Question Short **R**eflection Scale item.

- **CR2** We as a team often reflect on our work in order to improve it.
- **CR3** I think it is important to try to improve [specific work task].
- **CR4** I frequently reflect on [specific work task].
- **CR5** Reflecting on [specific work task] helps me to improve [the task].
- CR6 In team meetings we frequently talk about how we can improve [specific work task].
- **CR7** Outside of meetings, I often talk with my colleagues about [specific work task].
- **CR8** It is important to me to discuss frequently with others about [specific work task].
- **CR9** Conversations with colleagues help me to improve [specific work task].
- **CR10** Even a few days later, I can remember the [specific work task/event] well when I reflect on it by myself or with others.



4.3.3 Learning outcomes

On the outcome level of reflection (learning), we ask the two mandatory subjective questions below. Further questions are listed in Appendix A. **CL** stands for **C**ore Question **L**earning item.

- **CL1** I made a conscious decision about how to behave in the future.
- **CL2** I gained a deeper understanding of my work life.

4.4 Level 3: Behaviour

4.4.1 Subjective improvement in work

To keep the amount of questions as short as possible, we only ask participants a single item regarding the behaviour level as a core question. Though other possible questions are listed in Appendix A, this core question assesses the central aspect: Did the work behaviour improve?

The question is formulated in a way that if a specific work behaviour is the target of the reflection support (e.g., time management), the placeholder [work performance] is replaced by the specific work behaviour (here: "time management"). The default of "work performance" is only used if the app aims to improve work behaviour in general. By asking about the specific supported behaviour, we are more likely to detect changes. **CB** stands for **C**ore Question **B**ehaviour item.

CB1 [The app] helped me improve my [work performance].

Note that in the special case of RNHA, it is more important that a particular resident's care improves (as a result of employees' work behaviour changes) than that an employees work, *per se*, improves. To this end, employees can also be asked about their care of a particular resident by inserting "care of [identifier of Resident X]" in the [work performance] placeholder. This question could also potentially be used in other situations, for example, if a company is particularly interested in individual customers' satisfaction (rather than an average measure).

An employee's changes in work behaviour can also be assessed by his or her manager. When possible, managers' ratings of relevant behaviour should be requested. In the case of RNHA, managers can also be asked about the resident's care.

Some key performance indicators (KPIs) listed by the testbeds (see Appendix A, section 9.3.4.1) are relevant to employee behavioural change. For example, although time on task might be relevant for an individual employee, it can also be assessed on an organisational level with the aim of improving the organisation or to ensure its function (e.g., payment according to time spent on task). In many cases, these measures are already assessed by the organization. For Level 3, we are interested in these measures on an individual level, but they can also be applied to Level 4 on an organizational level.

4.5 Level 4: Results

Change over time in testbed-relevant organizational-level KPIs is assessed. Relevant KPIs provided by each testbed are listed here. Additional KPIs are listed in Appendix A, section 9.3.4.1. The KPIs should be measured starting at the first implementation until the end of (or even shortly after) all implementations. Care should be taken that only the relevant



unit(s)/personnel who used the apps are assessed, in order to more accurately detect changes. It is important to note that context factors that cannot be controlled or influenced by the apps during the project can dampen changes that could be observed; for this reason, precision in measurement is key.

We identified 8 general categories of KPIs derived from the important KPIs provided by testbeds:

- 1. Decreased number of negative events
- 2. Increased client satisfaction
- 3. Decreased staff turnover
- 4. Decreased time on task
- 5. Increased quality of work
- 6. Increased employee satisfaction
- 7. Matching external criteria (standards/accreditation)
- 8. Compatibility with existing procedures

To ease comparison across testbeds that operationalise KPIs differently, we use the categories and numbering above. In the following, $CK[Number]\{B|I|R|N|H\}$ stands for **C**ore Question **K**PI measure, with the number indicating the KPI category, and the last letter indicating the testbed.

4.5.1 BT

British Telecom (BT) is a large telecommunications company, serving customers in more than 170 countries. BT manages its 1500 large scale (often highly customized and individual) contracts by contract teams.

The target group for MIRROR are the members of these contract teams. Most employees work from home as teleworkers. Teams are dispersed all over the country, and they are managed virtually. Thus, knowledge management and experience exchange is a major issue at this testbed. Work is highly standardized on the project level, i.e. there is a standard business process for contract management. There is a huge range of formal training opportunities, many of them available in the company's e-learning system. Performance management is implemented to ensure quality of work.

Relevant KPIs are:

CK1B customer issue reduction

CK2B customer satisfaction increases

CK4B Right First Time

CK5B Individual / Team Performance (quarterly performance reviews, individual or team average: outstanding, very good, achieve standards, development needed, unsatisfactory)



4.5.2 Infoman

Infoman AG is an IT company that consults on, sells, and personalizes Microsoft Customer Relationship Management Software to help analyse and optimise the marketing, sales, and service processes of their customer companies.

People mainly work in small teams of two to three people. Altogether, the company has about 60 employees, most of them based at headquarters. However, they have a lot of meetings with customers at the customers' sites, which require internal preparation and post-processing. Daily work is heavily focused on customers' needs, which require a high degree of flexibility and the development of individual best practice. Consulting and sales thus involve a high degree of reflection on interaction with the customer. Therefore, knowledge management and sharing is considered to be a major challenge at this testbed.

Infoman aims for their sales and consultant personnel to do better work. Better work means that either they need less time for the same tasks or their work results are of better quality. Another relevant KPI is the employees' subjective work satisfaction.

Relevant KPIs are:

CK2I customer satisfaction increases

CK4al less time on task

CK4bl less time spent on searching for knowledge

CK4cl amount of reuse of PowerPoint slides

CK5I higher quality of work

CK8I smooth integration of app use in the normal work flow

4.5.3 NBN

The Neurological Clinic (NBN) is a large, modern hospital in Germany with approximately 400 full-time equivalent employees (FTE's) dealing with approximately 1000 strokes a year (approx. 2000 emergencies altogether, including severe neurological emergencies other than stroke). The Neurological Clinic has a strong interest in improving their daily business processes. They have both quality and error management processes in place to ensure quality of work.

The Stroke Unit was selected as department for the user studies because the work in this emergency unit is characterized by high time pressure and stress levels on the one hand, whereas on the other hand this department is widely regarded in the clinic – by clinic and management staff – as role model for other departments. Employees work in interdisciplinary teams of physicians, care staff, and therapists. Work is organized in shifts, and there are regular, well-structured handovers.

Relevant KPIs are:

CK1N reduction of complaints

CK2N customer satisfaction increases

CK4N identify duplication of processes/process error

CK6aN employee satisfaction



CK6bN mental and physical health of employees

4.5.4 Regola

Regola is a company that has been active for over fifteen years in the field of Information & Communication Technology (ICT). Thanks to its wide-ranging skills in both the IT and media industries, Regola is able to offer integrated solutions for highly complex projects. Regola's most important dimension is the area of computer solutions for the health and emergency management sector. Support of individual and collaborative (team) reflection in large amounts of volunteers could lead to huge learning gains that are previously untapped.

Relevant KPIs are:

CK3R employee retention/turnover (number volunteers in and out, adherence to assigned task)

CK4R time on task (average call time)

CK5aR critical incident performance (% deviation from degree of severity)

CK5bR service quality (operation time)

CK7R achievement of standards and accreditation

4.5.5 RNHA

The Registered Nursing Home Association (RNHA) is an association of nursing homes in the UK, committed to the provision of high quality services to residents in care homes. A specific characteristic of this application partner is that it comprises multiple, independent homes, which permits support for reflection to be implemented in different homes to test the applicability of different approaches. High levels of participation within MIRROR have been experienced from amongst the homes that have volunteered to be MIRROR test sites.

A growing challenge for nursing homes is the higher proportion of increasingly elderly residents suffering from dementia when admitted to the homes. This can lead to instances of challenging behaviour where the elderly people are confused and react, sometimes aggressively and irrationally, to their unfamiliar surroundings. This requires high amounts of reflection on the side of the carers and nurses working in the homes, as there is no one-size-fits-all solution when dealing with personalities approaching the end of their lives with their individual and complex life-histories.

Most of the care staff, except for recently qualified nurses, are not educated to degree level and only have National Vocational Qualifications. This means that staff without formal training can be confronted with complex situations to resolve. Work is organized in day and night shifts with handovers; protocols document every treatment and activity.

Relevant KPIs are:

CK1aH reduction in the number of incidents of challenging behaviour/difficult relations with residents

CK1bH reduction in the severity of incidents of challenging behaviour/difficult relations with residents

CK2H increase in happiness of the residents (as seen by their relatives)

CK3H decreased staff turnover



4.6 Participant Characteristics

We are interested in the effects of the apps on learning by reflection, and individual abilities and attitudes can influence these effects. To take these participant characteristics into account, we ask (potential) participants of the apps fill out a questionnaire that assesses personality traits and attitudes relevant for reflection. Given that traits are by definition stable over time, they only need to be measured once in the beginning of evaluation and once at the end of the project (to get a reliability estimate). Based on the psychological process of reflection, we think that the following traits, abilities, and attitudes are relevant:

Need for Cognition

One's need for cognition refers to the extent to which people enjoy and engage in challenging cognitive activities. As reflection requires participants to think about their past behaviour, this trait is very relevant. We use the short 18-item scale developed and validated by Cacioppo, Petty, and Kao (1984) to assess need for cognition.

• Ambiguity Tolerance

Ambiguity tolerance is the ability to perceive ambiguous information and behaviour neutrally and openly. Given that reflection requires participants to deal with often complex and ambiguous information, we think this trait is relevant. We use Budner's (1962) original scale, which is comprised of 16 items, to assess ambiguity tolerance.

Mindfulness

Mindfulness refers to paying attention to one's experience on a moment-to-moment basis. Reflection deals with past behaviour, and the memory of past behaviour might be influenced by the way participants perceive events they encounter. One common measure for mindfulness is the Kentucky Inventory of Mindfulness Skills (KIMS). We use a shortened version (KIMS-Short), containing 20 items, which was validated by Höfling, Ströhle, Michalak, and Heidenreich (2011). These items already exist in English and German.

Self-Efficacy

Self-efficacy refers to one's believe in one's own competence. As reflection should lead to changes in (work) behaviour (and thereby to organisational changes), which requires the belief that this is actually possible, we think this is an important aspect of reflection. We use the 10-item General Self-Efficacy Scale (GSE) by Schwarzer and Jerusalem (1995), which specifically measures perceived self-efficacy of dealing with daily hassles and stressful life events. It is available in many different languages, including the three languages used in the testbeds: English, German, and Italian.

IT Attitudes

The questionnaire is an abbreviated version of the IT Attitudes and Usage Questionnaire from the User Studies (see D1.1). To reduce workload of the participants, we focus only on the general attitudes towards computers and mobile devices.

The scales are found in Appendix A, section 9.2. If measured and connected to the other data via the participant ID (CD1), relevant individual differences can be statistically controlled for in order to provide greater statistical power for evaluating the effectiveness of the apps.



4.7 Additional Questions

MIRROR apps all perform essentially the same function – supporting learning by reflection – and can be similarly assessed and compared, but some apps have specific functions requiring very specific questions. App-specific questions for summative evaluation capture the consequences of formative issues (e.g., design, information visualization) for usage, learning (including reflection), behaviour, and organisation/community.

Individual extensions of the research methodology must be and are possible. For example, app developers might deem it necessary to assess for specific side effects or additional (positive) effects of the app not covered within this summative evaluation framework. If this is the case, these questions can be added – and if relevant for summative evaluation – will be reported in the summative evaluation Deliverable 10.3 and/or 1.7.



5 Evaluation Procedure

We now describe our evaluation procedure – how the Evaluation Toolbox is used by the developers and testbeds to evaluate the success of the MIRROR apps.

5.1 Rationale

We evaluate the apps after each significant implementation in the testbeds during the next two years. By "significant implementation" we mean cases in which there is contact with the actual target audience of the apps, the apps are actually used, and data is produced. While a longer test of the apps is planned for the last project year, this procedure allows us to assess improvement in the higher-order effects of the apps on usage, reflection/learning, work behaviour, and organisation/community early on. This allows us to evaluation the MIRROR project as a whole, in addition to evaluation the individual apps, and also gives us a good basis of comparison within the specific versions of the same app and between different apps over time.

We have designed the procedure to allow us to:

ensure a transparent evaluation process,

All project partners are able to see during the next two years which apps are developed, where they are implemented, and how well they achieve their purpose.

assist the developers and testbeds during the app testing,

All testbeds can see shortly after the implementation whether reflective learning was supported and how well, giving them leverage for internal communication (convincing key players in their organisation of the value of MIRROR apps and gaining access to additional participants proves to be an important aspect of the project for some testbeds). Voluntary data sharing ensures that other developers can learn as well.

provide developers and testbeds a high degree of freedom,

This evaluation approach can be used by developers and testbeds on their own, as long as the implementation is documented as described below.

allow for a strong summative evaluation at the end of the project.

This approach allows for both a within-app comparison over time and a comparison between different apps, which will be reported in D1.7/D10.3.

This procedure allows us to perform a summative evaluation with a data collection period of two years in a standardized way. The core questions provide a standardized basis of comparison without being a "straightjacket", as individual additions are possible.

The evaluation procedure was developed with input by and discussion with all project partners of MIRROR and agreed upon at the fourth General Assembly of MIRROR in Trondheim, Norway. We have already trialled this procedure with some developers and testbeds during the first months of 2012, well before the deliverable due date of June 2012.



5.2 Evaluation Procedure Steps

When an app is implemented in a testbed during the final two years of the MIRROR project (when there is contact with the actual target audience of the apps, the apps are actually used, and data is produced), the developers perform the following steps:

1. Negotiation with the testbeds regarding time and scope

Developers and testbeds negotiate the amount and kind of participants, the time frame, the data gathering, etc.

2. Clarification of expected effects and how to assess them

Developers refer to the CSRL model (see deliverable 1.4) and select the relevant questions from the App-specific Reflection Questions (see 4.3.1). They discuss with the testbeds the relevant KPIs to be assessed and ensure that the effects on this level can be tracked during the implementation phase. If KPIs are collected routinely (e.g., employee survey), they coordinate with the testbeds when it is best to introduce the App. The core questions (see Chapter 4) are assessed in any case.

These decisions are documented by the developers in the **App Evaluation Sheet** (see 5.3.1) and in the **Evaluation App Implementation List** (see 5.3.2). The former proves a detailed description of a single app implementation suitable for reviewing after two years, while the latter provides a tabular overview of all apps and implementations to keep in mind what was done during the two years.

3. Pre-Implementation Measures

Prior to the first implementation, a unique **Participant Identifier** (see 4.1) is assigned, and **Demographics** (also see section 4.1) and **Participant Characteristics** (see 4.6) questionnaires are completed by the participants. The **Short Reflection Scale** (see 4.3.2) is completed before every implementation.

4. Implementation of the App

The app is then implemented as intended by the developers and testbeds.

5. Post-Implementation Measures

The evaluation **Core Questions** (see Chapter4) (including the Short Reflection Scale, to test for changes) are assessed after the implementation, covering the effects of the app implementation on reaction, learning, behaviour, and results levels.

6. Documentation

Data collected within this framework are documented in the **Evaluation Data File Template** (see 5.3.3) by the developers and sent to KMRC. Data may also be shared bilaterally between the developers if agreed. Due to privacy reasons, testbeds will only receive aggregated data (identical to the User Studies specified in Deliverable 1.1).



5.3 Documents

5.3.1 App Evaluation Sheet

The **App Evaluation Sheet** contains all the necessary information regarding the specific implementation of a specific app.

It contains information about:

- responsible work package
- testbed implemented
- app version (number)
- time frame in which it is implemented
- times the app should be used
- · number of participants
- type of participants (e.g., nurses, therapists, etc.)
- whether data can be shared among developers

Furthermore, it contains sections for

- a short description of the app (or a link to a more detailed document),
- research questions (mainly: Which effects should the app have?),
- tools used (link to additional questionnaires implemented), and
- evaluation procedure (what was done when).

The sheet is made available to everyone prior to the introduction of the app into the testbed.

5.3.2 Evaluation App Implementation List

The Evaluation App Implementation List is a single spreadsheet that shows which app is implemented in which testbed and by whom. It is continuously updated and available to all project partners, ensuring transparency of the process. It gives a quick, tabular overview and will be used to keep a current history of the many implementations of apps within the next two years.

It contains columns for:

- work package(s) developing the app
- testbed in which it is implemented
- app name and version
- consecutive number (if the same app is tested multiple times in the same testbed)
- start date of implementation
- end date of implementation
- kind of participants
- number of participants
- type of testing (e.g., what was the goal)
- link to app evaluation sheet

Due to the advanced stage of the project, we have already started to document the app implementations in the testbeds.



5.3.3 Evaluation Data File Template

Data sharing of the core questions is done with the **evaluation data file template**. It is copied for each app implementation and sent to KMRC after each implementation of the apps, providing a standardized way to collect and compare the data. Additional questions from the toolbox or from the developers are also to be made available for the summative evaluation, if possible in the same evaluation data file.



6 Strategy for Overall Integration

Over the course of the next two years, data from the app implementations in the testbeds will be collected as used for the summative evaluation. The Core Questions will ensure comparability within and between the apps and allow us to assess the contribution of the project to supporting learning by reflection at work over this two-year timeframe. By collecting relevant data after each app implementation, we will be in a better position to carry out an intensive final summative evaluation of a longer term implementation during the last year of the project. This procedure will also allow us to relate the data of this final trial to data gathered over time to show the changes in reflective learning more clearly.

6.1 Documentation in D1.7 and D10.3

D10.3 will present the main findings from the summative evaluation. We will summarize our evaluation of the usage and effectiveness the each of the apps and their potential for supporting reflective learning in the different testbed.

D1.7 will revisit our evaluation framework and process, as well as summarize and integrate the results from D10.3.

However, centrepiece of both of these summative evaluation deliverables will be the change in reactions to apps over time, comparing within and between the apps, based on responses to the Core Questions. This information will be presented as a table that allows for a simple visual comparison, all in one place.

6.1.1 Intra- and Inter-App Comparisons

The Core Questions Table (Table 4) gives a succinct overview of the key features of the apps. Comparisons between apps can be made by comparing columns, and comparison within an app over time may be made by comparing single entries within a cell. To assist the developers and testbeds during the next two years (see section 5.1), it will be made available to the project partners and continuously updated.

Note that the App-specific Reflection Questions provide information on how well the each app performs its designated functions (see section 4.3.1). What exactly these functions are is selected by the app developers (see section 5.2). Thus, to allow a quick comparison, these values will be standardized before being entered into the table. Indices will show whether the apps support the same function or different functions (see Table 4 Notes). Results from the same app in different testbeds will also be indicated by separate indices.

Core Question Table

	ind	lividual refle	ction	collabo	rative reflection	on	organisational reflection					
App (WPs)	Task Detecti on App (WP4)	Mood Map (WP3/4)	Virtual Care Home (WP7)	TalkReflect (WP6)	Micro- blogging with Yammer (WP5)		ProcessThinking (WP8)					
Reaction												
Number of	//	// ^{Inf}	//	//	//	//	//	//	//			
times used		// ^{BT}										
Total time (minutes) used	//	// ^{Inf} //	//	//	//	//	//	//	//			
Average time	//	// ^{Inf}	//	//	//	//	//	//	//			
(minutes) used		// ^{BT}										
Number of	//	// ^{Inf}	//	//	//	//	//	//	//			



times each key		// ^{BT}			ı							
times each key function of App		//				•		•	•			
is used												
	, , a	, , Int, b	, , a	, , c	, , a		, , d					
[App Specific	// ^a	//	/ ^a	/ ^c	/ ^a		/ ^d					
Reflection		// ^{BT, b}										
Questions]		1.7										
[Short	//	//Inf	//	//	//	//	//	//	//			
Reflection		// ^{BT}				-						
Scale]												
Learning												
I made a	//	// ^{Inf}	//	//	//	//	//	//	//			
conscious		/ ^{BT}										
decision about												
how to behave												
in the future.												
I gained a	//	// ^{Inf}	//	//	//	//	//	//	//			
deeper		// ^{BT}										
understanding												
of my work life.												
Behaviour												
The app helped	//	//Inf	//	//	//	//	//	//	//			
me improve my		// ^{BT}										
[work												
performance].												
Results												
[KPIs]	//	// ^{Int}	//	//	//	//	//	//	//			
- -		// ^{BT}										

^a: [The app] helped me to collect information relevant to reconstructing experiences from work., ^b: [The app] helped me to reflect on experiences from work., ^c: [The app] helped me to collect data on behaviour before the reflection session., ^d: [The app] helped me to collect data on behaviour after the reflection session., ...

.../.../ => 1stversion / 2ndversion / 3rdversion / etc.

Table 4: Example of the Core Question Table Allowing Comparison of Core Questions

Taken as a whole, this table will provide us with a basis for comparison and show the overall success of our project. In addition to this overall perspective, we also consider each app in detail when we look at app-specific reflection support.

6.1.2 App-Specific Reflection Support

By requesting detailed information about the apps themselves, including data about which steps of the reflection process they support (see 5.3.1), and by using the summative evaluation results of each app implementation, we will also highlight the specific strengths and weaknesses of the each of the apps.

In D10.3 we will shortly describe each app with regard to how well it supports the process of reflection (from the CSRL model) it is intended to address, and how the participants in the testbed profit from using it. Informed by data that is collected in addition to the Core Questions, we will point to the specific advantages and challenges of the app in specific testbeds.

In D1.7 we focus more on the general effect of supporting reflection with the MIRROR apps and take a meta-perspective on the summative evaluation and the possibilities and limitations of supporting learning by reflection with digital apps.



7 Outlook

This summative evaluation deliverable allows the specification of clear success indicators and gives information on objective, reliable, and valid measurement of these indicators on different levels.

We have modelled the relevant aspects of learning by reflection at work in great detail with the *i** approach (Chapter 2) and derived summative evaluation criteria to construct our summative evaluation framework (Chapter 3). While the *i** provided a bottom-up view, our modification of the Kirkpatrick model allowed us to see the bigger picture.

Based on these criteria that determine the success of the MIRROR Apps, we developed tools to assess these measures. The Toolbox (the Core Questions in Chapter 4 and additional questions in Appendix A, section 9.3) allows us to assess relevant indicators for reflective learning and their impact for individuals and team learners, as well as the organisation as a whole (business impact). It also ensures comparability between the different apps and different testbeds. The evaluation procedure is described and agreed upon within the project (Chapter 5), and the documentation process allows us to conduct the overall integration of the results (Chapter 6).

In the next two years the developers and testbeds will work together to test the apps in the testbeds and support learning by reflection. We will closely follow the results assessed during the next two years and use it as basis to determine the success of MIRROR and its apps to support learning by reflection.



8 References

- Argyris, C., &Schön, D. (1978). *Organizational learning: A theory of action perspective*. Reading, Massachusetts: Addison-Wesley.
- Boud D., Keogh, R., & Walker, D. (1985). *Reflection: Turning experience into learning.* London, New York: Routledge.
- Budner, S. (1962). Intolerance of ambiguity as a personality variable. Journal of Personality, 30(1), 29-50.
- Cacioppo, J. T., Petty, R. E., & Kao, C. F. (1984). The efficient assessment of need for cognition. Journal of Personality Assessment, 48, 306–307.
- Daudelin, M. W. (1996). Learning from experience through reflection. *Organizational Dynamics*, *24*(3), 36–48.
- Dewey, J. (1933). How we think (Revised Edition). New York: D.C. Heath.
- Höfling, V., Ströhle, G., Michalak, J., & Heidenreich, T. (2011). A Short Version of the Kentucky Inventory of Mindfulness Skills. Journal of Clinical Psychology, 67(6), 639–645.
- Järvinen, A., & Poikela, E. (2001). Modelling reflective and contextual learning at work. Journal of Workplace Learning, 13(7/8), 282–290.
- Kirkpatrick, D. L., & Kirkpatrick, J. D. (2006). *Evaluating trainings programs: The four levels.* 3rd Edition. Berrett-Koehler Publishers, Inc.
- Kolb. D. A., & Fry, R. (1975): Toward an applied theory of experiential learning. In C. Cooper (Ed.), *Theories of group process*. London: John Wiley.
- Moon, J. A. (1999). *Reflection in learning and professional development: Theory and practice*. London: Routledge.
- Schwarzer, R., & Jerusalem, M. (1995). Generalized Self-Efficacy scale. In J. Weinman, S. Wright, & M. Johnston, Measures in health psychology: A user's portfolio. Causal and control beliefs (pp. 35-37). Windsor, UK: NFER-NELSON.
- The Net Promoter Score and System (2012). Retrieved June 25, 2012 from http://www.satmetrix.com/net-promoter/net-promoter-score-and-model/
- Yu E., & Mylopoulos J. M. (1994). Understanding "why" in software process modelling, analysis and design. In *Proceedings of the 16th International Conference on Software Engineering* (pp. 159-168). IEEE Computer Society Press.



9 Appendix A: Evaluation Toolbox

9.1 Core Questions

9.1.1 About Yourself

[CD1] Your Participant Code

Please write down your Participant Code. Your code consists of:

- 1. The first letter of your place of birth
- 2. Your own day of birth (two digits)
- 3. The first letter of your father's first name
- 4. The first letter of your mother's first name

Example: A person born in London on the 7^{th} of July, with parents named Jake and Sue and born would enter: "L" (for London) in the 1^{st} blank, and "07" (for a birthday on the 7^{th}) in the 2^{nd} blank, "J" (for Jack) in the 3^{rd} blank, and "S" (for Sue) in the 4^{th} blank. So, this person's code would be: L 07 J S.

If you don't know any of these, use the letter below the correct box shown below.

	1 st letter of your place of birth	ur place Your day or your father's			
Your Code					
If unknown, use:	X	00	Υ	Z	

[CD2] Team-ID (if applicable):	
[CD3] Current Date:	
[CD4] Gender: □ male□ female	
[CD5] Age Group: □≤19 □20-29 □30-39 □40-49 □50-59 □≥60	
Job [CD6] Scope: □ Full-time □ Part-time	
[CD7] Department:	
[CD8] Position:	
[CD9] Years in current position: years	
[CD10] Years in current team (if applicable): years	
[CD11] Years in similar positions (e.g., at another company):	years
9.1.2 Usage of the App	
[CU1] How many times have you used [the App]? Approximately times	mes.



Page 57

[CU2] H	ow long hav	ve you u	used [the	App] in	total? A	pprox	imately ₋	m	inutes.
[CU3] H minutes.	J	d you u	use [the /	App] on	averag	e ead	ch time?	Approximate	ely
[CU4] H	low many times.	times of	did your	usage	of the	арр	include	[functions]?	Approximately

9.1.3 App-Specific Questions

Note to evaluators: Use only the relevant questions!

ID	Question	yly ree	ree	y ree	la	y agree		strongly agree
		strongly disagree	disagree	slightly disagree	neutral	slightly	agree	stronç
CA1	[The app] helped me to collect information relevant to reconstructing experiences from work.							
CA2	[The app] helped me to reflect on experiences from work.							
CA3	[The app] helped me to collect data on behaviour before the reflection session.							
CA4	[The app] helped me to collect data on behaviour after the reflection session.							
CA5	[The app] helped me to collect information that could help me decide when to							
CA6	[The app] helped me to reconstruct a work experience.							
CA7	[The app] helped me by capturing my reflection outcomes.							
CA8	[The app] helped me by making reflection outcomes available for later use							
CA9	[The app] helped me by capturing information for evaluation of learning/reflection.							
CA10	[The app] helped me by reminding me to reflect.							
CA11	[The app] helped me by providing information relevant for the decision to reflect.							
CA12	[The app] helped me by providing accurate information about my work.							
CA13	[The app] helped me by providing information relevant for the framing of reflection.							
CA14	[The app] helped me by showing the availability of resources needed for reflecting.							
CA15	[The app] helped me to allocate or structure the resources needed for reflection.							
CA16	[The app] helped me by providing information about related experiences.							
CA17	[The app] helped me to remember and reconstruct the experience/situation.							



Page 58

CA18	[The app] helped me by providing access to data (e.g., simulations) relevant to the re-evaluation of					
CA19	experience. [The app] helped me by providing access to data					
0.1.00	relevant to the experience					
CA20	[The app] helped me by providing access to resources resulting from reflection sessions.					
CA21	[The app] guided me in capturing information					
O, (2.)	about my work experiences.					
CA22	[The app] guided me in deciding whether/when to	П			_	
2	reflect.					
CA23	[The app] guided me in finding the resources needed for reflection.					
CA24	[The app] guided me in allocating/structuring the					
	resources needed for reflection.					
CA25	[The app] helped me by supporting sharing of experiences.					
CA26	[The app] guided me in sharing experiences with others.					
CA27	[The app] guided me in reconstructing and					
	remembering the experience/situation.					
CA28	[The app] guided me in articulating the meaning of an experience.					
CA29	[The app] guided us in negotiating the meaning of					
	an experience.					
CA30	[The app] guided us in documenting different viewpoints on the experience.					
CA31	[The app] guided me in re-evaluating an		П	П		П
C 4 2 2	experience.					
CA32	[The app] guided me in reaching a resolution.					
CA33	[The app] guided me in making the reflection outcome applicable to my work.					
CA34	[The app] guided me in making the reflection		П	П		П
CASE	outcome applicable to further reflection.					
CA35	[The app] guided me in considering constraints of the reflection outcome.					
CA36	[The app] guided me in considering the option of not applying the reflection outcome.					
CA37	[The app] guided me in describing work scenarios]	
	that could lead to desired results.					
CA38	[The app] guided me in describing both "good practice" and "bad practice" work scenarios.					
CA39	[The app] provided help with collaboration.					
CA40	The app] provided relevant content for reflection.					
CA41	[The app] guided me through the reflection]	
	process.					
CA42	[The app] helped me by simulating the work process.					
CA43	[The app] helped me by providing me with virtual experience in my work domain.					

9.1.4 Reflection Scale

Use this scale prior to and after implementing the ap	se thi	this scale pr	or to and	d after imp	plementing	the a	app.
---	--------	---------------	-----------	-------------	------------	-------	------

ID	Question	strongly disagree	disagree	slightly disagree	neutral	slightly agree	agree	strongly agree
CR1	I often reflect on my work in order to improve it.							
CR2	We as a team often reflect on our work in order to improve it.							
CR3	I think it is important to try to improve [specific work task].							
CR4	I frequently reflect on [specific work task].							
CR5	Reflecting on [specific work task] helps me to improve [the task].							
CR6	In team meetings we frequently talk about how we can improve [specific work task].							
CR7	Outside of meetings, I often talk with my colleagues about [specific work task].							
CR8	It is important to me to discuss frequently with others about [specific work task].							
CR9	Conversations with colleagues help me to improve [specific work task].							
CR10	Even a few days later, I can remember the [specific work task/event] well when I reflect on it by myself or with others.							
9.1.5	Learning outcomes							
ID	Question	strongly disagree	disagree	slightly disagree	neutral	slightly agree	agree	strongly agree
CL1	I made a conscious decision about how to behave in the future.							
CL2	I gained a deeper understanding of my work life.							



9.1.6 Work Behaviour

ID	Question	strongly disagree	disagree	slightly disagree	neutral	slightly agree	agree	strongly agree
CB1	The app helped me improve my [work performance].							

9.1.7 Results Level

9.1.7.1 BT

CK1B customer issue reduction

CK2B customer satisfaction increases

CK4B Right First Time

9.1.7.2 Infoman

CK2I customer satisfaction increases

CK4al less time on task

CK4bl less time spent on searching for knowledge

CK4cl amount of reuse of PowerPoint slides

CK5I higher quality of work

CK8I smooth integration of app use in the normal work flow

9.1.7.3 NBN

CK1N reduction of complaints

CK2N customer satisfaction increases

CK4N identify duplication of processes/process error

CK6aN employee satisfaction

CK6bN well-being of employees

9.1.7.4 Regola

CK3R employee retention/turnover (number volunteers in and out, adherence to assigned task)

CK4R time on task (average call time)

CK5aR critical incident performance (% deviation from degree of severity)

CK5bR service quality (operation time)

CK7R achievement of standards and accreditation



9.1.7.5 RNHA

- **CK1aH** reduction in the number of incidents of challenging behaviour/difficult relations with residents
- **CK1bH** reduction in the severity of incidents of challenging behaviour/difficult relations with residents
- **CK2H** increase in happiness of the residents (as seen by their relatives)
- **CK3H** decreased staff turnover



9.2 Participant Characteristics

9.2.1 Need for Cognition

ID	Question	strongly disagree	disagree	neutral	agree	strongly
TN1	I would prefer complex to simple problems.					
TN2	I like to have the responsibility of handling a situation that requires a lot of thinking.					
TN3	Thinking is not my idea of fun.					
TN4	I would rather do something that requires little thought than something that is sure to challenge my thinking abilities.					
TN5	I try to anticipate and avoid situations where there is likely chance I will have to think in depth about something.					
TN6	I find satisfaction in deliberating hard and for long hours.					
TN7	I only think as hard as I have to.					
TN8	I prefer to think about small, daily projects to long-term ones.					
TN9	I like tasks that require little thought once I've learned them.					
TN10	The idea of relying on thought to make my way to the top appeals to me.					
TN11	I really enjoy a task that involves coming up with new solutions to problems.					
TN12	Learning new ways to think doesn't excite me very much.					
TN13	I prefer my life to be filled with puzzles that I must solve.					
TN14	The notion of thinking abstractly is appealing to me.					
TN15	I would prefer a task that is intellectual, difficult, and important to one that is somewhat important but does not require much thought.					
TN16	I feel relief rather than satisfaction after completing a task that required a lot of mental effort.					



Page 63

TN17	It's enough for me that something gets the job done; I don't care how or why it works.			
TN18	I usually end up deliberating about issues even when they do not affect me personally.			

9.2.2 Ambiguity Tolerance

ID	Question	Strongly disagree	Moderately disagree	Slightly disagree	Neither agree nor disagree	Slightly agree	Moderately agree	Strongly agree
TA1	An expert who doesn't come up with a definite answer probably doesn't know much							
TA2	I would like to live in a foreign country for a while							
TA3	There is really no such thing as a problem that can't be solved.							
TA4	People who fit their lives to a schedule probably miss most of the joy of living							
TA5	A good job is one where what is to be done and how it is to be done are always clear							
TA6	It is more fun to tackle a complicated problem than to solve a simple one							
TA7	In the long run it is possible to get more done by tackling small, simple problems rather than large and complicated ones							
TA8	Often the most interesting and stimulating people are those who don't mind being different and original							
TA9	What we are used to is always preferable to what is unfamiliar.							
TA10	People who insist upon a yes or no answer just don't know how complicated things really are.							
TA11	A person who leads an even, regular life in which few surprises or unexpected happenings							



Page 64

	arise really has a lot to be grateful for.				
TA12	Many of our most important decisions are based upon insufficient information.				
TA13	I like parties where I know most of the people more than ones where all or most of the people are complete strangers.				
TA14	Teachers and supervisors who hand out vague assignments give one a chance to show initiative and originality				
TA15	The sooner we all acquire similar values and ideals the better				
TA16	A good teacher is one who makes you wonder about your way of looking at things				

9.2.3 Mindfulness

ID	Question	Never or very rarely true	Rarely true	Sometimes true	Often true	Very often or always true
TM1	I'm good at finding the words to describe my feelings.					
TM2	I criticize myself for having irrational or inappropriate emotions.					
ТМЗ	When I'm doing something, I'm only focused on what I'm doing, nothing else.					
TM4	When I'm walking, I deliberately notice the sensations of my body moving.					
TM5	When I take a shower or bath, I stay alert to the sensations of water on my body.					
TM6	It's hard for me to find the words to describe what I'm thinking.					
TM7	I believe some of my thoughts are abnormal or bad and I shouldn't think that way.					
TM8	I have trouble thinking of the right words to express					



Page 65

	how I feel about things.			
TM9	When I do things, I get totally wrapped up in them and don't think about anything else.			
TM10	I make judgments about whether my thoughts are good or bad.			
TM11	I pay attention to sensations, such as the wind in my hair or sun on my face			
TM12	When I have a sensation in my body, it's difficult for me to describe it because I can't find the right words.			
TM13	I pay attention to sounds, such as clocks ticking, birds chirping, or cars passing.			
TM14	Even when I'm feeling terribly upset, I can find a way to put it into words.			
TM15	I tell myself that I shouldn't be thinking the way I'm thinking.			
TM16	I notice the smells and aromas of things.			
TM17	I tend to do several things at once rather than focusing on one thing at a time.			
TM18	I think some of my emotions are bad or inappropriate and I shouldn't feel them.			
TM19	I notice visual elements in art or nature, such as colours, shapes, textures, or patterns of light and shadow.			
TM20	I get completely absorbed in what I'm doing, so that all my attention is focused on it.			



9.2.4 Self-Efficacy

ID	Question	Not at all true	Hardly true	Moderately true	Exactly true
TS1	I can always manage to solve difficult problems if I try hard enough.				
TS2	If someone opposes me, I can find the means and ways to get what I want.				
TS3	It is easy for me to stick to my aims and accomplish my goals.				
TS4	I am confident that I could deal efficiently with unexpected events.				
TS5	Thanks to my resourcefulness, I know how to handle unforeseen situations.				
TS6	I can solve most problems if I invest the necessary effort.				
TS7	I can remain calm when facing difficulties because I can rely on my coping abilities.				
TS8	When I am confronted with a problem, I can usually find several solutions.				
TS9	If I am in trouble, I can usually think of a solution.				
TS10	I can usually handle whatever comes my way.				

9.2.5 IT Attitudes

In this questionnaire, we would like to ask you about your attitudes to and usage of computer technology. Please answer the following questions, as far as they apply to you.

ID	Please indicate your agreement with the following statements	strongly disagree	disagree	neutral	agree	strongly agree
TI01	Whenever I use something that is computerized, I am afraid I will break it.					
TI02	At home, I often use a computer.					



Page 67

T103	Usually, it was others who convinced me to use a computer.			
TI04	I would like to keep up with technological advances.			
TI05	I find it hard to understand computers.			
TI06	Using a computer is too time-consuming.			
TI07	I like to play computer games.			
T108	I have had bad experiences with computers.			
TI09	I do not feel I have control over what I do when I use a computer.			
TI10	I think computers and other technological advances have helped to improve our lives.			
TI11	I really enjoy learning new personal computer software.			
TI12	I give more computer advice to other people than I receive.			
TI13	There are many things that I can do more easily and quickly using a computer.			
TI14	I like using cell phones/smart phones.			
TI15	I feel confident using a cell phone/Smartphone.			

Thank you for your time and assistance



9.3 Further Questions beyond the Core Questions

9.3.1 Reaction Level

9.3.	1.1	Usage
------	-----	-------

9.3.1.1	Usaye								
ID	Question	strongly disagree	disagree	slightly disagree	neutral	slightly agree	agree	strongly agree	
USE01	I did not have the time to use [the app]								
USE02	I did not have the physical space (e.g., necessary privacy) to use [the app]								
USE03	I did not see an advantage in using [the app]								
USE04	I was not motivated to use [the app]								
USE05	I could find out how [the app] worked myself								
USE06	I need more formal training with [the app]								
USE07 What were barriers to using [the app]?									
	of time needed to understand the App: Approximate Usefulness/Satisfaction	ely	n	ninute	S				
ID	Question	strongly disagree	disagree	slightly disagree	neutral	slightly agree	agree	strongly agree	
SAT01	I am satisfied with the App								
SAT02	I think the app is useful for professional training								
SAT03	I think the app can be used to complement professional training								

Page 69

9.3.1.3 Inclination Long-Term Usage

ID	Question	strongly disagree	disagree	slightly disagree	neutral	slightly agree	agree	strongly agree
LT01	I see the long-term advantage of using the app in my work-life.							
LT02	I would like to use the app continuously as part of my work-life.							
LT03	It is practical for me to continue using the app in my work-life.							

9.3.2 Learning Level

9.3.2.1 Long Reflection Scale

ID	Question	strongly disagree	disagree	slightly disagree	neutral	slightly agree	agree	strongly agree
LRS01	I frequently reflect on my own practice at work.							
LRS02	I would like to have more time to reflect on my work while doing my job.							
LRS03	Reflecting on my work is part of my job.							
LRS04	I often reflect on my own activity to find out if I could have improved on what I did.							
LRS05	I often question the way I am doing something, and try to think of a better way of doing it.							
LRS06	If I feel bad about some experience at work, I take time to reflect on it.							
LRS07	In the case of some unexpected success, I take time for reflection							
LRS08	I like reflecting on my own work.							
LRS09	I do not think it is beneficial for me to reflect on my work.							
LRS10	I often feel the need to reflect on what I have been doing.							

Page 70

LRS11	Typically, I try to get things done without reflecting too much about better ways of doing them.				
LRS12	I often come up with ideas on how things could be organised differently here.				
LRS13	I often reflect on the way we work in this organisation.				
LRS14	I am not concerned with improving the organisation's work practice.				
LRS15	We as a team often come up with ideas on how things could be organised differently here.				
LRS16	We as a team sometimes make suggestions on how to improve our work practice.				
LRS17	I often talk about my work-related experiences with my colleagues.				
LRS18	Talking about my work-related experiences with my colleagues is in most cases helpful for me.				
LRS19	If I think I have done my work badly, I discuss this with my superior.				
LRS20	When I have changed my work practice to adapt to a new situation, I share this with my colleagues.				
LRS21	If I think I have done my work badly, I discuss this with my colleagues.				
LRS22	I often talk about work experiences with my friends or family.				
LRS23	I do not share my work experiences with my colleagues.				
LRS24	Talking about work-related experiences with my friends or family is in most cases helpful for me.				
LRS25	I often learn by observing others at work who do things differently from how I do them.				
LRS26	I often compare myself with my colleagues				



Page 71

	regarding their work practice.				
LRS27	Comparing myself with others makes me feel uncomfortable.				
LRS28	I can also benefit from discussions about work-related experiences of others.				
LRS29	As a team, we frequently reflect on our team's work practice.				
LRS30	Reflecting on our team work is part of our job.				
LRS31	We as a team often discuss our activities to find out if we could have improved on what we did.				
LRS32	We as team often question the way in which we do something and try to think of a better way of doing it.				
LRS33	If we experience discord in our team, we take time to reflect on it.				
LRS34	If we are successful as a team, we take the time to analyse how we achieved this.				
LRS35	As a team, we work out what we can learn from previous activities.				
LRS36	While working as a team, we have time to stop to reflect on our work practice.				
LRS37	If things don't work out as they should, we take the time as a team to find the possible cause of the problems.				
LRS38	If things don't work as planned, we consider what we can do about it.				
LRS39	We as a team benefit from reflecting on our work.				
LRS40	When we discuss our team work, we often come up with new ways of doing our job.				

Page 72

9.3.2.2 Re-Evaluation of Experience

Please take other aspects like Think-aloud protocols or audio/video-recordings of collaborative reflection sessions into account.

ID	Question	strongly disagree	disagree	slightly disagree	neutral	slightly agree	agree	strongly agree
REE01	I frequently gain a new perspective on my past behaviour.							
REE02	The app helped me to gain a new perspective on my past behaviour.							

9.3.2.3 General App Effects

3.3.Z.3	Deneral App Lifects							
ID	Question	strongly disagree	disagree	slightly disagree	neutral	slightly agree	agree	strongly agree
[The app] helped me to								
GAE01	come to a decision whether we should reflect.							
GAE02	find situations on which we should reflect.							
GAE03	understand my emotions better.							
GAE04	deal better with my emotions.							
GAE05	achieve a better understanding of the situation.							
GAE06	met my reflection objectives.							
GAE07	focus on the relevant aspects of reflection.							
GAE08	find a more satisfying reflection outcome.							
[The app] made								
GAE09	the reflection sessions more effective.							
GAE10	the reflection sessions more efficient.							
GAE11	the reflection sessions more successful.							



Page 73

GAE12	the reflection sessions more achievable in work constraints.										
GAE13	me more confident in the reflection outcome.										
GAE14	the reflection outcome more accessible for others.										
GAE15	the reflection outcome more comprehensible for others.										
GAE16	the reflection outcome more applicable for others.										
Data gathering with [the app] was											
GAE17	accurate										
GAE18	effortless										
GAE19	relevant										
GAE20	timely										
Collabora	atively, [the app}										
GAE21	made the reflection session more efficient.										
GAE22	increased our ability to reflect.										
GAE23	motivated us to reflect.										
GAE24	helped us to meet our reflection objective.										
GAE25	helped us to achieve our reflection outcome.										
GAE26	helped us to do the reflection session within the work constraints.										
GAE27	helped me maintain my privacy.										
GAE28	made my individual reflection outcomes more valued.										
GAE29	made my individual reflection outcomes more										

Page 74

	relevant.				
GAE30	allowed me to be heard.				
GAE31	made the reflection outcomes more socially acceptable.				
GAE32	helped us to reconstruct our experiences more comprehensively.				
GAE33	helped us to agree on our reconstructed experience.				
GAE34	helped us to deal with our emotions.				
GAE35	helped us to make sense out of the experience.				
GAE36	allowed us to re-evaluate the experience with a strong rationale.				
GAE37	helped us to focus on the more important aspects of the experience.				
GAE38	helped us to select an appropriate frame for re- evaluating the experience.				
GAE39	helped us to come to a valuable outcome.				
GAE40	helped us to select appropriate criteria for re- evaluating the experience.				
GAE41	helped us to critique the experience in a way that was fair for all participants.				
GAE42	helped us to get to a solution that was fair for all participants.				
GAE43	helped us to get to a useful solution.				
GAE44	helped us to get to a solution we were satisfied with.				
GAE45	helped us to get a feasible solution.				
Other rele	evant questions				
GAE46	With [the app], I was more motivated to reflect				



Page 75

	on work in general.							
9.3.2.4	Behavioural Intentions	·	•	·	•			·
ID	Question	strongly disagree	disagree	slightly disagree	neutral	slightly agree	agree	strongly agree
BI01	I made a conscious decision to change my behaviour							
BI02	The decision to change my behaviour was well-founded							
BI03	I made a conscious decision to continue with my behaviour							
BI04	The decision to continue with my behaviour was well-founded							
BI05	I am more confident that the decision I made is the right one							
BI06	I am motivated to actually change my behaviour							
BI07	I am confident that I can actually change my behaviour							
BI08	I made a conscious decision about how to behave in the future.							
9.3.2.5	Knowledge/Skills							
ID	Question	strongly disagree	disagree	slightly disagree	neutral	slightly agree	agree	strongly agree
KS01	I gained a deeper understanding of my work life							
KS02	I improved my understanding in the area that I wanted to improve in							
KS03	I did improve my work-related skills							
KS04	I did improve my work-related skills in the area I wanted to improve in							



Page 76

KS05	The app stimulated knowledge exchange				

9.3.3 Behaviour Level

These questions can be answered by the participant, but also by a manager about a participant.

9.3.3.1 Work

ID	Question	strongly disagree	disagree	slightly disagree	neutral	slightly agree	agree	strongly agree
WK01	I used my learning on the job							
WK02	The app helped me to improve my work experience.							
WK03	The app increased my work satisfaction.							
WK04	The app helped me to improve my [work performance]							
WK05	I kept up my change of behaviour							
WK06	The app helped to improve my performance							
WK07	The app helped to improve team performance							
WK08	The app helped me save time							
WK09	The app helped me to focus on my work tasks							
WK10	The app helped me to satisfy my customers faster							
WK11	The app helped me to tackle difficult work situations							

9.3.3.2 Other Work-Related Criteria

ID	Question	trongly lisagree	isagree	ightly sagree	eutral	ightly gree	gree	rongly gree	
		stro	dis	Slic dis dis	neı	slig	agr		ı

Page 77

WK12	Using the app made me more confident that I can succeed in my work-tasks.				
WK13	Using the app supported me to master my work-tasks.				
WK14	The app improved my work satisfaction.				

9.3.4 Results Level

9.3.4.1 KPIs

1. Decreased number of negative events

e.g., customer issue reduction, reduction of complaints, reduction in the number/severity of incidents of conflicts with customers

2. Increased client satisfaction

e.g., customer satisfaction increases (self-reports and/or reports from others, e.g., relatives when it comes to residents in care homes)

3. Decreased staff turnover

e.g., employee retention/turnover numbers (also works for volunteers, e.g., number volunteers in and out, adherence to assigned task)

4. Decreased time on task

e.g., less time on (specific sub-)task (e.g., searching for knowledge, reuse of PowerPoint slides, average call time), also related measures that strongly influence time on task, e.g., doing it Right First Time or identification of duplication of processes/process errors

5. Increased quality of work

e.g., critical incident performance (% deviation from degree of severity), service quality (operation time)

6. Increased employee satisfaction

e.g., employee satisfaction and well-being of employees

7. Matching External Criteria (standards/accreditation)

e.g., achievement of standards and accreditation

8. Compatibility with existing procedures

e.g., smooth integration of app use in the normal work flow



9.3.4.2 Community

9.3.4.2.1 MIRROR Adoption

Requests from other (similar) organizations to use the apps

9.3.4.2.2 Level of Participation

- Percentage of people willing to use the MIRROR apps (of those we offered to use the app)
- Requests from within the testbeds to use the apps

We can also determine the success of the MIRROR Apps and its chances to affect other members of the organisation and the potential uptake of the community itself by using a loyalty metric like the Net Promoter metric (see http://www.satmetrix.com/net-promoter/net-promoter-score-and-model/). By asking "How likely is it that you would recommend [the App] to a friend or colleague?" on a 0-to-10 point rating scale we can determine the number of promoters (score 9-10), passives (score 7-8), and detractors (score 0-6). The NPS is the percentage of promoters minus the percentage of detractors and allows an assessment of the possible success of the Apps.



10 Appendix B: The Kirkpatrick Evaluation Model

The original Kirkpatrick model of evaluation (Kirkpatrick & Kirkpatrick, 2006) was developed to assess formal training programs. As such, it includes four levels of evaluation that deal specifically with effects of a particular training session.

Level 1: Reaction

How favorably do participants react to the training? What did participants think and feel about all aspects of the training?

Assessed using "smile sheets" or "happy sheets" used to rate satisfaction

Level 2: Learning

To what degree do participants acquire the intended knowledge and skills based on the training event? Is there a change in attitude?

Assessed in training session with a knowledge test or demonstration.

Level 3: Behavior

To what degree does participants' job behavior change when they are back on the job, based on what they learned in the training?

Assessed via observation of participant over the 3 to 6 months following the training.

Level 4: Results

To what degree do targeted outcomes occur as a result of the training event and subsequent reinforcement?

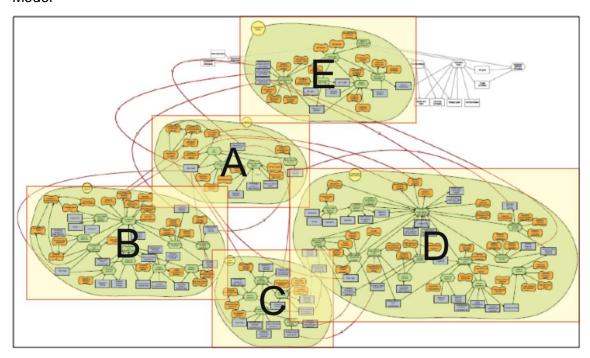
Can be assessed as monetary or performance-based outcomes.

We have adapted this model to serve the specific purposes of MIRROR. For a description of these adaptations, see section 3.2.



11 Appendix C: i* Model Graphics

Model



A: Worker

B: Individual Reflector

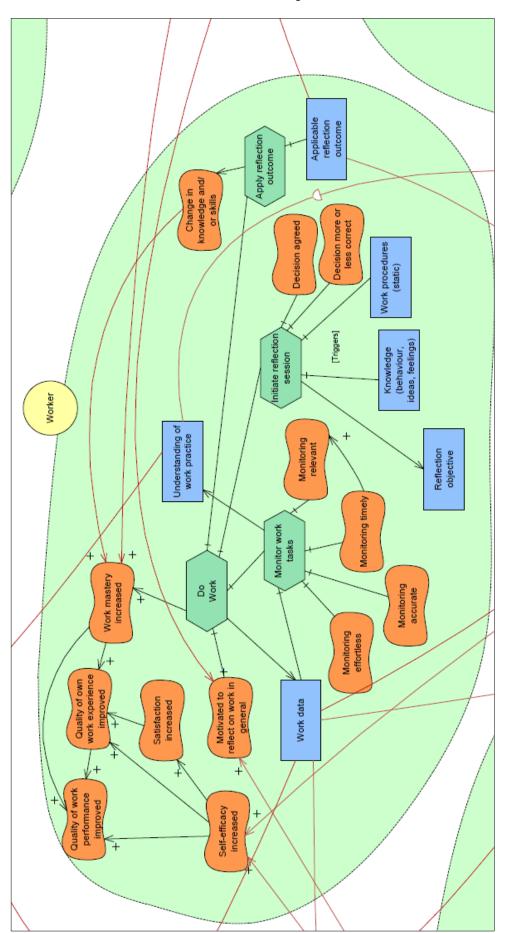
C: Individual Team Reflector

D: Collaborative Team Reflector

E: Organisational Reflector

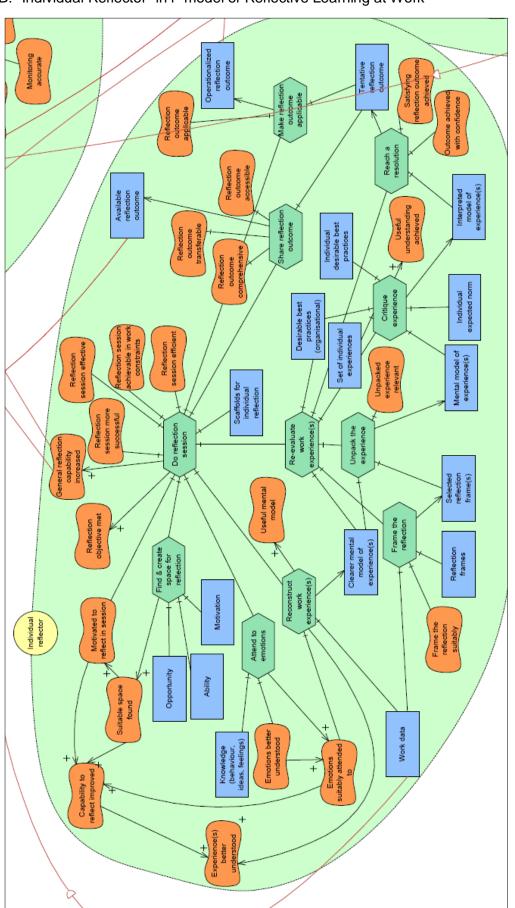


A: "Worker" in i*Model of Reflective Learning at Work



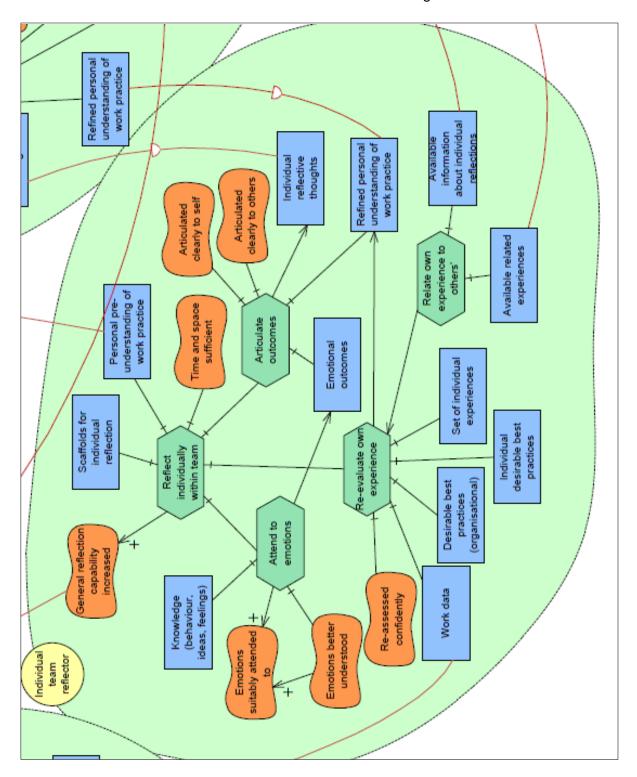


B: "Individual Reflector" in i* model of Reflective Learning at Work



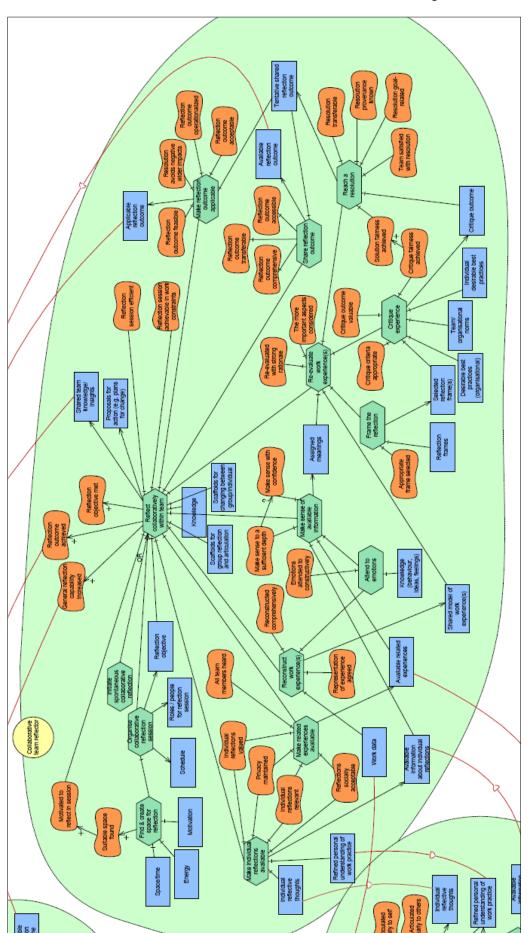


C: "Individual Team Reflector" in i* model of Reflective Learning at Work





D: "Collaborative Team Reflector" in i* model of Reflective Learning at Work





E: "Organisational Reflector" in i* model of Reflective Learning at Work

