Applying Recommender Systems in Collaboration Environments

Ernesto Damiani

Etisalat British Telecom Innovation Center
Abu Dhabi, UAE

Valerio Bellandi, Paolo Ceravolo, Fulvio Frati

Computer Science Department
Università degli Studi di Milano, Italy

Ronald Maier, Isabella Seeber, Gabriela Waldhart

Department of Information Systems
University of Innsbruck, Austria

Abstract

Team-based organizational structures are now widely adopted for activities such as product development, customer support and process-improvement initiatives due to their low overhead and minimal management. However, team collaboration often faces pitfalls like information overload or misunderstandings due to goal misalignment. In this paper we put forward the idea that computer-supported collaboration environments can have a positive impact on team collaboration, increasing team members awareness, focusing attention on task execution, and foster the frequency of interaction between team members. We study the impact of recommender systems on team processes in computer-supported collaboration environments, describing the results of two experiments that show how recommendations impact interactions in teams. Teams using recommendations spent less effort on information handling, engaged more in communication and shared their work more equally than teams without recommendations.

Keywords: Collaboration environment, experiment, interaction, recommender system, team collaboration, team process

---

1. Introduction

Organizations are characterized by an increasing share of knowledge work [1] and the corresponding transformation of organizational structures from work organized around individuals to team-based work structures. One central cornerstone of teams is that any decision, solution, or new idea represents a product that has emerged from the teams interactions and is not attributable to an individual alone [2]. The team members fruitful collaboration represents the basis for driving innovation and organizational success [3]. Many teams face obstacles in their collaboration, such as problems when and how to communicate, leading to poor communication; they might be unaware of the other team members knowledge hindering the synthesis of diverse knowledge that could be brought to solve problems or perform the task at hand [4]. They might experience information overload, which could lead to the breakdown of communication and higher time requirements for information handling [5].

Past research underlined the positive impact of computer-supported collaboration in that it can increase team members awareness [6] [7], orient their attention towards task execution [8] and increase their frequency of interaction [9]. When teams find ways to improve their communication, they can reduce time-consuming coordination activities [10] in favour of task-related information exchange and work on the tasks for improved team performance [11]. According to Feedback Intervention Theory, automated feedback has a guidance effect on the team members attention [12]. Feedback not only affects the behavior of individuals, but also impacts on the behavior of teams and consequently team performance [13]. General purpose recommender systems [14] are increasingly appreciated in collaborative settings as they aim to support information processing among team members [15] and decrease information overload by suggesting items likely to be relevant [16]. So far, research on team-based interventional feedback [17], [18], [15], [19] has mostly investigated the impact on team outcomes [20], but hardly considered the so often theorized effect on team processes [3].

In this paper, we focus on the impact of recommender systems [21] on team processes in computer-supported collaboration environments e.g., [15],[19], [3]. We aim at extending the current understanding of the effects of recommendations in such collaboration environments by investigating how feedforward recommendations impact interactions in teams. For this purpose, we conducted two laboratory experiments. The first experiment studied the difference of communication and coordination between a treatment group of teams receiving recommendations and a control group of teams receiving no recommendations when working on a decision-making task. The second experiment explored changes in communication and coordination after recommendations. The implemented recommender system (RS) is part of the Innovation Factory (IF), a computer-supported collaboration environment. The RS, visualised as a tag cloud, recommends knowledge elements to users based on the topic which they are currently writing about [22].

The paper is organized as follows: Section 2 discusses related work and Section 3 describes the Innovation Factory as the collaborative environment adopted in our experiments. Section 4 describes the study design and section 5 discusses the results of our laboratory experiments before Section 6 presents our conclusions.

2. Related Work

2.1. Recommender Systems

The term Recommender System (RS), was introduced for the first time in 1997 by Resnick and Varian [23]. A RS aims at rating resources according to the interest a specific user will show in them within a dedicated resource space. Typically, this prediction is made by considering implicit or explicit ratings expressed by other users on the same resources [24]. As studies on RS are relatively recent and interest in their applications is growing, the state of the art is rapidly evolving. At the present time we can distinguish between five main approaches for implementing recommender systems.

- *Content-based:* the RS rates resources based on their degree of similarity with other resources already rated by the same user.

- *Collaborative-filtering:* the RS rates resources for a user based on implicit or explicit ratings provided by other users. The final rate of a resource depends on the similarity between the users who rated the resource and the user querying the RS. It should be noted that collaborative-filtering is today the most popular technique for implementing RS.

- *Demographics-based:* the RS rates resources based on similarity between demographics (age, gender, country of residence) of users who rated the resource high and those of the user querying the RS. The benefit of a demographics-based approach is that it does not require a complete history of user ratings of the type needed by collaborative techniques.

- *Social Network-based:* the RS rates resources based on preferences expressed by users sharing a social relation with the user querying the RS. This approach is typically used in combination with collaborative filtering techniques.

- *Hybrid RS:* the system rates the items to be suggested based on a combination of the approaches described above. Robin Burke has written a complete classification of hybrid systems [25], listing a number of hybridization methods to combine pairs of recommender algorithms.

All the above mentioned techniques have a common drawback: the "cold-start" problem also known as the "early rater" or "sparse ratings" problem [26]. The RS requires a critical mass of ratings available in order to compute good quality ratings. While ratings are initialized manually in many systems, the need to address the"cold-start" problem has fostered research on *knowledge-based* RS that rate resources based on resource descriptions [27]. Using an inference engine, the RS computes the best match between a resource description specified by the user and description on resources available in its knowledge base.

Other studies [28] discuss the advantages and disadvantages of the different RS algorithms, comparing collaborative filtering to content-based or knowledge-based approaches. The main advantages of collaborative filtering are related to its simplicity: it is domain independent and can work with a relatively simple data structure. The main disadvantage is that collaborative filtering techniques cannot recommend resources when historical data are insufficiently available[1]. Content-based or knowledge-based techniques do not suffer from the "cold-start" problem and can work even with a limited data set; however, the process of content encoding and representation is not trivial, as it is highly domain-dependent and very expensive if it cannot be automated from independent organizational processes [29].

Empirical studies have shown that there is no "absolute best" among collaborative filtering, content-based and knowledge-based techniques. In [30] the listening data of approximately 360,000 unique users of the social radio *Last.fm*[2] were analyzed to compare the quality of the similarity scores obtained by classical collaborative filtering based on user preferences and by a knowledge-based approach based on folksonomy. In [31] the authors performed experiments on three datasets, namely Delicious[3], CiteULike[4] and BibSonomy[5] and compared a range of collaborative and content-based approaches with respect to item recommendation. The results showed that the combination of collaboration and content-based algorithms obtained the best performance.

2.2. Recommender Systems in Technology Enhanced Learning

Recommender systems have attracted much interest in the Technology Enhanced Learning (TEL) due to their high potential of eliciting relevant learning resources [32]. Since information retrieval, in terms of searching for relevant

[1] The above mentioned "cold-start" problem.
[2] http://www.last.fm
[3] http://delicious.com
[4] http://www.citeulike.org
[5] http://www.bibsonomy.org

learning resources, is a pivotal activity in TEL, RS for TEL applications (RS-TEL) have attracted much interest. Probably the most complete survey on RS-TEL is [33]. In the conclusions, the authors discuss the validation problem of RS-TEL, highlighting the fact that a systematic comparative evaluation of RS-TEL systems is still lacking.

Nevertheless, some interesting experimental results are available, even if the different studies do not allow for a systematic comparison, due the heterogeneity of the experimental designs that were adopted. The work [34] emphasizes the positive effect that RS have on calling the attention of users to other users accessing the same resources, e.g. via a message like "someone is looking at the same knowledge that you are looking at". By letting a user know that other users also access or have accessed the same resource, a certain level of justification of the item's relevance is given. This, in turn, positively affects the trust that users have in the RS suggestions. Moreover, several studies discuss the role that justification and explanation of recommendations have when improving the quality of the user interaction with the RS [35] [36]. Other works put forward the idea that combining content with ratings can improve the perceived quality of the suggestions [37]. Bobadilla *et al.* conducted experiments demonstrating that collaborative filtering RS give better results if the importance of the recommendations that each user generates is weighted based on their level of knowledge [38].

More recently, a consistent experimental methodology was proposed [39], and [40] conducted a series of experiments comparing the effectiveness of different collaborative filtering algorithms for RS-TEL. An interesting result of this study is that, in general, the quality of recommendations increases with the number of neighbors, i.e. of similar users. However, after a certain point, the improvement gain flattens out. The authors conclude that more experiments are needed, as evidence exists that the size of the user community may affect the performance of RS.

In [41] experiments with students proved that a classic e-learning environment can be improved by adding a content-based RS, not perhaps in terms of improving learning result, but in terms of attractiveness for students. If a TEL environment is enhanced by the addition of a collaborative filtering, RS-TEL, the increase in time spent by learners within the TEL environment is even higher than the one registered for content-based RS-TEL.[6]. It should be noted that these effects were observed for beginner and intermediate learners, rather than for advanced ones.

2.3. Recommender Systems and Collaborative Work

There is a long tradition of RS-relevant studies with respect to computer-supported cooperative work [42], [43]. More recently, the research domain of collaboration engineering [44] aimed at understanding how facilitation support needs to be design to transfer past practices of effective collaboration into new collaboration settings. Often the cause of failure of collaboration lies within problems such as, too much time spent on non-relevant activities, or problems with goals, e.g., conflicting expectation about meeting content [45]. In addition, the evolution of the Internet has shown the importance of technology, such as blogs, email, instant messaging, social networks, wikis and other instances of what is often called "social software" in recreating social contexts [46, 47]. Thus, employees have many communication channels at their disposals to send and receive important and relevant information causing additional challenges in terms of information overload. Among many other effects this could lead to the breakdown of communication and higher time requirements for information handling [48]. With the exception of some early studies, such as for instance [49], the implications of RSs on collaboration were insufficiently investigated. The study shows that the RS has positive effects on the achievements of objectives that can even be independent of the quality of the recommendations they provide. This implies that recommendations can stimulate collaboration [50]. In addition, a major challenge for developing a RS supporting collaboration lies in taking into account the goals of the whole group. The work [51] analyzes the

[6] However, no difference in perception between the RS-TEL paradigms could be established.

implication of a sequence of suggestions with respect to the perceived satisfaction of a group. An interesting finding is that a recommendation sequence is more appreciated when the contribution of each individual is highlighted.

The review of the literature shows that recommendations have effects on users, with positive impact on information seeking. However, there is a gap in our understanding why and how recommendations affect teams during collaboration. Progress in descriptive modeling of human-recommender interaction has not yet enabled reliable prediction of RS efficacy [52]. Taking into consideration that recommendations do not necessarily impact the outcome, it appears to be interesting to investigate the effects that recommendations have on stimulating collaboration in teamwork. In particular, this work focus on measuring the effects of RS on interaction frequency and work shares, as detailed in next sections.

3. The Innovation Factory

This section outlines the concept of our Innovation Factory, an extended collaborative environment supporting participatory design [22]. IF aims to facilitate the handling of collaborative tasks by facilitating coordination activities and making important resources detectable. IF reuses open source products for collaborative writing, instant group messaging, and document management. The IF can be described as composed by three main modules:

- A Collaborative Environment (CE) that provides functionalities such as chat, synchronous editor (pad), wiki and so on.
- A Recommender System (RS) that provides suggestions to stimulate the collaborative process.
- A Knowledge Base (KB) to store and manages documents.

The key contributions of the IF are (1) the integration of the different modules and (2) a *tag cloud* to visualise the recommendations computed by the RS

Figure 1: The IF interface.

providing team members with tags linked to relevant documents in the knowledge base (see Fig. 1).

Our RS incorporates collaborative filtering and content-based recommendation techniques. For the latter aspect, the RS relies on a KB that contains *(i)* common knowledge (e.g., the organization structure), and *(ii)* domain-specific knowledge. The RS is applied in reaction to three particular kinds of inputs: *(i)* a stimulus, describing a task to be performed or a set of goals to be achieved, *(ii)* a target, defining the set of employees over which suggestions must be applied, and *(iii)* a set of local configurations defining, e.g., the type and the similarity measures. The RS computes the *concept adequacy*, between the targets and stimulus, and returns recommendations on concepts connected to the stimulus.

The RS is composed of independent modules. The RS *Configuration Panel (CP)* defines the specific settings of any instantiation of the RS [7]. In particular, each stimulus is represented by a concept or by a pattern of concepts in the knowledge base. Stimuli represent new events or new knowledge that require the organization to improve its competences. The notion of target is also important to configure the IF. In fact, the target defines the resources to be recommended.

The sub-modules of the IF generate recommendations as lists of resources (represented in term of data objects of the knowledge base) that are considered

[7]The resources analyzed within IF are identified in the meta-model, but their specific properties and relations expressed depend on the domain and/or the organization. The structure of the stimuli can be configured in the CP.

useful to handle the stimulus.

The *Search Engine (SE)* and the *Similarity Machine (SM)* organize the resources in a similarity space. When a stimulus reaches the RS, the QE and the SM identify objects that are related to this stimulus.

In turn, the *Diversity Machine (DM)* searches out of the boundaries of the similarity space, in order to increase the diversity of the provided recommendations. The DM relies on the notion of expert in order to identify resources that, while dissimilar to the stimulus, can turn out to be useful in handling it. Roughly speaking, the criterion used used by the DM is the following: recommend resources that were visited/used by experts that succeeded in achieving goals related to the current stimulus in the past. The idea is that such resources may be related to latent goals that are not part of the stimulus.

At the end, the individually ranked lists provided by the SE, SM, and DM are merged and these recommendations are visualised as a tag cloud.

4. Experimental Design

To assess the overall quality of collaboration, it has been suggested to assess collaboration, among others, along the dimensions time management, effort, task division, and reaching consensus [53]. Along these dimensions and the research background presented in Section 2, we develop our hypothesis. Regarding time management, researchers found that teams struggle with the effort necessary for finding information and making it accessible for all team members [54]. Reasons for that may be that, team members have a large personal, team, and organisational knowledge base where it is difficult to keep an overview about new or updated information. Our tag cloud, i.e., the visualisation of recommendations for resources in the knowledge base, is deemed helpful to overcome this challenge. Therefore, we phrase the first hypothesis as follows:

H 1. Teams with recommendations will take on less effort for information handling in the knowledge base than teams with no recommendations.

Beside the effort needed to handle information in a team, sharing work and assigning sub-tasks is considered as challenging for teams, due to e.g., different competences of individual workers with respect to self-organizing their work and focussing on ongoing activities. Our recommender continuously provides recommendations based on the collaboratively developed content, which we assume can help, keeping an individuals attention and engagement task and lead to a more equal share of work. Therefore, our second hypothesis is the following:

H 2. Teams with high recommendation support will achieve better division of work than teams with low recommendation support.

While H2 focusses on the share of work among team members, we now assume that a valuable contribution in collaborative work can also be the participation in a discussion. Based on related work that found out that recommendations can stimulate collaboration [50] we are interested if this is also true for communication behaviour. Additionally, we aim to find out if there is an increased interaction after a recommendation was selected. Hence we define H3 in two parts as follows:

H 3. Teams with recommendations will (H3a) communicate more and (H3b) increase their interaction in subsequent collaboration.

Evaluating collaborative environments impact requires an understanding if and how a recommender system can lead to alterations of team behaviour. Previous studies on this topic limited their investigation to the analysis of results achieved at the end of the collaboration process [20], [41].

Our choice was to consider collaboration activities performed during the collaboration process. For this purpose, we designed our experiments to investigate collaboration in terms of interaction frequency.

4.1. Experimental Designs

We investigated the impact of recommendations on team processes with two laboratory experiments. The first experiment adopted a between-subjects

	Experiment 1	Experiment 2
Type of Design	Between-subjects	Within-groups
Task	Business Processes, Web Services, and Security Requirements	Business Processes and Cultural Heritage
Duration	90 minutes	90 minutes
Subjects	18 graduate students	39 graduate students
Teams	6	7
Team size	3 persons	3-6 persons
Gender	Male 100%	Male 75% Woman 25%
Age	20-25	21-30

Table 1: Experimental Design Synopsis

design for testing H1, H2, and H3a. The second experiment, which adopts a within-group design, was conducted to test H3b investigating the changes in interaction over time when teams receive recommendations.

Additional information on our experiments are detailed in the following.

4.2. Factor and Factor Levels

In the first laboratory experiment, the single factor or independent variable, i.e. recommendation, was tested in two conditions. In contrast to the control group, the treatment group or experimental group received recommendations. In the second laboratory experiment, all teams received recommendations. Both laboratory experiments adopted the same factor, i.e., feedforward recommendation. The relevance score was based on the recommendations of the RS. In the control condition, teams did not have any recommendations at their disposal. Team members could access the linked knowledge base from where they could start their information search. The knowledge base consisted of tagged documents and web-links.

4.3. Subjects

The first experiment involved graduate students from two universities; 18 students enrolled in the master program in Information Systems at the University of Innsbruck (Austria), and 7 students enrolled in the master program in Computer Science at the Università degli Studi di Milano (Italy). Students were placed into 7 teams based on their competence profiles. The sample consists of

3 teams in the experimental group and 3 teams in the control group, involving 18 students overall. One team was excluded from data analysis, as the involved students did not participate in the pre-survey one week prior to the experiment and therefore had no competence profile associated.

The second experiment involved 39 graduate students from the Università degli Studi di Milano. Teams were constructed, based on competence profiles, using the same methodology adopted for the first experiment.

As reported in Table 1 in both experiments we had a prevalence of males in the 20-30 age range.

4.4. Controlling for Differences in Competence

It was anticipated that students' skills varied in their extent of technical knowledge relevant to the task, since students had an undergraduate degree in different domains. A screening was deemed necessary to balance competences in the teams. For this purpose, a pre-survey was administered to ensure that competences of participating students were distributed equally. The survey collected competence information on task-related skills, i.e., understanding web technologies, process models, business models, and web services. Based on the students' self-evaluation, competence profiles were built and assessed as weak, normal or strong. In the first laboratory experiment, 7 students were associated with strong competence profiles. For the second laboratory experiment, 7 students were associated with strong competence profiles. In both laboratory experiments, these students were assigned into teams whose team members had normal or weak profiles.

4.5. Tasks

In the first laboratory experiment, the task described an extract of a BPMN-modelled credit management-process of a bank. The task description asked each team to extend the BPMN-model with a short description of the involved web services and their activities. Furthermore, team members had to discuss any potential security threads that could arise in this context. In the second laboratory experiment, the task required the students to define a new business model in the cultural heritage context, providing a value model, a marketing strategy and an implementation strategy to distribute the new services. In both experiment, the goal was to collaboratively create a written report incorporating the important topics that were discussed.

4.6. Collaboration Environments

The team members were placed in different rooms and were provided with the IF, the collaborative environment, conceptually described in Section 3. The two experiments took place at two different phases in the software development process of the IF. The heart of the collaborative environment, i.e., the visualisation of the recommendations from the RS, was identical in both performed experiments, including a consolidated communication log with time-stamped information on chat protocols, tag cloud clicks, and access to the knowledge base. Table 2 depicts these log events referred to as *Event Classes*. They include *Chat*, *Pad*, *KB*, *TC*, and *other*. These event classes are used to display general statistics on the time spent for each event class. In the first experiment, the Innovation Factory used the Mikiwiki platform [55] as an instant chat tool, Google Docs as a collaborative writing tool (pad), and a knowledge base incorporated in the Mikiwiki. In the second experiment, tools providing functionalities for collaborative writing, instant messaging, and document management were changed to fit the requirement of being open source where all provided by the IF environment. Now, the IF was built on LifeRay [56], that incorporates the chat tool and the knowledge base, and used an Etherpad [8] portlet as a collaborative writing pad. This new implementation is configurable with themes, portlets and gadgets and guaranty best performances in terms of concurrent access. Furthermore, the IF was equipped with the tag cloud, that was configured to visualise recommendations computed by the RS during the task execution.

[8] http://etherpad.org

Event Class	Description
Chat	The user read or writes messages on the chat
Pad	The user accesses, inserts, deletes and modifies contents by writing
KB	The user accesses and reads a document in the knowledge base
TC	The user elect in the tag cloud a recommendation from the RS
other	Other activities on the collaborative environment

Table 2: Definition of Event Classes

4.7. Measurements of Dependent Variables

All dependent variables were measured on the basis of information provided from the consolidated communication log of the collaboration environments.

Information handling. Information handling describes the time spent by each participant for information seeking using the knowledge base (KB). We operationalized this variable by measuring the amount of KB document accesses per team member, expressed in term of relative frequency. Lower values of information handling relate to less clicks in the KB whereas higher values of information handling relate to more clicks in the KB.

Equal work division. Equal work division describes the degree of the intensity of work-sharing among team members in a team. It is expressed by the following formula:

$$W_u = \sum_{t=1}^{t=n} \frac{A_{u,t}}{A_t} \frac{1}{n}.$$

Where W_u is the work-sharing rate of user u, t represents a single unit of time, n is the total number of time units generated by a team, A_t is the total number of activities (e.g., KB access, chat, etc.) performed at time t and $A_{u,t}$ is the total number of activities performed at time t by the individual user u[9]. Because low values of W correspond to a better sharing of the workload we rank this parameter inversely. Consequently, the metric describes the following: at a given time unit, if several users are working the value is low, indicating good work-sharing; otherwise if few users are working then the value is high, indicating

[9]To avoid null values we considered only time units were at least one activity was performed.

bad work-division.

Communication frequency. Communication frequency describes the extent of communication among team members. We operationalized this variable by measuring the amount of chat messages per team member. It should be noted that higher values of communication frequency relate to more communication among team members performed in the chat whereas lower values of communication frequency relate to less communication among team members.

Interaction trend. Interaction trend describes whether interaction at time interval t will increase (positive trend) or decrease (negative trend) at time interval t_w for a team. Unlike the dependent variables mentioned above, it is a group-level variable. For its operationalization, we organized data extracted from the IF communication log in time intervals of 1 *minute*. Figure 2 visualizes the interaction frequency per time interval for Team 2. Vertical arrows signal time intervals where interaction with the tag cloud took place.

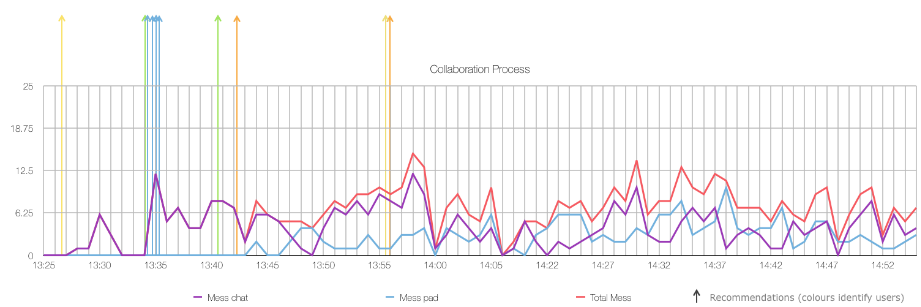

Figure 2: Temporal evolution of the communication process during team work.

For each time interval, the interaction frequency stemming from chat, pad, TC, KB, and other features was calculated. We first normalised the interaction frequency $\tilde{f}$ for each time interval t_i by dividing it by the maximum frequency $maxf$ observed in data. We then computed Δ as the difference between $\tilde{f}_{t_i}$ and the mean of frequencies in a time window w; calculating this mean as $\bar{x}_{(\tilde{f}_{t_i}, \tilde{f}_{t_i+w-1})}$. This value shows if the interaction frequency in the window of subsequent time intervals is following a *positive trend* (interaction increases), or a negative *negative trend* (interaction decreases). Table 3 illustrates the operations applied with $w = 5$ *minutes* and $max = 15$ to measure the *interaction trend* per team.

Team	Time	Total Messages	Mess Norm $\tilde{f} = \frac{t_i}{maxf}$	$\Delta(\tilde{f}_{t_i}, \bar{x}_{(\tilde{f}_{t_i}, \tilde{f}_{t_i+w-1})})$
T2	01:00	0	0.0	0.45
T2	02:00	2	0.0	0.55
T2	03:00	14	0.9	-0.29
T2	04:00	10	0.7	-0.15
T2	05:00	10	0.7	-0.16
T2	06:00	7	0.5	0.01
T2	07:00	7	0.5	0.07
T2	08:00	5	0.3	0.31
T2	09:00	9	0.6	0.08
T2	10:00	8	0.5	0.13
T2	11:00	11	0.7	0.00
T2	12:00	15	1.0	-0.35

Table 3: Trends of time intervals with $w = 5$ *minutes* and $maxf = 15$.

We then separated the communication log into 28 records describing time intervals where users clicked on concepts recommended in the tag cloud, referred by R, and 533 records describing all the other time intervals, referred by $\neg R$. These categories structure the independent variable. The dependent variable is then structured in categories related to the interaction trend differentiating between `positive` or `negative` orientation. Time intervals with trends less than the *mean* value fall in the category `negative` orientation. Time intervals with trends greater than *mean* value fall in the category `positive` orientation. In our case *mean* was computed taking the average between the $meanR$ (0.131) and the $mean\neg R$ (-0.002), resulting in 0.065.

4.8. Procedure and Data Analysis

The experiments were designed to last two hours, with the option of extending it for an additional 15 minutes in case of delays in the starting phase. The tasks were estimated to take one and a half hours, leaving about 30 minutes for management purposes (welcoming students, introducing the task, managing distribution of participants into separate rooms, and synchronising the remote groups). The IF was the sole collaborative environment allowed, with all other communication channels disabled. The setup was tested and improved based on a pre-study with nine volunteers.

For both experiments we adopted a non-parametric approach [57]. Mann-Whitney U tests were applied to test for significant differences between control and experimental teams. Chi-square tests of independence were used to test for differences between interaction when using recommendations and not. The Mann-Whitney U-test can be applied to small samples and requires neither normal distribution of data nor homogeneity of variance. In contrast to parametric tests based on variance (e.g., ANOVA), the Mann-Whitney U test calculates significant differences between two factors on the basis of a ranking order [57]. The Chi-square test of independence [58] is applied when comparing populations described by categorical variables. It is used is used to determine whether there is a significant difference between the expected frequencies f_e and the observed frequencies f_o in the distribution of the independent variables. This is achieved by comparing the occurrences of the categories describing the dependent variables between the categories of independent variables. Formally this can be expressed as:

$$\sum \frac{(f_o - f_e)^2}{f_e}.$$

5. Results

With the data collected in the two experiments we tested our hypothesis and will here present our results along the three hypothesis. Table 4 provides a detailed overview of log activities that the 3 teams in the experimental group

Teams with Recommendations			Teams without Recommendations		
Event Class G1	**Frequency**	**Relative Frequency**	**Event Class G4**	**Frequency**	**Relative Frequency**
Chat	190	49.61%	Chat	123	27.95%
Pad	87	22.72%	Pad	150	34.09%
KB	68	17.75%	KB	128	29.09%
TC	10	2.61%	TC	0	0%
other	28	7.31%	other	39	8.86%
Event Class G2	**Frequency**	**Relative Frequency**	**Event Class G5**	**Frequency**	**Relative Frequency**
Chat	243	51.81%	Chat	126	36.63%
Pad	113	24.09%	Pad	101	29.36%
KB	62	13.22%	KB	88	25.58%
TC	28	5.97%	TC	0	0%
other	23	4.90%	other	29	8.43%
Event Class G3	**Frequency**	**Relative Frequency**	**Event Class G6**	**Frequency**	**Relative Frequency**
Chat	127	34.23%	Chat	154	24.76%
Pad	93	25.07%	Pad	215	34.57%
KB	90	24.26%	KB	191	30.71%
TC	24	6.47%	TC	0	0%
other	37	9.97%	other	62	9.97%

Table 4: Activities performed by the two groups

and the 3 teams in the control group performed during the first laboratory experiment. The manipulation of the independent variable, i.e., recommendation, was effective, as all treatment groups used the tag cloud. In teams receiving recommendations, the highest share of activities relate to chat activities (45.79%) followed by writing activities in the pad (23.96%) and information seeking activities in the knowledge base (17.99%). In contrast to that, teams receiving no recommendations have the highest share of activities connected to writing activities in the pad (33.14%) followed by information seeking activities in the knowledge base (28.95%), and chat activities (28.66%). Table 5 provides the descriptive statistics for testing H1, H2, and H3a. For the within-subject designed second experiment it was impossible to provide data in summarised form but we are happy to provide log data and descriptive statistics upon request.

5.1. Information handling

In terms of *Information Handling* (H1), we investigated whether recommendations will have an influence on the effort required for searching for informa-

Treatment	Mean (SD) Information Handling	Mean (SD) Equal work division	Mean (SD) Communication Frequency
Teams receiving recommendations	.1767 (.0800)	.2364 (.0185)	.4456 (.1269)
Teams receiving no recommendations	.2833 (.0735)	.3453 (.0519)	.3078 (.0976)

Table 5: Descriptive Statistics of Laboratory Experiment 1

Treatment	Information Handling	Equal work division	Communication Frequency
recommendations vs	$\geq \neg H1$	$\geq \neg H2$	$\geq \neg H3a$
no recommendations	Z=-2.566, p=.010	Z=-3,488, p=.000	Z=-2,357, p=.017

Table 6: Results of Hypotheses Tests H1, H2, and H3a

tion in the knowledge base. Figure 3 illustrates how subjects were ranked in the Mann-Whitney U test. Results show that there is a significant difference between teams receiving recommendations (M=.1767; SD=.0800) and teams that did not receive recommendations (M=.2833; SD=.0735). Teams in the experimental condition, i.e., receiving recommendations, took significantly less effort for manually browsing the knowledge base in the search for relevant information (z=-2.566; p=.010). It can be assumed that the recommendations visualized by the tag cloud provided them with relevant document suggestions so that they did not feel the need to screen all documents in the knowledge base available. In contrast to that, it appears that teams with no recommendations still felt obliged to access the knowledge base. A reason could be that people feel uncomfortable making decisions when they could foster potentially relevant information. Nevertheless, our results show that teams with recommendations took less time for handling information from the knowledge base. This supports H1.

5.2. Division of work

In terms of *equal work division* (H2), we investigated whether recommendations will have an influence on the distribution of work among team members.

	Experimental Group			Control Group	
id	param	rank	id	param	rank
1	21%	8.5	10	34%	17
2	14%	4.5	11	30%	12.5
3	9%	1	12	33%	16
4	10%	2	13	29%	11
5	11%	3	14	15%	6
6	14%	4.5	15	17%	7
7	31%	14.5	16	36%	18
8	21%	8.5	17	30%	12.5
9	28%	10	18	31%	14.5
	Σ rank=	56.5		Σ rank=	115
	R1 = (Σ rank)/ 9	6.27777778		R2 = (Σ rank)/ 9	12.7222222

Figure 3: Subjects ranked based on relative frequency of variable *KB*.

Figure 4 illustrates how subjects were ranked in the Mann-Whitney U test. Our results show that there is a significant difference between teams receiving recommendations (M=.2364; SD=.0185) and teams that did not receive recommendations (M=.3453; SD=.0519). Teams in the experimental condition, i.e., receiving recommendations, had a significantly less distorted share of work than teams in the control condition (z=-3.488, p=.000). It appears that the provision of recommendations helped team members to engage more equally in the task. Particularly, during information search, team members usually require a lot of time to browse through potential resources that turn out to be irrelevant. In this case, it might happen that team members get demotivated to engage, as their contribution to solving the team task is weak. Recommendations help to guide the team members attention to potentially relevant sources. This could be the reason why team members receiving recommendations felt more equally motivated to further engage in the task and therefore share work effort.

5.3. Interaction behaviour

In terms of *communication frequency* (H3a), we investigated whether recommendations will have an influence on the frequency of communication among team members. Figure 5 illustrates how subjects were ranked in the Mann-Whitney U test. Our results show that there is a significant difference between teams receiving recommendations (M=.4456; SD=.1269) and teams that did not

	Experimental Group	
id	param	rank
1	0.22	1
2	0.225	4
3	0.228	5
4	0.243	7
5	0.257	9
6	0.275	10
7	0.226	3
8	0.231	6
9	0.223	2
	Σ rank=	47
	R1 = (Σ rank)/ 9	5.22222222

	Control Group	
id	param	rank
10	0.38	15
11	0.382	17
12	0.385	18
13	0.279	12
14	0.273	8
15	0.277	11
16	0.374	13
17	0.377	14
18	0.381	16
	Σ rank=	124
	R2 = (Σ rank)/ 9	13.7777778

Figure 4: Global development of activities, ranked with parameter: *work-sharing*.

receive recommendations (M=.3078; SD=.0976). Teams in the experimental condition, i.e., receiving recommendations, had a significantly higher communication frequency, relatively to other interactions, than teams in the control condition (z=-2.357, p=.017). This means that teams receiving recommendations communicated more with their peers using the chat functionality. This appears to be interesting since most past research on IT-supported collaboration environments fails to foster communication among their users. In contrast to that, our results suggest that recommendations embedded in collaboration environments can trigger communication among team members. For problem-solving and decision-making tasks, communication among team members is assumed to be of high importance since teams can tap into the diversity of expertise which consequently can be brought to the task.

5.4. Recommendations impact the interaction behavior of teams

As described in Section 4, the second laboratory experiment, which adopted a within-subject design, was conducted to further investigate what impact recommendations have on the interaction behavior over time.

The results of the Chi-Square Test of Independence show that when teams made use of the provided recommendations, they had a more positive outlook on increasing their interaction frequency in the subsequent interactions (z=8.667, df=1 and p=.003). Table 7 shows the occurrences for interactions with positive

	Experimental Group			Control Group	
id	param	rank	id	param	rank
1	48%	4	10	35%	10
2	49%	3	11	24%	15.5
3	40%	7.5	12	13%	18
4	40%	7.5	13	30%	13
5	66%	1	14	39%	9
6	24%	15.5	15	29%	14
7	59%	2	16	19%	17
8	41%	6	17	45%	5
9	34%	12	18	33%	11
	Σ rank=	58.5		Σ rank=	113
	R1 = (Σ rank)/ 9	6.5		R2 = (Σ rank)/ 9	12.5

Figure 5: Subjects ranked based on relative frequency of variable *Chat*.

or negative trends over R and $\neg R$. In other words, observed trends exhibit more positive orientation then expected, validating hypothesis H3. It can be assumed that recommendations sparked interaction in the team by accessing more documents, increasing communication among team members, editing the collaborative writing pad, etc.

	R	$\neg R$	**Total**
Negative trend (< 0.065) (observed/expected)	10/17.37	338/330.63	348
Positive trend (> 0.065) (observed/expected)	18/10.63	195/202.37	213
Total	28	533	561

Table 7: Observed interaction occurrences with positive or negative trends over R and $\neg R$

5.5. Limitations

The study was designed in such a way that subjects had to run through multiple collaboration phases which might have been interpreted differently by the teams. Both experiments are laboratory experiments, used an artificial decision-making task, employed students as proxies for members of small-sized IT-supported, distributed teams and therefore our findings are limited in their generalizability. The analyses in our current study focused on interactions with the computer-supported collaboration environment as surrogates for team processes. While this is a promising avenue for analyzing the effects of automated recommendations on team processes, future work might also take into account perception-based measures of team processes as well as an assessment of the team outcomes.

6. Conclusions

This paper reported the results of two experiments designed to explore the effects of recommendations on team processes in computer-supported collaboration environments. Our findings substantiate our claim that recommendations are associated with interaction frequency, regulating *information handling* (z=-2.566, p=.010), *work sharing* (z=3.488, p=.000), *communication frequency* (z=-2.357, p=.017) and *interaction trend* (z=-8.667, p=.003).

We showed that recommendations had a positive impact on lowering efforts for information handling, increasing communication as well as interactions with the computer-supported collaboration environment, and facilitating more equal sharing of work among team members. Research concerned with the investigation of IT-supported teams can benefit from our findings in multiple ways. The potential benefits of computer-supported collaboration environments are well received across disciplines (e.g., [59], [60], [13], [61]). However, there are hardly any well-established and validated theories that explain the role of automated facilitation in teams [62]. Our findings add to this stream of research and contribute to a better understanding of how team processes are affected by recommendations. Pentland demonstrated that how team members communicate is more important to team effectiveness than what they communicate. Our constructs information handling, equal division of work, communication frequency, and interaction trend can be considered to extend Pentlands collection of communication and interaction constructs, i.e., energy, engagement, and exploration.

Of course, many issues remain to be addressed. An important question is whether the support provided by recommendations not only affects the dynamics of team interactions, but it is also successful in improving the quality of task execution in respect to task goals. Future work might explore associations between patterns of team processes that are enabled by recommendations and corresponding outcomes in laboratory as well as company settings.

References

[1] E. N. Wolff, The growth of information workers., Communications of the ACM 48 (10) (2005) 37–42. doi:10.1145/1089107.108913.

[2] J. Keyton, S. J. Beck, M. B. Asbury, Macrocognition: a communication perspective, Theoretical Issues in Ergonomics Science 11 (4) (2010) 272–286. doi:10.1080/14639221003729136.

[3] J. Mathieu, M. T. Maynard, T. Rapp, L. Gilson, Team effectiveness 1997-2007: A review of recent advancements and a glimpse into the future, Journal of Management 34 (3) (2008) 410–476. doi:10.1177/0149206308316061.

[4] R. Huckman, B. Staats, The hidden benefits of keeping teams intact., Harvard business review 91 (12) (2013) 27–29.

[5] M. J. Eppler, J. Mengis, The concept of information overload: A review of literature from organization science, accounting, marketing, mis, and related disciplines, The information society 20 (5) (2004) 325–344.

[6] I. Seeber, R. Maier, G. Waldhart, V. Bellandi, P. Ceravolo, E. Damiani, F. Frati, J. Hrastnik, Computer-supported collaboration environments and the emergence of collaboration aspects, Presentation at the conference The changing nature of work: working smarter with ICT, Pre-ICIS2013 OSRA Workshop held in Milan (2013).
URL http://hdl.handle.net/2434/232448

[7] P. Dourish, Extending awareness beyond synchronous collaboration, Position Paper for CHI 97 (1997) 249–256.

[8] F. Siampou, V. Komis, N. Tselios, Online versus face-to-face collaboration in the context of a computer-supported modeling task, Computers in Human Behavior 37 (0) (2014) 369 – 376. doi:10.1016/j.chb.2014.04.032.

[9] J. I. Tutty, J. D. Klein, Computer-mediated instruction: A comparison of online and face-to-face collaboration, Educational Technology Research and Development 56 (2) (2008) 101–124.

[10] T. W. Malone, K. Crowston, What is coordination theory and how can it help design cooperative work systems?, in: Proceedings of the 1990 ACM Conference on Computer-supported Cooperative Work, CSCW '90, ACM, New York, NY, USA, 1990, pp. 357–370. doi:10.1145/99332.99367.
URL http://doi.acm.org/10.1145/99332.99367

[11] S. W. Kozlowski, D. R. Ilgen, Enhancing the effectiveness of work groups and teams, Psychological science in the public interest 7 (3) (2006) 77–124.

[12] A. N. Kluger, A. DeNisi, The effects of feedback interventions on performance: a historical review, a meta-analysis, and a preliminary feedback intervention theory., Psychological bulletin 119 (2) (1996) 254.

[13] B. C. Wheeler, J. S. Valacich, Facilitation, gss, and training as sources of process restrictiveness and guidance for structured group decision making: An empirical assessment, Information Systems Research 7 (4) (1996) 429–450.

[14] F. Ricci, L. Rokach, B. Shapira, Introduction to recommender systems handbook, Recommender Systems Handbook (2011) 1–35.

[15] M. Limayem, G. DeSanctis, Providing decisional guidance for multicriteria decision making in groups, Information Systems Research 11 (4) (2000) 386–401.

[16] L. Terveen, D. W. McDonald, Social matching: A framework and research agenda, ACM transactions on computer-human interaction (TOCHI) 12 (3) (2005) 401–434.

[17] T. Chenoweth, K. L. Dowling, R. D. S. Louis, Convincing dss users that complex models are worth the effort, Decision Support Systems 37 (1) (2004) 71–82.

[18] P. Todd, I. Benbasat, The use of information in decision making: an experimental investigation of the impact of computer-based decision aids, Mis Quarterly (1992) 373–393.

[19] Z. Wong, M. Aiken, Automated facilitation of electronic meetings, Information & Management 41 (2) (2003) 125–134.

[20] I. Seeber, R. Maier, B. Weber, Opening the black box of team processes and emergent states: A literature review and agenda for research on team facilitation, in: System Sciences (HICSS), 2014 47th Hawaii International Conference on, IEEE, 2014, pp. 473–482.

[21] F. Ricci, L. Rokach, B. Shapira, Introduction to recommender systems handbook, in: F. Ricci, L. Rokach, B. Shapira, P. B. Kantor (Eds.), Recommender Systems Handbook, Springer US, 2011, pp. 1–35. `doi:10.1007/978-0-387-85820-3_1`.
URL `http://dx.doi.org/10.1007/978-0-387-85820-3_1`

[22] V. Bellandi, P. Ceravolo, E. Damiani, F. Frati, J. Maggesi, L. Zhu, Exploiting participatory design in open innovation factories, in: Signal Image Technology and Internet Based Systems (SITIS), 2012 Eighth International Conference on, 2012, pp. 937–943. `doi:10.1109/SITIS.2012.139`.

[23] P. Resnick, H. R. Varian, Recommender systems, Commun. ACM 40 (3) (1997) 56–58. `doi:10.1145/245108.245121`.
URL `http://doi.acm.org/10.1145/245108.245121`

[24] G. Adomavicius, A. Tuzhilin, Toward the next generation of recommender systems: A survey of the state-of-the-art and possible extensions, IEEE Trans. on Knowl. and Data Eng. 17 (6) (2005) 734–749. doi:10.1109/TKDE.2005.99.
URL http://dx.doi.org/10.1109/TKDE.2005.99

[25] R. Burke, Hybrid recommender systems: Survey and experiments, User Modeling and User-Adapted Interaction 12 (4) (2002) 331–370. doi:10.1023/A:1021240730564.
URL http://dx.doi.org/10.1023/A:1021240730564

[26] B. M. Sarwar, J. A. Konstan, A. Borchers, J. Herlocker, B. Miller, J. Riedl, Using filtering agents to improve prediction quality in the grouplens research collaborative filtering system, in: Proceedings of the 1998 ACM conference on Computer supported cooperative work, ACM, 1998, pp. 345–354.

[27] M. De Gemmis, L. Iacquinta, P. Lops, C. Musto, F. Narducci, S. G., Learning Preference Models in Recommender Systems, Springer, 2010, pp. 387–408.

[28] D. Fleder, K. Hosanagar, Blockbuster culture's next rise or fall: The impact of recommender systems on sales diversity, Management Science 55 (5) (2009) 697–712.

[29] H. Drachsler, H. Hummel, R. Koper, Personal recommender systems for learners in lifelong learning networks: the requirements, techniques and model, International Journal of Learning Technology 3 (4) (2008) 404–423.

[30] E. Tacchini, Serendipitous mentorship in music recommender systems. ph.d. thesis., Ph.D. thesis, Computer Science Ph.D. School - Università degli Studi di Milano (2012).

[31] T. Bogers, A. van den Bosch, Collaborative and Content-based Filtering for Item Recommendation on Social Bookmarking Websites, in: Proceedings

of the ACM RecSys'09 Workshop on Recommender Systems & the Social Web, New-York, NY, USA, 2009, pp. 9–16.

[32] F. Wang, M. Hannafin, Design-based research and technology-enhanced learning environments, Educational Technology Research and Development 53 (4) (2005) 5–23.

[33] N. Manouselis, H. Drachsler, R. Vuorikari, H. Hummel, R. Koper, Recommender systems in technology enhanced learning, Springer, 2011, pp. 387–415.

[34] H. Ogata, Y. Yano, et al., Combining knowledge awareness and information filtering in an open-ended collaborative learning environment, International Journal of Artificial Intelligence in Education (IJAIED) 11 (2000) 33–46.

[35] R. Sinha, K. Swearingen, The role of transparency in recommender systems, in: CHI'02 extended abstracts on Human factors in computing systems, ACM, 2002, pp. 830–831.

[36] G. Adomavicius, R. Sankaranarayanan, S. Sen, A. Tuzhilin, Incorporating contextual information in recommender systems using a multidimensional approach, ACM Transactions on Information Systems (TOIS) 23 (1) (2005) 103–145.

[37] P. Symeonidis, A. Nanopoulos, Y. Manolopoulos, Providing justifications in recommender systems, Systems, Man and Cybernetics, Part A: Systems and Humans, IEEE Transactions on 38 (6) (2008) 1262–1272.

[38] J. Bobadilla, F. Serradilla, A. Hernando, et al., Collaborative filtering adapted to recommender systems of e-learning, Knowledge-Based Systems 22 (4) (2009) 261–265.

[39] G. Shani, A. Gunawardana, Evaluating recommendation systems, in: Recommender systems handbook, Springer, 2011, pp. 257–297.

[40] K. Verbert, H. Drachsler, N. Manouselis, M. Wolpers, R. Vuorikari, E. Duval, Dataset-driven research for improving recommender systems for learning, in: Proceedings of the 1st International Conference on Learning Analytics and Knowledge (LAK 2011), 2011, pp. 44–53.

[41] A. Cristea, F. Ghali, Towards adaptation in e-learning 2.0, New Rev. Hypermedia Multimedia 17 (2) (2011) 199–238. doi:10.1080/13614568.2010.541289.
URL http://dx.doi.org/10.1080/13614568.2010.541289

[42] G. Desanctis, R. B. Gallupe, A foundation for the study of group decision support systems, Manage. Sci. 33 (5) (1987) 589–609. doi:10.1287/mnsc.33.5.589.
URL http://dx.doi.org/10.1287/mnsc.33.5.589

[43] J. Grudin, Computer-supported cooperative work: History and focus, Computer 27 (5) (1994) 19–26. doi:10.1109/2.291294.
URL http://dx.doi.org/10.1109/2.291294

[44] G.-j. D. Vreede, R. O. Briggs, Journal of the association for information systems collaboration engineering : Foundations and opportunities : Editorial, Journal of the Association 10 (March 2009) (2009) 121–137.

[45] G. De Vreede, R. Davison, R. Briggs, How a silver bullet may lose its shine, Communications of the ACM 46 (8) (2003) 96–101.

[46] J. Hendler, N. Shadbolt, W. Hall, T. Berners-Lee, D. Weitzner, Web science: an interdisciplinary approach to understanding the web, Commun. ACM 51 (7) (2008) 60–69. doi:http://doi.acm.org/10.1145/1364782.1364798.

[47] B. Shneiderman, Web science: a provocative invitation to computer science, Commun. ACM 50 (6) (2007) 25–27. doi:http://doi.acm.org/10.1145/1247001.1247022.

[48] M. Eppler, J. Mengis, The concept of information overload: A review of literature from organization science, accounting, marketing, mis, and related disciplines., Information Society 20 (5) (2004) 325–344.
URL http://search.ebscohost.com/login.aspx?direct=true&db=aph&AN=15059847&site=ehost-live

[49] S. Minto, G. Murphy, Recommending emergent teams, in: Fourth International Workshop on Mining Software Repositories, MSR2007, IEEE, 2007, pp. 5–5.

[50] A. M. Rashid, K. Ling, R. D. Tassone, P. Resnick, R. Kraut, J. Riedl, Motivating participation by displaying the value of contribution, in: Proceedings of the SIGCHI conference on Human Factors in computing systems, CHI '06, ACM, New York, NY, USA, 2006, pp. 955–958. doi:10.1145/1124772.1124915.
URL http://doi.acm.org/10.1145/1124772.1124915

[51] J. Masthoff, A. Gatt, In pursuit of satisfaction and the prevention of embarrassment: affective state in group recommender systems, User Modeling and User-Adapted Interaction 16 (3) (2006) 281–319.

[52] S. M. McNee, J. Riedl, J. A. Konstan, Making recommendations better: an analytic model for human-recommender interaction, in: CHI'06 extended abstracts on Human factors in computing systems, ACM, 2006, pp. 1103–1108.

[53] A. Meier, H. Spada, N. Rummel, A rating scheme for assessing the quality of computer-supported collaboration processes, International Journal of Computer-Supported Collaborative Learning 2 (1) (2007) 63–86.

[54] M. Guitert, Time management in virtual collaborative learning: the case of the universitat oberta de catalunya (uoc), eLearn Center Research Paper Series (2011) 5–16.

[55] L. Zhu, I. Rossano Vaghi, Mikiwiki: A meta-design framework for collaboration, in: 2011 International Conference on Collaboration Technologies and Systems (CTS), IEEE, 2011, pp. 109–116.

[56] R. Sezov, Liferay in Action: The Official Guide to Liferay Portal Development, Manning, 2012.

[57] D. J. Sheskin, Handbook of Parametric and Nonparametric Statistical Procedures, 4th Edition, Chapman & Hall/CRC, 2007.

[58] P. E. Greenwood, A guide to chi-squared testing, Vol. 280, John Wiley & Sons, 1996.

[59] A. Pentland, The new science of building great teams, Harvard Business Review 90 (4) (2012) 60–69.

[60] G. J. de Vreede, R. O. Briggs, A. P. Massey, Collaboration engineering: foundations and opportunities: editorial to the special issue on the journal of the association of information systems, Journal of the Association for Information Systems 10 (3) (2009) 7.

[61] S. M. Fiore, K. A. Smith-Jentsch, E. Salas, N. Warner, M. Letsky, Towards an understanding of macrocognition in teams: developing and defining complex collaborative processes and products, Theoretical Issues in Ergonomics Science 11 (4) (2010) 250–271.

[62] E. Salas, N. J. Cooke, J. C. Gorman, The science of team performance: progress and the need for more..., Human Factors: The Journal of the Human Factors and Ergonomics Society.