Collaborations in Open Learning Environments

Citation for published version (APA):

Spoelstra, H. (2015). *Collaborations in Open Learning Environments: Team Formation for Project-based Learning*. Open Universiteit.

Document status and date:
Published: 24/04/2015

Document Version:
Publisher's PDF, also known as Version of record

Collaborations in Open Learning Environments

Team Formation for Project-based Learning

Howard Spoelstra

SIKS Dissertation Series No. 2015-04

ISBN/EAN: 978-94-91825-89-7

Cover design and illustrations by Paquita Perez Salgado and Howard Spoelstra
Printed by Datawyse, Maastricht, The Netherlands
Front page and back page Catweazle images courtesy of Paul Pert at The Cult Screen Archive: http://www.paulpert..com

Collaborations in Open Learning Environments

Team Formation for Project-based Learning

Proefschrift

ter verkrijging van de graad van doctor
aan de Open Universiteit
op gezag van de rector magnificus
prof. mr. A. Oskamp
ten overstaan van een door het
College voor de promoties ingestelde commissie
in het openbaar te verdedigen

op vrijdag 24 april 2015 te Heerlen
om 16.00 uur precies

door

Howard Anna Frederik Spoelstra
Geboren op 7 juni 1961 te Heerlen

Promotor
Prof. dr. P.B. Sloep
Open University of the Netherlands

Co-promotor
Dr. P.M. van Rosmalen
Open University of the Netherlands

Overige leden van de beoordelingscommissie
Prof. dr. C. Delgado Kloos
University Carlos III of Madrid, Spain

Prof. dr. Ph. Dessus
University Pierre Mendès of Grenoble, France

Prof. dr. M. Valcke
Ghent University, Belgium

Prof. dr. A.P.J. van den Bosch
Radboud University Nijmegen, The Netherlands

Prof. dr. J.E.M.M. Syroit
Open University of the Netherlands

To Annie, Folkert, and Paquita

CONTENTS

CHAPTER 1

Collaborations in Open Learning Environments

1.1. Introduction

The development of the Internet has changed the educational landscape dramatically, for both initial (compulsory) and further education. Learners can consult a myriad of web-based, often interactive, sources of knowledge, such as Wikipedia, YouTube, specialised fora and webpages, blogs, etc. Due to the social aspects of the Internet, learners who are so inclined, are also enabled to get into contact with peers. More specifically, nowadays learners can partake in online open learning environments, such as course-based Massive Open Online Courses (MOOCs) and open-ended Social Learning Networks. Due to their openness, everyone (with a computer and Internet connection) can join and benefit from the knowledge they have to offer, and many learners do, particularly professionals who seek continuous development opportunities. A recent study into the learner demographics of 43 MOOC-based courses (Christensen, Steinmetz, Alcorn, Bennett, Woods, & Emanuel, 2013) demonstrates that about 80% of these learners have a bachelor's degree or higher, while over 40% are under 30, and almost 50% are between 30 and 60. Most of them are male, and are fully (self) employed. Almost halve of these learners subscribe to satisfy their curiosity, while over 40% aim to gain skills to do their jobs better. This indicates that indeed MOOC learners endeavour to further their professional development. As Charles (2014) argues, such learners are best served with collaborative constructivist learning settings. Online learning facilities should therefore be designed to support close collaboration and social connection, because online learners can become isolated (Jones, Ferreday, & Hodgson, 2008, Kester & Sloep, 2009) and lose motivation (Kim, 2009). They should do the more so because collaborative learning in itself provides additional benefits over individual learning.

1.2. Benefits of collaborative learning

Collaborative learning can pride itself on a long list of affordances (Springer, Stanne, & Donovan, 1999; Felder, Felder, & Dietz, 1999; Johnson, Johnson, Stanne, & Garibaldi, 1990; Marin-Garcia & Lloret, 2008; Dahms & Stentoft, 2008; Fisher & Baird, 2005; Alvarez & López, 2010; Johnson & Johnson, 1999). Collaborative learning:

- increases motivation so that learners are more inclined to deal with hard, complex problems and spend more time studying,
- generates higher levels of academic performance, as individual and group learning processes feed-back into each other,
- improves retention of the content learned,
- fosters critical thinking,
- increases the diversity of the knowledge and experiences being acquired,
- creates realistic and inter-professional learning experiences,

- improves learner retention, which is beneficial with respect to the reportedly large learner drop-out rates from open learning environments.

How then is collaborative learning currently supported in open learning environments? Is support available, and if so, does it answer the question about how to find the right information and the right peers, thereby providing effective collaborative learning opportunities? We briefly discuss some of the collaboration supporting aspects in the designs of current open learning environments and some criticisms from educational designers and researchers.

1.3. Collaborative learning in open learning environments

Open learning environments come in various guises. We address three flavours, based on the commonly made distinction between xMOOCs, cMOOCs, and Social Learning Networks.

1.3.1. xMOOCs

xMOOCs are course-based open learning environments that currently draw most attention and learners often have adopted instructivism-type pedagogies. These teacher-centred approaches are primarily directed at the individual learner and often take the form of watching a video lecture and then answering some quiz questions about the lecture. Thereby collaboration is often supported by electronic fora aimed at peers supporting each other.

However, these fora become so crowded that learners can't find their way to collaborate. As McGuire (2013) puts it: *"Ironically, the biggest obstacle preventing MOOC students from forming relationships is the feature most relied on to encourage them. Discussion forums are the number one complaint."* Some design initiatives allow learners to form groups by suggesting peers based on simplistic criteria such as proximity of location or common language (Coursera, 2014). In general however, as Stacey (2013) remarks, these MOOCs: *"…are focused on objectivist and behaviourist methods of teaching and learning. Their pedagogy is based on an assumption that when there are tens of thousands of learners social learning isn't feasible."* Stacey goes on to note that: *"Students tend to find online behaviourist and objectivist learning pedagogies boring, impersonal, and not interactive or engaging".*

1.3.2. cMOOCs

This stands in contrast with the early course-based open learning environments, which explicitly expected learners to connect to each other and learn collaboratively (referred to as 'cMOOCs', Siemens, 2004; Downes, 2006). They employ some innovative pedagogical elements, sometimes referred to as "connectivist" pedagogy (Siemens, 2004; Verhagen, 2006; Kop & Hill, 2008). This pedagogy puts an emphasis on learner self-direction, collaboration and collective creation of knowledge and artefacts with tools of the learners' choice. Cormier, Siemens, Downes and Kop (2010) describe a course in a connectivism-based environment thusly: *"… is an unusual course. It does not consist of a body of content you are supposed to remember. Rather, the learning in the course results from the activities you undertake, and will be different for each person. In addition, this course is not conducted in a single place or environment. It is distributed across the web. We will provide some facilities. But we expect your activities to take place all over the internet. We will ask you to visit other people's web pages, and even to create some of your own."*

However, it seems not all learners were prepared or able to learn in these settings. Already in 2011, Kop, Fornier and Mak (2011) report on learner experiences in these MOOCs: *"Many participants realized the importance of connections with other learners and of relationship building to advance learning. However, in a MOOC, they found these things extremely hard."*

1.3.3. Social Learning Networks

Social Learning Networks (SLNs) provide opportunities for self-directed continuous professional development. They emerge both inside and across knowledge domains, using tools such as intranet or internet-based fora, etc. In them, professionals can gather information, form interpersonal links, create, and share knowledge (Koper & Sloep, 2002; Steeples & Jones, 2002; Goodyear, Banks, Hodgson, & McConnell, 2004; Sloep, Van der Klink, Brouns, Van Bruggen, & Didderen, 2011b; Rajagopal, 2013). According to Knowles (1975), self-directed learning occurs when learners themselves take responsibility for identifying learning needs, to develop learning goals, prepare a learning plan, locate learning resources and implement the plan, and afterwards evaluate the results and the process. However, not all learners score high on self-direction readiness scales (Guglielmino, 2013). Therefore, particularly related to providing support to SLN learners, much research has been carried out to develop tools for e.g., mutual recommendation of learning materials (Drachsler, Hummel & Koper, 2008), receiving peer-support (Van Rosmalen, Sloep, Kester, Brouns, De Croock, Pannekeet, et al., 2008a), to support community development (Kester & Sloep, 2009), to learn how to present yourself to promote trust between learners (Rusman, Van Bruggen, Sloep, Valcke, & Koper, 2012), and for providing social recommenders (Fazeli, Brouns, Drachsler, & Sloep, 2012). Nowadays, professionals can augment their networks by using social networks (SNs) such as Facebook, Google +, YouTube, and Linked-in. These provide rich additional sources of knowledge, social communication and sometimes also collaboration facilities.

However, Alvarez and Olivera-Smith (2013) remark about learning in SNs that: *"…on their own* [they] *are not learning environments per se, but they afford ample and potentially effective opportunities to improve student learning."* In these networks, learners do not have teachers or tutors (they are their own peer-teachers and peer-tutors), and are assumed to take responsibility for their own learning. Again, Alvarez and Olivera-Smith (2013), remark: *"… there is also a danger that, due to the vastness of resources available in the web, students may find themselves drifting in an "information ocean", straining to solve ill-structured problems with little idea of what concepts, rules and principles are required for the solution or of how to organise themselves and what is the best solution"*

From these observations a general picture with respect to support for collaborative learning emerges: i) In xMOOCs, collaborative learning receives limited attention; ii) In cMOOCs, the ample collaboration opportunities and ill-defined structure of the tasks can lead to learning settings in which learners get lost; iii) In Social Learning Networks, the wide range of (learning) materials makes it difficult for the learner to effectively define learning goals and find appropriate learning materials; iv) In all environments it appears that finding effective teams of peer learners (in contrast to randomly assembled groups) is not well supported. While initiatives are undertaken to remedy some of these issues (see e.g., the NovoEd MOOC environment (NovoEd, 2014), in which learners can receive recommendations for peer learners based on simplistic criteria, such as proximity of location or common language), we observe that none of the open learning environment's designs has ventured to select and implement a collaborative pedagogy in these environments.

1.4. Support for collaborative learning in open learning environments

In open learning environments collaborative learning processes can take form as suggested by e.g., Stahl (2006). Stahl's framework, in a cyclic process, describes phases in which individuals express learning goals, collaborate with peers and use and create learning materials, which are then again used to learn from. However, it leaves open issues with respect to a befitting ecology of learning. As such, the framework assumes that i) learner problem statements are related to the environments in which they are made, ii) collaboration takes place between suitable knowledgeable peers, iii) knowledge sources are available that fit the learners needs, iv) the interactions between learners are structured, not fleeting and shallow. For collaborative learning to be effective, one needs to make sure the processes actually take place, and not be left to chance. As Stahl (2013), notes: "*Group cognition… needs appropriate CSCL technologies, group methods, pedagogy and guidance to structure and support groups to effectively build knowledge that can be shown to be a group product not reducible to individual mental representations*". We propose that one flavour of collaborative learning pedagogy can fulfil these needs: *project-based learning* (PBL; Blumenfeld, Soloway, Marx, Krajcik, Guzdial, & Palincsar, 1991; Davies, de Graaff & Kolmos, 2011; Kolmos, 2012). In setting up effective PBL, teams of peers are formed fit to execute the learning projects. These projects define learning goals related to learning materials, the organisation and duration of activities, and lead to a joint product. As in current professional working conditions activities are often organised in projects, this pedagogy can almost seamlessly blend in with the experiences many learners already have. Our proposition is supported further by e.g., examples from formal education for oncoming professionals demonstrating its effectivity (Westera & Sloep, 1998). Research by Bell (2010) considers PBL to provide skills for 21st-century learning, thereby addressing a societal need. It is therefore that we aim to develop support for small-scale project-based collaborative learning settings inside large-scale open learning environments. However, as staff is short and learners are many, it is difficult to provide support for designing projects and forming effective teams (Wiley, 2004). We therefore research instruments designed *to automate the processes related to team formation for project-based learning in open learning environments*.

1.5. Toward team formation for project-based learning in open learning environments

This thesis thus presents our research into how to automate the process of team formation for project-based learning for use in open learning environments. As indicated above, for effective PBL to happen in open learning environments, projects need to be defined that relate to the environment (domain) and teams need to be formed that from their inception can perform well. This approach is supported by research from the field of computer-supported collaborative learning (CSCL), which has long since shown that collaborative designs for online learning should pay attention to the characteristics of the learner, the formation of the team, the structure of the task, and scaffolding (See e.g., Valcke, 2009). Particularly related to forming teams fit for their task, research indicates that team formation and PBL experts need to take into account factors such as the individual learner's abilities, personality traits, the curriculum area, the team size, and the project task at hand when setting up PBL and forming teams (Graf & Bekele, 2006; Martin & Paredes, 2004; Wilkinson & Fung, 2002). Based on such findings an *initial* model for team formation for PBL was defined. This version of the model is depicted in Figure 1.1.

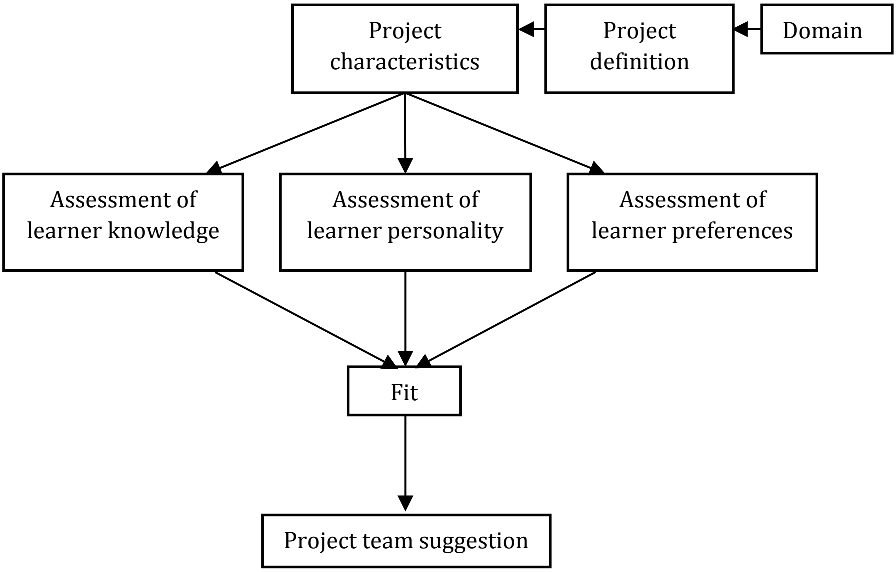


Figure 1.1. *The initial team formation for project-based learning model.*

1.6. How this thesis is organised

In Chapter 2 we provide the rationale behind our research: We discuss several shortcomings in current open learning environments with respect to supporting effective learner collaborations, and argue for introducing project-based learning as a solution to these shortcomings. Theoretical backgrounds of team formation for project-based learning are presented. From these we distil three categories of data (knowledge, personality and preferences) required to form teams, and develop principles for the formation of teams that can effectively aim at different outcomes of the team work process: fostering learning, enhancing creativity, or productively solving a project problem. From contrasting the setup of PBL in formal education with the specific circumstances one encounters in open learning environments, the question arises how the model can be effectuated in open learning environments. Therefore, the central research question addressed in Chapter 2 is:

- *Which principles and processes underlie the introduction of project-based learning and team formation in open learning environments, given the specific characteristics of open learning environments and their users?*

In Chapter 3 we go into the design of services that are capable of expertly performing the team formation task. We examine existing team formation support tools from various domains. From this examination we conclude that these are not well-suited for use in open learning environments, because open learning environments do not necessarily provide the data that are required for these tools to function. This observation leads to the research question addressed in Chapter 3:

- *How can one design a team formation service for project-based learning in social learning networks that optimises either learning outcomes, creative outcomes or productive team performance outcomes?*

In Chapter 4 we build on the team formation service instruments we defined in Chapter 3. We set out to validate both the team formation principles for the formation of productive, creative and learning teams, and the outcomes of the algorithms in which we implemented them. In Chapter 4, we therefore address the following research questions:

- *Are the team formation principles for forming productive, creative and learning teams in alignment with the opinions and experiences of practitioners from the educational field about how such teams should be formed?*
- *Are the results from the computer algorithms in alignment with the results of practitioners from the educational field performing the same task?*

In Chapter 5 we present our research into designing the remaining building blocks required to implement the team formation for PBL model. We report on the results of an empirical study into implementing the model, tailored to the formation of learning teams. The study makes use of latent semantic analysis technology (LSA, Landauer, Foltz, & Laham, 1998) to implement the assessments connected to the knowledge-related elements of the model. The research questions express the integrative nature of this chapter:

- *Using LSA, can we construct a knowledge domain in such a way that we can adequately determine the level of fit of projects to the domain?*
- *Can we adequately determine the extent to which prospective team members have different levels of prior knowledge?*
- *Can we determine a knowledge difference between learners ('zone of proximal development') at which learning is most effective?*
- *How does the personality factor 'conscientiousness' impact on learning and the interaction process between learners?*
- *Can we suggest learning materials from inside the knowledge domain to learners in such a way that learners consider these materials relevant to the project they work on?*

In the final Chapter 6 we review the results achieved and discuss methodological issues and study limitations. We describe the research valorisation opportunities and provide directions for future research.

CHAPTER 2

Toward Project-based Learning and Team Formation in Open Learning Environments[1]

[1] This chapter (with minor changes in terminology and lay-out) was previously published as: Spoelstra, H., Van Rosmalen, P., & Sloep, P.B. (2014). Toward Project-based Learning and Team Formation in Open Learning Environments. *Journal of Universal Computer Science, 20*(1), 57–76.

Abstract

Open Learning Environments, MOOCs, as well as Social Learning Networks, embody a new approach to learning. Although both emphasise interactive participation, somewhat surprisingly, they do not readily support bond creating and motivating collaborative learning opportunities. Providing project-based learning and team formation services in Open Learning Environment can overcome these shortcomings. The differences between Open Learning Environments and formal learning settings, in particular with respect to scale and the amount and types of data available on the learners, suggest the development of automated services for the initiation of project-based learning and team formation. Based on current theory on project-based learning and team formation, a team formation process model is presented for the initiation of projects and team formation. The data it uses is classified into the categories “knowledge”, “personality” and “preferences”. By varying the required levels of inter-member fit on knowledge and personality, the team formation process can favour different teamwork outcomes, such as facilitating learning, creative problem solving or enhancing productivity. The approach receives support from a field survey. The survey also revealed that in every-day teaching practice in project-based learning settings team formation theory is little used and that project team formation is often left to learner self-selection. Furthermore, it shows that the data classification we present is valued differently in literature than in daily practice. The opportunity to favour different team outcomes is highly appreciated, in particular with respect to facilitating learning. The conclusions demonstrate that overall support is gained for the suggested approach to project-based learning and team formation and the development of a concomitant automated service.

2.1. Introduction

More and more, learning takes place in open learning environments (OLEs) with geographically dispersed learners, such as Open Online Courses (OOCs) and their large-scale variants called MOOCs (Massive Open Online Courses). However, recent reports reveal that learning in MOOCs has its drawbacks: Dropout rates are massive, while the intended collaboration between learners is limited (Daniel, 2012; Edinburgh University, 2013). The following factors may contribute to this:

- Learners overestimating their abilities (learners subscribing who would otherwise not be allowed to do the course on the particular level),
- The novelty of the offerings (attracting subscribers who are mainly interested in the workings of the OOC), or
- Learners finding out during the course that they are not willing or able to commit to the course regime (learners and others subscribing who are not sufficiently motivated to follow through the course).

However, we argue that the dropout and limited collaboration might also – at least partly – be explained by a lack of motivating learning opportunities based on well-founded pedagogics. When OOCs were first suggested by Downes (2006) and Siemens (2004), they were based on the pedagogical vantage point of networked learning ("connectivism" is the term they coined to label these OOC settings), emphasizing learner self-direction and contribution. These ideas in themselves were not new, as e.g., work by Westera (1998), Wellman, Salaff, Dimitrova, Garton, Gulia and Haythornthwaite (1996), and Steeples and Jones (2002) already described learning settings based on such principles. But perhaps due to the success of MOOCs and the burden they place on teaching staff when supporting such large-scale settings, this vantage point has been abandoned. Current MOOCs have become learning environments primarily based on behaviouristic pedagogy, in contrast with the OOCs envisioned by Downes and Siemens (Daniel, 2012). However, an OLE may also be too open: Learners' self-direction suggests that learners are able to identify their learning needs and the resources required to fulfil these needs and have strategies to learn and assess their progress towards these needs. In such cases OOCs could consider to offer assistance, by providing e.g., an entry test or a trial lesson, enabling the learners to better estimate their abilities with regard to what the OOC requires of them.
In an effort to re-establish the earlier networked learning foundations in OOC-based learning, we argue that the results from earlier and on-going research into the development of Social Learning Networks (Koper & Sloep, 2002) can help understand the problems OOC learners and teachers face and can help to overcome motivational and dropout problems. These Social Learning Networks (SLNs) are defined as computer-supported, partially overlapping ensembles of *communities* of learners, in which support is provided for learning, sharing and developing

knowledge, with the help of technology (Sloep, Berlanga, Greller, Stoyanov, Retalis, & Van der Klink, 2011a). SLNs aim at supporting potentially large groups of distributed self-directed learners who can work and learn collaboratively in projects (for e.g., innovation, research or assignments), set up working groups, communities, discussions or conferences to acquire competences (Koper, 2009; Sloep & Berlanga, 2011c).
Two important observations about SLN learners' characteristics should be kept in mind:

- That groups of self-directed learners initially have only weak links between them: The learners have limited knowledge about other learners (Jones, Ferreday, & Hodgson, 2008), and
- That these learners can suffer from a lack in continuous self-motivation (Kim, 2009).

Therefore research into Social Learning Networks design focussed on various support methods to improve both the coherence between learners (in order to build up the network between learners) and the motivation of the learners (in order to retain learners and to improve learning outcomes). Designs have been developed, ranging from recommending resources to each other (Drachsler, Hummel & Koper, 2008), doing small activities together to get acquainted based on peer-support (Van Rosmalen et al., 2008a), to learning how to present yourself in the network to promote trust (Rusman et al., 2012). These designs, and others, have been successfully tested within the context of Social Learning Networks and are likely also applicable to MOOCS.
We suggest that performing collaborative learning activities together is another excellent opportunity to motivate learners and to change and anchor (loose) relationships (See Textbox 2.1 for an example of how we envision this to happen in OLEs). So far, however, this opportunity has only rarely been explored in SLNs and MOOCS. Two well-known collaborative learning strategies are problem-based learning and project-based learning. Both support collaboration, but the former primarily focuses on supporting the collaborative process (in particular on problem solving strategies), while the latter primarily focuses on supporting the creation of a collaborative product. Most learners, especially lifelong learners, will already have experience in collaborating in projects as current education habitually provides these settings, whereas problem-based learning is less common. Additionally, problem-based learning, with its focus on acquiring problem solving strategies, may require different criteria for team formation. In this chapter we will therefore concentrate on project-based learning.
Project-based learning (PBL) should fit very well to the issues outlined above. Literature lists several benefits to be had from PBL. They include improving the learners' motivation, so that learners are more inclined to deal with hard, complex

problems and spend more time studying (Johnson, Johnson, Stanne, & Garibaldi, 1990; Marin-Garcia & Lloret, 2008). Other benefits of PBL are found in the blend of learning and working and the realistic (inter-professional) learning experience (Springer, Stanne, & Donovan, 1999; Felder, Felder, & Dietz, 1999), which prepares learners for real life working conditions. Collaborative learning, when compared to individual learning, is also shown to lead to an increase in learning outcomes (Hsiung, 2010).
Therefore introducing PBL opportunities into OLEs, such as SLNs and MOOCS, would address the points mentioned above: It builds links between learners that learn together (which might enable the transformation of these loosely coupled learners into e.g., communities of practice (COPs) (Lave & Wenger, 1991; Sloep, 2013), and it provides motivating learning settings. It will introduce a well-known learning paradigm which fits into the original networked learning pedagogies for OLEs. As do the other support designs discussed above, supporting PBL has the potential to contribute to solving part of the problems outlined.

May 2013: Emma recently started her new job at the microelectronics department. For the first two months her main task was to strengthen her knowledge in this domain. She decided to follow a highly recommended MOOC course because this MOOC, besides the regular lectures and other materials, also contained a 4-week collaborative project work period. She fondly remembered studying in projects during her initial education and the relationships they helped to build. In the MOOC, the projects were presented on a “project wall”, offering the opportunity to apply. The project assignments varied between standard projects proposed by the MOOC, to projects defined by peer-learners, companies and research institutes. The application process followed an automated, open procedure to select the best applicants. Emma selected an interdisciplinary project on biochip design, which was to be performed by at least 4 persons. She could apply by sending in a brief summary of around 100 words on her knowledge and skills with regard to a list of topics address by the project, by filling out her preferences in a profile (on her preferred collaboration language, availability schedule, etc.) and by taking a personality test. Emma decided to give it a try and went through the intake procedure. A few days later she received an invitation to participate in the project and contacted her fellow project members to make arrangements.

Textbox 2.1. Project-based learning in an OLE.

However well PBL seems to fit the networked learning educational aims of OLEs, the introduction of PBL in an OLE is *not* a straightforward operation. This can become clear from the following comparison between the possibilities of setting up project-based learning in formal, teacher-led educational settings versus doing so in OLEs.

In formal education:

- A teacher likely will have the task to define projects that fit inside the formal educational curriculum,
- A teacher will be responsible for the formation of the project teams,
- A teacher can rely on personal knowledge about the learners and/or data sources (grades, prior courses) from e.g., a Learning Management System (LMS) to form teams,
- The learners learn in relatively small cohorts. These cohorts mostly show coherence with respect to place, time and collective progress in the curriculum and commit themselves to the formal educational regime.

In OLE learning settings, due to issues of scale and openness:

- A teacher will not be able to provide a sufficient amount of project proposals to accommodate all learners,
- A teacher will not be able to effectively form teams,
- Participation in an OLE does not amount to the same data gathered from the learner as in a formal educational setting. Therefore the data required to form effective teams are most likely not available. Furthermore, as learners can and do drop out of OLEs, the data that is available will often be incomplete or erroneous,
- Learners in OLEs can have a wide variety of knowledge backgrounds,
- Learners in OLEs originate from over the world, carrying with them characteristics such as language preferences, time zones, agendas, etc.

We therefore suggest to design an *automated project-based learning and team formation support service* which takes into account the OLE learning settings.

Since there is ample research on team formation principles for staffing projects from multiple disciplines (from education, human resource management, etc.), we take that research as a starting point for the introduction of PBL and team formation suitable for OLEs. The design will need to be able to accommodate:

- The number of learners,
- The burden on teachers for providing projects and forming teams in OLE settings,
- The characteristics of the OLE learners,
- The learners' probable lack of knowledge of effective team formation.

To address these issues, the design starts from the considerations that in OLE-based settings:

- Learners should be enabled to start projects, so teachers don't have to define them,
- That these projects are not necessarily positioned in well-defined curricula,
- That projects are staffed by learners who can have a wide variety of knowledge backgrounds and project-related preferences,
- That for an automated team formation process to be effective it should be based on current theory and practise (thereby mimicking team formation expertise as embodied in teachers).

Therefore, the main research question we address in this chapter is: *Which principles and processes underlie the introduction of project-based learning and team formation in open learning environments, given the specific characteristics of open learning environments and their users?*

This chapter is divided into 8 sections. After this introductory Section 2.1, in Section 2.2 we introduce team formation theory for project-based learning. It addresses which data should be considered when forming teams and how rules could be applied during team formation to form teams fit for a specific task. In Section 2.3 we present a team formation process model, which we derived from the theory examined. In Section 2.4 we describe the method we used to corroborate the team formation model with professional practitioners in project-based learning and team formation. Section 2.5 presents the results obtained. In Section 2.6 we discuss these results, while Section 2.7 draws conclusions. Section 2.8 presents directions for future research.

2.2. Team formation theory for project-based learning

In formal educational settings, organising PBL includes the definition of a project task and the formation of a project team around that task. Oakley, Felder, Brent and Elhajj (2004) and Obaya (1999) found that to form effective teams team formation expertise is required, thus discouraging unsupervised or self-selection-based team formation. When team formation is not based on team formation expertise, its results can be subject to pitfalls. Self-selection, for example, can affect the quality of the project outcome through: a) Team formation around pre-existing friendships, which hampers the exchange of different ideas; b) The tendency of learners with similar abilities to flock together, so strong and weak learners do not mix, thus limiting interactions and preventing weaker learners to learn how stronger learners would tackle problems. The stronger learners would also not benefit from the possibilities to teach their peers, and c) The problems under-represented minorities can experience. For example a woman in computer science can become

isolated in a team, which can lead to non-participation or adoption of a passive role, like the team's secretary. A non-native speaker of some language might become excluded from discussions (Oakley, Felder, Brent, & Elhajj, 2004). These findings are supported by an earlier study (Fiechtner & Davis, 1985), which reported that out of 155 students two-thirds indicated that their worst group work experiences were with self-selected groups, while their best experiences were with teams formed by their teachers.
As we need to define projects and form teams to execute them, we next need to investigate which data are required to perform this process.

2.2.1. Data to take into account when setting up PBL and team formation

Felder and Brent (2007) hold, that for a teacher to form effective teams, the teacher requires data about the prospective team members and the project task. Research by Graf and Bekele (2006), Martín and Paredes (2004), Wilkinson and Fung (2002) and Slavin (1989) provides an overview as to which data should be taken into account when PBL is set up and teams are formed. We present these data in two categories:

a) Knowledge related data: The curriculum area in which the project task will be positioned; the project task, and its characteristics (such as collaboration language, duration and suggested team size) and the individual learner’s abilities and prior learner achievements.
b) Personality related data: The individual learner’s personality traits, and motivational orientation.

Depending on the characteristics of the OLE learners, we might have different ways to gather these data: When the learners are students enrolled at the educational institution offering the OLE, a large part of the data needed might be mined from the educational administrative systems. However, when the OLE-based course primarily attracts external learners these data will have to be gathered from the learners themselves directly or by asking for access, if available, to e.g., their e-portfolio (Penalvo et al., 2012). Next, in order to be able to suggest ways to fit learners to projects, we examine team formation principles.

2.2.2. Fitting learners to a project, each other and possible team work outcomes

Judge and Ferris (1992) and Kristof (1996) consider the process of project team formation to be an optimisation process for finding an optimal fit between a person and team. Werbel and Johnson (2001) and Werbel and Gilliland (1999) qualify the concept of fit as containing complementary elements (providing to the team something which other members lack) in some respects, while containing supplementary elements (sharing something with other members) in other respects. In an example aimed at improving learning in a team, Werbel and Johnson (2001) suggest a rule to form a team for that purpose: a team formed to foster learning should consist of team members that provide complementary fit in knowledge background, but who at the same time show supplementary fit in personality. Teams formed in this way allow their members to learn from each other's different knowledge backgrounds while the team shows high levels of cohesiveness and faster decision making (Muchinsky & Monahan, 1987; Kristof, 1996).

Vygotsky (1978) provides a quantifier for differences in knowledge backgrounds: They should be within the zone of proximal development for learners to be bridgeable. As for the personality aspect, Goldberg (1990) and Jackson (Jackson, Wood, Bogg, Walton, Harms, & Roberts, 2010) consider the personality aspect "conscientiousness" (which measures learner carefulness, thoroughness, sense of responsibility, level of organization, preparedness, inclination to work hard, orientation on achievement, and perseverance) to be the predominant indicator for future success in project work.

However, teamwork can have multiple aims. If the teamwork aim is e.g., to provide a creative solution for a problem, then too much complementary fit in knowledge could lead to a loss of creativity (West, 1997). This suggests that multiple team formation rules can be designed, depending on the envisioned outcome of the project work.

2.3. A project-based learning and team formation process model

In the introduction we set out with the challenge how to provide OLE-learners with motivating, network-strengthening collaborative learning opportunities. We argued that providing support for project-based learning and team formation could answer this question. In this section we therefore introduce a generic process model for the initiation of project-based learning and team formation. In order to be widely applicable, this process model aims to fit into both OLE and formal educational settings. The process model takes into account the team formation theory introduced in the previous section. We first categorise the data to be taken into account when setting up tasks and forming team, as introduced in Section 2.2.1, and then argue that for team formation for OLE-learners, a third category of data is of utmost relevance. Next, we describe how PBL and team formation can be initiated and then we present the process model we derived.

2.3.1. Process model data categories

In Section 2.1, we introduced two categories of data needed to set up project-based learning and perform the project team formation. These were:
a) Knowledge-related data
b) Personality-related data

However, learners in OLEs can carry with them characteristics that distinguish them from regular curriculum-bound learners (such as their geographical distribution, mastering different languages, having jobs and different time schedules, families, etc. (Fetter, Berlanga, & Sloep, 2010). These characteristics can pose practical problems in collaborations, even prohibiting collaboration. So for them, obviously a third category of data is relevant to take into account when forming teams:

c) Preferences-related data (such as preferred collaboration language, availability, time zone, etc.).

2.3.2. The project based learning and team formation process model

In our view, the initiation of PBL and team formation starts with the definition of a project related to (a part of) a knowledge domain (such as a curriculum, a topic in a SLN, or the MOOC topic). Its characteristics are defined (such as preferred duration, team size, etc.). Depending on the level of openness of the OLE, these can be defined by the learner, by the teacher, or partly by both. In order to ensure the appropriateness of the project definition in relation to the domain, a level of fit between the project definition and the domain can be determined. In our related work (Spoelstra, Van Rosmalen, Van de Vrie, Obreza, & Sloep, 2013) we suggest that language technologies can help to assess the overlap between the suggested project definition and the OLE domain. Alternatively, the learners themselves can assist in controlling the quality. PeerScholar, a system for assessing writing assignments, shows that peer assessment can be a valid quality control alternative in large classes (Paré & Joordens, 2008). Furthermore, a framework with clear rubrics could guide both the students who propose or execute a project and the peers who assess the quality of the project definition or (once the project is started) the project's (intermediate) results.
In the next step in the process model, the learner's knowledge, personality and preferences are assessed. The outcomes of these assessments are compared with the requirements the project and characteristics set forth, and the other members' assessments outcomes. This then leads to a measure of fit between the project and the prospective team members. The team formation process ends with a suggestion for a project team when one set of project-suitable members can be found that shows optimum fit (the best fit team solution), or when all members are dispersed over teams (the best possible average solution). Figure 2.1 depicts the process model.

The proposed process model introduces an important feature: By providing the ability to qualify what should be considered a good fit between a project and its members (see Section 2.2.2), it becomes feasible to direct the project formation process outcome towards forming teams that are specifically suited for a particular project aim.

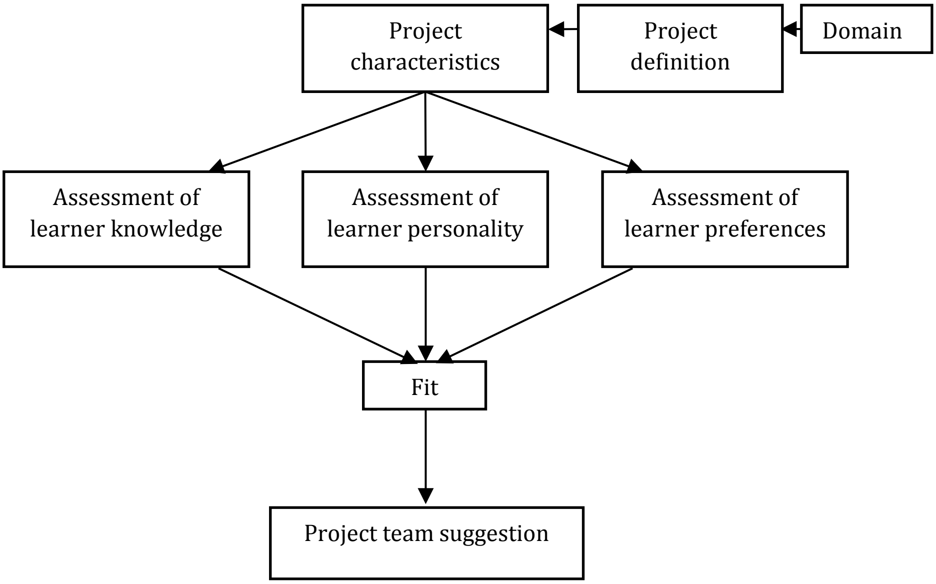


Figure 2.1. *The project-based learning and team formation process model.*

2.3.3. Qualifying fit

By allowing qualification of the fit between learners with respect to knowledge and personality, the process model opens up the opportunity to form teams aimed at different project work outcomes. These qualifications can be described in team formation rules. One such rule was already introduced above: Teams aimed at facilitating learning should be comprised of members with comparable personalities but with different knowledge backgrounds. Other rules might aim at forming teams that are targeted at other well-known project aims, such as creative problem solving, or expertly and productively solving a problem.

The principles outlined in the process model are based on the project-based learning and team formation theory introduced above, and are aimed at project definition and team formation in OLE settings. However, in order make sure we developed a process model which also receives support from project-based learning and team formation practitioners in the formal educational field, we conducted a field survey. The survey existed of interviews and a questionnaire. These had a two-fold aim: Firstly, ascertaining whether the team formation principles identified in Section 2.2 and the process model presented in Section 2.3 are aligned with PBL and team formation practice in formal learning settings, and secondly, to identify how the process model' affordance of differentiating team formations for different project work outcomes is valued. The method applied and the results obtained are presented in the Sections 2.4 and 2.5.

2.4. Method

In order to gain insight into team formation practise and feedback on the team formation process model, a field survey was conducted by means of an open access web-based questionnaire. Its initial setup was discussed in four semi-structured interviews with teachers and designers of project-based learning, who were also team formation practitioners. They worked at three different universities in the Netherlands. The interviews followed a predefined two-part schema. In the first part questions addressed team formation theory, practitioner experience, data used to form teams, team formation methods, and the recognition of team formation risks. The second part asked questions related to the proposed team formation process model with respect to the data categories we discerned, the desired teamwork outcomes, and whether it would be acceptable in practise to use outcomes of an automated team formation tool. In order to ensure a broad range in the experts' backgrounds in team formation, 2 interviewees were chosen from a distance teaching university and 2 interviewees were chosen from regular teaching universities (one technical university and one medical university, with an educational focus on problem-based learning). Of each type of institution, one interviewee primarily worked in supporting PBL on a daily basis and the other interviewee primarily worked in the development of PBL settings. The interviewees worked at three different universities. All had multiple years of experience in either supporting or designing PBL settings.

Following the feedback from the interviews, a web-based questionnaire was created with a list of 30 questions. The questions were split into four parts. The first part inquired into demographics, such as gender and current work place. The second part contained questions addressing team formation data and theory used in practice, the current team formation methods, how respondents dealt with strong and weak learners, differences in learner's background knowledge, learner personalities, minority aspects, and how learners were prepared for team-based activities. In the third part of the questionnaire the questions were related to the proposed team formation process. They addressed the principles of supplementary fit and complementary fit with respect to knowledge and personality, the categorisation of data suggested in Section 2.3, the relative and absolute importance of the categories knowledge, personality and preferences in the team formation process. The relative and absolute importance could be indicated on a 5-point scale (1=not important, 5=very important). Respondents were also asked to indicate the importance of the proposed target outcomes of project-based learning on a 5-point scale (1=not important, 5=very important). The fourth part consisted of two open questions in which the respondents were asked whether and, if any, under which conditions they would accept a team formation suggestion from an automated tool, and whether respondents had general remarks on the suggested team formation approach. The questionnaire could be answered anonymously, and did not force respondents to answer all questions. Before respondents were invited to participate, 2 colleagues at the Open University in the Netherlands tested the questionnaire for intelligibility and logical correctness. Finally, the respondents were invited from international groups working and teaching in project-based learning settings, using an open invitation. The questionnaire was open for responses for 2 months.

2.5. Results

We present the results in accordance to the 4 parts into which the questionnaire was divided.

2.5.1. Demographics and PBL settings

The in total 26 respondents stemmed from 8 different European countries. Of the respondents 29% were female, while 71% were male. 73% worked at a university, 15% worked at a university for professional education, while 4% worked in vocational training. No respondent reported to be working in the private sector. The respondents indicated that their students' team-based activities most often lasted between 3 and 6 months, while the extremes were 1 to 2 weeks and a whole year. The respondents indicated to be mainly active in project-based learning settings (40%) and problem-based learning settings (32%).

2.5.2. Project team formation practice

The respondents most often (41%) reported optimum team sizes to be between 3 and 6. When asked which team formation methods were in use, our respondents reported 12 unique team formation methods in total. Besides 9 teacher-driven team formation methods, they also reported 3 criteria for team formation based on learner self-selection. In order to compare these team formation methods with the data categories we identified above, in Table 2.1 we present the teacher driven team formation methods sorted into these categories. Please note that respondents could select more than one method (and even conflicting ones) since they would not necessarily use the same method in all team formation situations.

Table 2.1. *Teacher driven team formation methods reported, related to the data categories knowledge, personality and preferences.*

Category	Methods
Knowledge related	Group students with the same background knowledge (58%) Mix strong and weak students (27%) Heterogeneity in knowledge background (11%) Group strong students together (8%)
Personality related	Spread learners with similar personalities (21%) Group learners with similar personalities (8%) Group learners belonging to certain minorities (27%), such as form teams with only female members
Preferences related	Check for overlapping calendars (19%)

The respondents reported 3 methods, which were not categorisable: “Learner preferences for specific projects” (42%), “Allow students to self-select teams” (50%), and “Randomly select team members” (37%).
The respondents also reported on activities undertaken to prepare learners for successful teamwork. These were all aimed at preparing for self-selection: “Organizing joint meetings before team formation takes place” (33%), “Pointing to other students' prior track records” (28%), “Pointing to online profiles of other students in social networks” (17%), and “Providing training in giving/receiving feedback and conducting negotiations” (67%).

2.5.3. The proposed team formation process model

The results presented in the tables 2.2 and 2.3 express the respondent’ ratings on the (relative) values of the model’ data categories when used in OLEs. Table 2.4 expresses the respondent’ general opinions on desirability of the different project work outcomes we suggested.
Table 2.2 shows the results of our respondents’ ratings of the overall importance of the individual data categories knowledge, personality and preferences for the team formation process. The combined scores of “rather important” and “very important” on knowledge are 64%, while the combined scores on preferences and personality are 60% and 12%, respectively. This score on personality is surprising, as it stands in contrast with the emphasis team formation theory puts on this category.

Table 2.2. *The importance of the categories knowledge, personality and preferences in the team formation process, on level of importance.*

Importance	Knowledge	Personality	Preferences
Very important	28%	0%	16%
Rather important	36%	12%	44%
Important	28%	24%	28%
Somewhat important	8%	44%	12%
Not important	0%	20%	0%

The respondents also rated the importance of the categories knowledge, personality and preferences in relation to each other. These results are shown in Table 2.3.

Table 2.3. *The importance of the categories knowledge, personality and preferences related to each other.*

Importance	Knowledge versus Personality	Knowledge versus Preferences	Personality versus Preferences
1st most important	8%	8%	4%
1st more important	50%	40%	20%
Equal importance	29%	28%	28%
2nd more important	13%	20%	44%
2nd most important	0%	4%	4%

The importance of the knowledge category was rated above the personality category. The importance of the knowledge category was also rated over the preferences category. The importance of the preferences category was rated over the personality category. This suggests a relative order of importance of the categories: (1) knowledge, (2) preferences and (3) personality in the team formation process. This outcome suggests that for future implementations of the team formation service, the different data categories should be allowed to have different weights in the team formation process. None of the respondents indicated any other category of data to be relevant to the team formation process.
The respondents showed clear views on their preferred target outcomes of teamwork. The combined scores on “Very important” and “Rather important” for the outcome “Improved learning” scored highest with 76%, while the same combined scores for “Enhanced creativity” and “Improved productivity” scored 64% and 48% respectively (see Table 2.4).

Table 2.4. *Preferred target outcomes of project-based activities, on level of importance.*

Importance	Improved learning	Enhanced creativity	Improved productivity
Very important	48%	20%	8%
Rather important	28%	44%	40%
Important	24%	16%	36%
Somewhat important	0%	20%	12%
Not important	0%	0%	4%

2.5.4. Accepting team formation suggestion from a team formation service

In the fourth part of the questionnaire the respondents were asked whether and under which conditions they would accept a team formation suggestion from an automated tool, and whether respondents had general remarks on the suggested team formation approach. Of the 11 responses to the first question, 5 express acceptance of automated team suggestions. Another 5 responses express acceptance with some reservations, while 1 response expresses declination of automated team suggestions. (Some text has been translated from Dutch to English, or edited for reasons of readability.)

Responses expressing acceptance

“It would solve for us a problem with the formation of complete teams.”
“My students sometimes already use a program (written by some of our students as a design exercise) "find study buddy" in which they can vary criteria like location distance or number of same courses taken to find a buddy.”
“Following a suggestion is always better than entering a team completely blank.”
“I would. You don't have to spend time to be on a team. Everybody is in a team. A disadvantage is that a student with negative experience with another student will not accept the result of a computer when this particular student is selected in his team.”

” Yes, it will provide me the possibility to discuss with new people and I’m expecting that it will be a good collaboration.”

Responses expressing conditional acceptance
“I would accept, simply because it takes time to take into account multiple criteria to form teams, and also because it will only be a suggestion.”
“I would accept such a suggestion and also the students would as a basis for further investigation in forming an optimal team.”
“Maybe ... if students already know each other, they might know better with whom they might work better than an automated system. Otherwise, one such system would manage to group them better then they or the teacher can.”
“Suggestions are always helpful by providing a deeper insight into the team selection process. The final decision should be taken by the tutor, but an automatic system could present valuable facts and recommendations (mostly about prior knowledge, previous teams, and other member’s preferences).”
“Yes, if I agree with the selection”.

Response expressing declination
“No, I would like to experience different teams myself, to learn more about different competences. An automatic system leaves no space for experiments in your own personal competence growth.”

General remarks
The respondents gave the following general remarks on the research presented (responses not directly related to the current research have been omitted):
“The suggested approach to team formation seems much more thoughtful when compared to my current practice”.

2.6. Discussion

The primary research question we addressed in this chapter was: “Which principles and processes underlie the introduction of project-based learning and team formation in open learning environments, given the specific characteristics of open learning environments and their users?” Before we present our conclusions we first discuss the results from the interviews and questionnaire:
When we compare the team formation methods reported to be in use in practice with the categories knowledge, personality and preferences, we find that some methods (learner preferences for specific projects, allow students to self-select teams, randomly select team members) cannot be related to the categories suggested, as they relate to learner-self-selection-based team formation. The team formation theory we examined considers the use of such criteria to be detrimental to the quality of the outcomes of teamwork. They are therefore discarded from our design.

Some criteria seem to be in contradiction to each other. In the knowledge category, the criteria “group students with the same background knowledge” and “heterogeneity in knowledge background” and the criteria “mix strong and weak students” and “group strong students together” look contradictory. However, this may well be explained by the respondent’s focus on facilitating learning or enhancing creativity. Contradictions in the personality category can be explained in a similar way.

There is only one mention of a preferences-related criterion (“let students themselves check for overlapping student calendars”), which can be explained by the fact that our respondents work with cohorts of learners in traditional educational settings that show homogeneity in, for example, the preferred language or available time slots for project-based collaborations.

The overall relevance of the categories knowledge, personality and preferences for the team formation process show a low value for personality. Where “knowledge’ receives a joint score of 64% on “important to very important” and “preferences” receives a joint score of 60%, “personality” only scores 12%. This outcome is somewhat surprising given the emphasis team formation theory puts on personality as an import factor in a team formation process. This might be explained from the fact that practitioners in team formation from the educational field might not be able, or do not have the instruments to assess personality easily. This score can also reflect the lack of respondents from the private sector, where tests related to personality aspects are a more mainstream part of e.g., job application procedures.

We found the relative order of importance of categories of data in the team formation process to be: (1) knowledge, (2) preferences and (3) personality. However, fit in preferences indicate “condiciones sine qua non”, as without overlapping preferences collaboration cannot take place. The fit in preferences therefore precedes the fit in knowledge and personality, which are the important factors when forming teams targeted at specific outcomes.

The practitioners put an emphasis on “improving learning” as a desirable project aim. Given that the respondents all stem from a background in education, this can hardly come as a surprise. However, the almost equally strong emphasis on “enhance creativity” suggests that “improve learning” and “enhance creativity” should both be supported target outcomes in a team formation process for educational purposes. The private sector might put more emphasis on “improving productivity” as a desirable outcome. To allow the team formation tool to be used in a wide range of settings, we aim to support all three of the suggested target outcomes.

From the remarks we received about the acceptance of team formation suggestions from an automated tool, we get the distinct impression that such a tool would be welcomed. This welcome is in some respects conditional. Some teachers would like to be able to use the tool for input into a team formation process they can oversee themselves. The reservations mostly apply to traditional educational settings and have less bearing on the possible benefits a tool can have for setting up project-based learning and team formation services in OLEs, where no other support is available. A reservation about the possibility to create a meaningful personality profile is duly noted. However, from the team formation theory we conclude that inclusion of a personality profile improves the team formation process beyond the current practice, also in traditional settings.

2.7. Conclusions

Learners in OLEs have to show continuous self-motivation to learn in relatively anonymity. As this is inherently difficult, the practice-oriented, motivational, and coherence-creating affordances of project-based learning can support these learners. There are, however, significant differences between setting up PBL in formal, teacher-led learning settings and OLE settings, in which teachers play only a small part. In traditional settings practitioners (e.g., teachers) would normally initiate PBL, using their expertise related to learner knowledge and personality, the curriculum and the task to be designed. Due to the number of learners in OLEs, we assume no support will be available to start the PBL and team formation process. Because of the learners' different backgrounds we also assume that not all data will be available that would otherwise be available in formal learning settings. We suggested a solution to these problems by allowing the process of starting PBL and team formation to be carried out by learners themselves. But these learners probably lack the knowledge to perform the team formation process. We therefore need to design support, with learner self-direction and self-organisation in mind. The PBL and team formation theory introduced suggested that data is required on the project and on the learner' knowledge and personality. The OLE context required the inclusion of a third category of data: Preferences. From these categories we constructed a process model, aimed at the introduction of PBL and team formation support for OLE learners. The model describes "fit" as the result of an optimisation process, which matches prospective team members into a team for a specific project. We further introduced team formation rules that can influence the team formation process toward setting the stage for mutual learning and teaching, enhancing the possibility of a creative project outcome or to improve productivity. We expect that the different team formation methods mentioned in the survey results in Section 2.5.2 can be translated into these team formation rules.

The data we identified from theory as playing a role in team formation process largely overlap with the data used by team formation practitioners; therefore we conclude that theory and practise at large are aligned. The exception is the data on personality, which receives more emphasis in theory than in practise. We assume this is due to the inherent difficulty in measuring and taking into account this data in the team formation process. There is a strong tendency to focus merely on knowledge as a general indicator for success, despite studies that indicate that other factors are more predicative of success (Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007). As the inclusion of the category personality gets support from team formation theory, we consider it to be an important factor in team formation. Our respondents find knowledge to be the most important category of data to be used in the team formation process, over preferences and finally personality. This result indicates that being able to give these categories different weights in the team formation process would be an important asset for a team formation service. The respondents express strong support for the possibility to direct team formations toward the outcomes we suggested. They indicated a clear order in which they prefer the different targets; from the most preferred "Improve learning", to "Enhance creativity" to the least preferred "Improve productivity". Nevertheless, all possible outcomes suggested receive high importance rates. From this we conclude that it will be important to also provide OLE learners or teachers with the opportunity to indicate the preferred project aim when they use the envisaged PBL and team formation service.

Our final conclusion is that the question "Which principles and processes underlie the introduction of project-based learning and team formation in open learning environments, given the specific characteristics of open learning environments and their users?" can, in principle, be answered by our process model. The model receives support from the educational field, although support for the use of the category personality is limited. However, as team formation theory values this category highly, we think we should design a PBL and team formation service taking personality into account. This would also provide the opportunity to implement the team formation rules, which depend on variations in both knowledge and personality fit. An implementation of the service including this category can also improve the team formation practice in traditional learning settings, even when it is only used to provide team formation suggestions and leaves the final team formation decision to the expert.

2.8. Future work

In the next step of our research, we will focus on the technical aspects of transforming teacher-based PBL and team formation into an implementation of service-based PBL and team formation. It will take into account the data to be gathered upon which to base a PBL and team formation service, and how these data can be gathered and analysed so they can be mapped to knowledge, personality and preferences. When the learners are students enlisted at an educational institution and follow an official curriculum, some of the data needed could be mined from the educational administrative systems. However, when an OLE primarily attracts self-directing lifelong learners (links to) these data will have to be provided by the learners themselves. Our future work will also address the definitions of fit (expressed in team formation rules) with respect to the different project work outcomes. For the knowledge category we firstly plan to use learner self-reported levels of knowledge on the project task. Later on we envision this method of knowledge assessment will be replaced by a means to relate both project descriptions and learners' project applications (or CV's, e-porfolio's or materials studied earlier) to materials available in the domain (cf. Laham, Bennett, & Landauer, 2000). The learner personality will be assessed with the help of available tests, such as the Big Five test (Barrick & Mount, 1991). The preferences will be based on a learner profile. Each step of the development of the service will be evaluated with students and teachers, taking into account both the ease of use and the quality of team formation advice generated.

CHAPTER 3

A Team Formation and Project-based Learning Support Service for Social Learning Networks[2]

[2] This chapter (with minor changes in terminology and lay-out) was previously published as: Spoelstra, H., Van Rosmalen, P., Van de Vrie, E., Obreza, M., & Sloep, P.B. (2014). A Team Formation and Project-based Learning Support Service for Social Learning Networks. *Journal of Universal Computer Science, 19*(10), 1474–1495.

Abstract

The Internet affords new approaches to learning. Geographically dispersed self-directed learners can learn in computer-supported communities, forming social learning networks. However, self-directed learners can suffer from a lack of continuous motivation. And surprisingly, social learning networks do not readily support effective, coherence-creating and motivating learning settings. It is argued that providing project-based learning opportunities and team formation services can help overcome these shortcomings. A review of existing team formation tools evidences that a new design for team formation and the initiation of project-based learning is required before these can be supported in social learning networks. A design is proposed which identifies 'knowledge', 'personality' and 'preferences' as categories in which data is needed to form teams, and it specifies how the required data are gathered and assessed. The design defines rules deduced from team formation principles from prior team formation research to optimise team formations towards increased productivity, creative solutions or higher learning outcomes. The rules are implemented in three team formation expressions each calculating one of the desired team formations. The expressions are deployed on a set of test data, demonstrating the effectiveness of the team formation service design. The chapter includes a discussion of the results and provides indications for future research.

3.1. Introduction

The 21st century requires new approaches to innovation and learning. More and more, learning takes place in geographically dispersed networks, which we call social learning networks (SLNs). SLNs are defined as computer-supported networks of informal (non-formal) learners. In these networks, people can learn, share and develop knowledge and technology helps them to do so (Sloep, 2009). They aim at supporting potentially large groups of distributed self-directed learners, who can, in their efforts to acquire competences, work and learn collaboratively (e.g., for innovation, research or assignments) or set up working groups, communities, discussions or conferences (Koper & Sloep, 2002; Koper, 2009; Sloep et al., 2011b). However, some of the characteristics of these groups of self-directed learners are that there are only weakly linked (they initially have limited knowledge about other learners) (Jones, Ferreday, & Hodgson, 2008) and that they may find it difficult to remain motivated (Kim, 2009).

There are various ways to improve the coherence and the motivation of learners in a network, ranging from recommending resources to each other (Drachsler, Hummel & Koper, 2008), doing small activities together (Van Rosmalen et al., 2008b), to actively working together (Goodyear, 2005). For SLNs in particular, the introduction of project-based learning (PBL) opportunities should fit very well. It would enable self-directed learners to engage in focused and motivating learning activities in close collaboration with other learners. The benefits of PBL are found in that it improves the learners' motivation, so that learners are more inclined to deal with harder problems and spend more time studying (Johnson, Johnson, Stanne, & Garibaldi, 1990; Marin-Garcia & Lloret, 2008). Furthermore, it blends learning and working and thus creates a realistic (inter-professional) learning experience (Westera & Sloep, 1998; Springer, Stanne, & Donovan, 1999; Felder, Felder, & Dietz, 1999). Recent research by Hsiung (2010) shows that collaborative learning leads to an increase in learning outcomes, when compared to individual learning.

However, introducing PBL in a SLN is not a straightforward operation. In traditional, formal educational settings, teachers have the expertise to define projects that fit in a formal educational curriculum and are responsible for the formation of the project teams. Teachers might rely on personal knowledge about the learners and/or data sources (grades, prior courses taken) from e.g., a Learning Management System (LMS) to form teams. In traditional educational settings, the learners learn in cohorts (with respect to place, time and collective progress in the curriculum) and commit themselves to the formal educational regime. Such an educational context stands in stark contrast with a SLN learning context. In a SLN there is no teacher with curriculum knowledge and team formation expertise. Furthermore, the data as mentioned above required to form teams are not readily available, while the learners exhibit self-directing and self-organising behaviour. The learners most probably do not know each other. When designing PBL and team

formation support for SLNs, we therefore have to consider that in SLNs, projects will be started by a learner (or a stakeholder connected to the network), that these projects are not necessarily positioned in a well-defined curriculum and that prospective team members can have a wide variety of knowledge backgrounds, personalities and project-related preferences.
In earlier work we introduced a team formation process model (see Figure 3.1) (Spoelstra, Van Rosmalen, & Sloep, 2012) for use in SLN contexts. The model describes the assessment of learner *knowledge*, *personality* and *preferences*, in order to determine a *fit*-value for a team of learners for a specific project. We demonstrated that there is support for this approach to PBL and team formation from the educational field. By allowing variations in the strength and weight of the learners' *knowledge* and *personality* in the suggested teams, the model also introduced the ability to direct the team formation process towards specific project outcomes (such as facilitating learning from other team members while solving a project problem, coming up with creative solutions for the project problem, or expertly and productively solving a project problem). These variations in knowledge and personality are defined in *team formation rules*. The collection of learner *preferences*, however, denote 'condiciones sine qua non' for collaboration and thus determine whether a project can take place at all with a particular team of learners. So the preferences serve as constraints on the application of the team formation rules.

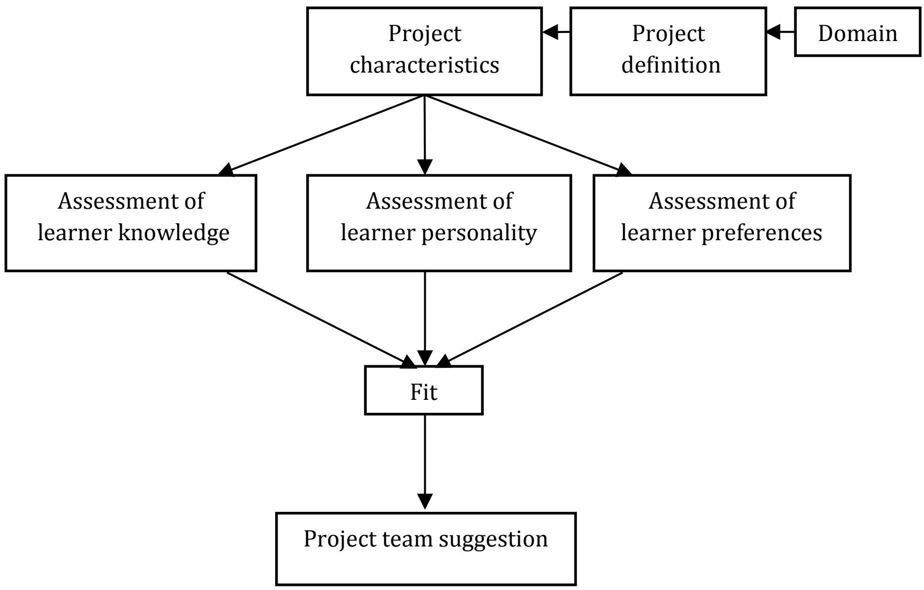


***Figure 3.1**: The model for the team formation process.*

In the design of a PBL and team formation service for use in SLNs we take into account the differences between traditional educational settings and SLN settings as introduced above. As indicated, the most important differences are that there is no team formation expert (teacher) available, that the learners themselves should be enabled to start projects, and that the data used to start PBL and team formation has to be derived from different sources than in traditional learning settings. Therefore, in this chapter we address the following the question: *How can one design a team formation service for project-based learning in social learning networks that optimises either learning outcomes, creative outcomes or productive team performance outcomes?*

The remainder of the chapter is devoted to answering this question. It is organised as follows: Section 3.2 provides an overview of prior research on team formation systems, the approaches to team formation they take, their aims and the data they rely on for forming teams. The section concludes with our assessment of their usefulness in SLN contexts. Section 3.3 presents the design of our PBL and team formation service as well as the principles through which the data that feed into it are gathered. It also presents the definitions of the team formation rules and the formalisation of these rules into formal expressions. The expressions allow the calculation of team compositions from the data gathered. Section 3.4 reports on the outcomes of a team formation exercise using the expressions on a set of simulated data. In Section 3.5 we discuss the outcomes, draw conclusions and indicate directions for future research.

3.2. Existing team formation approaches and systems

Team formation is a very active research area. Initially, this research was started in the human resource management (HRM) domain. However, as learning in teams also is considered to be preparatory for real life working conditions, team formation has also become an important topic in the educational research field. More recently, team formation is also being researched in the social network domain, using social network analysis (SNA) techniques. Team formation can be studied from different perspectives, such as competence, cultural, or personality perspectives. It can be performed for different aims and can be based on a multitude of different kinds of data. It can be studied as a separate entity, but also as being embedded in e.g., the management of international teams. An example of the latter is e.g., the People-Capability-Maturity-Model as applied to the area of Global Software Development (Colomo-Palacios, Casado-Lumbreras, Soto-Acosta, Misra, & Garcia-Penalvo, 2012).

The research outlined above resulted in a variety of team formation systems that are currently available. They use different data and various teaming criteria, support different aims and contexts and sometimes use multiple technologies to team up people. In the subsections 3.2.1 through 3.2.3, we provide a review of which data these systems use to form teams, sorted by the application domains:

Human Resource Management, Social Networks and Education. The question we aim to answer from this review is whether these systems and the data sources they use to form teams for the goals they support can also be used in social learning networks. In subsection 3.2.4 we will argue that these systems all have drawbacks, prohibiting their use in the context of SLNs, thereby further strengthening our case for the design of a new team formation service.

3.2.1. Systems developed for use in Human Resource Management

1. Knowledge and collaboration habits (Wi, Mun, Oh, & Jung, 2009). The system suggested provides grouping based on project keywords and data in knowledge repositories, keyword search in reports, paper, patents, and books. It also uses SNA techniques for finding co-authors of publications.

2. Competences mined from employee publications (Rodrigues, Oliveira, & De Souza, 2005). The proposed system aims at facilitating collaboration and knowledge sharing, dissemination and creation in scientific organizations. Terms from user publications have to be manually connected to competences (and level of mastering) by a user in the role of 'knowledge manager'. After a project manager creates a project model, the system can mine the best suited project members through the required competences.

3. Knowledge, personality and working relationships (Chen & Lin, 2004). From a representation of knowledge, teamwork capability (experience, communication skills, and flexibility in job assignment) and collegiality (using the Myers-Briggs type indicator test), teams are suggested.

3.2.2. Systems developed for use in Social Networks

4. Type of relationship, subject, institution, geographic location, time (Monclar, Oliveira, De Faria, Ventura, de Souza, & Campos, 2011). The analysis aims at discovering emerging groups in Social Networks.

5. Co-authors and related research papers (AFRL, 2006). The research uses co-authorship information to create a network of relations in combination with user concept maps to enable ad-hoc team formation.

6. *References in scientific papers* (Sie, Drachsler, Bitter-Rijpkema, & Sloep, 2012). The proposed system creates a network from user publications, using the measures of "betweenness" and keyword similarity. The system can either recommend authors for future publications, including prior co-authors (to strengthen current bonds between authors and strive for acceptance of a certain research topic), or recommend new co-authors (to foster creativity).

3.2.3. Systems developed for use in Education

7. Gender, nationality, age, previous marks, team role, and learning style (Ounnas, Davis, & Millard, 2009). The authors suggest a system in which the grouping constraints and their strengths are ranked by an instructor, who also sets the project to be staffed. The system aims to increase the satisfaction of grouping constraints and to overcome the orphans' problem (learners not assigned to a team after the team formation process has ended).

8. Learner knowledge related to a task knowledge model represented in learning objects (Pollalis & Mavrommatis, 2009). The system proposed keeps track of learner knowledge and aims to group learners with comparable knowledge backgrounds to the knowledge required to perform a defined task. It is aimed at distance learners but disregards grouping criteria outside 'knowledge' as the authors suggest group formation in distance learning has less use for criteria such as gender, age, nationality, or religion.

9. Creativity score and rating of ideas (Ardaiz-Villanueva, Nicuesa-Chacón, Brene-Artazcoz, Sanz de Acedo Lizarraga, & Sanz de Acedo Baquedano, 2011). The system calculates a creativity value for a user, based on the number and the length of the user provided responses to a generated idea. It uses user ratings given to ideas gathered in a brainstorm, combined with the creativity value, to suggest teams. An instructor can change the team formations. The project topics are already set, as the system works inside a PBL setting.

10. Thinking styles (Wang, Lin, & Sun, 2007). The system proposed is a teacher-based tool, called DIANA. It uses data on psychological variables from questionnaires on thinking styles. It can form heterogeneous groups with respect to these styles.

11. Learner characteristics (Tobar & De Freitas, 2007). The system uses data as defined in IMS LIP (which defines both set data, such as ID, name, address, phone, email, web-address, physical, technical and cognitive characteristics, and variable data, such as goals, learning plans, learning preferences). These data are contained in a learner database and can be used by a teacher to form groups.

12. Knowledge and learning styles (Christodoulopoulos & Papanikolaou, 2007). An instructor can form heterogeneous and homogenous groups from enrolled students, based on 3 criteria (knowledge, and two axis of learning style test results). Learners take a test to determine their learning style. Unfortunately, we could not determine how the authors derived the score on knowledge.

13. Performance in previous work, activity in collaboration (Soh, Khandaker, & Jiang, 2008). A system called I-MINDS can form buddy groups for unstructured collaborations and teams for structured cooperative learning activities. It uses computer-based agents to model the learners or the groups. The user model is gradually filled, based on learner activities. Structured cooperative learning follows a model with a teacher-predefined set of activities.

14. Performance and personality traits (Graf & Bekele, 2006). This research is aimed exclusively at forming heterogeneous groups, based on group work attitude, interest for the subject, achievements motivation, self-confidence, shyness, level of performance in the subject, and fluency in the language of instruction. The data on the users is represented in a vector space.

3.2.4. Assessment of the usability of existing systems and approaches for team formation in SLNs

The above overview of systems, aims and contexts for forming teams also describes what data these systems and approaches use to form teams. It might suggest there is a considerable overlap with the data our approach suggests to use to form teams. There are, however, distinct differences between the aims and implementation contexts in which these systems can be used and the SLN aims and implementation context:

- The systems for use in human resource management (systems 1, 2 and 3) rely on the availability of data in enterprise repositories,
- The systems for use in social networks (systems 4,5 and 6), while not relying on e.g., users filling out questionnaires and taking interviews, do expect the availability of detailed logs of interactions between users,
- The systems for use in education are sometimes constrained to learning situations where specific team formations are required (Systems 10 and 14), or are sometimes based on data contained in, e.g., a LMS (Systems 7, 8, 11, and partly, 12),
- Often the systems reviewed require users (administrators, teachers, tutors or instructors) to define projects, to start the team formation process or to solve team formation problems (Systems 7, 8, 9, 10, 11, 12, and 13).

However, as explained in Section 3.1, a SLN does not necessarily provide the data on which these existing systems can operate. Therefore, alternative approaches have to be explored, such as asking the learners to submit specific evidence on the required knowledge or having them point to relevant entries in their e-portfolio (Penalvo et al., 2012). SLNs also have no users in the specific roles required to run these systems. And while most of the systems examined from the educational domain only support curriculum-based activities, SLNs support self-directing learners in potentially wider knowledge domains. These differences, combined with the fact that SLN learners currently cannot easily benefit from focussed and motivating collaborative learning opportunities, warrant that we design a new approach to forming teams for project-based learning in these social learning networks.

3.3. PBL and team formation service design for use in SLNs

The team formation model presented in Section 3.1 might readily be recognized as belonging to traditional educational settings. PBL theory and team formation theory (Obaya, 1999; Oakley, Felder, Brent, & Elhajj, 2004) suggest that in such settings a team formation expert (e.g., a teacher) should initiate projects, while using knowledge about the curriculum to define an appropriate task. This expert uses knowledge (which can be both implicit and explicit) about the learners to form teams. However, as explained above, in SLN settings, these data nor teachers, are available. As SLN learners self-direct and self-organise we need to design a support service that enables learners themselves to perform the chain of activities required to initiate PBL and team formation.

Following the model introduced earlier, our service is designed to gather three categories of data for initiating PBL and team formation:

I) *Knowledge,* contained in: a) the collective learning materials available in the SLN, which make up the *domain*, b) *projects* and their *characteristics* (such as preferred team size, duration etc.) as defined by learners or other stakeholders in the network, c) *knowledge* available from possible team members, as evidenced by learners submitting materials for that purpose

II) *Personality*: data on the learners' personalities

III) *Preferences*: data on the learners' preferences with respect to project activities.

In order to perform the assessments depicted in the model, these data are handled by different *experts-by-proxy*. We differentiate between a *knowledge proxy*, a *personality proxy* and a *preferences proxy*.

3.3.1. The proxy designs

The aim of the *knowledge proxy* is three-fold: 1) to create a representation of the knowledge contained in all the topics in the learning materials present in the SLN, 2) to deduce which of these topics are addressed in the project task, and 3) to assess whether and how much knowledge learners have available on the topics addressed.

The knowledge proxy operates on a) the collective learning materials available in the SLN that make up the domain, b) descriptions of projects by learners (or other stakeholders in the network), c) knowledge available from possible team members, as evidenced by learners submitting materials for that purpose. It is important to notice that we assume that these sources are all explicitly available in a textual form.

The assessment of knowledge through the analysis and comparison of data in a textual form is a complex task. However, prior research demonstrated the successful application of a textual analysis method, called Latent Semantic Analysis (LSA), to match *people* to *jobs* and *learning materials* (Laham, Bennett, & Landauer, 2000; Landauer, Foltz, & Laham, 1998; Landauer, 2007). In our knowledge proxy design, these entities translate to *learners*, *projects* and the *collective learning materials* in the SLN domain. Figure 3.2 depicts an example of a simplified version of the process the *knowledge proxy* performs: It creates a representation of the knowledge in the domain (containing topics 1 through 6) and it analyses a project description, which is shown to relate to 3 topics in the domain (Topics 1, 3, and 5). After learners submit knowledge evidence on these topics the proxy analyses the degree to which the learner' knowledge overlaps the knowledge in the domain topics by using the domain topics as reference points. In Figure 3.2 the results of these analyses are depicted as percentages.

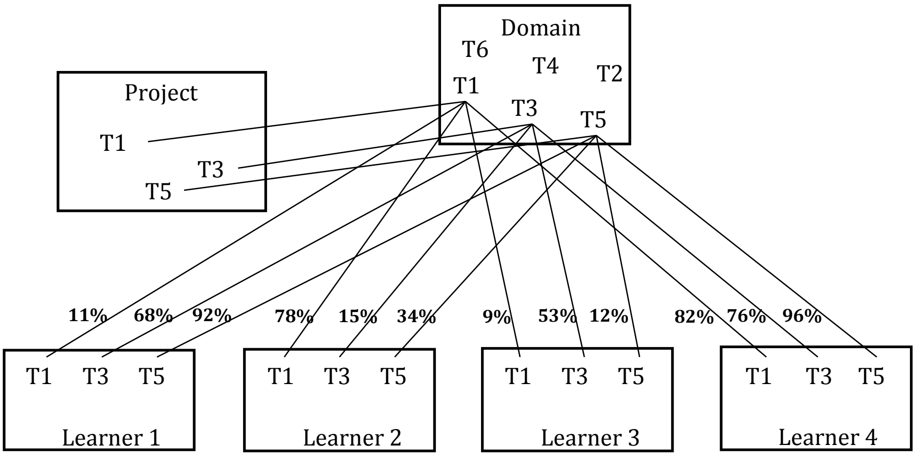


Figure 3.2: *An example of the knowledge proxy process:*

In Figure 3.2 a project refers to 3 topics in the domain (T1, T3, and T5). Learner submitted knowledge evidence (from Learners 1 to 4) for these topics is compared with the topic materials in the domain. The percentages indicate the degree of the knowledge overlap.

The *personality proxy* takes a different approach in that it uses data on learner personality, which is gathered through a personality test. We specifically chose to assess learners on the personality construct "conscientiousness" (which measures learner *carefulness, thoroughness, sense of responsibility, level of organization, preparedness, inclination to work hard, orientation on achievement,* and *perseverance*) because it predicts a person's future performance in a team

(Goldberg, 1990; Jackson et al., 2010). The learner conscientiousness score is established by using the Big Five personality test (Barrick & Mount, 1991).
The *preferences proxy* establishes a learner preferences profile, in which learners enter data on such variables as availability, time zone, possible collaboration languages and preferred tools. The proxy then determines the overlap with respect to the project characteristics mentioned above and the learners' project work related preferences. When preferences do not overlap at all, this fully blocks user inclusion in a team. (E.g., when one learner indicates to be available only on Mondays, while another learner indicates to never be available on Mondays, their calendars are mutually exclusive and thus these two learners will never be matched in a team). We currently envision the learner to enter this data in the profile.
From this it follows that the *first step* in the team formation process is finding overlapping sets of preferences by comparing the project characteristics and learner preferences. By doing this, the proxy's result limits the number of learners from which teams can be formed. The team formation process then continues with the data on knowledge and personality.
It is important to notice that the data gathered on learners is *not of a static nature*, but can be refreshed every time a learner re-enters knowledge evidence for a project, retakes the personality test, or updates preferences. Furthermore, future iterations of the team formation service might be enabled to connect to user data already available in such e-portfolios as described in e.g., the TRAILER project (Penalvo et al., 2012). For the remainder of this chapter we assume that the results of the assessments are available.

3.3.2. Definition of the team formation service and rules for targeting productive, creative or learning outcomes

The proxies' data gathering designs presented above provide the data to the team formation service. The service combines the two separate sets of data by following *team formation rules.* We discern three possible teamwork target outcomes and indicate three sets of rules, one for each outcome. The rules are based on existing research:

(1) Productive problem solving:

- Forming teams from learners who have different conscientiousness scores impedes their task negotiations after the project team has been formed, which would then hinder the team task execution (Gevers & Peeters, 2009)
- Members of productive teams should be capable and conscientious and must have domain knowledge (Isaksen & Lauer, 2002)

The general team formation rule we infer is: *Productivity is fostered when team members have high scores on knowledge of the project topics and show high levels of conscientiousness*.

(2) Creative solutions:

- Too much complementary fit in knowledge can lead to a loss of creativity and to group thinking (West, 2002)
- People with high conscientiousness scores tend to be less creative (George & Zhou, 2001; Wolfradt & Pretz, 2001)
- Groups with members that possess different knowledge backgrounds will be more innovative because they contribute from different perspectives (Paulus, 2000)
- Successful research teams are heterogeneous (Dunbar, 1997)

We infer as general team formation rule: *Team creativity is fostered when team members have highly differentiated scores on knowledge of the project topics and show low levels of conscientiousness*.

(3) Facilitating learning:

- Learning is fostered when team members provide a complementary fit in knowledge backgrounds and show a supplementary fit in personalities (Werbel & Johnson, 2001).
- Mutual teaching and learning are among the most important activities in defining and solving problems (Paulus, 2000).
- There is a maximum 'distance in knowledge' (the zone of proximal development) that can be bridged when learning with more capable peers (Vygotsky, 1978).

From these findings we infer as general team formation rule: *Learning in a team is facilitated when knowledge on the project topics is distributed over the members (allowing each member to learn and teach) and the differences in the levels of project topic knowledge between the members are not too high and the members' conscientiousness levels all are high*.

Table 3.1 provides an overview of the team formation rules (with respect to learner knowledge and conscientiousness) and the target outcomes. In the table, the terms "supplementary" and "complementary" are used to denote "sharing knowledge with other members" and "providing knowledge to the team which other members lack", respectively.

***Table 3.1:** Research basis, the kind and level of knowledge rule and conscientiousness rule for specific target outcomes.*

Research basis	Kind and level of knowledge rule	Conscientiousness rule	Target outcome
Gevers & Peeters, Isaksen & Lauer	Supplementary and high	All high	Productive problem solving
George & Zhou, Wolfradt & Pretz, West, Paulus, Dunbar	Complementary and high	All low	Creative solutions
Werbel & Johnson, Vygotsky, Paulus	Complementary and high, but within limits	All high	Facilitating learning

3.3.3. Team formation expressions

Based on the target outcomes defined in Table 3.1, we devised three mathematical team formation expressions that can be applied to the data gathered. They suggest formations of productive, creative, or learning teams, respectively. Applying the expressions results in measures of fitness calculated for all possible teams of a chosen size, recruited from a given set of learners. For each possible team, the team fitness value is represented in a value between "0" and "1", with "1" indicating the highest possible fit for that outcome. This allows for comparing teams with respect to fitness over their different target outcomes. Weights can be used to indicate the importance of e.g., knowledge over conscientiousness in the team formation process. In the expressions below all weights are equal and sum up to 1. Other weight distributions are likely of relevance but have not been systematically explored. In all expressions, for demonstration purposes, the maximum score on knowledge (*Max_K*) is set to 10 and the maximum score on conscientiousness (*Max_C*) is set to 5. Both the desired team size (n) and number of topics (k) the project refers to are arbitrarily set to 4.

Productive teams

The team formation expression for the outcome "productive problem solving" (see Table 3.1 and Figure 3.3) describes teams whose members have the highest average score on knowledge and the highest average score on conscientiousness.

$$FitP_i = W_K * \frac{Avg\_K_i}{Max\_K} + W_C * \frac{Avg\_C_i}{Max\_C}$$

***Expression 3.3:** Team formation expression for productive teams.*

Explanation of Figure 3.3: In the first part, the average score on knowledge of all members of team *i* over all topics is calculated (Avg_K_i) and divided by the maximum knowledge score (*Max_K)*. In the second part the average score on conscientiousness over all members is calculated (Avg_C_i) and divided by the maximum conscientiousness score (*Max_C*). These two scores are multiplied by their weights (W_K, W_C) separately and then summed. As the two parts each result in a value between 0 and 1 and the sum of the weights always is 1, this results in a measure of fitness ($FitP_i$) for each team considered between 0 and 1. In Table 3.2 we present an example of a score set leading to a *FitP* of 1.

Table 3.2: *Example of scores on topic knowledge and conscientiousness leading to a FitP of 1.*

Member	Topic 1	Topic 2	Topic 3	Topic 4	Cons
L01	10	10	10	10	5
L02	10	10	10	10	5
L03	10	10	10	10	5
L04	10	10	10	10	5

Creative teams

The mathematical team formation expression for the outcome "creative solutions" (See Table 3.1 and Figure 3.4) maximises when team members have a maximum difference in knowledge between their best score and their second best score *over their own topic scores*, and when there is a maximum difference in knowledge between the best score and the second best score *inside a topic*. It minimises the average conscientiousness score in the team.

$$FitC_i = W_K * \frac{\sum_j DifK_j}{TeamSize * Max\_K} + W_E * \frac{\sum_t DifK_t}{NumTop * Max\_K} + W_C * \frac{Max\_C - Avg\_C_i}{Max\_C}$$

Expression 3.4: *Team formation expression for creative teams.*

Explanation of Figure 3.4: In the first part the expression calculates the differences for each team member *j* between their highest score on a topic and the next best score on a topic ($DifK_j$) and sums these differences up over all team members. The result is divided by the product of the team size (*TeamSize*) and maximum score on knowledge (*Max_K*). In the second part, the differences for each topic *t* between the highest score on that topic and the next best score on that topic ($DifK_t$) are summed up. The result is divided by the product of the number of topics (*NumTop*) and the maximum score on knowledge (*Max_K*). Finally, in the third part, from the maximum conscientiousness score (*Max_C*) the all-member average conscientiousness score (Avg_C_i) is subtracted. The result is divided by the maximum conscientiousness score (*Max_C*).
All three scores are multiplied by their weights (W_K, W_E, W_C) separately and then summed. As all three parts each result in a value between 0 and 1 and the sum of

the weights always is 1, this results in a measure of fitness ($FitC_i$) for each team considered between 0 and 1. In Table 3.3 we present an example of a score set leading to a *FitC* of 1.

Table 3.3: *Example of scores on topic knowledge and conscientiousness leading to a FitC of 1.*

Member	Topic 1	Topic 2	Topic 3	Topic 4	Cons
L01	10	0	0	0	0
L02	0	10	0	0	0
L03	0	0	10	0	0
L04	0	0	0	10	0

Learning teams

The team formation expression for the outcome "facilitating learning" (see Table 3.1 and Figure 3.5) mathematically describes teams whose members can teach and learn to and from each other inside each knowledge topic, while having a high score on Conscientiousness. It optimises the match between possible teachers and learners in the team by using Vygotsky's 'zone of proximal development' (Vygotsky, 1978) as a parameter (*zpd*) to calculate teaching and learning effectiveness for the team over all project topics.

$$FitL_i = W_K * \frac{\sum_t \sum_l \sum_j \left| score_{t,j} - score_{t,l} \right|}{d_{jt} \cdot zpd \cdot n \cdot k} + W_C * \frac{Avg\_C_i}{Max\_C}$$

Figure 3.5: *Team formation expression for learning teams.*

Explanation of Figure 3.5: In the first part, every topic score of a member is compared to the other member's topic scores (|$score_{t,j}$, $score_{t,l}$|). When there is no difference between the scores, the members cannot teach to each other, nor learn from each other. If the difference is inside the parameter *zpd* (currently set to be between 0 and 3), then that member becomes a teacher to the other member. The member' teaching effectiveness depends on the difference from the set *zpd*. For example when member 1 scores 8 on topic 1 while member 2 scores 6 on topic 1, then the difference is 2. With a *zdp* set to 3, the teaching effectiveness between these members is calculated as 2/3. In the same manner, learning effectiveness is calculated. This is repeated for all other members. For each member the teaching and learning effectiveness scores are summed up and then divided by that member's summed number of times being a teacher and number of times being a learner in the topic (d_{jt}). We define the result as that member's effectiveness in the team. This process is repeated for all members (n) inside the topic. Finally, all teaching scores are added, all learning scores are added and all effectiveness scores are summed. With the multiplication of the sum of the effectiveness scores with the

sum of the sum of all learning scores and the sum of all teaching scores, we arrive at a score for that topic, which is then normalised. This process is repeated over all topics (k), and all topic scores are summed. This final sum represents the teams learning capability.
In the second part, the average team conscientiousness score (Avg_C_i), divided by the maximum conscientiousness score (Max_C) is calculated. The two scores are multiplied by their weights (W_K and W_C) separately and then summed. As the two scores each result in a value between 0 and 1 and the sum of the weights always is 1, this results in a measure of fit for each team considered ($FitL_i$) between 0 and 1. There are two exemptions to the rule: If the difference between two topic scores is higher than the parameter zdp, or when a teacher has a score on a topic lower than a set minimum score (currently set to 6), teaching and learning effectiveness for that teacher/learner pair is set to be ~0. In Table 3.4 we present an example of a score set leading to a *FitL* of 1 when the zone of proximal development is set to 3 and the minimum teacher topic score is set to 6.

Table 3.4: *Example of scores on topic knowledge and conscientiousness leading to a FitL of 1.*

Member	Topic 1	Topic 2	Topic 3	Topic 4	Cons
L01	10	9	10	9	5
L02	7	6	10	6	5
L03	10	6	7	9	5
L04	7	9	7	6	5

We anticipate that the application of the three expressions to the same data set will result in differentiated team formation suggestions for each of the three outcomes, and that the results indicate which outcome fits best to any of the teams possible.

3.4. Results of the application of the team formation expressions on a test data set

For the simulation we used a set of test data on 10 learners (see Table 3.5). The test data set presupposes that the project description had already been analysed and was found to refer to knowledge on 4 topics in the domain. It further presupposes that the analysis of knowledge evidence on these 4 topics, as submitted by 10 learners, had already been performed. This is reflected in the numerical scores under the topics 1 through 4 (ranging from 1 to 10, where 10 indicates the highest possible score on a topic). The scores on Conscientiousness (Cons) in Table 3.5 are the simulated results of a personality test (ranging from 1 to 5, where 5 indicates the highest level). The team size of the teams to be formed was arbitrarily set to 4 learners per team.

Table 3.5: *The test data set.*

Member	Topic1	Topic2	Topic3	Topic4	Cons
L01	9	8	8	9	5
L02	4	6	4	5	4
L03	4	3	4	9	1
L04	5	4	6	8	5
L05	3	4	10	2	1
L06	8	9	8	5	4
L07	4	9	5	3	2
L08	8	9	8	7	3
L09	5	8	7	8	3
L10	4	5	3	4	1

3.4.1. Application of the expressions

When the expressions above are applied to the test data set, all fitness values for the 210 unique combinations [Number_of_learners! / ((Number_of_learners – team_size)! * team_size!)] of 4 learners are calculated. The output we receive lists all possible teams and their scores on FitP, FitC and FitL, totalling to 630 values. In Table 3.6 we present only the 3 highest scores per outcome, and the lowest score (all results are truncated to 3 decimals). In the three columns FitP, FitC and FitL the scores are sorted from high to low.

Table 3.6: *Team formations for 4 teams of 4 learners, sorted by FitP, FitC or FitL.*

Team members	FitP	Team members	FitC	Team members	FitL
L01,L04,L06,L08	0.797	L03,L05,L07,L10	0.500	L02,L04,L06,L09	0.660
L01,L04,L06,L09	0.784	L03,L05,L06,L10	0.442	L04,L06,L08,L09	0.609
L01,L02,L04,L06	0.781	L03,L05,L08,L10	0.442	L02,L04,L06,L08	0.598
~	~	~	~	~	~
L03,L05,L07,L10	0.363	L01,L02,L04,L06	0.092	L03,L05,L07,L10	0.126

The individual team members and their scores on Topics 1 to 4 and conscientiousness for the teams with the highest scores on FitP, FitC and FitL are presented in Table 3.7.

For FitP, a team comprised of learners L01, L04, L06, and L08 receives the highest score (0.797), while the lowest score (0.363) is for a team comprised of learners L03, L05, L07, and L10. For FitC, a team formed from learners L03, L05, L07, and L10 receives the highest score (0.500), while a team of learners L01, L02, L04, and L06 receives the lowest score (0.092). As for FitL, a team with learners L02, L04, L06, and L09 scores highest (0.660). A team with learners L03, L05, L07, and L10 scores lowest (0.126).

Table 3.7: *Individual learner scores (M1 to M4) on topic knowledge (T1 to T4) and conscientiousness (Cons) for the teams with the highest fit values on FitP, FitC and FitL from Table 3.6.*

	Team with highest FitP					Team with highest FitC					Team with highest FitL				
	T1	T2	T3	T4	Cons	T1	T2	T3	T4	Cons	T1	T2	T3	T4	Cons
M1	9	8	8	9	5	4	3	4	9	1	4	6	4	5	4
M2	5	4	6	8	5	3	4	10	2	1	5	4	6	8	5
M3	8	9	8	5	4	4	9	5	3	2	8	9	8	5	4
M4	8	9	8	7	3	4	5	3	4	1	5	8	7	8	3

3.4.2. Differentiations in team formation suggestions

When sorted for FitC, the highest scoring team on FitP is found on position 208 and when sorted for FitL that team is found on position 6. Both when sorted for FitP and for FitL, the highest scoring team on FitC is found on position 210. When sorted for FitP, the highest scoring team on FitL is found on position 16 and when sorted for FitC, it is found on position 196. The differentiation is not only relevant with respect to rank in the results, but also with respect to actual fitness value calculated. Table 3.8 allows for comparing the teams with the highest fitness values on a particular outcome (these fitness values are highlighted in the table) with how well they fit to any of the other outcomes.

Table 3.8: *Team fitness values on FitP, FitC and FitL for the highest scoring teams on FitP, FitC and FitL, respectively.*

Team of members	FitP	FitC	FitL
L01,L04,L06,L08	0.797	0.100	0.569
L03,L05,L06,L10	0.363	0.500	0.126
L02,L04,L06,L09	0.713	0.142	0.660

The results also indicate which kind of team could preferably be formed from these learners: the highest overall fitness-value (0.797) is received for a team (consisting of the learners L01, L04, L06, and L08) that aims at the outcome “productive problem solving”. An interesting find is that a team consisting of learners L02, L04, L06, and L09, while receiving the highest fitness value for the outcome “facilitating learning” (FitL = 0.660), would likely do better if it were to aim for “productive problem solving” as an outcome (FitP = 0.713).

3.5. Discussion and conclusion

We set out to answer the question: *How can one design a team formation service for project-based learning in social learning networks that optimises either learning outcomes, creative outcomes or productive team performance outcomes?*

Our perspective was that social learning networks currently do not readily support effective, coherence-creating and motivating learning settings. We therefore suggested to provide these learners with a project-based learning and team formation service. As a starting point we took our team formation model (Spoelstra, Van Rosmalen & Sloep, 2012). A survey of existing team formation tools and techniques revealed that these are not easily applicable in a "team formation for project-based learning in social learning networks"-approach. They assume data and user roles that are not available in SLNs. For this reason we proposed a design which allows project-based learning and team formation to be based on data that can be acquired directly from the SLN and its learners. The design puts learners in control of the process of defining and staffing projects, thus honouring these learners' self-directing and self-organising behaviour. The design uses the data categories 'knowledge', 'personality', and 'preferences' (as defined in the team formation model) and describes the ways in which the data can be gathered and processed to suggest team formations. A benefit of the design is that it is also based on personality characteristics, which is rarely the case in existing tools, but which – according to literature (Roberts, Kuncel, Shiner, Caspi, & Goldberg, 2007) – is highly relevant.

The team formation and project-based learning service deploys three different proxies to gather and assess data: 1) To assess both required and available knowledge, the *knowledge proxy* analyses textual data; 2) To assess learner personality, the *personality proxy* determines a learner's conscientiousness by using a personality test; 3) To determine project work preferences, the *preferences proxy* determines whether collaborative project work can happen at all.

In order to determine how learners should be teamed up based on knowledge and personality we analysed existing research on team formation principles. The outcomes led to the definition of team formation rules for forming productive, creative, or learning teams, respectively. These rules were formalised in team formation expressions.

The application of the expressions to a set of simulated test data demonstrates their ability to form teams and to suggest different teams based on the desired teamwork outcomes. The results provide both team rank on all three possible outcomes and the absolute fitness values for those outcomes. The results further allow us to suggest which outcome would fit best to any of the teams that could be formed. We believe these results clearly show the ability of the expressions to differentiate between teams fit for any of the proposed teamwork outcomes.

Future research can introduce further differentiation in the results: when one primary outcome is selected for a team, its fitness scores on the other outcomes

might act as qualifiers for that outcome. This would provide a method for closer selection of teams, based on how the primary outcome will likely be achieved.
With its strong base in PBL and team formation research, we believe our approach addresses important issues in team formation. However, one could argue that knowledge might also be contained in other forms of evidence currently not taken into account, and that even though 'conscientiousness' is very important predictor of a learner' success in future project work, it is not the only personality aspect playing a role in team work. Furthermore, research by e.g., Kirton (2003) indicates that the more diverse a team is, the greater its potential for problem solving will be, but the more difficult it becomes to manage. This might be of particular interest in the case of creative teams, where the favoured low average conscientiousness, combined with highly diverse knowledge could lead to teams having difficulty working together. Future research will determine whether the introduction into the expression for creative teams of additional personality factors such as 'Extravertness' (Barrick & Mount, 1991) are necessary to mitigate this effect.

However, our premise was that social learning network learners only have limited knowledge of other learners, and that these networks do not have historic data on learner performance. We therefore believe that our team formation service offers an import first step in supporting project-based learning and team formation in such networks. Nevertheless, (parts of) the team formation service can have a wider application in settings where the required data is already partly available. When data on prior knowledge and preferences are available (e.g., in a classroom setting), the service only requires the addition of personality data to be usable. The preliminary analyses of the results from a survey about whether and under which conditions teachers would accept team formation suggestions from an automated system based on the proposed design indicate that of 11 responses, 5 express acceptance of automated team suggestions, while 5 responses express acceptance with some reservations. These reservations are mostly concerned with aspects such as who has the final say in team formation. As our tool delivers team formation suggestions from which users can deviate, we feel convinced that a team formation tool based on the principles outlined above will be welcomed.
Another area of application might be found in the context of Massive Open Online Courses (MOOCs), where the use of team formation tools could be a way to enhance the currently rather limited interaction between students.

In our future research we will report on an implementation of the knowledge, personality and preferences proxies using real student reported data in a large-scale experiment. A next step will then be to further implement the knowledge proxy, for which we suggest to use the LSA textual analysis method to match knowledge from learners to knowledge required by a project.

CHAPTER 4

Team Formation Instruments to Enhance Learner Interactions in Open Learning Environments[3]

[3] This chapter (with minor changes in terminology and lay-out) was previously published as: Spoelstra, H., Van Rosmalen, P., Houtmans, T., & Sloep, P.B. (2015). Team Formation Instruments to Enhance Learner Interactions in Open Learning Environments. *Computers in Human Behavior, 45*, 11-20, http://dx.doi.org/10.1016/j.chb.2014.11.038.

Abstract

Open learning environments, such as Massive Open Online Courses (MOOCs), often lack adequate learner collaboration opportunities; they are also plagued by high levels of drop-out. Introducing project-based learning (PBL) can enhance learner collaboration and motivation, but PBL does not easily scale up into MOOCS. To support definition and staffing of projects, team formation principles and algorithms are introduced to form productive, creative, or learning teams. These use data on the project and on learner knowledge, personality and preferences. A study was carried out to validate the principles and the algorithms. Students (n=168) and educational practitioners (n=56) provided the data. The principles for learning teams and productive teams were accepted, while the principle for creative teams could not. The algorithms were validated using team classifying tasks and team ranking tasks. The practitioners classify and rank small productive, creative and learning teams in accordance with the algorithms, thereby validating the algorithms outcomes. When team size grows, for practitioners, forming teams quickly becomes complex, as demonstrated by the increased divergence in ranking and classifying accuracy. Discussion of the results, conclusions, and directions for future research are provided.

4.1. Introduction

Open learning environments, such as Massive Open Online Courses (MOOCs), currently attract large bodies of learners. Initially these environments were envisioned to provide learning settings based on the pedagogical vantage point of networked learning, with a strong emphasis on learner self-direction and learner contribution. Downes (2006) and Siemens (2004) coined the term “connectivism” to label such learning settings. In parallel a different kind of MOOC rose to attention, one that builds on behaviourist, rather than social-constructivist educational principles. Reports, however, from both learners and MOOC providers indicate that drop-out rates from both kinds of MOOCs are massive, and that in particular the latter kind offers limited opportunities for learner collaboration. (Daniel, 2012; Edinburgh University, 2013; Morrison, 2013; McGuire, 2013). While there are many reasons for drop-out rates to be high, these effects can at least partly also be explained by learning settings that do not motivate learners. In the up till now small-scale connectivist MOOCs learners are expected to be self-directing, which can present learners with difficulties related to insufficient task structure (Kop et al., 2011). In the large-scale behaviourism-based MOOCs, scaffolding, teacher-learner contacts and collaborative learning opportunities are limited, which leads to sub-optimal learning (Daniel, 2012; Edinburgh University, 2013). Some MOOCs (Stanford University, 2012; NovoEd, 2014) address this by allowing self-selection into teams or by providing relatively simplistic grouping criteria such as by proximity of geographic location or by language(s) mastered.

In general, collaborative learning processes in open learning environments can take shape as suggested in the theoretical computer-supported collaborative learning (CSCL) framework of Stahl (2006). Stahl describes that, in a cyclic process, individuals express problems, collaborate with peers to develop shared understanding, use and create learning materials, which are then again used by others to learn from. While learning can be instigated by individuals and whole communities can benefit from its outcomes, Stahl places the actual learning process in the context of the small group. However, with regard to implementing the framework, Stahl (2013) also notes it: “… needs appropriate CSCL technologies, group methods, pedagogy and guidance to structure and support groups to effectively build knowledge…”. In this chapter we investigate a particular approach to forming teams for collaborative learning in open learning environments. We surmise this is a specific operationalization of Stahl’s framework. Hence we ensure that i) learner problem statements are related to the learning settings in which they are made, ii) collaboration takes place in teams with suitable knowledgeable peers only, iii) only knowledge sources are available that fit the learners needs, iv) the interactions between learners are structured, not fleeting and shallow.

Our approach promises to unleash the powers of constructivist learning and to implement the well-researched team-based learning settings of project-based learning (PBL; Blumenfeld, Soloway, Marx, Krajcik, Guzdial, & Palincsar, 1991;

Davies, de Graaff, & Kolmos, 2011) in MOOCs (Sloep, Berlanga, & Retalis, 2014). Implementing PBL provides several well-known benefits. First, it improves the learners' motivation, so that learners are more inclined to deal with hard, complex problems and spend more time studying (Johnson, Johnson, Stanne & Garibaldi, 1990; Marin-Garcia & Lloret, 2008). Second, and related to improving motivation, PBL plays a role in learner retention (Dahms & Stentoft, 2008; Fisher & Baird, 2005). Third, PBL blends learning and working, thereby creating realistic (inter-professional) learning experiences (Springer, Stanne & Donovan, 1999; Felder, Felder & Dietz, 1999), which prepare learners for real-life working conditions (Haines, 2014). Forth, generally speaking, collaboration between learners as envisioned in PBL has been shown to lead to an increase in learning outcomes compared to individual learning (Hsiung, 2010). Fifth, it can prevent knowledge sharing issues learners encounter when trying to use e.g., social media as open learning environments. Ma & Chan (2014) found that in social media only a tiny proportion of users engage in a type of knowledge exchange that is ultimately beneficial to them.

Implementing PBL in traditional educational settings requires expertise from teachers for defining project tasks and staffing them. However, as in large scale MOOCs staff time expenditure needs to be kept low, we propose that learners themselves play an active role in defining projects for PBL. Learners who are enabled to *self-define tasks* develop a motivating sense of ownership and responsibility for their learning processes. At the same time, however, *self-selection of teams* ought to be discouraged. Fiechtner and Davis (1985) and Oakley, Felder, Brent and Elhajj (2004) hold that for teams to be effective, team formation should be performed by experts. These experts use knowledge of the project tasks and of the prospective team members to form teams (Graf & Bekele, 2006; Martín & Paredes, 2004; Wilkinson & Fung, 2002; Obaya, 1999; Slavin, 1989). In large-scale MOOCs however, a complicating factor is that these experts will most probably not be available. Therefore we argue that if large groups of learners in MOOCs are to be enabled to self-define project tasks and to receive effective team formation suggestions, we need to develop automated support services. These mimic expert behaviour in assessing whether projects relate to the MOOC's learning materials and form teams based on task and team member characteristics (beyond language and geographical location). The services provide intelligent team formation principles, for which we build on extensive preparatory research. In this research we inferred several *team formation principles* from team formation literature and developed the corresponding *team formation algorithms* (Spoelstra, Van Rosmalen, Van de Vrie, Obreza, & Sloep, 2013; Spoelstra, Van Rosmalen, & Sloep, 2014). It is our future goal that our instruments will be able to assess whether suggested projects qualify for execution inside MOOCs and to form effective project teams. In this chapter, however, we focus on the *validation* of the set of team formation instruments we developed, based on important factors in team formation, such as knowledge, personality, and preferences. First, we aim to validate the team formation principles we inferred. Second, we aim to validate their implementation in algorithms, using real-world learner data for their input. This validation will be based on practitioner agreement with the team formation principles and by comparing practitioner outcomes on team formation tasks to the outcomes of the computer algorithms. The remainder of the chapter is structured as follows: In Section 4.2 we present a team formation model, which uses learner knowledge, personality and preferences to suggest teams fit for executing a project. In section 4.3, we present the research questions and hypotheses, on the basis of which we aim to validate the team formation instruments. Section 4.4 describes the materials and methods we used to test the hypotheses. In Section 4.5, the results are presented. Sections 4.6 provides an extensive discussion of these results, while in Section 4.7 we draw conclusions and suggest future research.

4.2. A team formation model

The automated service builds on earlier work in which we introduced a team formation model for use in open learning environments, as well as in more traditional learning settings. The model was constructed based on a review of PBL and team formation literature. It aims to mimic the behaviour of team formation experts (i.e., use knowledge on task and team members to form teams fit for various tasks)
(Spoelstra et al., 2013). An updated version of the model is presented here, which explicitly adds the assessment of fit of a project in a knowledge domain. It also puts the assessment of learner preferences logically before the assessments of knowledge and personality (see Figure 4.1)
The model describes the definition of a project (a task addressing multiple topics carried out by multiple learners) in a knowledge domain. This definition is assessed for fit in the knowledge domain. Next, learner preferences (such as available time slots or languages spoken) are compared to the project characteristics (such as duration, preferred number of team members, preferred language). This comprises the *first* step in the chronology of the team formation process, which limits the number of learners from which teams can be formed. In the *second* step, the assessment of knowledge is used to match the knowledge required for executing the project to the knowledge the prospective team members can provide. The assessment of personality is aimed at predicting team member performance (Jackson et al., 2010; George & Zhou, 2001; Barrick, Mount, & Strauss, 1993). For this, personality can be represented by the personality trait "Conscientiousness", which can be assessed with e.g., the Big Five personality test (Barrick & Mount, 1991). In the *third* step the resulting data are combined, based on a team formation principle, to determine the fit of teams of learners to a task and to suggest project teams. By using just two dimensions (knowledge and conscientiousness) we can form teams fit for various project tasks. We discern between three common types of project tasks: 1) expertly and productively working on a project, 2) creatively

solving a project problem, and 3) sharing knowledge (teach and learn) with fellow team members while solving a project problem. Based on team formation theory, in earlier work (Spoelstra et al., 2013) we inferred three team formation principles that vary on the aforementioned dimensions. Each principle is directed at optimising the team formation process toward one of these three types of tasks. In the next subsection we present these team formation principles.

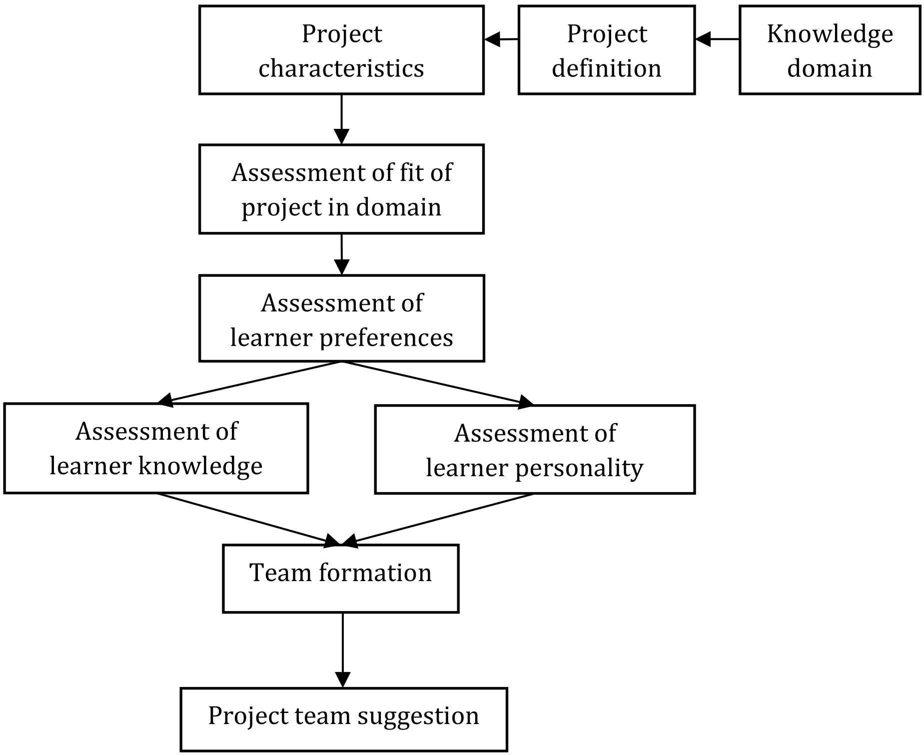


Figure 4.1: *The team formation model.*

4.2.1. Team formation principles

The team formation principles we aim to validate are the following:

- The team formation principle for productive teams: "Productivity in a team is fostered when team members have high scores on knowledge of the project topics and the team members show high, homogeneous levels of conscientiousness".
- The team formation principle for creative teams: "Creativity in a team is fostered when team members have differentiated scores on knowledge of the project topics and the team members show low levels of conscientiousness."

- The team formation principle for learning teams: "Learning in a team is facilitated when knowledge on the project topics is distributed over the members (allowing each member to learn and teach). However, the differences in knowledge should not be too high, and the team members should show high levels of conscientiousness."

These team formation principles, through their corresponding expressions (see Appendix A), were implemented in computer algorithms. The implementation of the algorithms for productive and creative teams was rather straightforward: the algorithm for productive teams favours teams with members that all have high knowledge scores and high conscientiousness scores. The algorithm for creative teams favours teams in which topic knowledge is maximally diversified over team members, who also have low conscientiousness scores.
The team formation algorithm for learning teams, however, is more complex as:

- It models one of the aspects from Vygotsky's principle of "zone of proximal development": difference in knowledge between learners (Vygotksy, 1978). This aspect is expressed in a parameter "zpd", which puts a limit on the knowledge differences allowed between team members. From this difference it calculates teaching and learning effectiveness between each team member inside each project topic. It currently follows a 10-point grading system (grades range between 1 and 10, with 10 being the highest possible grade while 6 is considered to be the passing mark). The value of the parameter "zpd" is currently set to 3.
- It implements a minimum knowledge level, below which teaching is assumed to be undesirable, as the member considered for the teaching role is assumed not master the topic sufficiently well. This value is set to 6.

Please note that this means that the algorithm assumes that the learners in the peer-tutor/learner pairs with topic knowledge grades of 10 and 7, 9 and 6, 8 and 5, 7 and 4, and 6 and 3 can all learn effectively, provided they also have high scores on conscientiousness. In practice pairs with a smaller knowledge differences may be selected if the optimal peer-tutor/learner pairings are not available. In the next section we present our research questions and hypotheses.

4.3. Research questions and hypotheses

As indicated in the introduction, our focus is on the validation of the team formation principles and the outcomes of their implementations in algorithms. These principles and algorithms fit in how we suggest to operationalise Stahl's CSCL framework regarding team formation for PBL in the context of open learning environments. Our two main research questions are:

(R1) *Are the team formation principles for forming productive, creative and learning teams in alignment with the opinions and experiences of practitioners from the educational field about how such teams should be formed?*

(R2) *Are the results from the computer algorithms in alignment with the results of practitioners from the educational field performing the same task?*

To answer research question 1 (R1), we put forward the following hypothesis:

(H1) *Practitioners from the educational field agree that the three individual team formation principles for productive, creative and learning teams lead to the formation of teams fit for their associated tasks.*

We consider H1 to be accepted when the practitioners agree that each of the three individual team formation principles lead to the formation of teams fit for their associated tasks (i.e. when practitioners answer “agree” or “strongly agree” on a five-point Likert scale).

Research question 2 (R2) will be answered by evaluating the results (i.e. team formation suggestions) of an implementation of the team formation expressions in algorithmic form, applied to real world learner data. We draw up the following connected hypotheses:

(H2a) *Given the same data as used by the algorithms, practitioners from the educational field classify the teams in accordance with the algorithms.*

(H2b) *Given the same data as used by the algorithms, practitioners from the educational field rank the teams in accordance with the algorithms.*

H2a will be accepted when the practitioners classify teams in accordance with the team formation principles, while H2b will be accepted when the practitioners rank teams in accordance with the team formation principle. Due to the complexity of the tasks, we assume that human performance will be effected negatively when the tasks get more complex. Hence, related to R2, we explore whether the performance of practitioners on classification and ranking is effected when we increase both the size of the teams and the numbers of topics the project addresses.

4.4. Materials and Method

For the experiment a representative set of real world learner data on knowledge, conscientiousness and preferences was required. This set was gathered by means of an online survey which is described in subsection 4.4.1. The survey was conducted among learners from the School of Psychology and from the Master

Educational Sciences of the Open University of the Netherlands. It was carried out in the Dutch language. The data gathered was then processed with the team formation algorithms. Their output is described in subsection 4.4.2. The method applied for the experiment is described in subsection 4.4.3.

4.4.1. Learner data

Part 1 of the learner inquiry gathered learner demographics (gender, age, etc.). In total, 168 complete responses were gathered. Of the participants, 31 were male, 137 female. Stratification over age groups was as follows: 20-29 (24), 30-39 (46), 40-49 (58), 50-59 (37) and 60-69 (3). Of these, 121 learners studied at the Psychology faculty, while 47 learners studied Learning Sciences.

In part 2 we determined learner *conscientiousness* scores*.* To that end we presented the learners with a Big Five personality test (Barrick & Mount, 1991), containing 44 questions. We used the Dutch translation of the test (Denissen, Geenen, Van Aken, Gosling, Samuel, & Potter, 2008). The test assessed all Big Five personality aspects (Extraversion, Neuroticism (vs. Emotional Stability), Conscientiousness, Agreeableness, and Openness to Experience). The learners' conscientiousness scores varied between 2.00 and 4.56. The reliability scores (rounded to two significant decimals) for the five factors of the test were: Extraversion (.81), Agreeableness (.74), Conscientiousness (.84), Neuroticism (.85), and Openness (.86). These results are fully in line with an earlier validation of the BFI in the Dutch language.

In part 3 of the survey we asked learners to self-rate their *knowledge* on four topics that were addressed in courses on research methods and techniques. The topics were: 1) Defining research questions and theoretical designs for a study, 2) Gathering data, 3) Analysing data, and 4) Discussing and concluding on results. Following a 10-point grading system (with grades ranging from 1-10, with 10 as highest grade), the self-reported scores on the topics ranged between 3-9, 1-10, 1-10 and 1-9, respectively.

Part 4 asked the learners about their project work *preferences*, such as their preferred collaboration languages, the time slots in which they were available for collaboration (in the morning, and/or in the afternoon, and/or in the evening for every week day and the weekend as a whole) and the total number of hours they had available for collaboration weekly. Additionally, learner time zone information was collected to be able to adjust for time zone related availability mismatches. As indicated in Section 4.2, learner preferences effectively filter the number of possible team members for any project. For the current experiment we filtered using the data on availability. As criterion we used learner availability on the separate days of the week and on the weekend as a whole. This resulted in 8 groups of learners. The numbers of learners available in these groups were as follows: Monday (27), Tuesday (23), Wednesday (30), Thursday (34), Friday (29), Saturday (33), Sunday (29), and the whole weekend (29). Please note that learners could be available on multiple days.

4.4.2. Team formation algorithms output

The knowledge and conscientiousness scores of the 8 groups of learners were processed by the team formation algorithms to form project teams with 2 members covering 2 topics (using the topic knowledge grades on topics 1 and 2), to form project teams with 3 members covering 3 topics (using the topic knowledge grades on topics 1, 2 and 3), and to form project teams with 4 members covering 4 topics (using the topic knowledge grades on topics 1, 2, 3 and 4). This resulted in a total of 8 (one for each availability slot) times 3 lists (one for each team size). Each of the resulting 24 lists contained the fit values for the task types "productive", "creative", and "learning". Please note that for the remainder of this chapter we will refer to these teams as *2x2 teams*, *3x3 teams*, and *4x4 teams*, respectively.
The number of team members and the number of project topics were chosen for the purpose of the present experiment only. The team formation algorithms did not impose these choices. The robustness of the algorithms was tested by inputting the data of all learners (n=168) and calculating fit values for all possible 2x2, 3x3 and 4x4 teams. This resulted in text files containing team formations and fit values for 14,028, 776,216, and 32,018,910 unique teams respectively, with file sizes of 1Mb, 73Mb, and 3.3Gb. On a machine with an Intel i7 CPU and with 4Gb of internal memory, the algorithms completed successfully.

4.4. Method

The validation of the hypotheses described in Section 4.2 was conducted by means of an online survey. We invited all members (n=405) of the teaching staff of our university to participate on a voluntary basis. In total 56 participants completed the survey. Of these, 26 were female, while 30 were male. The distributions over age groups was: 20-29 (4), 30-30 (9), 40-49 (10), 50-59 (20), 60-69 (12), 70-79 (1).
In order to test hypothesis 1 (H1: team formation principles), in three separate questions the participants were presented with the three team formation principles. They were asked if they agreed whether applying the principle would lead to the formation of teams fit for the type of task the principles described. (cf. R1). The questions could be answered on a 5-point Likert scale, with the answer options "strongly disagree", "disagree", "neither agree nor disagree" "agree", and "strongly agree".
In order to test hypothesis 2a (H2a: classifying teams) we first presented the participants with a preparatory question with 3 near perfect 2x2 teams, each adhering to one of the team formation principles. These were shown in isolation from each other. The examples elaborately explained the application of the team formation principles and asked the participants to classify the teams. Next, from each of the calculated lists of teams for each task type (productive, creative, and learning) and each team size (2x2, 3x3, and 4x4) we selected the three highest scoring teams. We randomly grouped the teams of equal size into 3 sets of 3 teams. The 9 sets (each containing 3 samples) were presented sequentially to the

participants in order of increasing team size. The participant's task was to classify the teams as examples of either productive, creative or learning teams by applying the team formation principles and selecting to which principle the team fitted best. The participants were instructed to only give "No answer" for their answer if they could not decide on any one type.

In order to test hypothesis 2b (H2b: ranking teams), we first presented the participants with an example of the task. Next, from each of the calculated lists of teams for each task type (productive, creative, and learning) and each team size (2x2, 3x3, and 4x4) we selected the teams with the three highest, three most average and three lowest fit values on productivity, creativity, and learning. We randomly ordered the teams of equal task type and equal size into 9 sets of 3 teams. The 9 sets were presented separately to the participants. We started with all 2x2 teams of the individual task types (productive, creative, and learning) and then, while keeping this order, increased team size to 3x3, and finally to 4x4. The participant's task was to rank the teams in accordance to their assessment of the level of adherence the teams showed to the current formation principle. The final question in the survey invited the participants to comment on the survey and their tasks. (See Appendix B for an example of the team formation algorithms output and the ranking and classifying task.)

4.5. Results

The results of the survey among teaching staff (n=56) are presented in the order of the research questions as stated in Section 4.3.

4.5.1. Team formation principles

Our participants were asked to indicate their level of agreement on whether each of the team formation principles would lead to the formation of teams fit for the task type. The results are presented in Figure 4.2.

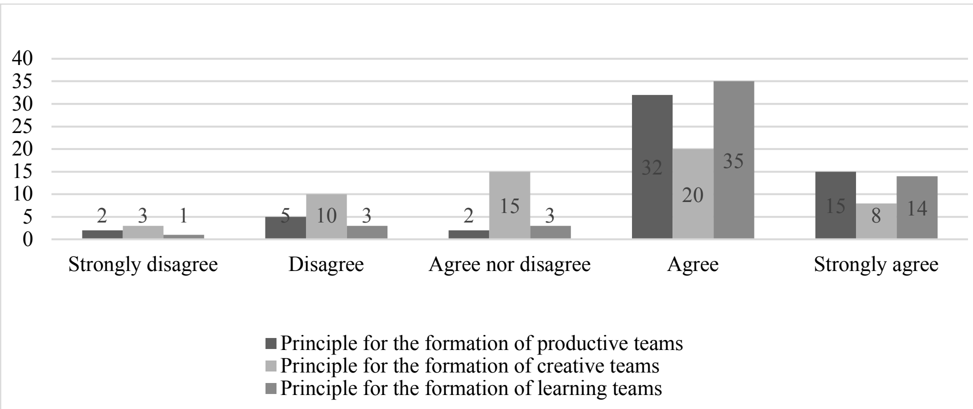


Figure 4.2: *Levels of agreement to the three team formation principles on a 5-point Likert scale.*

The number of participants indicating agreement and strong agreement to the principles for the formation of productive, creative, and learning teams totalled to 47 (84%), 28 (50%) and 49 (88%), respectively.

4.5.2. Classifying teams

As indicated in subsection 4.4.3 above, the actual task of classifying teams was preceded by a preliminary question. It presented constructed examples adhering well to the team formation principles. (See Tables 4.1, 4.2 and 4.3.)

Tables 4.1, 4.2, and 4.3*: Examples of productive, creative and learning teams with two members (L1 and L2), each having a score on 2 knowledge topics (T1 and T2) and one score on conscientiousness (Cons).*

Productive	T1	T2	Cons
L1	8	9	4.32
L2	9	8	4.78

Creative	T1	T2	Cons
L1	8	4	1.67
L2	3	9	2.11

Learning	T1	T2	Cons
L1	9	6	4.33
L2	7	9	4.45

For these teams the team formation algorithms calculated fit values of 0.880, 0.493, and 0.939, respectively.

The numbers of participant's classifying these team in line with the algorithms were: 52 (93%), 54 (96%), and 48 (86%), respectively. In 4 cases the productive team was alternatively classified as a learning team. In 2 cases the creative team was alternatively classified as a learning team. In 2 cases the learning team was alternatively classified as a creative team, and in 6 cases as a productive team.

The next 3 questions asked participants to classify teams of size 2x2. The cumulative results from these tasks are shown in Figure 4.3. The labels on the vertical axis indicate for which classification these team had the highest fit values. As all three types of teams were shown 3 times, the total number of answers on any type of team is 168 (3x56).

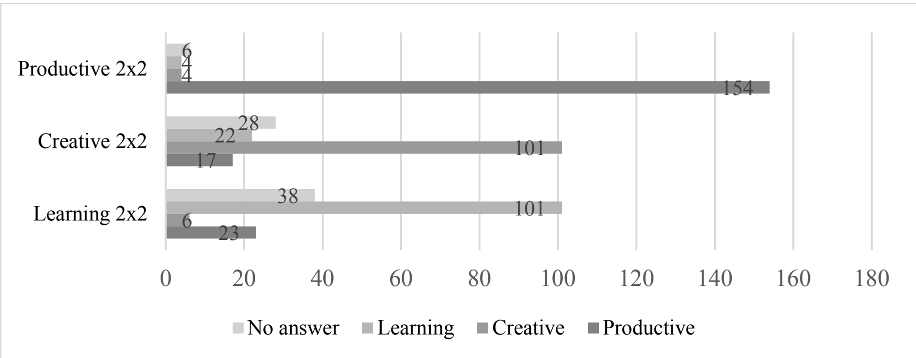


Figure 4.3: *Cumulative results of classifying three 2x2 teams of each type of team.*

The participants classified these 2x2 teams in accordance with the team formation algorithms as follows: Productive 154 (92%), Creative 101 (60%), and Learning 101 (60%).

Additionally, we explored to what extent this task becomes more complex when both team size and number of topics addressed in the project increased. Therefore the next 6 questions we asked to classify 3x3 and 4x4 team respectively. Figures 4.4 and 4.5 depict the results of these tasks.

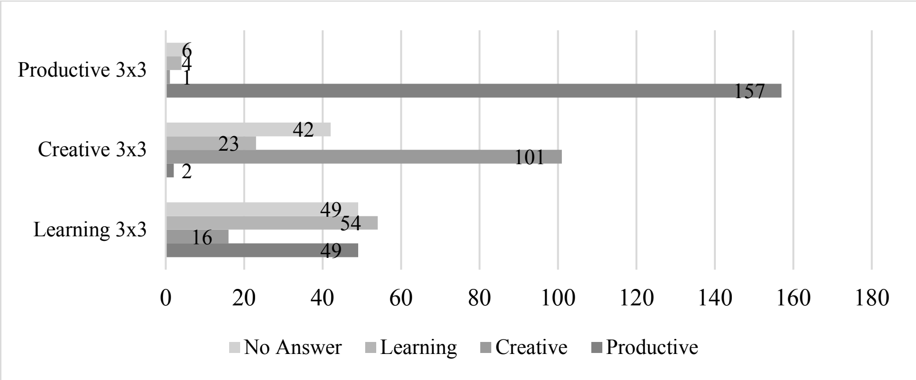


Figure 4.4: *Cumulative results of classifying three 3x3 teams of each type of team.*

The participants classified these 3x3 teams in accordance with the team formation algorithms as follows: Productive 157 (93%), Creative 101 (60%), and Learning 54 (32%).

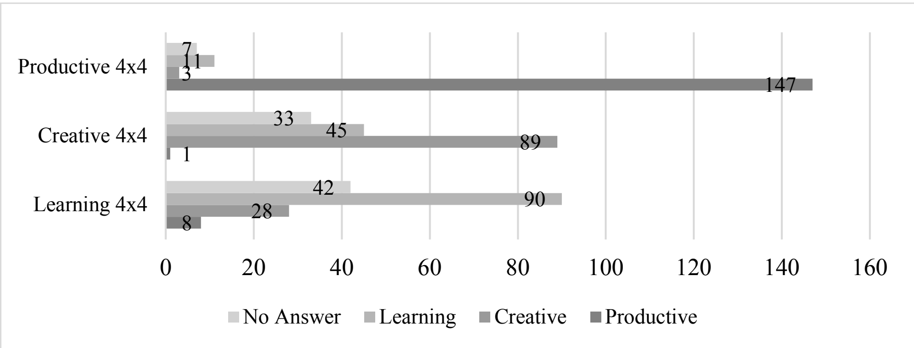


Figure 4.5: *Cumulative results of classifying three 4x4 teams of each type of team.*

The participants classified the 4x4 teams identical to the team formation algorithms as follows: Productive 147 (88%), Creative 89 (53%), and Learning 90 (54%). To allow for easy comparison, the collective results of the 2x2, 3x3 and 4x4 classifying tasks are displayed in Table 4.4.

Table 4.4: *Percentages of classifications identical to the team formation algorithms for three types of teams of sizes 2x2, 3x3, and 4x4, including "no answers".*

Team size	Productive	Creative	Learning
2x2	92 %	60 %	60 %
3x3	93 %	60 %	32 %
4x4	88 %	53 %	54 %

In Table 4.5 we present a breakdown of these results into team sizes, numbers of classifications identical to the algorithm results and numbers and kinds of alternative classifications, including the number of no-answers. The cells in the diagonal from upper left to lower right for each team size represents the classification that aligns with the one calculated by the algorithm, while the other cells in each row represent the type and number of the alternative classifications.

Table 4.5: *Numbers of identical and alternative classifications, and no answers for the three types of teams of size 2x2, 3x3, and 4x4.*

n=168	Productive	Creative	Learning	No answer
Productive 2x2	154	4	4	6
Creative 2x2	17	101	22	28
Learning 2x2	23	6	101	38
Productive 3x3	157	1	4	6
Creative 3x3	2	101	23	42
Learning 3x3	49	16	54	49
Productive 4x4	147	3	11	7
Creative 4x4	1	89	45	33
Learning 4x4	8	28	90	42

Table 4.6 shows the percentages of identical classifications, excluding the "no answers".

Table 4.6: *Percentages of classifications identical with the team formation algorithms for three types of teams of sizes 2x2, 3x3, and 4x4, excluding "no answers".*

Team size	Productive	Creative	Learning
2x2	92 %	72 %	78 %
3x3	97 %	80 %	45 %
4x4	91 %	66 %	71 %

4.5.3. Ranking teams

Our participants ranked teams based on how well they adhered to each individual team formation principle. They did this for 9 sets of 3 teams in the order productive, creative, and learning and with increasing team size. Figure 4.6 shows the results of these ranking tasks.

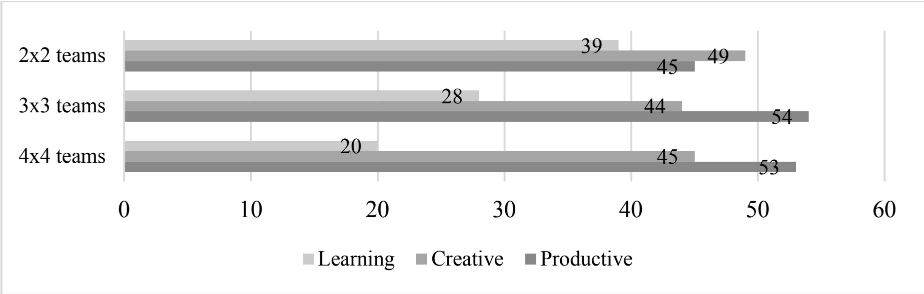


Figure 4.6: *Numbers of rankings of team of sizes 2x2, 3x3, and 4x4 in accordance with the ranking from the team formation algorithms for each type of team.*

The numbers and percentage of rankings matching the ranking from the team formation algorithms for the 2x2 productive teams were 45 (80%), for the creative teams 49 (88%), and for the learning teams 39 (70%).
The collective results are displayed in Table 4.7, which allows easy comparison.

Table 4.7: *Numbers and percentages of rankings of three types of teams in line with the ranking from the team formation algorithms for teams of sizes 2x2, 3x3 and 4x4.*

n=56	Productive	Creative	Learning
2x2 teams	45 (80 %)	49 (88 %)	39 (70 %)
3x3 teams	54 (96 %)	44 (79 %)	28 (50 %)
4x4 teams	53 (95%)	45 (80 %)	20 (36 %)

Our exploration into the complexity of the team formation task showed that the numbers of rankings identical to the results from the team formation algorithms for the 3x3 productive teams were 54 (96%), creative teams 44 (79%), and learning teams 28 (50%). The results for the 4x4 productive teams matching the rankings from the team formation algorithms were 53 (95%), creative teams 45 (80%), and learning teams 20 (36%). Both these sets of results are also shown in Figure 4.6 and Table 4.7. We observe that while the numbers of rankings of productive and creative teams in accordance with the rankings from the team formation algorithms remained roughly at the same level, the number of matching rankings of learning teams showed a considerable drop (from 70% to 36%).

4.5.4. Comments on the survey

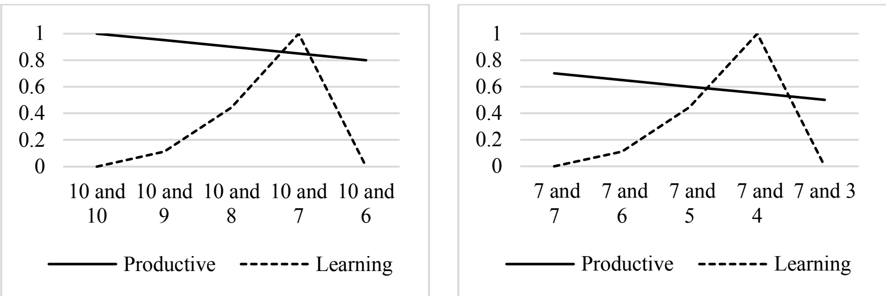
The comments most relevant were as follows: One participant noted that using only conscientiousness as a personality factor would underrepresent personality in the team formation principles. More specific, in line with the low acceptance rate of the principle for forming creative teams, some participants remarked that creativity cannot effectively be described with only knowledge and conscientiousness as factors. Another participant remarked that the survey was difficult to answer, mostly because of having to take into account the use of Vygotsky' zone of proximal development in the team formation principle for learning teams. More specifically, several participants remarked on the heavy mental load the survey put on them.

4.6. Discussion

Given the results of the validation of our first hypothesis (84%, 50% and 88% of the participants agreed with the proposed principles for the formation of productive, creative, and learning teams, respectively), the hypothesis was accepted for productive and learning teams. The principle for the formation of creative teams was not accepted. Likely, this is due to the use of only two factors in the team formation principles: knowledge distributions and conscientiousness levels. This can also be surmised from the comments on the survey. Even though there is evidence for a relationship between conscientiousness and creativity (see e.g., Robert & Cheung, 2010, who show that there is a significant negative relationship between group conscientiousness and group performance on a creative task), this apparently is too narrow a basis for the formation of creative teams. It may be necessary to include additional personality factors, for instance based on Barrick and Mount (1991), who find a relation between creativity, openness to experience (one of the Big Five personality factors), and job performance. One may also have to take into consideration the sub-factors of which conscientiousness is made up. Research by e.g., Barrick, Mount and Strauss (1993) and Reiter-Palmon, Illies and Kobe-Cross (2009) indicates that conscientiousness consists of two components: an achievement component (consisting of the facets competence, achievement striving and self-discipline) and a dependability component (consisting of the facets order, dutifulness and deliberation). Reiter-Palmon et al. (2009) and Kaufman (2011) argue that the achievement component is related to creative job performance, while the dependability component is not. While some researchers suggest that all humans have creative ability, but with different styles and levels (e.g., Kirton, 2003), others search to define factors above and beyond conscientiousness and openness to experience that determine a person's creativity (Kaufman & Sternberg, 2010). Similarly, Sie, Bitter-Rijpkema, Stoyanov & Sloep (2014) show that experts they consulted list open communication, a positive attitude, trust, keeping appointments, and personality as influential factors for cooperation networks for creative innovation. These approaches can inform future research for a better delineation of what makes a team creative and, in case, how to form a creative team.

The validation of the implementation of the principles in algorithms was conducted by means of classifying and ranking tasks. The classifying task aimed at validating hypothesis 2a (Given the same data as used by the algorithms, practitioners from the educational field classify the teams in accordance with the algorithms). The results from the preparatory question with well-formed examples show that participants were able to perform these tasks well when isolated from each other. As the principle for the formation of creative teams was rejected, we focus this part of the discussion on the classification and ranking of productive and learning teams. The results obtained from the actual classifying tasks show that participants could classify productive teams of all sizes successfully. With regard to the learning teams, the results show a fair amount, i.e. 60%, of classifications align with the algorithm for 2x2 teams. At the same time it shows that the participants find this task difficult, and even more so for the 3x3 and 4x4 teams. Table 4.5 shows that both the numbers of alternative classifications and missing answers increase when the team size and the number of topics addressed in the project rise. This can be explained by e.g., the phenomenon of bounded rationality, in which time constraints and limited human power of abstraction hinder the rational decision making process (Gigerenzer & Selten, 2002). Several aspects of the classifying task relate to this phenomenon:

- The participants had to mentally apply the variables from all three team formation principles to the teams shown and consider each result to come to a classification. For both productive and creative teams only two basic rules applied, while for learning teams considerably more rules had to be taken into account to decide on a classification.
- The number of team members and topics addressed in the project increased. This had the effect that overall the numbers of classifications in accordance with the team formation algorithm results declined (see Table 4.1).
- The participants were allowed to indicate that they could not arrive at a conclusion by selecting "no answer". As can be seen in Table 4.5, the numbers of "no answers" were highest for learning teams of any size. From Table 4.5 it can also be observed that overall the numbers of "no answer" increased as soon as the team size was larger than 2x2.
- In specific circumstances the difference in fit to a principle between productive and learning teams is minimal: The productive teams receive high fit values when knowledge scores are high, while the learning team receive high fit values when knowledge scores have an optimum difference, so also when knowledge scores are high. This effect is demonstrated in the Figures 4.7 and 4.8.



Figures 4.7 and 4.8: *Fit values from the productive and learning algorithms for pairs of knowledge scores of two team members on one topic, for high and for average knowledge scores.*

For the algorithms even a very small difference is sufficient to make a distinction, obviously this is not the case for the participants. It then depends on the data available whether the examples drawn from it provide sufficient basis for clearly distinctive teams. As is shown in the Figures 4.9, 4.10, and 4.11 the differences between productive and learning teams were often limited, specifically with the 3x3 teams.

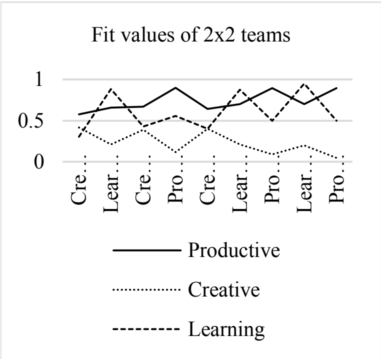


Figure 4.9: *Fit values for productive, creative and learning teams of size 2x2 shown in the classifying part of the survey.*

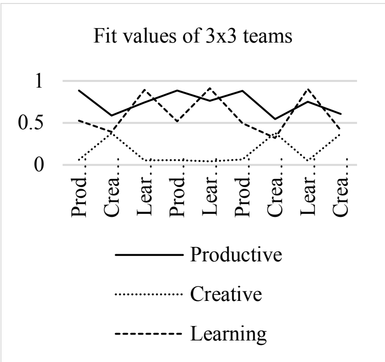


Figure 4.10: *Fit values for productive, creative and learning teams of size 3x3 shown in the classifying part of the survey.*

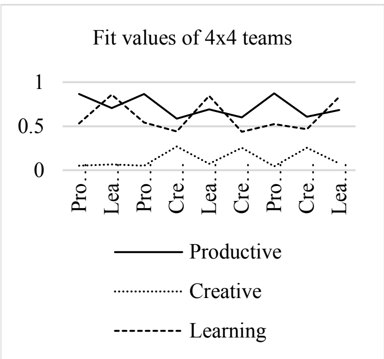


Figure 4.11: *Fit values for productive, creative and learning teams of size 4x4 shown in the classifying part of the survey.*

This is reflected in the classifications given by the participants for 3x3 learning teams (see Tables 4.4, 4.5, and 4.6). With respect to hypothesis 2a this indicates that by and large human assessors do classify teams in ways identical to the classifications from the algorithms.
The ranking task showed similar, but not as many, complicating aspects. In this task only one team formation principle had to be considered at any time, but the team size and numbers of topics increased. The results presented in Table 4.7 reflect the complexity of the tasks, as the numbers of rankings identical to the ranking from the team formation algorithms for the principle for productive and creative teams remained roughly on the same level, while the numbers of rankings of the learning teams dropped from 70% to 50% to 36% when the team size went up. With respect to hypothesis 2b (Given the same data as used by the algorithms, practitioners from the educational field rank the teams in accordance with the algorithms) this indicates that human assessors rank teams in accordance with the team formation principles. But again, for learning teams, human performance drops considerably when the task complexity rises.

4.7. Conclusions and directions for future research

From the perspective of improving collaborative learning opportunities in open learning environments, we researched a possible operationalisation of Stahl's CSCL framework in open learning environments, such as MOOCs. We did this by investigating and discussing an automated team formation service for project-based learning. We provided several reasons why the provision of such a service can be beneficial for both learners and support staff. Among them are the benefits of collaborative learning with respect to motivation, and therewith for lowering drop-out. As providing support for such learning settings in open learning environments can be resource-intensive for staff, we introduced several team formation instruments. These exist of team formation principles and algorithms that set the stage for the formation of effective productive, creative, and learning teams. The algorithms use data on learner knowledge, personality and preferences to form teams. It was stressed that first and foremost any implementation requires validated instruments. We therefore presented an experiment aimed at validating the principles and the results from the team formation algorithms. Participants were recruited from practitioners in the educational field. The results from an acceptance test of the three team formation principles demonstrated that the principles for the formation of productive teams and for the formation of learning teams received wide support. We discussed the possible steps to take to further refine the principle for the formation of creative teams. In order to make sure the team formation algorithms could perform in settings with large groups of learners we gathered a large amount of real world data on learner knowledge, personality and preferences (n=168). The processing of these data proved the algorithms' robustness. The validation of the results of the team formation algorithms followed a dual approach. Based on their understanding of how the team formation principles work, practitioners first classified teams into three types (productive, creative and learning teams). Next, they ranked teams within one type of team. The results of these classifying and ranking tasks showed that the participants classified small productive, creative and learning teams largely in accord with how our algorithms judged these teams. However, with increased task complexity, especially when learning teams were concerned, increased divergence occurred between classifications and rankings resulting from the algorithms and classifications and rankings performed by human assessors.

The team formation principles for productive and learning teams were accepted and thus form a validated basis for team formation. The results of human application of the principles and the results of automated application of the principles largely overlap, but only when small teams are concerned. We take this as an indication of the usefulness of automating the team formation task rather than as a sign of their breakdown at larger team sizes. Information overload and bounds to people's rational capabilities make it difficult to deal with this kind of complexity (Sie et al., 2014).
While our team formation principles are based on the important aspects of knowledge and personality, this doesn't exclude that other aspects could be taken into account to form teams. For example, Zhang, De Pablos, & Xu (2014) examine the effects of cultural differences on the knowledge sharing processes in multi-national virtual classes. As a result of focussing on team formation prior to actual team work, we currently do not take into account research into actual collaborations inside teams (see e.g., Zhang, X., de Pablos, Zhang, Y., 2012).

Our overall conclusion, however, is that we found clear support for both the current team formation principles for productive and learning teams and the correct implementation of these principles in our team formation algorithms.

In order to support the learning aspects of open learning environments, our future research will focus on the implementation of project-based learning and team formation instruments. Personality tests suitable for our purpose are publicly available and we can easily implement a preferences filter. We therefore focus on the automated assessments of both the fit of projects to the knowledge domain, and of the knowledge available with learners. As we aim to implement the service in learning settings, we will restrict ourselves to forming learning teams only. We will investigate whether the formation of teams following the principle for learning teams can demonstrably foster learning.

CHAPTER 5

LSA-based Project Team Formation in MOOCS[4]

"The meaning of that which by its nature is understandable, as has been demonstrated repetitiously, can only be grasped from the context of the utterings."

(Benedictus De Spinoza, Tractatus theologico-politicus, 1670)

[4] This chapter is based on: Spoelstra, H., Van Rosmalen, P., Houtmans, T., Van Bruggen, J., & Sloep, P. (2015). LSA-based Project Team Formation in MOOCS. (Submitted)

Abstract

After introducing a model for team formation for project-based learning, this chapter discusses a set of automated services which allow project-based learning to scale up to large, open learning environments, such as MOOCs. The services, by using a team formation principle for the formation of learning teams, also take care of placement in a team. In an experimental setting, the knowledge contained in a domain of study is modelled with latent semantic analysis. This affords one i) to assess projects for their fit in the domain, ii) to form teams with members selected based on differences in prior knowledge, and iii) to recommend learning materials. Results show that assessing project fit and prior knowledge is attainable and that learning does indeed occur in teams formed according to the team formation principle. Furthermore, learners highly value the recommended learning materials. The findings are discussed and we conclude that it is feasible to implement team formation services for project-based learning in open learning environments. Suggestions for future research are included.

5.1. Introduction

Large numbers of learners are attracted to open learning environments, such as Massive Open Online Courses (MOOCs). These initially were intended for *connectivist* learning (cMOOCs), which emphasises collaboration, knowledge co-creation, and learner self-direction (Siemens, 2004; Downes, 2006). Many of the later MOOCs were built on principles of *behaviourist* or *mastery learning* (xMOOCs) (Coursera, 2014), which focusses more on knowledge reproduction. Both flavours, however, make it hard for learners to collaborate. With respect to connectivist MOOCs, Kop, Fournier, and Mak (2011) noted that: “Many participants realized the importance of connections with other learners and of relationship building to advance learning. However, in a MOOC, they found these things extremely hard.” Regarding behaviourism-based MOOCs, Daniel (2012) and Edinburgh University (2013) report limited opportunities for learner collaboration. Moreover, in general, in MOOCs drop-out rates are high, sometimes up to 90%[5]. These problems did not go unnoticed, and thus recent initiatives seek to improve support for collaboration between learners. With respect to group formation in particular, initiatives often only allow self-selection of co-learners (The Open University, 2013; NovoEd, 2014). Several researchers, however, hold that self-selection should be discouraged when effective teams are to be formed (Fiechtner & Davis, 1985; Oakley, Felder, Brent, & Elhajj, 2004). A basis for the formation of effective learning teams can be found in Clarà and Barberà (2013). They propose that: “internalization [of a representation of knowledge] is possible only if the learner has the opportunity of using the representation jointly with others within a zone of proximal development”. Team formation theory emphasises that besides knowledge, *personality* is an important factor in the formation of effective teams (Barrick, Stewart, Neubert, & Mount, 1998; Obaya, 1999; Oakley, Felder, Brent, & Elhajj, 2004; Humphrey, Hollenbeck, Meyer, & Ilgen, 2007).

These team formation principles fit well into social constructivist approaches to collaborative project-based learning (PBL) (Bell, 2010). Some of PBL’s benefits are:

- a positive effect on lowering drop-out (Fisher & Baird, 2005; Dahms & Stentoft, 2008)
- improvements in the learners’ motivation, so that learners are more inclined to deal with hard, complex problems and spend more time studying (Johnson, Johnson, Stanne & Garibaldi, 1990; Marin-Garcia & Lloret, 2008)
- the creation of realistic (and possibly inter-professional) learning experiences (Felder, Felder & Dietz, 1999; Springer, Stanne & Donovan, 1999), which fit well with many current work practices.
- optimised learning outcomes when team formation is based on knowledge differences between learners within each member’s “zone of proximal

[5] Kathy Jordan maintains an expanding data set that graphs completion rates versus MOOC size: http://www.katyjordan.com/MOOCproject.html

development" (zpd) (Vygotsky, 1978; Murray & Arroyo, 2002; Chihaia, 2007) and personality (Obaya; 1999; Oakley, Felder, Brent, & Elhajj, 2004).

It is because of these considerations that we argue that large-scale open learning environments, such as MOOCs, can benefit from implementing PBL design principles and team formation theory. In such environments, however, doing so would rapidly overburden staff (an instance of the teacher bandwidth problem, Wiley and Edwards, 2002). After all, staff who possess the required domain knowledge and team formation expertise needs a fixed amount of time per project to assess differences in knowledge and personality aspects between learners to form teams. As we have argued elsewhere, team formation is an inherently difficult and thus time-consuming task for humans to carry out (Spoelstra, Van Rosmalen, Houtmans, & Sloep, 2015). Alternatively, the learners themselves or third parties could suggest projects. As these also require assessment for their fit in the knowledge domain and formation of teams, selecting suitable learning projects and teams again involves a lot of staff input. In this chapter we therefore investigate how one might automate these processes. To this end, Spoelstra, Van Rosmalen, Van de Vrie, Obreza, and Sloep (2013) modelled the PBL setup and team formation process. An updated version, tailored specifically to the formation of learning teams is presented in Figure 5.1. In this figure, the solid arrows describe the steps one takes to put together a team for a particular project in a knowledge domain and recommend learning materials. The dotted arrows refer back to the elements of the model against which assessments are made.
In our earlier work (Spoelstra, Van Rosmalen, & Sloep, 2014; Spoelstra, Van Rosmalen, Houtmans, & Sloep, 2014) several components of this model were developed and tested. They consist of the definition of learner preferences filter [4] on e.g., preferred collaboration language, availability, etc., the assessment method for learner personality [5] by means of the Big Five personality test (Barrick & Mount, 1991), the team formation principle for the formation of learning teams [7] (see Appendix A), and the algorithm implementing this principle. To complete the suit of instruments required to implement the model, in this chapter we focus on the knowledge-related aspects of the model:

i) the development of a knowledge domain model [1], which is constructed from the learning materials in a domain of study
ii) the definition of projects [2] for students to work on (the defining element of PBL), which contain a general description of the project task and descriptions of the topics addressed in the project
iii) the assessment of these project topics for their fit to the knowledge domain [3]
iv) the assessment of learner-provided evidence for prior knowledge on these topics [6] so the team formation principle can be applied.

In order to investigate whether the model can be fully applied and indeed leads to learning gains, we additionally investigate:

v) what are the learning effects from team work in teams formed on the basis of knowledge differences [9]
vi) whether the personality aspect of conscientiousness has an effect on learning and/or the learning process [5]
vii) whether we can suggest learning materials to learners [10].

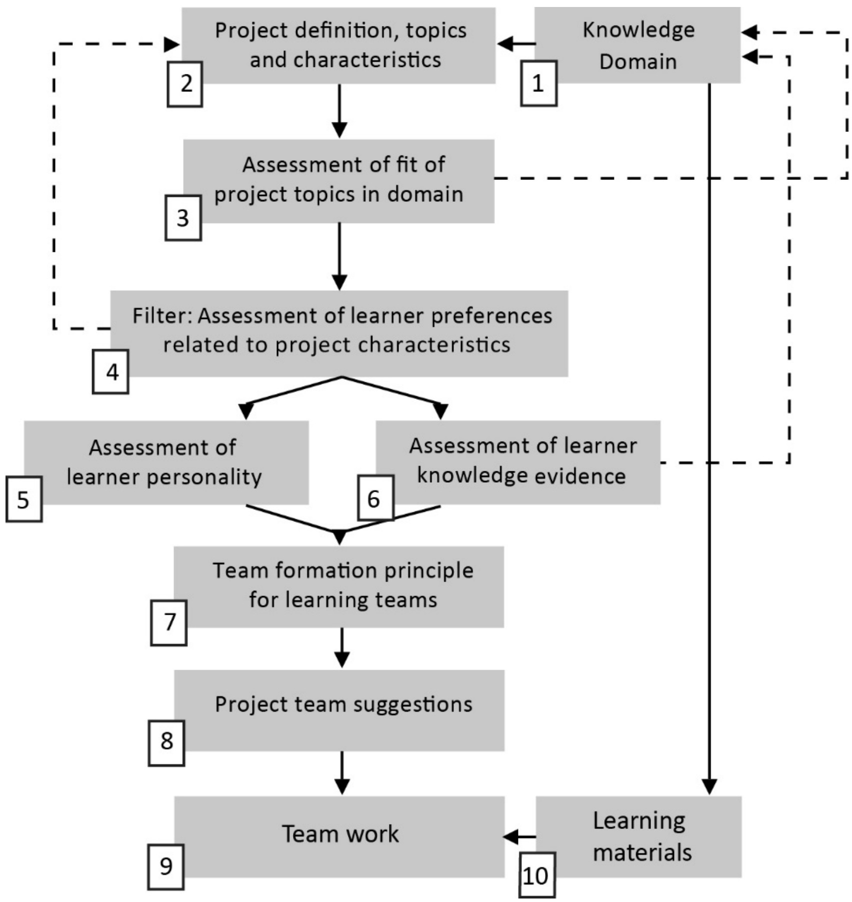


Figure 5.1: *Team formation for project-based learning model.*

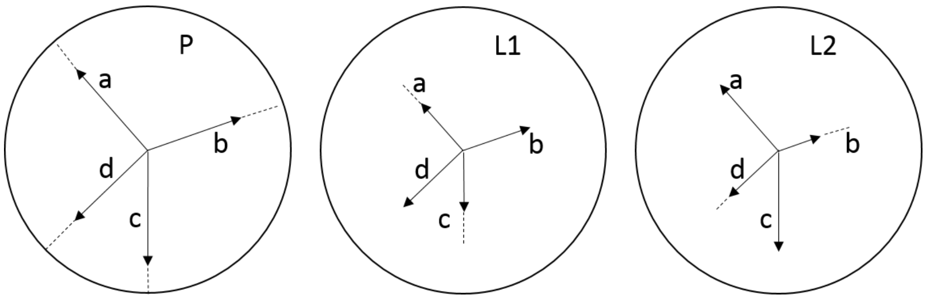
Our research heavily relies on the method of latent semantic analysis (LSA; Landauer, Foltz, & Laham, 1998), which has shown to be quite capable of approximating human assessments of knowledge. Therefore, in the next section, we first discuss prior research deploying LSA and our intended use of LSA. Thereafter, the chapter is structured as follows: In Section 5.3 we present our research questions with respect to setting up an experimental environment, assessing learner knowledge, the effects on learning from collaboration in teams formed based on knowledge differences, and the possibility of recommending learning materials. Section 5.4 describes the creation of a knowledge domain, the creation of project descriptions, and the research methods we applied. In Section 5.5, we present the results we obtained. These are discussed in Section 5.6. Finally, Section 5.7 draws conclusions and presents directions for future research.

5.2. LSA

The applicability of LSA for assessing knowledge and learning gains related to a knowledge domain has been demonstrated extensively in various application domains. *For HRM purposes*, Laham, Bennett and Landauer (2000) matched individuals' work track records with job descriptions. Texts about people's education, occupations, and task-experience were extracted from text-containing databases. These texts were used to create a semantic space. The system could match or compare any of these texts with one or more of the others. In an experiment with three job descriptions, they measured the fit of each worker to each task described and it was estimated how well each worker could replace another. They also showed LSA's potential for matching knowledge needed for new jobs with knowledge contained in training materials and with knowledge already possessed by individual workers. They concluded that LSA can successfully characterize tasks, occupations and personnel, and measure the overlap in content between instructional courses covering the full range of tasks performed in many different occupations. With respect to *learning text selection*, Dessus, Lemaire, Loiseau, Mandin, Villiot-Leclercq and Zampa (2011) described RAFALES, a system aimed at providing learning texts to learners based on an analysis of their current proficiency on a topic (i.e., which texts they had previously studied). They introduced the principle of Optimal Proximity for Acquisition (OPA) as an implementation of Vygotsky's principle of "zone of proximal development". The OPA determines the optimal semantic difference between a learner's current level of advancement through the learning materials and the next step to be taken. This allows the learner to benefit from optimal learning. The system used an expert text as reading target for the learner to reach. *To find and select peer-tutors*, Van Rosmalen, Sloep, Kester, Brouns, De Croock, Pannekeet, et al. (2008) used LSA to map learner questions onto knowledge available in a network and to find learners capable of answering these questions. The system they developed, called A Tutor Locator (ATL), maps questions onto a collection of text fragments representative of the knowledge domain. ATL returns correlations between the question and text fragments. It then selects peer-students who already studied the topics related to these text fragments, assuming they are the most suited to provide an answer. Besides being based on knowledge, the tutor selection is also based on tutor competency, availability and eligibility. *To measure prior knowledge and learning effects*, Wolfe, Schreiner, Rehder, Laham, Foltz, Kintsch, et al. (1998) and Rehder, Schreiner, Wolfe, Laham, Landauer and Kintsch (1998) demonstrated that LSA can assess that learning results differ based on the difference between the level of prior knowledge and the level of difficulty of texts presented to learners (they coined the term "zone of learnability" to indicate effective knowledge differences).

5.2.1. Our intended use of LSA

Figures 5.2a, b and c show schematically how we intend to use LSA to provide data for setting up the team formation process for PBL. In these figures, the circle perimeters represent documents containing the full textual knowledge on the various topics addressed in a knowledge domain. The four vectors (a, b, c, and d) in Figure 2a indicate that a project P addresses four of these topics (LSA computes such vectors). The *directions* of the vectors refer to the contents of four particular *sets* of (an arbitrary number of) documents containing the knowledge on those topics in the domain. The *length* of the vectors indicates how closely project topic descriptions resemble the knowledge available in the domain. The dotted lines extending from the vector endpoints indicate the topic knowledge available in the domain not addressed by the topic descriptions. The sets of documents to which the four vectors refer form the basis on which we assess learner-provided prior knowledge evidence. In the Figures 2b and 2c the four vectors indicate how the topic knowledge evidence provided by learners L1 and L2 relates to the four topics depicted in Fig. 2a. Again, the length of the vectors indicates to which extent the learners have knowledge of the content of the set of topic documents in the domain.



Figures 5.2a, b and c: *The relations of project topic descriptions to the knowledge domain (5.2a), learner's L1 and L2 knowledge of the project topics and mutual learning opportunities (5.2b and 5.2c).*

In Figure 5.2b, the dotted lines extending from the vectors "a" and "c" for L1 indicate the knowledge difference between L1 and L2 on those topics. In Figure 5.2c, the dotted lines extending from the vectors "b" and "d" for L2 indicate the knowledge difference between L2 and L1 on those topics. According to the team formation principle for learning teams (see Appendix A), these knowledge differences determine whether learners fit together well (i.e., their knowledge differences are within each other's zones of proximal development). Additionally, since neither learner shows full knowledge on the project topics (their vectors, even when combined, do not reach the outer circle), the topic related domain documents may be recommended to learners as learning materials.

5.3. Research questions

As the application of LSA requires preparations, the questions we aim to answer are divided into a preliminary question and several research questions. The preliminary question (PQ1) addresses setting up a knowledge domain environment in which project topic descriptions can be assessed for their fit in the domain, while the research questions (RQ1, RQ2, RQ3, and RQ4) address team formation, the learning effects in teams formed according to the team formation principle, and the recommendation of learning materials to learners.

PQ1: Using LSA, can we construct a knowledge domain in such a way that we can adequately determine the level of fit of projects to the domain?

This question will be answered by using standard procedures in preparing materials for LSA and experimentally finding adequate LSA processing settings at which the knowledge domain, upon being queried with project topic descriptions, returns a number of sufficiently related documents. Domain experts will be asked whether the retrieved document sets do indeed show sufficient relevance to the topics.

RQ 1: Can we adequately determine the extent to which prospective team members have different levels of prior knowledge?

This question will be answered by using learner-provided knowledge evidence on project topics as LSA queries and asking domain experts whether LSA has adequately classified texts according to their extent of topical knowledge.
If both the PQ1 and RQ1 can be answered affirmatively, we have a sufficient basis on which to build a PBL and team formation service (assuming we bring into the mix the outcomes of our earlier research on the principle and implementation of a team formation algorithm).
Three issues related to the implementation of the model remain open, though. First, the team formation principle for learning teams (see Figure A3 in Appendix A) utilises a parameter “zpd”. This parameter takes as its value a knowledge difference between learners at which (also adult) learning (Bonk & Kim, 1998; Huang, 2002) would be most effective. In the LSA- research into learning to which we referred above some form of “golden standard” is utilised against which an optimum knowledge difference for all individual learners is measured. In our approach to team formation, however, we explore a scenario in which members in teams are also each other’s teachers. Therefore, there is no single golden standard against

which we can assess an optimum knowledge difference. We want to find out if we can find indications for the location (in terms of knowledge difference) of the "zpd" in such settings. Second, besides prior knowledge, the team formation model utilises learner personality (in the form of learner conscientiousness levels) as the second important factor in team formation. We do not so far have evidence of the influence learner conscientiousness has on team performance; thus we intend to investigate effects conscientiousness levels have on learning and the collaboration process. Third, in Section 5.2.1 we considered the possibility to provide learners with learning materials. Therefore we want to find out whether recommended project-related learning materials are indeed appreciated by the learners. This leads to the formulation of the following additional research questions:

RQ 2: Can we determine a knowledge difference between learners ('zone of proximal development') at which learning is most effective?

This question will be answered by relating the learners' knowledge gains to the difference in knowledge between them and their peer-teachers. To this end, we pair learners in such a way that knowledge differences between them show a steady decline. This enables us to determine if and at which knowledge difference value the highest learning gains are achieved. This difference can then be used as the value for the parameter "zpd" in the team formation algorithm.

RQ 3: How does the personality factor 'conscientiousness' impact on learning and the interaction process between learners?

This question will be answered by relating conscientiousness scores to learner knowledge and knowledge gains and possible effects on the collaboration process during the experiment.

RQ 4: Can we suggest learning materials from inside the knowledge domain to learners in such a way that learners consider these materials relevant to the project they work on?

We answer this question by asking learners whether the learning materials we suggest are valuable learning sources for the topics in their projects.

5.4. Method

5.4.1. Construction of the knowledge domain and the project definitions

The preparatory steps of construction the knowledge domain and construction the project definitions are closely intertwined. It is important to notice that the parameters chosen for the construction of the domain influence the results one can

achieve when the domain is queried. The analysis of results of queries can be used to improve the parameter settings with which the domain is created. This process of creating a knowledge domain and adequate project topic descriptions is thus one of mutual optimisation. For the experiment's knowledge domain we selected the entry course from the Psychology curriculum "Introductie in de Psychologie" [Introduction into Psychology]. This course contained 18 chapters, from which we extracted the textual elements. These texts were processed with LSA, using Text-to-Matrix-Generator (TMG) 6.07 (Zeimpekis & Gallopoulos, 2006), implemented in Matlab 2007b (The MathWorks, 2007).
We defined two projects within the knowledge domain of introductory psychology. One was concerned with "Eyesight" and focussed on four topics related to eyesight: "The brain", "The construction and workings of the eye", "Solving problems with focusing" and "Seeing depth". These topics were mainly addressed in the chapters 2 and 3 of the course materials. The other project was concerned with "Mental disorders", and focussed on four topics related to mental disorders: "What are mental disorders", 'Factors related to mental disorders", "The DSM" (Diagnostic and Statistical Manual of Mental Disorders), and "The DSM, the anti-social personality and psychopathy". These topics were addressed mainly in the chapters 12 and 13 of the course materials. For reasons of brevity, in the remainder of this chapter we will often refer to these topics as Brain, Eye, Focussing, Depth, Disorders, Factors, DSM, and Psychopaths, respectively. For each of the eight topics we drafted descriptions (typically around 200 words each) by paraphrasing relevant sentences from the corpus documents. We then used these topic descriptions as LSA queries into the knowledge domain. For each query this resulted in a list of semantically related documents from the domain. As our aim was to find as many related (which we defined as stemming from the corresponding book chapters) documents as possible, we experimented with the number of documents in the corpus, the LSA weighting scheme and dimensionality reduction settings to reach this optimum. In order to confirm whether LSA had indeed found relevant documents to our topic descriptions, two randomly chosen topic descriptions with the documents that were retrieved based were presented to four staff members (teachers) from the Psychology faculty. They could indicate their relevance assessments on a 7-point Likert scale. We determined whether the raters consistently gave higher relevance to documents with higher LSA-determined relevance scores by calculating interclass correlation coefficients (ICC).

5.4.2. Experiment set up

A group of 2678 learners was invited to participate. They were told that, by way of assignment, they would have to produce an information leaflet for one of the two projects we defined. The group consisted of all learners currently active in the course and all learners who had studied the course in the previous year (either passing the exam or not). Participation in a survey acted as enrolment into the experiment. The survey noted: gender, the number of course chapters studied, and which chapter they studied last. In order to determine the participants' conscientiousness levels it also contained a full Big Five personality test (Barrick & Mount, 1991), validated for the Dutch language (Denissen, Geenen, Van Aken, Gosling, Samuel, & Potter, 2008). In total 158 participants completed the enrolment survey. Of these, 124 participants followed through the experiment till its conclusion (eventual response rate 4.6%). The experiment ran from the 30th of March 2014 to the 9th of May 2014, during which period three assignments were given. The participants had seven days to hand in assignment 1 (before the 7th of April 2014), ten days to hand in assignment 2 (before the 27th of April), and ten days to hand in assignment 3 (before the 9th of May, 2014).

After their enrolment, we randomly assigned the participants to one of the two projects. Next, we sent both groups of participants their project definitions. In both cases the introductions suggested that the participants were workers in a consultancy firm (advertising itself to its workers as a learning organisation), which had received a request to write an information leaflet. The commissioner, however, wanted a proof of knowledge before the project would actually be granted. In order to find the best teams for the job, all workers were asked to provide written evidence of their knowledge on the project topics.

Before *assignment 1* of the experiment we selected between four and six keywords representing the central concepts addressed in each project's four topic descriptions. These were presented to the corresponding groups of participants to act as primers on which to base their prior knowledge evidences. For example, for the topic "Brain" of the project "Eyesight" we selected: "Central nervous system", "Peripheral nervous system", "Neurons", "Neurotransmitters", and "All-or-none law". The participants were asked to provide evidence of their knowledge on their four project topics, each of which they based on the keywords we provided. The participants were instructed to limit themselves to 200 words per topic. By way of pre-test, we processed all knowledge evidence texts we received with LSA, using the texts as queries. To calculate the learners' *knowledge scores* on their topics, we compared the document numbers of their 15 highest LSA results with the document numbers of the 15 highest LSA results from the topic descriptions, and divided the average of the LSA results of the documents occurring in both results by the number of documents in the domain document set. Table 5.1 shows an example of such a calculation for one topic.

Table 5.1: *A domain document set (Doc set) and the LSA results (LSA) on the documents inside this set to which a participant's text referred (truncated to 2 decimals), and the knowledge score (KS) (see main text) calculated for the participant.*

Doc set	103	104	105	106	109	110	115	119	130	131	132	134	373	664	1308	KS
LSA	0.30	0.34	0.34	-	0.33	0.36	0.35	0.39	0.34	0.34	-	-	-	0.31	0.31	0.25

In this table, the domain document set shows the numbers of the documents in the LSA space to which a project topic description referred. A learner's text was then

used as query in the LSA domain. The empty cells in the second row of Table 5.1 show that the learner's text did not match on those documents in the topic document set. These indicate gaps in the participant's knowledge on the topic. A participant's knowledge score was calculated as the average of the LSA results divided by the maximum number of matches. The truncated result is: 3.71/15 = 0.25. All learner knowledge scores were calculated in this way. A learner's knowledge score thus reflects knowledge at the aggregate level of the topic, not at the level of the underlying documents.
For *assignment 2* we aimed to mimic the learner/peer-teacher relations between members in learning teams. The participants were informed that the fictitious consultancy firm aimed at refining the proof of knowledge, while allowing workers to learn from each other. In order to investigate learning effects at the level of the individual (which is at the core of the team formation principle for learning teams), we formed duos of participants. We coupled participants with lower knowledge scores on topics with participants with higher knowledge scores in such a way that, over duos, the knowledge score differences between its members gradually declined. This allowed us to observe the effects of variation in differences in knowledge scores between learners/peer-teachers (sizing up the "zone of proximal development"). To ensure that all participants could be part of a duo in the role of learner, we formed these duos across the four topics each project addressed. Thus everybody acted as both learner and peer-teacher. We returned to the participants (in their role as learner) their own text on a topic and the text by their duo partner (in their role as peer-teacher) on the same topic. As intervention, we asked the participants to rewrite their initial text based on what they thought could be improved from reading their peer-teacher's text and then to send in their new knowledge evidences. We then calculated the knowledge scores for the new knowledge evidences. These acted as post-test. To calculate *knowledge gains*, the old knowledge scores were subtracted from the new knowledge scores. Whether knowledge gains were significant was determined by performing paired t-tests on these dyads of knowledge scores.
At two moments we sought confirmation of the LSA results: 1) In order to confirm whether the LSA-based knowledge scores had indeed allowed us to form duos in which one member had less topic knowledge than the other, we presented four staff members of the Psychology faculty with all eight sets of couples of texts with the highest knowledge score differences, one set from each topic. Each set was ordered on knowledge score (lower scoring text first, higher scoring text last). The staff members were asked whether they agreed with the proposition "the second text exhibits more knowledge of the topic at hand than the first". The answer options used a 7-point Likert scale (with options ranging from strongly disagree to strongly agree). 2) In order to confirm whether a leaner's second text exhibited more knowledge on the topic than their first text, we presented the staff members with the pairs of first and second texts from the learners with the highest knowledge gains; we did so for all eight topics. They were again asked whether they

agreed with the proposition “the second text exhibits more knowledge of the topic than the first”. Again, the answer options existed of a 7-point Likert scale (with options ranging from strongly disagree to strongly agree).
To find out whether the LSA retrieval results from the topic descriptions could be used to recommend learning materials, in *assignment 3*, we sent the two groups of participants four sets of five documents related to the four topics in their assigned projects (five documents per set, consisting of the documents with the highest LSA scores from the domain document sets) and asked them whether they thought the document were relevant to the topics on which they had provided knowledge evidence in assignment 1. The answer options used a 5-point Likert scale which ranged from 1 (not relevant) to 5 (highly relevant).
Finally, we wanted to find out whether the learner’s conscientiousness levels had any effect on their knowledge scores, or the collaboration process. We therefore performed paired t-tests on their conscientiousness levels, combined with the knowledge scores from their first texts, their second texts, and knowledge gains, respectively. We also compared their conscientiousness levels with the dates at which assignments were handed in.

5.5. Results

The results below are presented in the order of the research questions from Section 5.3. For the reader’s convenience, the research questions are repeated in the title of each subsection below. The results use either *LSA scores* (expressed in a cosine value between vectors, where 1 indicates a 100% semantic similarity between texts and 0 indicates a 0% semantic similarity between texts) or *knowledge scores* (which were calculated as explained above).

5.5.1. Using LSA, can we construct a knowledge domain in such a way that we can adequately determine the level of fit of projects to the domain?

As already mentioned in Section 5.4.1, the preparation of a knowledge domain with LSA requires calibration. For this calibration we draw on both common practice described in literature (Landauer, McNamara, Dennis, & Kintsch, 2013) and our own experiences with setting up LSA knowledge domains (Van Rosmalen, Sloep, Brouns, Kester, Koné, & Koper, 2006; Kalz, Van Bruggen, Giesbers, Waterink, Eshuis, & Koper, 2014). LSA results are influenced by a number of settings. Some of these are the number of documents from which the knowledge domain is constructed, the weighting of local terms and global terms, and the number of dimensions in which the knowledge domain documents are represented. Our aim was to retrieve the largest number of documents stemming from the book chapters from which we derived our topic descriptions (chapters 2 and 3 for project “Eyesight”, and chapters 12 and 13 for project “Mental disorders”). We pursued this by querying the LSA space with a topic description and empirically determining the number of texts into which the learning materials would be decomposed, the most

promising weighting scheme and number of dimensions to represent our domain documents. A combination of decomposing the learning materials (consisting of 18 chapters, written in the Dutch language) into a corpus consisting of 2257 in themselves meaningful text documents; the TF*IDF (term frequency * inverse document frequency) weighting scheme and a setting of 250 dimensions provided the best results: After about the 15th result, the relevance of the documents retrieved decreased (the documents retrieved started originating from chapters outside the chapters 2-3 and 12-13). Hence, we set each of our topic descriptions to refer to a set of 15 domain documents. As our projects each addressed four topics, a project as a whole thus referred to a maximum of 60 domain documents (assuming documents referred to over the four topics do not overlap). (For a more elaborate overview of the LSA settings we used and of some additional post-processing corpus characteristics, see Appendix C.)
Figures 5.3 and 5.4 depicted the LSA cosine values of the 15 highest ranking documents related to the project description of the projects Eyesight and Mental disorders, respectively.

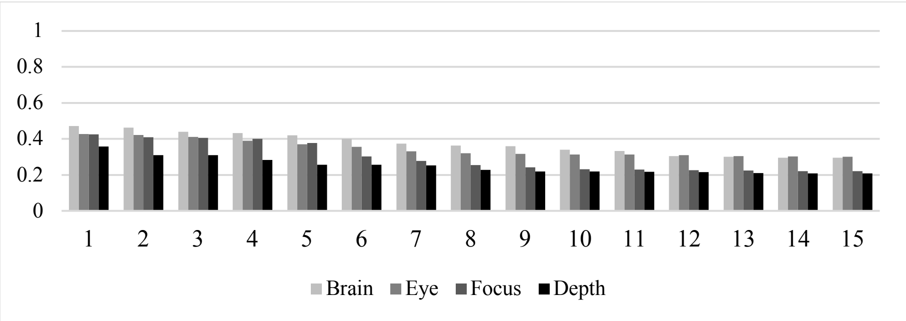


Figure 5.3: *LSA scores of the four topic descriptions on the 4 sets of 15 domain documents for the project "Eyesight".*

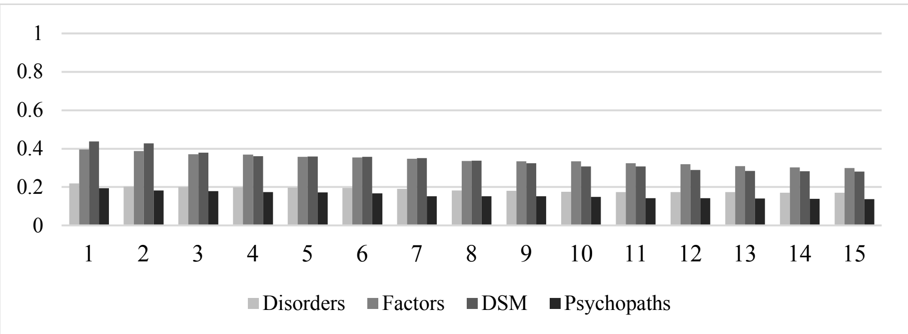
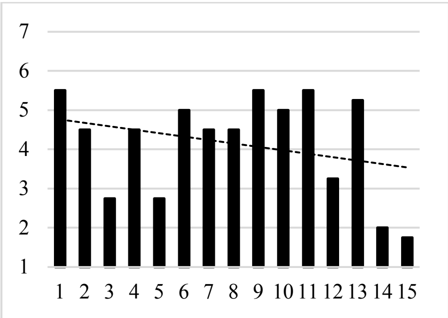
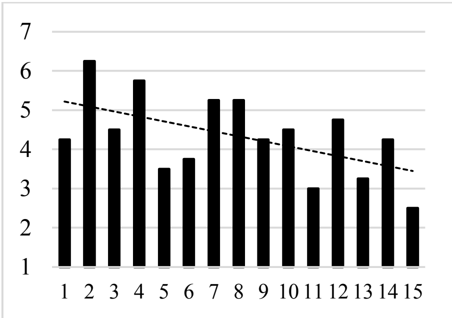


Figure 5.4: *LSA scores of the four topic descriptions on the 4 sets of 15 domain documents for the project "Mental disorders".*

The topic descriptions of project Eyesight resulted in the following average LSA scores: Brain: 0.373, Eye: 0.346, Focus, 0.300, Depth: 0.250. The topic descriptions of project Mental Disorders resulted in the following average LSA scores: Disorders: 0.197, Factors: 0.343, DSM: 0.339, and Psychopaths 0.159.
To test whether LSA had indeed found relevant documents, two randomly chosen topic descriptions with their domain document sets ("Brain" and "Factors") were presented to four domain experts from the Psychology staff. Of each of the 15 documents in each set they were asked whether they agreed to the statement: "This text is related to the topic description". They could answer on a 7-point Likert scale, ranging from strongly disagree to strongly agree. The results are shown in Figures 5.5 and 5.6, respectively.



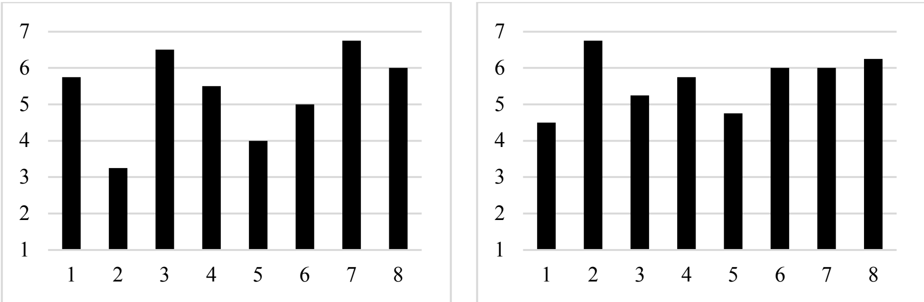


Figures 5.5 and 5.6: *Average staff-attributed relatedness of the set of 15 retrieved documents from the topic descriptions "Brain" (n = 4, sd = 1.00) and "Factors" (n = 4, sd = 1.12) on a 7-point Likert scale. The dotted lines represent linear regression lines fitted to the assessment data.*

These results show that for the topic descriptions presented to the staff, both domain document sets contained 10 documents that, on average, were rated above 4 (agree nor disagree). The linear regression line in each figure indicates that the average teacher-attributed relevance showed a slow decline. This accords with our expectation, as the documents resulting from the LSA-queries were presented in descending order of cosine value (and thus strength of semantic relation to the topic description). The interclass correlation coefficients (ICC) for both "Brain" and "Factors" were calculated. The results were .762, with 95% reliability interval between .477 and .911 (indicating good reliability) and .594, with a 95% reliability interval between .109 and .848 (indicating reasonable reliability), respectively.

5.5.2. Can we adequately determine whether prospective team members have different levels of prior knowledge?

From the learner's knowledge evidences we calculated the learner's knowledge scores on the four topics their project addressed. On these we based the formation of the learner/peer-teacher duos. To confirm whether the knowledge scores adequately represented different knowledge levels, we presented four staff members of the Psychology faculty with eight sets of learner/peer-teacher documents. We selected those eight pairs of texts that showed the highest learner/peer-teacher knowledge score differences. Staff experts were asked whether they concurred that the peer-teacher text showed more knowledge on a topic than the first learner text (see Figure 5.7). We also selected the eight pairs (one per topic) of learner first and second texts with the highest knowledge score difference between the learners' first and second texts and asked the staff experts whether they agreed that the second texts showed more knowledge than the first texts (see Figure 5.8).



Figures 5.7 and 5.8: *Staff expert answers to the question whether in a duo the peer-teacher text showed more knowledge on a topic than the learner's text 1 (n = 4, sd = 0.78) and whether a learner's text 1 showed more knowledge than the learner's text 2 (n = 4, sd = 0.66). Answers were given using a 7-point Likert scale ranging from strongly disagree (1) to strongly agree (7).*

These results indicates that LSA can indeed adequately distinguish between texts showing more or less knowledge, as is required for the formation of teams of learners/peer-teachers with knowledge differences.

5.5.3. Can we determine a knowledge difference between learners ('zone of proximal development') at which learning is most effective?

All but four learners in the 124 learner/peer-teacher duos we formed were sent a text on one topic that received a higher knowledge score than their own text. To provide insight into the overall learning effects, Table 5.2 shows the average knowledge scores of the learner's first texts, their peer-teachers texts, and the learner's second texts. These are reported per project, and per project topic.
Table 5.2 shows that in general the learner's second texts exhibited more knowledge, that is, after they read their peer-teacher's texts. As both the underlying average LSA results and the number of domain documents to which the learner texts referred increased, this indicates that both the width (with respect to on how *many* documents from the topic's domain document set the learners showed knowledge) and the depth (with respect to how *much* knowledge the learner showed on the domain documents) of knowledge had increased.

Table 5.2: *Average knowledge scores from learner first texts, peer-teacher texts and learner second texts, truncated to 2 significant decimals.*

	Average knowledge score of learner text 1	Average knowledge score of peer teacher text	Average knowledge score of learner text 2
Project Eyesight (*n* = 64)	0.08	0.20	0.13
Project Mental disorders (*n* = 60)	0.10	0.22	0.13
Project Eyesight: Brain (*n* = 18)	0.13	0.22	0.17
Project Eyesight: Eye (*n* = 18)	0.13	0.23	0.17
Project Eyesight: Focussing (*n* = 13)	0.02	0.24	0.09
Project Eyesight: Depth (*n* = 15)	0.02	0.13	0.08
Project Mental disorders: Disorders (*n* = 27)	0.09	0.16	0.12
Project Mental disorders: Factors (*n* = 7)	0.12	0.30	0.16
Project Mental disorders: DSM (*n* = 6)	0.14	0.30	0.17
Project Mental disorders: Psychopaths (*n* = 20)	0.03	0.13	0.08

When we define *possible knowledge gain* as the difference in knowledge score in a duo between a learner's first text and the peer-teacher's text, we observe that for project "Eyesight", on average, about 40% of the possible knowledge gain was realised. For project "Mental disorders", this average was about 28%.
Figure 5.9 presents an overview of how learner's first texts, pear-teacher's texts and learner's second texts relate to each other with respect to knowledge scores. The data points 1-18, 19-36, 37-49, and 50-64 stem from the topics "Brain", "Eye", "Focusing" and "Depth" of project "Eyesight", respectively. The data points 65-91, 92-98, 99-104, and 105-124 stem from the topics "Disorders", 'Factors", "DSM", and "Psychopaths" of project "Mental disorders", respectively.

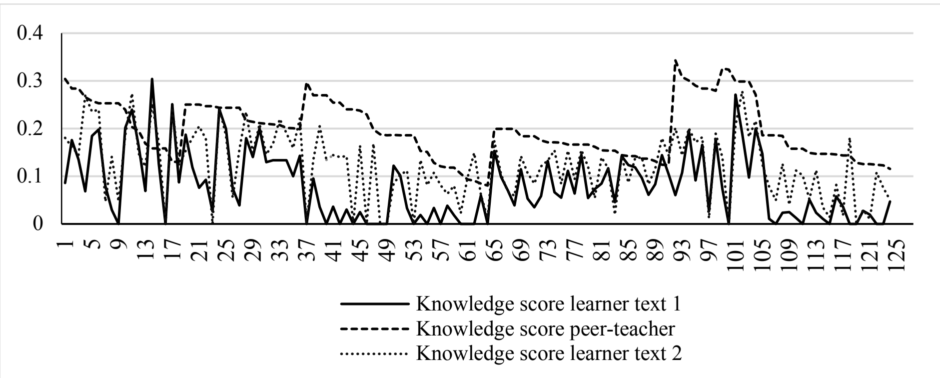


Figure 5.9: *Knowledge score of learner's texts 1, peer-teacher knowledge scores and knowledge scores of learner's texts 2.*

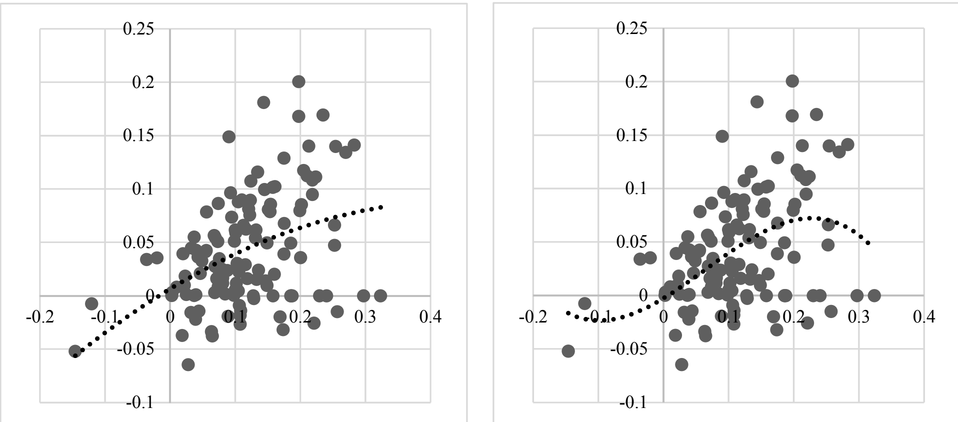
The average knowledge score from the learner's texts 1 was 0.06, while the average knowledge score from the learner's texts 2 was 0.11. The average knowledge score of their peer-teachers was 0.20.

We determined the significance of the knowledge gains with paired t-tests on the individual project topics. The results are depicted in Table 5.3.

Table 5.3: *Results of paired t-tests and p-values for knowledge gains per project topic.*

	t-test	p-value
Project Eyesight: Brain (n = 18)	2.304	.034
Project Eyesight: Eye (n = 18)	3.388	.003
Project Eyesight: Focussing (n = 13)	3.923	.002
Project Eyesight: Depth (n = 15)	3.804	.002
Project Mental disorders: Mental disorders (n = 27)	4.399	.000
Project Mental disorders: Factors (n = 7)	1.919	.103
Project Mental disorders: DSM (n = 6)	1.314	.246
Project Mental disorders: Psychopaths (n = 20)	3.700	.002

To gain insight into the knowledge difference between learner/peer-teacher at which most learning takes place, Figures 5.10 and 5.11 show the learners knowledge scores on their second texts set against the difference between the peer-teacher knowledge score and learner knowledge score on their first text.
The dotted line in Figure 5.10 represents a 2nd order polynomial regression line fitted to the knowledge score gains, the dotted line in Figure 5.11 represents a 3rd order polynomial regression line fitted to the knowledge score gains. The 2nd order polynomial regression line in Figure 5.10 shows a diminishing knowledge gain when the difference between the peer-teacher knowledge score and the learner's first text knowledge score gets smaller. The 3rd order polynomial regression line in Figure 5.11, however, shows an optimum learning effect when the difference in knowledge is around 0.22. The small negative knowledge gains observable to the left of the origin can be explained by the fact that we sent 4 learner texts that had received a lower knowledge score than their own first texts.



Figures 5.10 and 5.11: *Knowledge score difference between the peer-teacher and learner text 1 (horizontal axis) versus the knowledge score gain between the first and second learner texts (vertical axis) over both projects combined (n = 124).*

This result provides an indication of the level of knowledge difference at which learning becomes most effective.

5.5.4. How does the personality factor 'conscientiousness' impact on learning and the interaction process between learners?

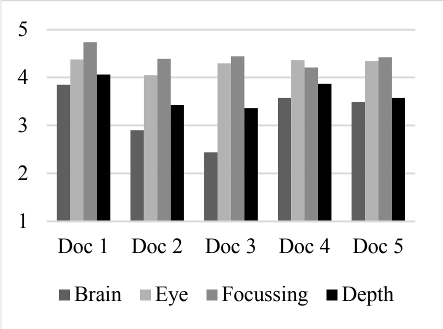
As our team formation principle is also based on the personality aspect of 'conscientiousness', we wanted to find out what effect, if any, the learner conscientiousness levels had on their knowledge scores. We didn't find any significant relationships between conscientiousness and knowledge score on the learner's first texts (r = .147; p = .102), nor on their second texts (r = .081; p = .371). We also didn't find a significant relationship between learner conscientiousness and knowledge score gain (r = -.091; p = .317). And there was no significant relation between the knowledge difference between the peer-teacher's texts and the learner's texts, and the ultimate knowledge score gains achieved by the learners (r = -.088; p = .331). However, when we divided the students into 4 intervals of equal size (30), we observed that the level of conscientiousness had an effect on their timeliness of handing in assignments (see Table 5.4).
In the fourth interval, the percentages of participants missing deadlines rapidly increased. This finding is in accord with findings from Gevers and Peeters (2009), who report a negative influence on the team temporal processes when conscientiousness scores are low.

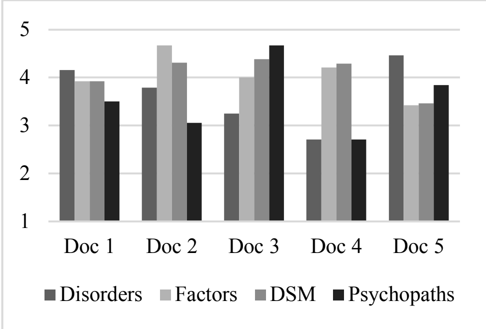
Table 5.4: *Conscientiousness scores of numbers and percentages of participants missing deadlines, split into intervals of 30 participants.*

Consc. between	missed deadline 1	missed deadline 2	missed deadline 3	average percentage	n =
5-4.33	3 (10.0%)	1 (3.3%)	3 (10.0%)	7.76 %	30
4.33-4.00	1 (3.3%)	1 (3.3%)	3 (10.0%)	5.53 %	30
4.00-3.56	2 (6%)	1 (3.3%)	5 (16.6%)	8.83 %	30
3.56-1.56	6 (20.0%)	3 (10.0%)	6 (20.0%)	16.66 %	30

5.5.5. Can we suggest learning materials from inside the knowledge domain to learners in such a way that learners consider these materials relevant to the project they work on?

We asked the participants to indicate the relevance of learning materials we suggested to them after they handed in their 2nd assignment. The learning materials (per project) consisted of the texts with the five highest ranking LSA results from the analysis of the project topic descriptions. Answers could be given on a 5-point Likert-scale (1 = not relevant, 5 = highly relevant). Figures 5.12 and 5.13 show the results.





Figures 5.12 and 5.13.: *The average student-attributed relevance of the 5 texts with the highest LSA scores over the 4 topics, of projects "Eyesight" (n = 61, sd = 0.94) and "Mental disorders" (n = 52, sd = 1.097), on five-point Likert-scales.*

The relevance indications from domain experts from the Psychology staff on the same five documents related to the topics "Brain" and "Factors" from Figures 5.5 and 5.6 above indicate that there is considerable overlap in teacher and learner appreciation of the relevance of the documents. When we split the results in above, neutral, or below the average, we find the following overlap in the pairs of appreciations from teachers and learners for the documents for "Brain": above/above, above/above, below/below, above/above, and below/neutral. For "Factors", we find: above/above, above/above, above/above, above/above, and neutral/above. These results indicate that it is thus feasible to base suggestions of learning materials relevant to the topics on which learners work based on LSA.

5.6. Discussion

We started off with making the case for introducing team formation and PBL design principles in large-scale open learning environments, such as MOOCs. We believe our approach to automate support for PBL and team formation processes can scale up to match the size of the majority of the MOOCs (75% of MOOCs have fewer than 1000 participants[6]). Further up-scaling might be possible when learners were to be enabled to provide their own learning projects (hence decreasing the burden on staff to provide the projects).

The LSA results of both project's topic descriptions show some interesting differences. In the project "Eyesight" (see Figure 5.3) the topic descriptions show a uniform LSA result profile (slow decline) and a uniform starting point (around 0.4). In project "Mental disorders" (see Figure 5.4) the LSA results of particularly the topic descriptions of "Disorders" and "Psychopaths" start out relatively low and show a flat profile. These two topic descriptions do not reference domain

[6] *fide* Katy Jordan, http://nogoodreason.typepad.co.uk/no_good_reason/2013/12/completion-data-for-moocs.html

documents that are ostensibly more relevant than others. This might indicate that these topic descriptions relate more evenly to the domain documents referenced, whereas the other descriptions favour documents containing more specialist knowledge. But in general it might also indicate that they do not correlate high enough to any domain documents. As these documents also form the basis for the assessments of prior knowledge (see below) and for our ability to recommend relevant documents as learning materials, topic fit profiles need to be chosen carefully. Therefore, topic descriptions should yield document domain sets with LSA values that are roughly similar and of intermediate height (by being similar, we prevent favouring knowledge on only a few documents; by being not too low, we prevent over-estimating the on-topic-ness, by being not too high the domain documents would keep their value as learning materials). However, in situations where learners or others were to be allowed to provide their own learning projects, further study is required to consolidate the rule proposed above and to determine a threshold below which topics are likely ‘off-domain’. Nevertheless, when we consider the fact that both staff and learners appreciate the relevance of the top five documents on these topics in ways more or less similar to the other topics, also the topics “Disorders” and “Psychopaths” seem to fit well to the domain.
The method used to calculate knowledge scores provided satisfactory results, as was evidenced by the staff, and indirectly by the learning results (see also below). The data (see Figure 5.9) indicated that our participants had relatively little prior knowledge on the topic “Depth” of project “Eyesight”, compared to the other topics of that project. They also show the learners had relatively more knowledge on the topic “DSM” from project “Mental disorders” compared to the other topics of that project. Overall, the knowledge scores on the topics in project “Eyesight” are higher than the knowledge scores on the topics in project “Mental Disorders”. This fits very well with the suggested order of study of the book chapters related to our projects, as part of the student body did not yet study, or had only just begun studying these chapters.
The learning gain was clearly significant. Several precautions were taken to make plausible that learning had indeed occurred from reading the peer-tutor text: participants were allowed only a short period of time to rewrite their initial texts; they were instructed to write in their own words (no copy-paste of peer-teacher text was found when we inspected the texts); and to stay within the bounds of the predefined maximum text size. At the topic level, we found significant knowledge gains for all but the topics “Factors” and “DSM” of project “Mental disorders. The small number of duos on these topics limited the power of the paired t-test, so that no firm conclusion, either way, could be drawn. Overall, of the 124 learners, only 29 showed no (9) or negative knowledge gains (20). The average negative knowledge gains of these 20 learners was small: 0.024, as was the average of the differences between their knowledge scores and the knowledge scores of their peer-teachers: 0.076.

With respect to optimal knowledge score differences leading to the highest knowledge gains, our data proved to be inconclusive (see Figure 5.10), or indicative at best (see Figure 5.11). This can be partly due to the (relative) homogeneity in knowledge backgrounds in our population. Furthermore, because we didn't use a golden standard against which to measure knowledge differences and knowledge gains, there was a limited number of cases on which we could build to find an optimal knowledge difference. With this, our research took a different approach from the more controlled environments in which an "optimal proximity for acquisition" (Dessus et al., 2011) or "zone of learnability" (Wolfe et al., 1998) against a pre-set "knowledge target" were assumed. However, by making the peer-teacher the "knowledge target", we believe our approach fits best in learning settings in which no golden standard is available and knowledge is co-constructed by means of collaborations between knowledgeable and less knowledgeable learners.
Regarding the ability to recommend learning materials based on the LSA results from the topic descriptions, the learners clearly valued these recommendations. This indicates that we can indeed recommend learning materials. Furthermore, the five recommended documents per topic were the same as the five documents with highest LSA-scores in the topic-related domain document sets. Our data show that there is considerable overlap in learner and staff appreciation of the relevance of these materials. Therefore this can be seen as additional support for the "on topic-ness" of the topic descriptions (cf. Section 5.5.1), this time from the perspective of the learner.

5.7. Conclusions and directions for future research

This chapter made the case for providing automated support for PBL and team formation in open learning environments. We built further on our own earlier work, in which we developed a team formation model and team formation principles and algorithms. Taking these as starting point, we identified the creation of a knowledge domain and the automation of knowledge-related assessments as the parts still required for the implementation of the automated service. For the settings involved in knowledge domain creation we built on existing work with LSA. We defined two projects, one on "Eyesight" and one on "Mental disorders". For each project we described four topics on which learners would evidence their prior knowledge. Using LSA, these topic descriptions were compared to documents in the knowledge domain for their fit. This resulted in a suggestion for a general rule to determine this fit: To ensure both a wide scope and high level of on-topic-ness of the projects to be run, analysing project topic descriptions with LSA should yield documents domain sets which receive LSA values that are both similar and of intermediate height. The results of the automated assessments of project topic fit to the knowledge domain, the learner prior knowledge and the knowledge after learning from the peer-teachers were all put before teaching staff. Their

assessments provided general support for the applicability of LSA for such assessments.
We additionally researched whether an optimum knowledge difference between learners and peer-teachers, (as in a zone of proximal development, zpd) could be found. We did not arrive at conclusive results, although our data tentatively suggest that a numerical value for the zpd (in terms of a knowledge difference) can be determined. We also aimed to find out whether the LSA analysis of topic descriptions would provide the possibility of recommending topic-related learning materials from the knowledge domain. Participants showed clear views on the documents we recommended: they were classified as relevant sources of knowledge for the task we put before them. Our investigation into the effects of learner conscientiousness on learning and/or collaboration process revealed an interesting effect: It was not the knowledge gain that was effected by conscientiousness, but rather the dependability of learners in keeping appointments (when conscientiousness scores were low).
All in all, we think we have successfully developed the building blocks for the implementation of an automated team formation service for PBL in open learning environments we set out to design.

5.7.1. Directions for future research

While we believe our results convincingly demonstrate the applicability of our approach in teacher extensive open learning environments such as MOOCs, several issues require further research.
The assessment of project topic descriptions for their fit in a knowledge domain currently has limitations. Although we formulated a general rule for this fit, this does not provide clear upper or lower bounds for when projects over-fit of under-fit the domain. To determine these bounds, we suggest live team-work experiments based on project topic descriptions of various levels of fit.
Our research into the knowledge difference at which the most learning takes place remained inconclusive. Homogeneity in knowledge backgrounds, and a “moving knowledge target” in the form of a peer-teacher led to a limited number of cases on which to build an argument for where the “zpd” can be found. We therefore suggest that additional, large-scale, research is required in knowledge domains with less homogeneous learners to find this optimum.
With respect to the effects of learner collaborations, our evidence of learning was gathered based on limited and mediated interactions between duos of learners. An interesting line of inquiry would be to analyse learning effects based on longer-term team interactions and the development of a joint project product. As a starting point we suggest to use LSA to continuously assess learner contributions to a joint product over time, as was demonstrated by e.g., Dong (2005).

CHAPTER 6

General discussion

6.1. Introduction

Our current societies stimulate continuous professional development. Professionals can find ample and variously organised (and non-organised) learning resources on the Internet. For example, they can learn by subscribing to different flavours of MOOCs, more or less organised along formal educational models; or they can engage in Social Learning Networks (Koper & Sloep, 2002). In Chapter 1 we identified that such learners are served best when provided with collaborative learning opportunities. From this vantage point we noticed several issues regarding the support for collaborative learning in open learning environments (such as xMOOCs, cMOOCs, and SLNs): these learning environments are either based on individual learning, provide insufficiently structured tasks, or leave the learner adrift in a sea of learning resources. The general issue, we noted, is that they do not provide an answer to the question of how to identify the best peers to learn with and what are the best learning materials. As envisioned by Stahl (2006), collaborative knowledge building processes go through several phases, such as expressing problems, collaborating with peers, and using and creating learning materials. In Chapter 1 we asked the question of how to tie these processes to a fitting pedagogy, and how to implement them, so effective learning environments can be developed. We proposed that implementing *team formation for project based learning* in open learning environments fits to the learners and addresses several issues found in current open learning environments (without suggesting to exhaust all the issues there are, of course). However, implementing team formation for PBL requires extensive knowledge and effort from staff. Therefore, our research first analysed these processes and then addressed automating them so they could be implemented in open learning environments. This chapter first presents a review of our main findings and then addresses some methodological issues. Next, we consider the contributions we believe we made to several research fields. We end this chapter with suggestions for future research.

6.2. Review of the main findings

In our *first* study (see Chapter 2) we determined that there are significant differences between setting up project-based learning in formal (teacher-led) learning settings and open learning environments. In open learning environments (such as tutor-led MOOCs), due to scale, the expertise to form teams is likely to be scarce, while the data required to start PBL and team formation are probably unavailable. To form effective teams, an expert requires data about the prospective team members and the project task (Felder & Brent, 2007). Graf and Bekele (2006), Martín and Paredes (2004), Wilkinson and Fung (2002), and Slavin (1989) suggested which data should be taken into account. We sorted these into two categories. First, the curriculum area in which the project task will be positioned, the project task, and its characteristics (such as collaboration language, duration

and suggested team size), and the individual learner's abilities and prior learner achievements were categorised under *knowledge*. Second, the individual learner's personality traits and motivational orientation were categorised under *personality*. To cater for learner-specific constraints (such as language, location, availability, etc.) we included of a third category of data: *preferences*.
With respect to how to form teams, research shows that complementary and supplementary knowledge and personality among members are important factors in team formation (Muchinsky & Monahan, 1987; Werbel & Gilliland, 1999; Werbel & Johnson, 2001). Based on e.g., West (1997), we suggested that varying the combinations of knowledge and personality along the complementary/supplementary dimension allows one to set the stage for different project work outcomes. We distinguished between improving learning outcomes, enhancing the possibility of a creative project outcome, or improving productivity. This allowed us to define a first iteration of a model of the processes involved in putting expert team formation in the context of PBL for open learning environments (see Figure 2.1). An online survey among several educational practitioners from eight different countries (n=26) showed that the most important category of data to be used in the team formation process is *knowledge*, next come *preferences* and finally *personality*. In the same survey, the practitioners indicated a clear order in which they preferred team-work outcomes; first comes "Improve learning", then "Enhance creativity", and last "Improve productivity". Upon comparing the respondents' practises in team formation with suggestions from literature, we noticed that:

- in educational practice in 50% of the cases, team formation is left to the learners, which is strongly discouraged in the team formation literature (Oakley, Felder, Brent & Elhajj, 2004), and is proven to be detrimental to learner satisfaction and learning outcomes (Fiechtner & Davis, 1985)
- in educational practise the aspect of learner personality is seldom used during team formation, while literature puts an emphasis on the inclusions of personality aspects in the team formation process (Roberts, Kuncel, Shiner, Caspi & Goldberg 2007).

Responses on an open question about acceptance of team formation suggestions from an automated team formation system showed that 10 out of 11 respondents would accept such suggestions. We concluded that the team formation model fits to principles underlying team formation for PBL, largely overlaps with principles used in practise, and receives support from practise. While not supported fully by practitioners, we believe that data on personality would be an important asset to include in a team formation instrument.

In the *second* study (see Chapter 3) we researched the principles underlying an automated team formation service for PBL to be used in open learning environments. A review of 12 existing team formation tools and techniques originating from different application fields revealed that these assume data and user roles that are most likely not available in open learning environments. For this reason we proposed a design based on data that can be acquired directly from the learners and the learning environment. Based on the data categories in the team formation for PBL model (knowledge, personality and preferences), and due to the differences in the nature of the data required, we designed three different "experts by proxy" to gather and assess these data:

- For the analysis of both required and available knowledge, based on prior research, we selected Latent Semantic Analysis (LSA) as a suitable method (Rehder, Schreiner, Wolfe, Laham, Landauer, & Kintsch, 1998).
- For the analysis of personality, we selected the personality aspect of "conscientiousness", as literature suggests that this aspect has the most predictive value for a person's future performance in a team (Barrick & Mount, 1991).
- With respect to preferences we proposed to include availability schedule, languages mastered, and preferred tools. Such preferences determine whether collaborative project work can happen at all between prospective team members. Therefore an assessment of overlap in preferences precedes the assessments of knowledge and personality.

Based on the distinction between productive, creative, or learning team outcomes and supplementary and complementary knowledge and personality (see Chapter 2) we performed a literature review into the effects of such combinations. This allows us to define team formation principles for the formation of such teams, as depicted in Table 6.1.

Table 6.1: *Team work outcomes in combination with kind and levels of knowledge and conscientiousness levels.*

Project outcome	Kind and level of knowledge	Conscientiousness
Productive problem solving	Supplementary and high	All high
Creative solutions	Complementary and high	All low
Facilitating learning	Complementary and high, but within limits	All high

These principles were formalised in team formation expressions, which were implemented in algorithms. When applied to a set of test data, these algorithms demonstrated that they are able to form teams and to suggest different teams based on the preferred teamwork outcomes.

In the *third* study (see Chapter 4) we validated both the team formation principles and the results from the team formation algorithms. Results from a survey among teaching staff (n=56) showed that participants accepted the principles for the formation of productive and learning teams, but did *not* accept the principle for the formation of creative teams. The same survey also aimed to validate the results of the team formation algorithms. A prior survey among learners (*n*=168) provided knowledge self-assessments and conscientiousness scores, which our algorithms successfully processed into team suggestions for productive, creative and learning teams. Based on their understanding of how the team formation principles work, practitioners classified teams into three classes (productive, creative and learning teams) and ranked teams within one type of team. The results of these tasks showed that the participants classified and ranked *small* productive, creative and learning teams largely in accord with how our algorithms judged these teams. The principled for the formation of productive and learning teams were thereby validated. However, with increasing task complexity, especially when the formation of learning teams was concerned, increased divergence occurred between human and computer classifications and rankings. The ranking and classifying task for learning teams were considerably more difficult in that they included taking into account more variables and rules (see Chapter 3, Section 3.3). These outcomes were explained from the perspective of "bounded rationality", which suggests that rational decision making is limited by time constraints and human power of abstraction. From the fact that the principles for productive and learning teams were accepted, and the judgement of respondents and algorithms (when small teams were concerned) showed considerable overlap, we concluded that our instruments perform adequately, also for larger teams (where human performance degrades). This contributed to the conclusion that it would be useful to automate the team formation task.

In the *fourth* study (see Chapter 5) we reported on the implementation of the team formation for PBL model in an *experimental learning context exclusively* and presented the final iteration of the team formation for PBL model for the formation of learning teams (see Figure 6.1). In this figure, the solid arrows describe the steps one takes to put together a team for a particular project in a given knowledge domain and recommend learning materials. The dotted arrows refer back to the elements of the model against which assessments are made.

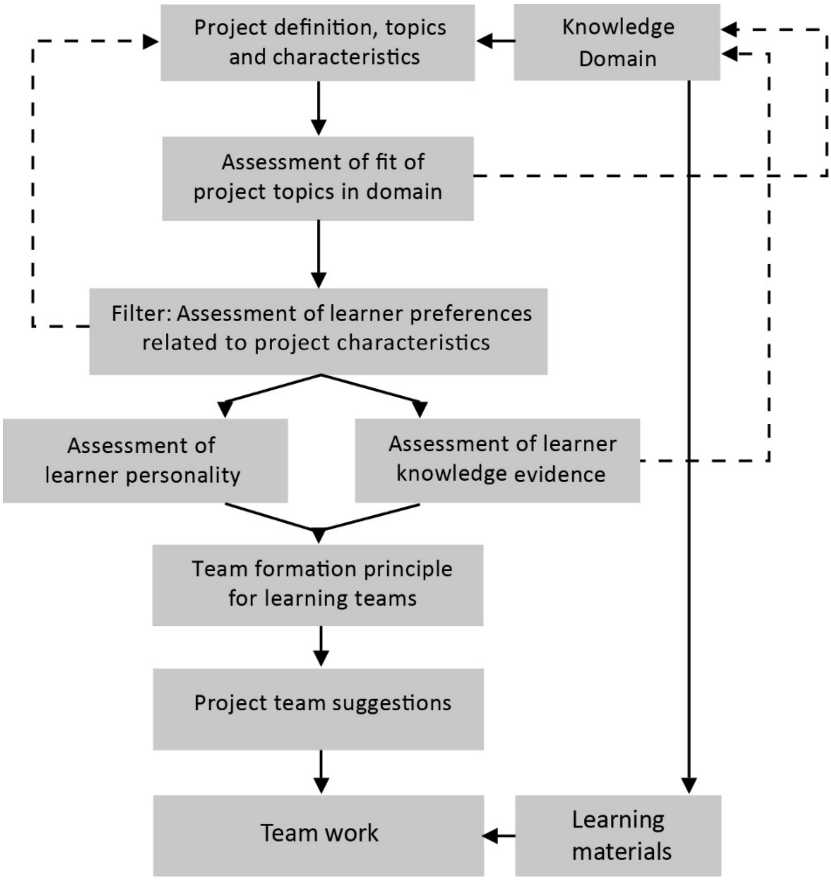


***Figure 6.1**: Team formation for project-based learning model.*

Following the requirements set out by our model, we applied LSA to successfully prepare a representation of a knowledge domain in which projects could be run. We defined two projects by describing their general tasks and the topics which they addressed. An LSA analysis of the topic descriptions determined the fit of these descriptions to the knowledge domain. Educational staff confirmed that we were indeed able to determine this fit. Subsequently, the empirical research in study 4 consisted of:

- Using LSA to assess learners' prior knowledge on the project topics. The results (as confirmed by teaching staff) showed that LSA assessed learners' prior knowledge in line with the teacher's assessments and could thus be used to form teams.
- Mimicking team learning by forming duos of learners such that one partner of the duo had more prior knowledge than the other partner, and providing the less knowledgeable members with a learning task. The LSA analysis of the

results from the learning task (again confirmed by teaching staff) showed that LSA is capable of assessing whether learning had taken place.
- Investigating the effects of the learner personality aspect of "conscientiousness" levels on learning outcomes and/or the collaboration processes. This revealed an interesting effect: we found that it was not the knowledge gain that was affected by conscientiousness, but rather the reliability of learners in keeping appointments (in brief: higher conscientiousness scores imply higher reliability).
- Investigating the learners' appreciation of automatically suggested learning materials. The learners confirmed that LSA is capable of suggesting learning materials from the domain of study related to the project topics at hand.

Due to the fact that we formed the learning duos based on declining differences in prior knowledge, we could provide an indication of the "location" of our participant's zone of proximal development in terms of the knowledge difference between learner and peer-teachers in our duos at which most learning took place. This indication, however, was not conclusive.
Overall, however, the most important result from our research is that it is feasible to implement a team formation for PBL model with the instruments we researched and developed.

6.3. Methodological considerations and study limitations

Our research necessarily was confined. We address several methodological issues and study limitations at some length.

6.3.1. Representative value

In our first study almost 90% of the respondents (*n*=26) worked at either a university or a university for professional education. 40% were mainly active in project-based learning settings and 32% in problem-based learning settings. Therefore the results are indicative of team formation practice in education only. In our second study we conducted a literature review of existing team formation tools. While we reviewed 12 tools from various application domains, this review was not exhaustive. Moreover, due to the origins of some of these tools their affordances were not directly aimed at educational contexts. Therefore the conclusions we drew are limited in scope. In the third study, our student-respondents (*n*=168) were either studying at the Psychology faculty (*n*=121) or, studying Learning Sciences (*n*=47). They used an online survey to self-report knowledge of topics addressed in four courses. These data were used to feed into our team formation algorithms. As both self-reporting and using surveys are known to have limitations, the teams formed might not be based on the actual knowledge levels of the respondents. The staff-respondents (*n*=56) asked to agree with the team formation principles all stemmed from our distance teaching university, which favours an individual learning setting. Therefore our staff-respondents might not be fully representative of team formation experts when expressing their opinions on the team formation principles. The fourth study involved learners (*n*=124) from the psychology faculty. They had either finished the course “Inleiding in de Psychologie” (Introductory Psychology) in the previous year or were still studying the course. Although we gathered data on when they absolved the course, and on which chapters they had studied, these were not taken into account during the evaluation of the results. While we took several precautions to minimized possible time-on-task effects, such effects could not be fully excluded. During this study we relied on the otherwise well-researched method of LSA for various knowledge-related assessments, therefore these were validated by a limited group of teachers only (*n*=4).

6.3.2. Proxy designs

In designing the team formation for PBL model, three “experts by proxy” were developed. These proxies aim to mimic expert behaviour in assessing knowledge, personality and preferences. To elaborate the knowledge proxy, we restricted ourselves to the most often used form of expression of knowledge: text. For the personality proxy, we selected the single personality trait “conscientiousness”, which, according to literature, is the personality factor most related to achievement in team work. As was demonstrated by the results from our third study, the implementation of the proxies for the creation of creative teams showed to be insufficient. In the preferences proxy (for which we suggested to assess availability, time zone, possible collaboration languages and preferred tools), which was used in our third study (see Chapter 4), we only used the preference relevant for this study i.e. availability. In many cases, the preferences proxy likely will be situation dependent, so it may require adaptation from case to case. Finally, while we identified several pitfalls (related to gender, belonging to a minority, etc.) to be avoided when forming teams, these were not implemented in the proxies. Other aspects, such as trust between learners (Rusman, Van Bruggen, Sloep, Valcke, & Koper, 2012) also play a role in team formation. These can be examined in future research.

6.3.3. Running collaborative projects

We implemented team formation and project-based learning in an experimental environment, that is, in isolation from curricula or course timeframes. During the setup, project descriptions were drawn up by an expert with knowledge of the technical demands of the system. Therefore we have to investigate how our findings extrapolate to situations in which teachers or learners provide fully self-defined project descriptions. The evidence on learning was gathered from time-limited and mediated interactions between duos of learners. While our setup touches the core of learning from a more knowledgeable peer, it did not reflect normal practise in PBL, where projects last longer, often involve more members, and show more social interactions.

6.3.4. Learning in the "Zone of proximal development"

One of the main aims of the fourth study (see Chapter 5) was to find evidence with respect to an optimum knowledge difference between team members at which learning is most effective in the elusive "zone of proximal development" (Bonk & Kim, 1998; Huang, 2002). However, as we could only explain a small part of the variance in knowledge gains from learner-teacher knowledge differences, we could only provide a suggestion about where the optimum that is characteristic of such a zone may be found. Possibly, it was hard to identify a clear-cut optimum as in the experiment aimed at finding the optimum knowledge difference we partnered learners and peer-teachers into duos with gradually decreasing knowledge differences. In that situation, every peer-teacher becomes a knowledge-target for the associated learner, but no knowledge target is the same (no golden standard is thus used). This way, one can experimentally determine an optimum knowledge difference at which most learning takes place. However, it requires a large number of cases to find that optimum. Importantly, however, most learners learned from their peer-teachers. This indicated that the knowledge differences between learners and peer-teachers appeared to be bridgeable. Perhaps, also it is due to our population that the optimum difference could not be unequivocally determined. The relative homogeneity of our group of learners (all studied the same course) may not be representative for situations in which learners from different knowledge backgrounds subscribe to e.g., a MOOC. In such situations knowledge differences may be bigger, and the optimal "zone of proximal development" may be easier to find.

With respect to our findings in study 4, we see that learning happens on "both sides" of the suggested optimum of knowledge difference. Therefore our research question with respect to determining a specific knowledge difference leading to optimum learning effects, might have been better phrased toward finding a range in knowledge differences between which learning is demonstrably effective. This suggestion also has repercussions for the team formation algorithm for the formation of learning teams, as it implements the zone of proximal development as a set knowledge difference above which learning is assumed not to happen.

6.3.5. General limitations

The use of LSA as an assessment method requires fine-tuning and pre-processing. A practical implication is that the affordances of this tool can only be fruitfully exploited in long-lasting relatively stable environments; alternatively, an approach has to be developed to automate this. The feasibility of the development of such environments (also for commercial purposes) is demonstrated by e.g., Summary Street (Wade-Stein & Kintsch, 2004). The design of our final experiment required preparations, which are likely to differ from context to context. So, while our results hold true in the settings we developed, strictly speaking it does not necessarily follow that a 1-on-1 transfer to other settings or domains will be equally successful.

6.4. Contributions to research fields and valorisation opportunities

Our research contributes to several research fields and opens up opportunities for practical application, the most important one in the domain of open learning environments. *Designers of open learning environments* should be able to profitably use our principles for team formation, based on the assessment of prior knowledge (Moos & Azevedo, 2008) and personality (Barrick & Mount, 1991). Societal developments put an emphasis on continuous professional development. Current open learning environments such as xMOOCS, cMOOCs and Social Learning Networks, all provide learning settings professionals can engage in. However, as we argued, such environments do not allow them to do so effectively (Kop, Fornier, & Mak, 2011; McGuire, 2013; Stacey, 2013; Alvarez & Olivera-Smith, 2013). As project-based learning provides excellent collaborative learning opportunities (Davies, de Graaff, & Kolmos, 2011; Kolmos, 2012), it would be a *valuable additional pedagogy* to any open learning environment. By providing the instruments for assessing project proposals and forming teams fit to execute these projects, we provide this opportunity. Implementing team formation for project-based learning expands the number of available pedagogical paradigms to choose from when designing open learning environments. To the field of *team formation* we contribute in several ways, notably of course the team formation model for PBL. By providing validated principles for the formation of productive and learning teams we add to the development of *team formation theory*. By using and evaluating *LSA* for assessing project fit to knowledge domain, for knowledge and learning assessments and for recommending learning materials (Laham, Bennett, & Landauer, 2000), we add to the knowledge base already available on its applicability for such assessments.

In small-scale settings the team formation principles can be applied manually, and therewith form an addition to the instruments available to e.g., teachers in classroom settings. The instruments developed can also have a *wider application,* in settings where the required data is already wholly or partly available. When e.g., data on prior knowledge is available (from, say, a learning management system) and preferences do not play a limiting role (such as in a classroom setting), the services only require the addition of personality data to be used. For teachers developing learning projects, the assessment of fit of projects to the desired parts of a course or curriculum can provide valuable feedback. Learners can be provided *with self-assessment opportunities* during a course of their study. In a wider sense, *team formation processes are found in many areas*: in professional environments communities of knowledge workers define and execute (international) projects. In such settings our team formation service can provide guidance for team formation. However, conditions in such settings are different from conditions in learning

situations. Subsequently, the team formation criteria would have to be reassessed for their exact fit to professional settings. By asking learners to provide prior knowledge evidence on central topics in a course, our instruments can *support pre-entry assessments*. This could add to the prevention of drop-out and disillusion when a learner finds out the level of the course is too high (or too low). Through *entry-level assessments*, learners can be directed toward learning materials better fitting their knowledge level.
Overall, however, our research results add a practical angle to the often theoretical reflections on how self-directed learners *should* learn instead of how these learners *can* learn.

6.5. Recommendations for future research

Our research aimed to lay the foundation for effective collaborative learning in open learning environments. We identified several pitfalls (related to gender, knowledge differences, etc.) to be avoided when forming teams (see Chapter 2, Section 2). These all merit future research. In our recommendations, however, we focus on the following.

6.5.1. Running collaborative projects

The knowledge and personality assessment methods provide both a theoretical and practical approach to team formation for project based learning. However, the results from study 4 were obtained from small, mediated interactions based on knowledge differences. While we expect that these also have effects when projects run for longer periods and include social aspects, testing these effects on a longer term would increase the validity of our findings. Furthermore, we ultimately focussed on implementing our model in learning settings. Therefore the effects of team formations based on the principles for the formation of productive and creative teams still have to be investigated.

6.5.2. Principles for the formation of teams

Our research did not provide a satisfactory principle for the formation of creative teams. Given the societal emphasis on the importance of creativity as a 21st century skill (Bell, 2010), future research should provide a well-supported principle for the formation of such teams (which should then be expressed in an updated algorithm for the formation of creative teams). We already provided some pointers to approach this issue: factors above and beyond conscientiousness and openness to experience (Kaufman & Sternberg, 2010) could be included in the personality test to fortify the principle for the formation of creative teams. In this case it would be interesting to not just look at knowledge gain but also to study if the team formation would positively impact 'learning to be creative' as a separate additional learning goal. Assessing this kind of learning, however, is not straight-forward. The research of assessment of and support for creativity is still very much itself a topic of research (Jordanous, 2012; Van Rosmalen, Boon, Bitter-Rijpkema, Sie, & Sloep, 2014).

6.5.3. *Definition of projects*

The projects defined during study 4 were drawn up by the researchers. They consisted of a description of the goals of the projects and contained topic descriptions paraphrasing documents from the knowledge domain. Their fit to the domain was assessed based on the topic descriptions, and not on the descriptions of the goals and the expected outcomes of the project. Therefore there currently is a dependency on knowledge of the domain to successfully map projects to the domain. Enabling learners to initiate projects themselves would entail supporting them in defining interesting and realisable projects. Another issue to take up is finding a moment during a course (e.g., at mid-term, or at the end of the course as part of the final assessment) at which learners have gathered sufficient knowledge to successfully execute a project.

6.5.4. *Finding the "ZPD"*

Our investigation into finding an optimum in the zone of proximal development was limited by the number of cases. Therefore large-scale research would be required to arrive at a firm conclusion about which knowledge differences are most effective. However, as discussed above, it might be more interesting to find a range in knowledge differences in which learning is effective. Such a range would perhaps reflect other differences between learners that affect their ability to collaborate effectively.

6.5.5. *Measuring the effects of team formation during project work*

Project-based learning normally entails the finalisation of a product, which is then presented to the commissioner. And it often entails (peer) reviews of intermediate results to ensure the project is on the right track. We only superficially addressed the joint creation of a product. Therefore an interesting line of research would be to continuously analyse learning effects based on longer-term team interactions and the development of a joint project product. As a starting point we suggest using LSA to continuously assess learner contributions for convergence toward a jointly agreed end product, as was researched by e.g., Dong (2005). Language, besides being used for the externalisation and internalisation of knowledge, also mediates the social interaction processes in teams. A foray into e.g., sentiment analysis during collaborations could help with timely interventions in the learning process (Siemens & Baker, 2012).

References

REFERENCES

AFRL, A. (2006). Application of social network analysis to collaborative team formation.

Alvarez, I.M., & López, D. (2010). Behaviour regulation during joint knowledge construction in cooperative tasks within virtual written and asynchronous learning environments. *Cult. Educ., 22*, 419–438.

Alvarez, I.M. & Olivera-Smith, M. (2013). Learning in Social Networks: Rationale and Ideas for Its Implementation in Higher Education. *Educational Science, 3*, 314-325

Ardaiz-Villanueva, O., Nicuesa-Chacón, X., Brene-Artazcoz, O., Sanz de Acedo Lizarraga, M., & Sanz de Acedo Baquedano, M. (2011). Evaluation of computer tools for idea generation and team formation in project-based learning. *Computers & Education, 56*(3), 700-711.

Barrick, M. R., & Mount, M. K. (1991). The big five personality dimensions and job performance: A meta-analysis. *Personnel Psychology, 44*, 1–26.

Barrick, M. R., Mount, M. K., & Strauss, J. P. (1993). Conscientiousness and performance of sales representatives: Test of the mediation effects of goal setting. *Journal of Applied Psychology, 78*, 715–722.

Barrick, M. R., Stewart, G. L., Neubert, M. J., & Mount, M. K. (1998). Relating member ability and personality to work-team processes and team effectiveness. *Journal of applied psychology, 83*(3), 377.

Bell, S. (2010). Project-Based Learning for the 21st Century: Skills for the Future. *The Clearing House: A Journal of Educational Strategies, Issues and Ideas, 83*(2), 39-43

Blumenfeld, P. C., Soloway, E., Marx, R. W., Krajcik, J. S., Guzdial, M., & Palincsar, A. (1991). Motivating project-based learning: Sustaining the doing, supporting the learning. *Educational psychologist, 26*(3-4), 369-398.

Bonk, C. J., & Kim, K. A. (1998). Extending sociocultural theory to adult learning. Adult learning and development: Perspectives from educational psychology, 67-88.

Charles, J-M. (2014). Constructivism, Direct Instruction and the Free-energy Optimization Principle: Cognitive Perspectives on Learning. *Canadian Journal for New Scholars in Education, 5*(2).

Chihaia, D. (2007). Facilitating Collaboration between Learners in an Online Environment. Technical Report, National College of Ireland. Last accessed on June 27, 2014 from https://www.academia.edu/765510/Facilitating_Collaboration_between_Learners_in_an_On-Line_Environment

Chen, S. J., & Lin, L. (2004). Modeling team member characteristics for the formation of a multifunctional team in concurrent engineering. *Engineering Management, IEEE Transactions on, 51*(2), 111-124.

Christensen, G., Steinmetz, A., Alcorn, B., Bennett, A., Woods, D., & Emanuel, E. J. (2013). The MOOC Phenomenon: Who Takes Massive Open Online Courses and Why? (November 6, 2013). Available at SSRN: http://ssrn.com/abstract=2350964 or http://dx.doi.org/10.2139/ssrn.2350964

Christodoulopoulos, C.E., & Papanikolaou, K.A. (2007). A Group Formation Tool for e-Learning Environments. In: Proc.19th IEEE International Conference on Tools with Artificial Intelligence, 2, pp. 117–123.

Clarà, M., & Barberà, E. (2013). Learning online: massive open online courses (MOOCs), connectivism, and cultural psychology. *Distance Education, 34*(1), 129-136.

Colomo-Palacios, R., Casado-Lumbreras, C., Soto-Acosta, P., Misra, S., Garcia-Penalvo, F. (2012). Analyzing Human Resource Management Practices Within the GSD Context. *Journal of Global Information Technology Management, 15*(3), 30–54.

Cormier, D., Siemens, G., Downes, S., & Kop, R. (2010). How this course works. Retrieved online on September 23, 2014 from http://connect.downes.ca/how.htm

Coursera. (2104). Last accessed on August 6, 2014 from https://www.coursera.org/about/

Dahms, M-L., & Stentoft, D. (2008). Problem Based Learning in Engineering Education: a Development Option for Africa? Paper presented at The 4th African Regional Conference on Engineering Education (ARCE-2008): 'Capacity Building in Engineering Education for Sustainable Development', Dar es Salaam, Tanzania, United Republic of.

Daniel, J. (2012). Making Sense of MOOCs: Musings in a Maze of Myth, Paradox and Possibility. Journal of Interactive Media in Education. Last accessed on June 3rd, 2013, from http://www-jime.open.ac.uk/jime/article/viewArticle/2012-18/html

Davies, J., de Graaff, E., & Kolmos, A. (Eds.), (2011). *PBL across the disciplines: Research into best practice*. Aalborg: Aalborg Universitetsforlag.

Denissen, J. J. A., Geenen, R., van Aken, M.A., Gosling, G., Samuel, D., & Potter, J. (2008). Development and Validation of a Dutch Translation of the Big Five Inventory (BFI). *Journal of Personality Assessment, 90*(2), 152–157.

Dessus. P., Lemaire, B., Loiseau, M. Mandin, S., Villiot-Leclercq, E., & Zampa, V. (2011). Automated Free-Text Assessment: Some Lessons Learned. *Int. J. Cont. Engineering Education and Life-Long Learning, 21*, 2/3.

Dong, A. (2005). The latent semantic approach to studying design team communication. *Design Studies, 26*(5), 445-461,

Downes, S. (2006). Learning networks and connective knowledge. Last accessed on June 3rd, 2013, from http://it.coe.uga.edu/itforum/paper92/paper92.html

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

Dunbar, K. (1997). How scientists think: On-line creativity and conceptual change in science. In T. B. Ward, S. M. Smith, and J. Vaid (Eds.), *Creative thought: An investigation of conceptual structures and processes*, pp. 461–493. Washington, DC: American Psychological Association.

Edinburgh University. (2013). MOOCs @ Edinburgh 2013 – Report #1. Last accessed on June 5th, 2013, from http://www.era.lib.ed.ac.uk/bitstream/1842/6683/1/Edinburgh%20MOOCs%20Report%202013%20%231.pdf

Fazeli, S., Brouns, F., Drachsler, H., & Sloep, P. B. (2012). Exploring social recommenders for teacher networks to address challenges of starting teachers. In V. Hodgson, C. Jones, M. de Laat, D. McConnell, T. Ryberg, & P. Sloep (Eds.). In: Proceedings of the Eighth International Conference on Networked Learning 2012 (pp. 74–80). Retrieved from http://dspace.ou.nl/handle/1820/4129

Felder, R.M., & Brent, R. (2007). Cooperative Learning. In P.A. Mabrouk, (Ed.), *Active Learning: Models from the Analytical Sciences*, ACS Symposium Series 970, Chapter 4. Washington, DC: American Chemical Society.

Felder, R. M., Felder, G. N., & Dietz, E. J. (1999). A longitudinal study of engineering student performance and retention v. comparisons with traditionally-taught students. *Journal of Engineering Education, 87*(4), 469–480.

Fetter, S., Berlanga, A. J., & Sloep, P. B.. (2010). Fostering Social Capital in a Learning Network: Laying the Groundwork for a Peer-Support Service. *International Journal of Learning Technology, 5*(4), 388–400.

Fiechtner, S. B., & Davis, E. A. (1985). Why some groups fail: A survey of students' experiences with learning groups. *The Organizational Behavior Teaching Review, 9*(4), 75–88.

Fisher, M., & Baird, D.E., (2005). Online learning design that fosters student support, self-regulation, and retention. *Campus-Wide Information Systems, 22*(2), 88–107.

George, J. M., & Zhou, J. (2001). When openness to experience and conscientiousness are related to creative behavior: An interactional approach. *Journal of Applied Psychology, 86*(3), 513–524.

Gevers, J. M. P. & Peeters, M. (2009). A pleasure working together? The effects of dissimilarity in team member conscientiousness on team temporal processes and individual satisfaction. *Journal of Organizational Behavior, 30*(3), 379–400.

Gigerenzer, G., & Selten, R. (2002). *Bounded Rationality*. Cambridge: MIT Press.

Goldberg, L. R. (1990). An alternative description of personality: The big-five factor structure. *Journal of Personality and Social Psychology, 59*(6), 1216–1229.

Goodyear, P. (2006). Technology and the articulation of vocational and academic interests: reflections on time, space and e-learning. *Studies in Continuing Education, 28*(2), 83–98.

Goodyear, P. Banks, S. Hodgson, V. and McConnell, D. (eds). (2004) *Advances in Research on Networked Learning*. London: Kluwer Academic Publishers.

Graf, S., & Bekele, R. (2006). Forming heterogeneous groups for intelligent collaborative learning systems with ant colony optimization. In: Proc. 8th International Conference on Intelligent Tutoring Systems (ITS 2006), Springer Berlin / Heidelberg, pp. 217–226.
Guglielmino, L. M. (2013). The Case for Promoting Self-Directed Learning in Formal Educational Institutions. *SA-eDUC, 10*(2).
Haines, R. (2014). Group development in virtual teams: An experimental reexamination. *Computers in Human Behavior, 39*, 213-222.
Hsiung, C. M. (2010). An experimental investigation into the efficiency of cooperative learning with consideration of multiple grouping criteria. *European Journal of Engineering Education, 35*(6), 670–692.
Huang, H. M. (2002). Toward constructivism for adult learners in online learning environments. *British Journal of Educational Technology, 33*(1), 27-37.
Humphrey, S. E., Hollenbeck, J. R., Meyer, C. J., & Ilgen, D. R (2007). Trait configurations in self-managed teams: A conceptual examination of the use of seeding to maximize and minimize trait variance in teams. *Journal of Applied Psychology, 92*, 885–892.
Isaksen, S. G., & Lauer, K. J. (2002). The Climate for Creativity and Change in Teams. *Creativity and Innovation Management, 11*, 74–86.
Jackson, J., Wood, D., Bogg, T., Walton, K., Harms, P., & Roberts, B. (2010). What do conscientious people do? Development and validation of the Behavioral Indicators of Conscientiousness (BIC). *Journal of Research in Personality, 44* (4), 501–511.
Johnson, D. & Johnson, R. (1999). *Learning Together and Alone: Co-Operative, Competitive and Individualistic Learning*. Allyn & Bacon: Needham Heights, MA, USA.
Johnson, D.W., Johnson, R., Stanne, M.B., & Garibaldi, A. (1990). Impact of group processing on achievement in cooperative groups. *The Journal of Social Psychology, 130*(4), 507–516.
Jones, C.R., Ferreday, D., & Hodgson, V. (2008). Networked learning a relational approach: weak and strong ties. *Journal of Computer Assisted Learning, 24*, 90–102. doi: 10.1111/j.1365-2729.2007.00271.x
Jordanous, A. (2012). A standardised procedure for evaluating creative systems: Computational creativity evaluation based on what it is to be creative. *Cognitive Computation, 4*(3), 246-279.
Jorge-Botana, G., León, J.A., Olmos, R., & Escudero, I. (2010). Latent Semantic Analysis Parameters for Essay Evaluation using Small-Scale Corpora. *Journal of Quantitative Linguistics, 17*(1), 1–29.
Judge, T.A., & Ferris, G.R. (1992). The elusive criterion of fit in human resource staffing decisions. *Human Resource Planning, 15*(4), 47–67.
Kalz, M, van Bruggen, J., Giesbers, B., Waterink, W., Eshuis, J., & Koper, R. (2014). A study about placement support using semantic similarity. *Educational Technology & Society, 17*(3), 54–64.
Kaufman, J. C., & Sternberg, R. J. (Eds.), (2010). *The Cambridge handbook of creativity*. Cambridge University Press.
Kaufman, S.B. (2011). Is conscientiousness compatible with creativity? Psychology Today. Last accessed on November, 19th, 2013 from http://www.psychologytoday.com/blog/beautiful-minds/201108/is-conscientiousness-compatible-creativity
Kester, L., & Sloep, P. B. (2009). Knowledge Dating and Knowledge Sharing in Ad-Hoc Transient Communities. In R. Koper (Ed.), *Learning Network Services for Professional Development* (pp. 43–55). Berlin, Heidelberg: Springer.
Kim, K.J. (2009). Motivational Challenges of Adult Learners in Self-Directed e-Learning. *Journal of Interactive Learning Research, 20*(3), 317–335.
Kirton, M. (2003). *Adaption—Innovation in the context of diversity and change*. London: Routledge.
Knowles, M. S. (1975). *Self-directed learning*. New York: association Press.
Kolmos, A. (2012). Changing the Curriculum to Problem-Based and Project-Based Learning. *Outcome-Based Science, Technology, Engineering, and Mathematics Education: Innovative Practices,* 50-61.
Kop, R., Fournier, H., & Mak, J.S.F. (2011). A pedagogy of abundance or a pedagogy to support human beings? Participant support on massive open online courses. International Review of Research on Open and Distance Learning (IRRODL), 12(7). Last accessed on June 9th 2014 from http://www.irrodl.org/index.php/irrodl/article/view/1041/2025

Kop, R., & Hill, A. (2008). Connectivism: Learning theory of the future or vestige of the past? International Review of Research in Open and Distance Learning, 9(3), Article–9. Retrieved from http://www.irrodl.org/index.php/irrodl/article/view/523/1103

Koper, R. (2009). (Ed.) *Learning Network Services for Professional Development*. Berlin and Heidelberg: Springer.

Koper, R., & Sloep, P. B. (2002). Learning Networks connecting people, organizations, autonomous agents and learning resources to establish the emergence of effective lifelong learning. RTD Programme into Learning Technologies 2003-2008. More is different. Heerlen, The Netherlands: Open Universiteit Nederland.

Kristof, A. L.. (1996). Person-organization fit: An integrative review of its conceptualizations, measurements, and implications. *Personnel Psychology, 49*, 1–49.

Laham, D., Bennett, W. Jr., & Landauer, T.K. (2000). An LSA-based software tool for matching jobs, people, and instruction. *Interactive Learning Environments, 8*(3), 171-185.

Landauer, T. K. (2007). *LSA as a theory of meaning*. Chapter in the Handbook of Latent Semantic Analysis. Pp. 3–34. Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). Psychology Press. Lawrence Erlbaum; 1st edition.

Landauer, T. K., Foltz, P. W., & Laham, D. (1998). Introduction to latent semantic analysis. Discourse Processes, 25, 259-284. Last accessed on June 10, 2014 from http://lsa.colorado.edu/papers/dp1.LSAintro.pdf

Landauer, T. K., McNamara, D. S., Dennis, S., & Kintsch, W. (Eds.). (2013). *Handbook of latent semantic analysis*. Psychology Press. Lawrence Erlbaum.

Lave, J., & Wenger, E., (1991). *Situated learning: legitimate peripheral participation*. Cambridge: Cambridge University Press.

Lennon, J. & McCartney, P. (1969). Lyrics to "*The End*", Abbey Road.

Ma, W. W., & Chan, A. (2014). Knowledge sharing and social media: Altruism, perceived online attachment motivation, and perceived online relationship commitment. *Computers in Human Behavior, 39*, 51-58.

Marin-Garcia, J.A., & Lloret, J. (2008). Improving teamwork with university engineering students: the effect of an assessment method to prevent shrinking. WSEAS Transactions on Advances in Engineering Education, 5, 1, 1–11.

Martín, E., & Paredes, P. (2004). Using learning styles for dynamic group formation in adaptive collaborative hypermedia systems. In Matera, M., Comai, S. (Eds.), Engineering Advanced Web Applications. Proceedings of Workshops in Connection with 4th International Conference on Web Engineering, 188–197. Munich: Rinton Press, Inc.

McGuire, R. (2013). Building a senses of community in Moocs. Last accessed on November, 9th, 2013 from http://online.qmags.com/CPT0813/default.aspx?pg=31&mode=1#pg31&mode1

Monclar, R. S., Oliveira, J., de Faria, F. F., Ventura, L., de Souza, J. M., & Campos, M. L. M. (2011, June). Using social networks analysis for collaboration and team formation identification. In *Computer Supported Cooperative Work in Design (CSCWD), 2011 15th International Conference on* (pp. 562-569). IEEE.

Moos, D. C., & Azevedo, R. (2008). Self-regulated learning with hypermedia: The role of prior domain knowledge. *Contemporary Educational Psychology, 33*(2), 270-298.

Morrison, D. (2013). How NOT to Design a MOOC: The Disaster at Coursera and How to Fix it. Last accessed on November, 9th, 2013 from http://onlinelearninginsights.wordpress.com/2013/02/01/how-not-to-design-a-mooc-the-disaster-at-coursera-and-how-to-fix-it

Muchinsky, P.M., & Monahan, C.J. (1987). What is person-environment congruence? Supplementary versus complementary models for fit. *Journal of Vocational Behavior, 31*, 268–277.

Murray, T., & Arroyo, I. (2002). Toward Measuring and Maintaining the Zone of Proximal Development in Adaptive Instructional Systems. In: Proceedings of the 6th International Conference on Intelligent Tutoring Systems, 749-758.

NovoEd (2014). Last accessed on July 10, 2014 from https://novoed.com/how-it-works

Oakley, B., Felder, R.M., Brent, & R., Elhajj, I. (2004). Turning student groups into effective teams. *Journal of Student Centered Learning, 2*(1), 10–34.

Obaya, A. (1999). Getting cooperative learning. *Science Education International, 10*(2), 25–27.
Ounnas, A., Davis, H. C., & Millard, D. E. (2009). A Framework for Semantic Group Formation in Education. *Educational Technology & Society, 12*(4), 43–55.
Paré, D.E. & Joordens, S. (2008), Peering into large lectures: examining peer and expert mark agreement using peerScholar, an online peer assessment tool. *Journal of Computer Assisted Learning, 24*, 526–540. doi: 10.1111/j.1365-2729.2008.00290.x
Paulus, P. (2000). Groups, Teams, and Creativity: The Creative Potential of Idea-generating Groups. *Applied Psychology, 49*, 237–262.
Penalvo, F. J. G., Zangrando, V., Holgado, A. G., Gonzalez, M. A. C., Pardo, A. M. S., Griffiths, D., Forment, M. A., Mykowska, A., Janssen, J., Alves, G. R., Minović, M. (2012). TRAILER project overview: Tagging, recognition and acknowledgment of informal learning experiences. In: Proc. International Symposium on Computers in Education (SIIE), 1–6.
Pollalis, Y.A., & Mavrommatis, G. (2009). Using similarity measures for collaborating groups formation: A model for distance learning environments. *European Journal of Operational Research, 193*(2), 626–636.
Rajagopal, K. (2013). Networking for Learning: The role of Networking in a Lifelong Learner's Professional Development. Doctoral Dissertation. Heerlen: Open Universiteit.
Rehder, B., Schreiner, M. E., Wolfe, M. B. W., Laham, D., Landauer, T. K., & Kintsch, W. (1998). Using Latent Semantic Analysis to assess knowledge: some technical considerations. *Discourse Processes, 25*(2-3), 337-354.
Reiter-Palmon, R., Illies, J. J., & Kobe-Cross, L. M. (2009). Conscientiousness is not always a good predictor of performance: The case of creativity. *The International Journal of Creativity & Problem Solving, 19*(2), 27–45.
Roberts, B. W., Kuncel, N., Shiner, R., Caspi, A., & Goldberg, L. (2007). The power of personality: A comparative analysis of the predictive validity of personality traits, SES, and IQ. *Perspectives in Psychological Science, 2*, 313 – 345.
Rodrigues, S., Oliveira, J., & De Souza, J. M. (2005). Competence Mining for Team Formation and Virtual Community Recommendation. In: Proc. Ninth International Conference on Computer Supported Cooperative Work in Design, 1, 44 –49.
Rusman, E., Van Bruggen, J., Sloep, P. B., Valcke, M., & Koper, R. (2012). Can I Trust You? Profile Elements that Inform First Impressions of Trustworthiness in Virtual Project Teams. *International Journal of Information Technology Project Management (IJITPM) 3*(1), 15-35.
Sie, R. L. L., Drachsler, H., Bitter-Rijpkema, M., & Sloep, P. B. (2012). To whom and why should I connect? Co-author Recommendation based on Powerful and Similar Peers*. International Journal of Technology Enhanced Learning, 4*(1), 121–137.
Sie, R. L. L., Bitter-Rijpkema, M. E., Stoyanov, S., & Sloep, P. B. (2014). Factors that influence cooperation in networks for innovation and learning. Computers in Human Behavior, doi:10.1016/-j.chb.2014.04.033
Siemens, G. (2004). Connectivism: A learning theory for the digital age. International Journal of Instructional Technology and Distance Learning. Last accessed on June, 3rd, 2013 from http://www.itdl.org/Journal/Jan_05/article01.htm
Siemens, G., & Baker, R. S. (2012). Learning analytics and educational data mining: towards communication and collaboration. In Proceedings of the 2nd international conference on learning analytics and knowledge (pp. 252-254). ACM.
Slavin, R.E. (1989). Research on cooperative learning: Consensus and Controversy. *Educational Leadership, 47*(4), 52–54.
Sloep, P. B., & Berlanga, A. J. (2011c). Learning Networks, Networked Learning Redes de Aprendizaje, Aprendizaje en Red. *Comunicar, XIX, 37*, 55–63.
Sloep, P. B., Berlanga, A. J., Greller, W., Stoyanov, S., Retalis, S., & Van der Klink, M. (2011a). Educational Innovation with Learning Networks: some pertinent tools and developments. Paper presented at the 2nd International Conference on Technology Enhanced Learning, Quality of Teaching and Reforming Education: Learning Technologies, Quality of Education, Educational Systems, Evaluation, Pedagogies (TECH-EDUCATION 2011). May, 18-20, (2011), Corfu, Greece.

Sloep, P.B. (2009). Fostering Sociability in Learning Networks through Ad-Hoc Transient Communities, In M. Purvis, B. T. R. Savarimuthu (Eds.). Computer-Mediated Social Networking. Proceedings of the First International Conference, ICCMSN 2008. Dunedin, New Zealand, June 2008, revised selected papers. (62–75). Heidelberg, Germany: Springer.

Sloep, P., Van der Klink, M., Brouns, F., Van Bruggen. J., & Didderen, W. (eds.). (2011b). *Leernetwerken*. Bohn Stafleu van Loghum.

Sloep, P.B. (2013). Learning in networks and in communities of practice. Retrieved on September, 9th 2013, from http://pbsloep.blogspot.nl/2013/09/learning-in-networks-and-in-communities.html

Sloep, P. B., Berlanga, A. J., & Retalis, S. S. (2014). Introduction to the special issue on Web-2.0 technologies in support of team-based learning for innovation. *Computers in Human Behavior*, doi:10.1016/j.chb.2014.04.031

Soh, L., Khandaker, N., & Jiang, H. (2008). I-MINDS: A Multiagent System for Intelligent Computer-Supported Collaborative Learning and Classroom Management. *International Journal of Artificial Intelligence in Education, 18*(2), 119–151.

De Spinoza, B. *Tractatus theologico-politicus* (1670). Citation translated by the author from the Dutch language edition delivered by Klever, W., Eburon, Delft, 2010, (pp. 223).

Spoelstra, H., Van Rosmalen, P., & Sloep, P. (2012). Project team formation support for self-directed learners in social learning networks. In: Proc. IADIS International Conference on Web Based Communities and Social Media (ICWBC & SM 2012), Lisbon, 2012, 89–96.

Spoelstra, H., Van Rosmalen, P., Van de Vrie, E. Obreza, M. & Sloep, P.B. (2013). A Team Formation and Project-based Learning Support Service for Social Learning Networks. *Journal of Universal Computer Science, 19*(10), 1474–1495.

Spoelstra, H., Van Rosmalen, P., & Sloep, P.B. (2014). Toward Project-based Learning and Team Formation in Open Learning Environments. *Journal of Universal Computer Science, 20*(1), 57–76.

Spoelstra, H., Van Rosmalen, P., Houtmans, T., & Sloep, P.B. (2015). Team Formation Instruments to Enhance Learner Interactions in Open Learning Environments. *Computers in Human Behavior, 45*, 11-20, http://dx.doi.org/10.1016/j.chb.2014.11.038.

Springer, L., Stanne, M.E., & Donovan, S.S. (1999). Effects of small-group learning on undergraduates in science, mathematics, engineering, and technology. *Review of Educational Research, 69*(1), 21–51.

Stacey, P. (2013). The Pedagogy of MOOCs. Retrieved online on September 23, 2014 from http://edtechfrontier.com/2013/05/11/the-pedagogy-of-moocs/

Stahl, G. (2006). *Group cognition: Computer support for building collaborative knowledge*. Cambridge, MA: MIT Press.

Stahl, G. (2013). Theories of Cognition in Collaborative Learning. Available from http://gerrystahl.-net/pub/clhandbook.pdf

Stanford University. (2012). Last accessed on November, 29th, 2013 from http://news.stanford.-edu/news/2012/september/venture-lab-platform-091712.html

Steeples, C. & Jones, C. (eds) (2002) *Networked Learning: Perspectives and Issues*. London: Springer.

The MathWorks, Inc. (2007). http://www.mathworks.nl/

The Open University. (2013). OLDsMOOC: Open Learning Design MOOC. http://www.olds.ac.uk/the-course/week-1/team-up

Tobar, C.M., & Luís de Freitas, R. (2007). A Support Tool for Student Group Definition. In: Proc. Frontiers In Education Conference - Global Engineering: Knowledge Without Borders, Opportunities Without Passports. FIE '07. 37th Annual 10-13 Oct. 2007. T3J-7 - T3J-8.

Valcke, M. (2009). Presentation at IAMSE 13th Annual Meeting - Leiden, the Netherlands June 29-July 3, 2009. Last accessed on July, 30th, 2014 from http://users.ugent.be/~mvalcke/CV/IAMSE_MVA.PPT

Van Bruggen, J., Sloep, P., Van Rosmalen, P., Brouns, F., Vogten, H., Koper, R., & Tattersall, C. (2004). Latent semantic analysis as a tool for learner positioning in learning networks for lifelong learning. *British Journal of Educational Technology, 35*, 729–738.

Van Rosmalen, P., Sloep, P., Kester, L., Brouns, F., De Croock, M., Pannekeet, K., & Koper, R. (2008a). A learner support model based on peer tutor selection. *Journal of Computer Assisted Learning, 24*(1), 74-86.

Van Rosmalen, P., Sloep, P.B, Brouns, F., Kester, L., Berlanga, A., Bitter, M., & Koper, R. (2008b). A model for online learner support based on selecting appropriate peer tutors. *Journal of Computer Assisted Learning, 24*, 483–493.

Van Rosmalen, P., Boon, J., Bitter-Rijpkema, M. E., Sie, R. L. L., & Sloep, P. B. (2014). Supporting co-creation with software, the idSpace platform. *Computers in Human Behavior, 37*, 413–422. doi:10.1016/j.chb.2014.04.036

Van Rosmalen, P., Sloep, P., Brouns, F., Kester, L., Koné, M., & Koper, R. (2006). Knowledge matchmaking in Learning Networks: Alleviating the tutor load by mutually connecting Learning Network users. *British Journal of Educational Technology, 37*, 881–895.

Verhagen, P. (2006). Connectivism: A new learning theory? Utrecht. Retrieved from http://elearning.surf.nl/e-learning/english/3793

Vygotsky, L. S. (1978). *Mind in society: the development of higher psychological processes.* Cole, M., John-Steiner, V., Scribner S. & Souberman, E. (Eds). Harvard University Press, Cambridge.

Wade-Stein, D., & Kintsch, E. (2004). Summary Street: Interactive computer support for writing. *Cognition and Instruction, 22*(3), 333-362.

Wang, D.Y., Lin, S.S., & Sun, C.T. (2007). Diana: A computer supported heterogeneous grouping system for teachers to conduct successful small learning groups. *Computers in Human Behavior, 23*, 1997–2010.

Wellman, B., Salaff, J., Dimitrova, D., Garton, L., Gulia, M., & Haythornthwaite, C. (1996). Computer Networks as Social Networks: Collaborative work, telework and virtual community. *Annual Review of Sociology, 22*(1), 213–238.

Werbel, J.D., & Gilliland, S.W. (1999). Person-Environment fit in the selection process. *Research in Personnel and Human Resources Management, 17*, 209–244.

Werbel, J.D., & Johnson, D.J. (2001). The use of person-group fit for employment selection: A missing link in person-environment fit. *Human Resource Management, 40*, 227–240.

West, M. (1997). *Developing Creativity in Organizations*. Leicester: British Psychological Society.

West, M. (2002). Sparkling fountains or stagnant ponds: An integrative model of creativity and innovation implementation in work groups. *Applied Psychology: An International Review, 51*, 355–424.

Westera, W. (1998). Paradoxes in open, networked learning environments: Toward a paradigm shift. *Educational Technology, 39*, 17-23.

Westera, W., & Sloep, P. B. (1998). The Virtual Company: Toward a Self-Directed, Competence-Based Learning Environment in Distance Education. *Educational Technology, 38*(1), 32–37.

Wi, H., Oh, S., Mun, J., & Jung, M. (2009). A team formation model based on knowledge and collaboration. *Expert Systems with Applications, 36*(5), 9121–9134.

Wiley, D. A. (2004). *Scalability and Sociability in Online Learning Environments*. In: H. F. O'Neil & R. S. Perez (Eds.), Web-Based Learning: Theory, Research, and Practice (pp. 295–306). Mahwah, NJ: Erlbaum.

Wiley, D. A., & Edwards, E. K. (2002). Online self-organizing social systems: the decentralized future of online learning. *Quarterly Review of Distance Education, 3*(1), 33.

Wilkinson, I., & Fung, I. (2002). Small-group composition and peer effects. *International Journal of Educational Research (Special issue), 37*, 425–447.

Wolfe, M. B. W., Schreiner, M. E., Rehder, B., Laham, D., Foltz, P., Kintsch, W., & Landauer, T.K. (1998). Learning from text: Matching readers and texts by Latent Semantic Analysis. *Discourse Processes, 25*(2-3), 309-336.

Wolfradt, U., & Pretz, J. E. (2001). Individual differences in creativity: personality, story writing, and hobbies. *European Journal of Personality, 15*, 297–310.

Zeimpekis, D., & Gallopoulos, E. (2006). TMG: A MATLAB toolbox for generating term document matrices from text collections, Grouping Multidimensional Data. Recent Advances in Clustering. J. Kogan, J., Nicholas, C. and Teboulle, M. (eds.), Springer, Berlin, 187-210.

Zhang, X., De Pablos, P. O., & Xu, Q. (2014). Culture effects on the knowledge sharing in multi-national virtual classes: A mixed method. *Computers in Human Behavior, 31*, 491-498.

Zhang, X., De Pablos, P., & Zhang, Y. (2012). The Relationship between Incentives, Explicit and Tacit Knowledge Contribution in Online Engineering Education Project. *International Journal of Engineering Education, 28*, 1341-1346

Summary

Nowadays learners looking for professional development opportunities can partake in online open learning environments, such as course-based Massive Open Online Courses (xMOOCs and cMOOCs) and open-ended social (learning) networks. A recent study showed that about 80% of these learners already have a bachelor's degree or higher, while over 40% are under 30 and almost 50% are between 30 and 60 years of age. Such online learners, this thesis maintains, are best served with collaborative constructivist learning settings because individual online learners can easily become isolated and lose motivation. However, support for collaborative learning in current open learning environments shows several deficiencies: i) in xMOOCs, collaborative learning receives limited attention; ii) in cMOOCs, the sheer number of collaboration opportunities and often ill-defined structure of the tasks quickly leads to learners getting lost; iii) in social (learning) networks, the wide range of (learning) materials makes it difficult for the learner to effectively define learning goals and find appropriate learning materials; iv) in any one of these environments it appears that finding effective teams of peer learners (in contrast to randomly assembled groups) is hardly supported, if at all (see Chapter 1, Section 3). We observe that none of the open learning environment's designs to date have ventured to select and implement a collaborative pedagogy in these environments. In response to this need, we decided to support project-based learning (PBL) in such environments. Among many other affordances, it prepares for important 21st-century skills (see Chapter 1, Section 4). From this perspective we start research to implement an automated team formation service for project-based learning in open learning environments (see Chapter 1, Section 5).

To form effective teams fit to execute a project, a team formation expert requires data about the prospective team members and the project task. Literature from the educational research domain mentions data such as the curriculum area in which the project task will be positioned, the project task itself and its characteristics (such as collaboration language, duration and team size), the individual learner's abilities, prior learner achievements, the individual learner's personality traits, and motivational orientation (see Chapter 2, Section 2.1). These we categorise under "knowledge" and "personality". A third category of data was included to cater for specific learner constraints: "preferences". Research shows that complementary and supplementary knowledge and personality are important factors in team formation. We suggest that varying combinations of knowledge and personality along the complementary/supplementary dimension can prepare teams for different project work outcomes: improving learning outcomes, enhancing the possibility of a creative project outcome, or improving productivity. Educational practitioners indicated that the most important category of data is *knowledge*, next comes *preferences* and finally *personality*. These practitioners also indicated a clear order in which they preferred team-work outcomes; first comes "Improve learning", then "Enhance creativity", and last "Improve productivity" (see Chapter 2, Section 3.3). As we surmise that in open learning environments staff is in short supply, and expertise to form teams might be unavailable, we research the possibility to automate the process of team formation for project-based learning. Figure 1 depicts the model of the process we aim to implement.

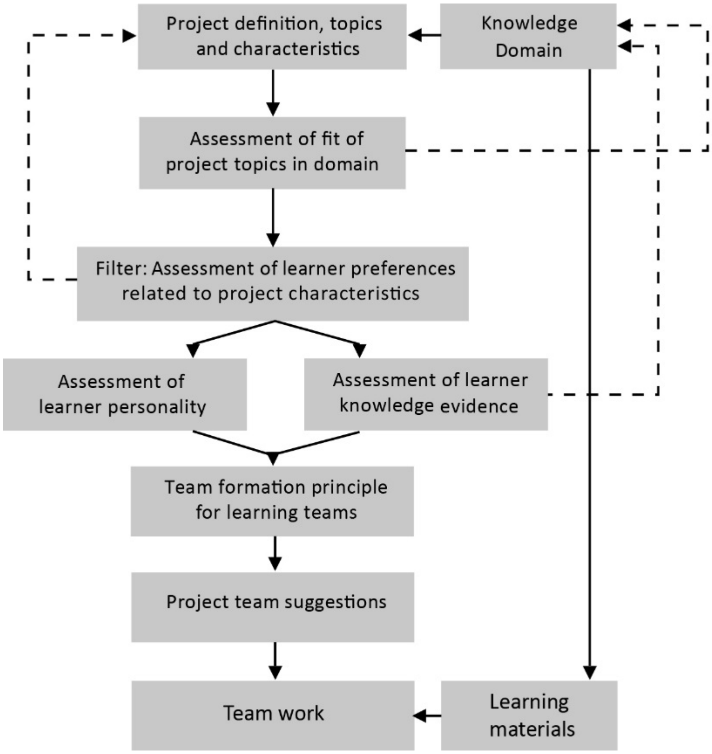


Figure 1. *Team formation for project-based learning model.*

The view it espouses is that projects are executed in knowledge domains, and are described in terms of the topics they address. The level of fit of the project topics to the domain determines whether and to what extent the project is executable in the domain. After filtering available project members on preferences, we assess learner knowledge and personality and, based on a specific team formation principle, determine whether a team of learners shows optimal fit to execute the project. Chapter 3 discusses automating the team formation process. A review of 14 existing team formation tools and techniques reveales that these assume data and user roles that are most likely not available in open learning environments (see Chapter 3, Section 2). Therefore we propose a design based on data that indeed can be acquired directly from the learners and the learning environment:

- For the analysis of both required and available *knowledge* we select Latent Semantic Analysis (LSA) as a suitable method.

- For the analysis of *personality*, we select the personality aspect of “conscientiousness”, as literature suggests that this aspect has the most predictive value on a person’s future performance in a team.
- With respect to *preferences* we propose to include availability schedule, languages mastered, and preferred tools. Such preferences determine whether collaborative project work can happen at all among prospective team members. Therefore an assessment of overlap in preferences precedes the assessments of knowledge and personality.

Based on the distinction between productive, creative or learning team outcomes as well as between supplementary or complementary knowledge and personality we performed a literature review of the effects of such combinations. This allows us to define team formation principles for the formation of such teams, as Table 1 shows.

Table 1: *Team work outcomes in combination with kind and levels of knowledge and conscientiousness levels.*

Project outcome	Kind and level of knowledge	Conscientiousness
Productive problem solving	Supplementary and high	All high
Creative solutions	Complementary and high	All low
Facilitating learning	Complementary and high, but within limits	All high

We thus arrive at the following principles for the formation of teams (see Chapter 3, Section 3.2):
1) "*Productivity in a team is fostered when team members have high scores on knowledge of the project topics and the team members show high, homogeneous levels of conscientiousness*".
2) "*Creativity in a team is fostered when team members have differentiated scores on knowledge of the project topics and the team members show low levels of conscientiousness*."
3) "*Learning in a team is facilitated when knowledge on the project topics is distributed over the members (allowing each member to learn and teach). However, the differences in knowledge should not be too large, and the team members should show high levels of conscientiousness.*"
These principles are translated into expressions (the details of which can be found in Appendix A) and put into algorithmic form.

Both the team formation principles and the outcomes of the team formation algorithms were validated. The results from a survey among teaching staff shows that participants accept the principles for the formation of productive and learning teams, but do *not* accept the principle for the formation of creative teams (see Chapter 4, Section 5.1). The same survey also aimed to validate the results of the team formation algorithms. From the fact that the principles for productive and learning teams were accepted, and respondents and algorithms (when small teams were concerned) show considerable overlap in classifying and ranking various teams, we conclude that our instruments perform adequately.

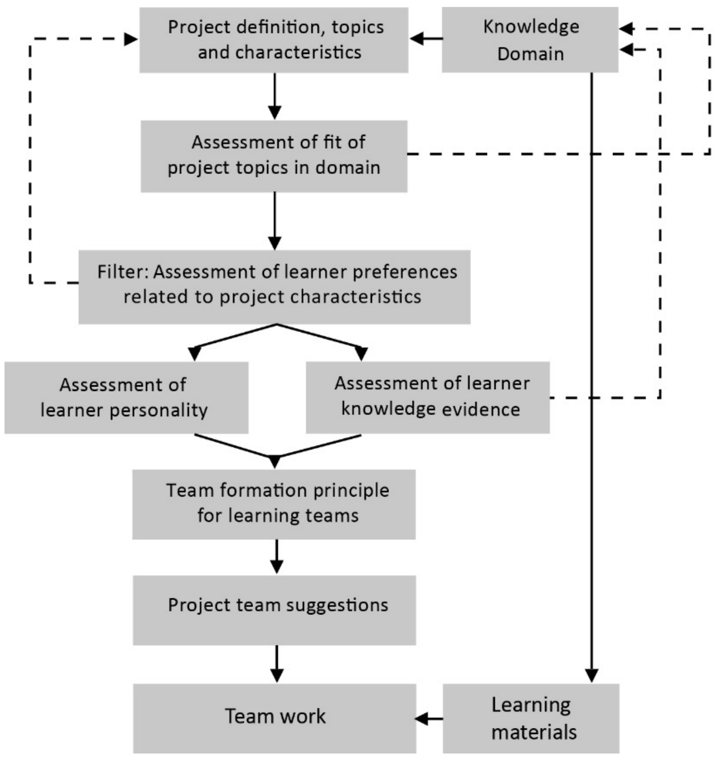
We implemented the final version of the team formation for PBL model (see Section 5.1) in an *experimental learning context.* We successfully applied LSA to prepare a representation of a knowledge domain in which projects could be run (see Chapter 5, Section 4.1). Two projects were described by their general tasks and by 4 topics they addressed. The LSA results of the topic descriptions analysed for their fit to the knowledge domain were validated by educational staff (see Chapter 5, Section 5.1). We then used LSA to assess learner prior knowledge on the project topics. The results (as confirmed by teaching staff) showed that LSA had assessed learner prior knowledge in line with the teacher's assessments and could thus be used to form teams (see Chapter 5, Section 5.2). Next, we mimicked team learning by forming duos of learners such that one partner of the duo had more prior knowledge than the other partner, and providing the less knowledgeable members with a learning task. The LSA analysis of the results from the learning task (again confirmed by teaching staff) showed that LSA is capable of assessing whether learning had taken place (see Chapter 5, Section 5.3).We found that both the width (with respect to on how *many* documents from the topic's domain document set the learners showed knowledge) and the depth (with respect to how *much* knowledge the learner showed on the domain documents) of knowledge had increased. We investigated the effects of the learner personality aspect of "conscientiousness" levels on learning outcomes and/or the collaboration processes. This revealed that it was not the knowledge gain that was affected by conscientiousness, but rather the reliability of learners in keeping appointments (when conscientiousness scores were low). We finally suggested learning materials to learners for them to assess the appropriateness of these materials. The learners confirmed that LSA is capable of suggesting valued learning materials from the domain of study, related to the project topics at hand. Our investigation into the location of our participant's zone of proximal development (in terms of the knowledge difference between learner and peer-teachers in our duos at which most learning took place) could not render conclusive results as the unexplained variance remained high.

In the final chapter we review the main findings of our study, address several methodological issues and limitations, describe our contributions to several research fields, and close off with recommendations for future research.

Samenvatting

Open en online leeromgevingen bieden professionals mogelijkheden voor hun verdere ontwikkeling. Daaronder vallen cursus-gebaseerde Massive Open Online Courses (zowel xMOOC's als cMOOC's) maar ook Sociale (leer)netwerken. Een recente studie toont aan dat 80% van de lerenden in MOOC's al een bachelor-diploma of hoger heeft, dat 40% van de deelnemers onder de 30 is, en bijna 50% tussen 30 en 60 is. Omdat individueel online leren gemakkelijk kan leiden tot isolatie en demotivatie, en omdat samenwerkend leren aansluit bij de werkpraktijken van veel van deze lerenden, stelt dit proefschrift dat zij het best gediend zijn met leeromgevingen die mede gebruik maken van samenwerkend leren. De huidige open leeromgevingen vertonen echter, daar waar het de ondersteuning van samenwerkend leren betreft, diverse tekortkomingen: i) in xMOOC's krijgt samenwerkend leren beperkte aandacht; ii) in cMOOC's zijn het juist de overvloed aan samenwerkingsmogelijkheden en de vaak slecht gestructureerde taken die al snel leiden tot "dwalende" lerenden; iii) in Sociale netwerken maakt de breedte van het aanbod van (leer)materialen het de lerenden vaak onmogelijk zelf effectief leerdoelen te formuleren en daar leermaterialen bij te vinden; iv) in deze omgevingen is ondersteuning voor het samenstellen van effectieve *teams* van peer-lerenden (in contrast met bij toeval gevormde *groepen*) vrijwel non-existent (Zie Hoofdstuk 1, Sectie 3).
De huidige open leeromgevingen worden vrijwel nooit ontworpen op basis van een pedagogiek gericht op samenwerkend leren. Ons onderzoek richt zich specifiek op het ondersteunen van projectgebaseerd samenwerkend leren (PBL) in deze leeromgevingen. Een belangrijke eigenschap van PBL, is dat het lerenden voorbereidt op het gebruik van 21ste-eeuwse vaardigheden. We gaan er vanuit dat docenten in open leeromgevingen (voor zover aanwezig) het aan tijd en kennis ontbreekt om PBL en het bijbehorende proces van teamformatie gedegen uit te voeren. Vanuit dat perspectief werd ons onderzoek naar het automatiseren van team formatie voor project-gebaseerd leren in open leeromgevingen gestart (Zie Hoofdstuk 1, Sectie 5).
Effectieve teams worden bij voorkeur gevormd door teamformatie-experts. Deze gebruiken daarbij hun kennis van de mogelijke teamleden en de projectopdracht. Specifiek met betrekking tot onderwijs noemt de literatuur data zoals het deel van het curriculum waar het project betrekking op heeft, de inhoudelijke projectopdracht en haar karakteristieken (zoals de taal waarin samengewerkt kan worden, de duur van het project, en de grootte van het team), de kennis van de individuele lerende, eerder behaalde resultaten, persoonlijkheidskenmerken, en motivatie (zie Hoofdstuk 2, Sectie 1). Deze data sorteerden we onder de categorieën "kennis-gerelateerd" en "persoonlijkheid-gerelateerd". Om rekening te kunnen houden met specifieke, aan de individuele lerende gebonden, eigenschappen, voegden we "voorkeuren" als derde categorie toe. Onderzoek toont verder aan dat de principes van complementariteit en supplementariteit met betrekking tot kennis en persoonlijkheid belangrijke aspecten zijn bij de formatie van teams. Op basis van de literatuur stelden we drie combinaties van kennis en persoonlijkheid voor, met

als doel de formatie van teams geschikt voor: verbetering van het leren, verhogen van de creativiteit, of verbetering van de productiviteit. Professionals werkzaam in het hoger onderwijs gaven aan dat “kennis” de belangrijkste categorie is om rekening mee te houden tijdens het formeren van teams. Daarna volgt “voorkeuren”, en tenslotte “persoonlijkheid”. De professionals gaven ook aan dat wat hen betreft er een duidelijke volgorde is waarin projectuitkomsten gewaardeerd worden: ten eerste “verbeteren van het leren”, ten tweede “verhogen van de creativiteit”, en ten derde “verbeteren van de productiviteit” (Zie Hoofdstuk 2, Sectie 3.3). Omdat we aannemen dat staf in open leeromgevingen beperkt beschikbaar is, en teamformatie-expertise waarschijnlijk ontbreekt, onderzochten we mogelijkheden om het teamformatieproces ten behoeve van projectgebaseerd leren te automatiseren. Figuur 1 toont het model dat we ontwikkeld en geïmplementeerd hebben.



Figuur 1: *model voor teamformatie voor project-gebaseerd leren.*

Dit model gaat er vanuit dat projecten uitgevoerd worden in een kennisdomein en beschreven worden aan de hand van onderwerpen in dit domein. Elk projectvoorstel wordt eerst gecontroleerd of het voldoende past binnen het gekozen domein. Mogelijke projectleden worden hierna gefilterd op hun

voorkeuren en indien passend wordt hun kennis en de persoonlijkheid beoordeeld. Tot slot wordt met behulp van de teamformatieprincipes het optimale team samengesteld uit de beschikbare kandidaten.
In Hoofdstuk 3 wordt het automatiseren van het teamformatieproces bediscussieerd. Een analyse van 14 bestaande teamformatiegereedschappen toonde aan dat deze aannamen doen over de beschikbaarheid van data en gebruikersrollen die in open leeromgevingen zeer waarschijnlijk niet beschikbaar zijn (zie Hoofdstuk 3, Sectie 2). Daarom stelden we een implementatie van het model voor, gebaseerd op data die direct van de lerenden en uit de leeromgeving verkregen kan worden, die gebruik maakt van:

- Latente Semantische Analyse (LSA) als methode om te analyseren hoe goed het project past binnen het domein en om de bij projectleden beschikbare kennis vast te stellen.
- het Big Five persoonlijkheidsaspect "Conscientieusheid" voor de analyse van de persoonlijkheid van het projectlid, omdat dit aspect volgens de literatuur de grootste voorspellende waarde heeft met betrekking tot toekomstige prestaties in een team.
- voorkeuren met betrekking tot beschikbaarheid, talenkennis en voorkeursgereedschappen. Zulke voorkeuren bepalen of het überhaupt tot samenwerking kan komen. Daarom gaat in het model het bepalen van een overlap in voorkeuren vooraf aan de bepaling van kennis en persoonlijkheid.

Een literatuuronderzoek toonde aan hoe (met betrekking tot complementaire en supplementaire kennis en persoonlijkheid), productieve, creatieve en lerende teams gevormd kunnen worden (Zie Tabel 1).

Tabel 1: *Uitkomsten van projectwerk gecombineerd met soort en niveau van kennis, en conscientieusheid.*

Projectwerkuitkomst	Soort en niveau van kennis	Conscientieusheid
Productief problemen oplossen	Supplementair en hoog	Alle hoog
Creatieve oplossingen	Complementair en hoog	Alle laag
Ondersteunen van leren	Complementair en hoog, maar binnen grenzen	Alle hoog

Hieruit werden de volgende principes voor de formatie van teams (zie Hoofdstuk 3, Sectie 2.2) opgesteld:
1) "*Productiviteit in een team wordt ondersteund wanneer de teamleden hoge kennisscores hebben op de projectonderwerpen en de teamleden hoge, homogene niveaus van conscientieusheid vertonen.*"
2) "*Creativiteit in een team wordt ondersteund wanneer de teamleden gedifferentieerde kennisscores hebben op de projectonderwerpen en de teamleden lage niveaus van conscientieusheid vertonen.*"
3) "*Leren in een team wordt ondersteund wanneer de kennis over de projectonderwerpen verdeeld is over de projectleden (waardoor alle leden zowel van*

elkaar kunnen leren als elkaar onderwijzen). De verschillen in kennis mogen echter niet te hoog zijn. De teamleden vertonen hoge niveaus van conscientieusheid."
Deze principes werden vertaald naar expressies (zie Appendix A) en in algoritmen geïmplementeerd.
Het onderzoek valideerde vervolgens zowel de teamformatieprincipes als de uitkomsten van de teamformatiealgoritmen. Onderwijsprofessionals gaven aan dat ze de principes voor de formatie van productieve en lerende team onderschrijven. Het principe voor de formatie van creatieve teams werd echter niet geaccepteerd (zie Hoofdstuk 4, Sectie 5.1). Uit het feit dat de principes voor de formatie van productieve en lerende teams geaccepteerd werden, en dat de deelnemers en de algoritmen (daar waar het kleine teams betreft) grotendeels overlappende resultaten vertoonden op classificeer- en rangschiktaken, concludeerden we dat onze teamformatie-instrumenten adequaat functioneren.

De uiteindelijke versie van het teamformatie voor PBL model (zie Hoofdstuk 5, Sectie 1) werd geïmplementeerd in een *experimentele leercontext*. LSA werd daarbij succesvol ingezet om een representatie van een kennisdomein te realiseren waarmee projecten op hun uitvoerbaarheid beoordeeld konden worden (zie Hoofdstuk 5, Sectie 4.1). Twee projecten werden beschreven met betrekking tot de opdracht (het schrijven van een informatie-leaflet) en de vier onderwerpen die daarin aan de orde moesten komen. Daarna gebruikten we LSA om de voorkennis van studenten op de projectonderwerpen te bepalen. De LSA-resultaten over de toepasselijkheid van de projectvoorstellen binnen het kennisdomein en de voorkennismeting van de studenten werden door inhoudsdeskundigen gevalideerd (zie Hoofdstuk 5, Sectie 5.1). Dit toonde aan dat de LSA-resultaten grotendeels overeenkomen met de beoordelingen van de inhoudsdeskundigen en dat deze daarmee gebruikt kunnen worden om teams te formeren (zie hoofdstuk 5, sectie 5.1 en 5.2). Vervolgens bootsten we de basis van teamleren na door duo's te formeren waarin een lid meer kennis van een onderwerp had dan het andere lid. Het lid met de mindere kennis kreeg een leertaak. De LSA-analyse van de resultaten uit de leertaak (weer door inhoudsdeskundigen gevalideerd) toonde aan dat dit instrument ons in staat stelt te beoordelen of er leren heeft plaatsgevonden (zie Hoofdstuk 5, Sectie 5.3). De analyse toonde aan dat zowel de breedte (met betrekking tot *hoeveel* domeindocumenten de lerende kennis vertoonde) als de diepte (met betrekking tot hoe *goed* deze kennis de van de domeindocumenten was) van de kennis toegenomen was. Omdat we ervan uitgaan dat we teams samenstellen zowel op basis van kennis als persoonlijkheid, onderzochten we ook het effect van conscientieusheid op de leeropbrengsten en het samenwerkingsproces. We ontdekten dat de betrouwbaarheid met betrekking tot het zich houden aan afspraken significant slechter wordt bij lage conscientieusheidswaarden. Tenslotte boden we de studenten leermaterialen aan uit het kennisdomein, met de vraag deze te beoordeelden op hun toepasselijkheid voor de projecttaak. Hieruit konden we concluderen dat we door gebruik van LSA in staat zijn waardevolle, op projectonderwerpen toegespitste, leermaterialen te suggereren. We onderzochten verder of we een “zone van naaste ontwikkeling” konden vinden (in termen van het kennisverschil tussen lerende en peer-docent waarbij het grootste leereffect optrad). Dit onderzoek leverde geen eenduidige resultaten op. De onverklaarde variantie bleef hoog.

Het laatste hoofdstuk vat de belangrijkste resultaten uit het onderzoek samen. We bespreken diverse methodologische aandachtspunten en beperkingen. Daarnaast beschrijven we onze bijdragen aan verschillende onderzoeksgebieden. En ten slotte geven we suggesties voor toekomstig onderzoek.

Dankwoord

Na een zachte aansporing van mijn latere promotor, begon ik in 2009 aan het schrijven van het projectplan dat tot dit proefschrift zou leiden. Daarbij niet wetende wat er in het verschiet lag. Achteraf bezien staat het effect van deze keuze op gelijke hoogte met het effect van de keuze die ik als 23-jarige maakte om weer te gaan studeren: er gaat een nieuwe wereld voor je open. Daarom dank ik de Open Universiteit voor de gelegenheid om dit onderzoek te doen. In het bijzonder dank ik mijn promotor Peter Sloep en mijn dagelijks begeleider, Peter van Rosmalen. Veel dank ben ik ook verschuldigd aan mijn collega's van het Welten-instituut en de faculteit PenOW. Dit proefschrift kon alleen tot stand komen door de welwillende medewerking van velen van jullie. Zonder anderen te kort te willen doen, richt ik graag een bijzonder woord van dank aan:

Jasper von Grumbkow. Je advies om tijdens het promotietraject vooral tijd-kapers op de kust te mijden, was heel wijs. Gelukkig heb ik deze raad ter harte genomen.

Mieke Haemers. Je ondersteuning betreffende te volgen procedures en je fenomenale accuratesse hebben ervoor gezorgd dat dit proefschrift niet alleen op tijd af kwam, maar ook een bijzonder niveau van correctheid heeft gehaald.

Wim van der Vegt. Je raad en daad bij het programmeren van tool op tool (tot welke versienummer zijn we gekomen?), zelfs vanaf je ziekbed, zijn onmisbaar geweest bij het verwerken van de vele verzamelde data.

Jan van Bruggen. Je bent altijd bereid een boom op te zetten over meer dan LSA en methodologische kwesties. Gelukkig zag jij wel potentie in een ooit bijna uitgerangeerd groepje niet-gepromoveerde medewerkers.

Tilly Houtmans. Op elk moment, ook al heb je al afscheid genomen van de Open Universiteit, sta je terzijde als de statistiek weerbarstig lijkt.

Evert van de Vrie. Een bron van inspiratie waar het de mogelijkheden van praktische toepasbaarheid van het onderzoek betreft. En een docent die matrices en expressies tot iets begrijpelijks kan transformeren (voor mij deed je vaak het omgekeerde!).

Peter Sloep. Je vermogen overzicht te houden heeft tot inspirerende wendingen geleid. Je uitgesproken waardering en vertrouwen leidden tot werklust en kwaliteitsbesef.

Peter van Rosmalen. Je nooit aflatende kritische blik heeft me veel geleerd. Het oog van de meester maakt de tomaat immers vet. Er kan veel komen uit de creatieve botsing tussen wiskundigen en filosofen, dat hebben we hopelijk samen bewezen.

Paquita. Het leven is zoveel leuker geworden sinds ik je ken! Altijd verbazend en nooit saai. Je belichaamt een unieke combinatie van rustpunt en startpunt van actie. Onze levendige uitwisseling heeft immaterieel en materieel zoveel bijgedragen. Mijn allerliefste dank voor alle liefde, discussies, steun en begrip. Zonder jou was dit proefschrift niet tot stand gekomen. Pero ahora: venga, vamos!

Ook bijzondere dank aan mijn ouders, Annie en Folkert, die een ongekend vertrouwen uitstraalden in de goede afloop, ook als ik dat zelf niet altijd had. Daardoor ervoer ik onvoorwaardelijke emotionele steun. Op dat vlak bestaat echter ook één grote leegte: Pa, ik had zo graag gezien dat je ook aan het eind nog aanwezig zou zijn geweest. Ik mis je.

In een longitudinale studie (n = 1), lopend van 2009 tot 2015, heb ik uit eerste hand mogen ervaren wat het betekent project-gebaseerd een probleem in een open omgeving op te lossen. Er werd veel gecommuniceerd (in teams van wisselende samenstellingen) én geleerd. Of dat in mijn “zone van naaste ontwikkeling” gebeurde?

> *“And in the end, the love you take is equal to the love you make”*
>
> Lennon & McCartney, 1969.

Curriculum Vitae

Howard A.F. Spoelstra (1961) works at the Welten Institute Research Centre for Learning, Teaching and Technology of the Open University of the Netherlands. He studied Philosophy, Cognitive Science and Knowledge Management and teaches various courses. His early research activities included the development of the Virtual Company educational design, which is based on the integration of learning and working in small teams. He participated in several multi-year EU-funded projects, such as the LTfLL (Language Technologies for Lifelong Learning) project, with a focus on the use of natural language analysis tools to assess learner conceptual development and determining learner prior knowledge in order to offer effective learning materials; the Cooper project, which developed a workflow-based distributed collaborative learning environment; the BioAPP project which aims to enhance creative problem solving skills among learners. In his PhD thesis he investigated the automation of team formation processes for project-based learning to facilitate collaborative learning processes in large scale open learning environments.

Appendices

Appendix A. The team formation expressions for the formation of productive, creative or learning teams

1. The team formation expression for productive teams:

$$FitP_i = W_K * \frac{Avg\_K_i}{Max\_K} + W_C * \frac{Avg\_C_i}{Max\_C}$$

Expression A1: *Team formation expression for productive teams*

Explanation of the terms used in Expression A1:
$FitP_i$: The level of adherence of team$_i$ to the team formation principle for productive teams
Avg_K_i : The average of the knowledge scores of all team members over all topics addressed in the project.
Avg_C_i : The average of the conscientiousness values of all team members.
Max_K: The maximum knowledge score of a team member. Following a 10-point grading system, this value set to 10.
Max_C: The maximum conscientiousness score, calculated from the Big Five test. The maximum is 5.
W_K : The weight of the factor knowledge in the team formation expression.
W_C : The weight of the factor conscientiousness in the team formation expression. Both weights add up to 1.0, which guarantees that the $FitP_i$ value always varies between 0 and 1. For the experiments weights were set to 0.5 each, so knowledge was of equal importance in the calculation of $FitP_i$ as was conscientiousness.

2. The team formation expression for creative teams:

$$FitC_i = W_K * \frac{\sum_j DifK_j}{n * Max\_K} + W_E * \frac{\sum_t DifK_t}{k * Max\_K} + W_C * \frac{Max\_C - Avg\_C_i}{Max\_C}$$

Expression A2: *Team formation expression for creative teams*

Explanation of the terms used in Expression A2:
$FitC_i$: The level of adherence of team$_i$ to the team formation principle for creative teams
$DifK_j$: The sum of the differences between the highest and next highest score over all members inside the respective project topics.
n : The number of members in the team.
Max_K : The maximum knowledge score of a team member. Following a 10-point

grading system this is set to 10.
$DifK_t$: The sum of the difference between the highest and next highest score over all topics inside the respective member' scores.
k : The number of topics in the project
Max_C : The maximum conscientiousness score from the Big Five test. The maximum is 5.
Avg_C_i : The average of the conscientiousness values of the members of team$_i$.
W_K : The weight of the factor knowledge in the team formation expression.
W_E : The weight of the factor expertise in the team formation expression.
W_C : The weight of the factor conscientiousness in the team formation expression.

The three weights add up to 1.0, which guarantees that the $FitC_i$ value always varies between 0 and 1. For the experiments they were set to 0.33 each, so that knowledge inside a topic, knowledge over topics and conscientiousness were of equal importance in the calculation of $FitC_i$.

3. The team formation expression for learning teams:

The team formation principle for learning teams was defined as: *"Learning in a team is facilitated when knowledge on the project topics is distributed over the members (allowing each member to learn and teach). However, the differences in knowledge should not be too high, and the team members should show high levels of conscientiousness."* (Spoelstra, Van Rosmalen, Van de Vrie, Obreza, & Sloep, 2013). By requiring that knowledge differences between learners should not be too high, it aims to team up learners in such a way that knowledge differences can be bridged. This models one of the aspects from Vygotsky's principle of "zone of proximal development": *difference in knowledge between learners*. By requiring team members to have high conscientiousness scores it favours teams with members who exhibit conscientious behaviour (which exists of carefulness, thoroughness, sense of responsibility, level of organization, preparedness, inclination to work hard, orientation on achievement, and perseverance). This particular trait was chosen because of its ability to predict job performance (Jackson, Wood, Bogg, Walton, Harms, & Roberts, 2010). This principle is formalised into the following expression:

$$FitL_i = W_K * \frac{\sum_t \sum_l \sum_j \left| DifK_{tjl} \right|}{d_{jt} \cdot zpd \cdot n \cdot k} + W_C * \frac{Avg\_C_i}{Max\_C}$$

Expression A3: *Team formation expression for learning teams.*

Explanation of the terms used in Expression A3:
| $DifK_{tjl}$ | : The absolute difference between two learners' (j and l) knowledge score's inside a topic. These are summed up over all pairs of learners $_{j,l}$ inside a topic and over all topics $_t$ in the project. Topic scores can vary between 1 and 10, following a 10-point grading system.
d_{jt} : the difference between the number of times a member has a higher score and a lower score when compared to other members (i.e., the number of times a member can act as a peer-tutor or as a learner).
zpd : (zone of proximal development) (Vygotsky, 1978). The maximum difference in knowledge between learners. This value is set to 3 grade points for the current experiment.
n : the number of members in the team
k : the number of topics in the project
Avg_C_i : The average of the conscientiousness values of the members of team$_i$.
Max_C : The maximum conscientiousness score, calculated from the Big Five test. The maximum is 5.
W_K : The weight of the factor knowledge in the team formation expression.
W_C : The weight of the factor conscientiousness in the team formation expression.
The expression describes teams whose members can teach and learn to and from each other inside each topic, while having a high score on conscientiousness. It optimises the match between peer-tutors and learners in the team by modelling one of the aspects of Vygotsky's principle of "zone of proximal development": *difference in knowledge* between learner and peer-tutor. The parameter "zpd" is used to calculate teaching and learning effectiveness for the team on a topic. The algorithm implementing this expression adds two exemptions to the rule: If the difference between two topic scores is higher than the value of the parameter *zdp*, or when a peer-tutor has a score on a topic lower than a set minimum grade (currently set to 6), teaching and learning effectiveness for that peer-tutor/learner pair is set to be 0.
The weights W_K and W_C (both set to 0.5 in the current study) can be set to stress the importance of knowledge over conscientiousness in the team formation, or vice versa. The scores from the first and second part are multiplied by their weights (W_K and W_C) separately and then summed. As the two scores each result in a value between 0 and 1 and the sum of the weights should always be 1, this results in a measure of fit for each team considered ($FitL_i$) between 0 and 1.

Appendix B. Examples of team formation algorithm output, team classifying, and team ranking tasks

Team members	Productive	Creative	Learning
{Student114,Student120,Student67,Student8}	0.761625	0.03325	0.513361111
{Student114,Student120,Student178,Student67}	0.730875	0.036125	0.703815177
{Student120,Student219,Student67,Student8}	0.747875	0.040125	0.499611111
{Student114,Student178,Student219,Student8}	0.707	0.04025	0.636537399
{Student114,Student67,Student74,Student8}	0.744875	0.041625	0.496611111
{Student116,Student49,Student61,Student67}	0.826125	0.041625	0.534921296
{Student178,Student67,Student8,Student84}	0.70425	0.041625	0.736194444
{Student120,Student178,Student219,Student67}	0.717125	0.043	0.690065177
{Student103,Student120,Student67,Student8}	0.739	0.043	0.486627315
{Student116,Student49,Student61,Student8}	0.810875	0.043	0.653988426
{Student114,Student120,Student219,Student67}	0.751875	0.044375	0.435555556
{Student120,Student178,Student67,Student8}	0.726875	0.044375	0.710613788
{Student40,Student61,Student67,Student8}	0.814375	0.044375	0.657719907
{Student114,Student178,Student67,Student74}	0.714125	0.0445	0.687065177
{Student120,Student67,Student74,Student8}	0.73975	0.04575	0.484888889
{Student116,Student178,Student49,Student61}	0.780125	0.045875	0.462397352
{Student103,Student120,Student178,Student67}	0.70825	0.045875	0.660009621
{Student116,Student40,Student49,Student67}	0.82075	0.045875	0.533655093
{Student114,Student116,Student49,Student61}	0.814875	0.04725	0.534377315
{Student116,Student211,Student40,Student61}	0.821125	0.04725	0.492710648
{Student103,Student114,Student120,Student67}	0.743	0.04725	0.440222222
{Student114,Student120,Student158,Student67}	0.718	0.04725	0.47007908
{Student114,Student120,Student32,Student67}	0.708625	0.04725	0.681565177
{Student116,Student40,Student49,Student8}	0.8055	0.04725	0.652722222
{Student116,Student40,Student61,Student67}	0.83675	0.04725	0.545199074
{Student178,Student40,Student61,Student67}	0.783625	0.04725	0.636094126
{Student114,Student120,Student42,Student67}	0.74175	0.047875	0.460176288
{Student219,Student67,Student74,Student8}	0.731125	0.0485	0.482861111
{Student114,Student219,Student67,Student8}	0.753	0.0485	0.504388889
{Student114,Student40,Student61,Student67}	0.818375	0.048625	0.567043981

Figure A1: *Excerpt of team formation algorithm output for teams consisting of 4 members, showing team fit values on Productive, Creative and Learning outcome.*

"Productivity in a team is fostered when team members have higher scores on knowledge of the project topics and the team members show higher levels of conscientiousness"

"Creativity in a team is fostered when team members have differentiated scores on knowledge of the project topics and the team members show lower levels of conscientiousness."

"Learning in a team is facilitated when knowledge on the project topics is distributed over the members (allowing each member to learn and teach). However, the differences in knowledge should not be too high, and the team members should show higher levels of conscientiousness."

Team 1:

	Onderwerp 1	Onderwerp 2	Onderwerp 3	Onderwerp 4	Conscientieusheid
Student 1	7	8	8	7	4.44
Student 2	8	8	8	8	4.67
Student 3	9	8	8	8	5.00
Student 4	7	9	8	7	4.78

Team 2:

	Onderwerp 1	Onderwerp 2	Onderwerp 3	Onderwerp 4	Conscientieusheid
Student 1	8	8	8	8	4.33
Student 2	8	8	8	8	4.67
Student 3	5	5	6	5	3.22
Student 4	5	5	5	5	2.89

Team 3:

	Onderwerp 1	Onderwerp 2	Onderwerp 3	Onderwerp 4	Conscientieusheid
Student 1	7	7	8	8	4.44
Student 2	8	8	8	8	4.67
Student 3	9	8	8	8	5.00
Student 4	7	9	8	7	4.78

	Productief	Creatief	Lerend	Geen antwoord
Welk soort team is team 1?	○	○	○	○
Welk soort team is team 2?	○	○	○	○
Welk soort team is team 3?	○	○	○	○

Figure A2: *Example of the team classifying task for teams of 4 members, addressing 4 topics (in Dutch).*

Productieve teams:

"Productivity in a team is fostered when team members have higher scores on knowledge of the project topics and the team members show higher levels of conscientiousness"

Team 1:

	Onderwerp 1	Onderwerp 2	Conscientieusheid
Student1	7	6	3.67
Student 2	8	8	3.78

Team 2:

	Onderwerp 1	Onderwerp 2	Conscientieusheid
Student 1	4	4	3.44
Student 2	5	5	2.89

Team 3:

	Onderwerp 1	Onderwerp 2	Conscientieusheid
Student 1	9	8	5.00
Student 2	7	9	4.78

Aan u de vraag deze teams te rangschikken op volgorde van het voldoen aan het teamformatieprincipe voor productieve teams (1= past het best, 2 past minder goed, tot 3 = past het slechtst).

Klik een optie uit de lijst links. Begin met de optie die het meest toepasselijk is en ga door tot de minst toepasselijke optie.

Uw keuzes:

Team 1
Team 2
Team 3

Uw rangschikking:

1:
2:
3:

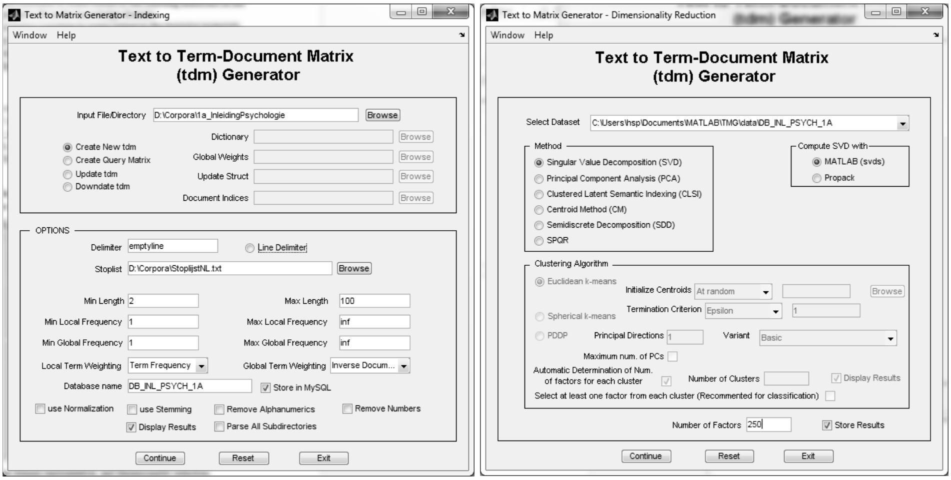
Klik op de schaar naast elk item om de laatst ingevoerde gegevens te verwijderen

Figure A3: *Example of the team ranking tasks for teams of 2 members, addressing 2 topics (in Dutch).*

Appendix C: LSA and settings used with Text-Matrix-Generator

The setup of a semantic space (the knowledge domain) can entail several steps. Often the source documents are stripped from occurrences of terms bearing no meaning for LSA (by means of a stop word list containing those terms) and of words below or above word-length thresholds. Then a term-document matrix is constructed, in which the cells contain the frequency of occurrences of a term in a document. Terms can be given a weight depending on how frequent they occur. An often used weighting scheme is Term Frequency * Inverse Document Frequency (TF*IDF). This scheme adds weight to terms appearing only infrequent in the source texts. As not all terms occur in all documents, the term-document matrix often has a high level of sparsity (cells containing a value of 0), but also contains noise (words occurring infrequently in a few documents. With large source document collections, the term-document matrix can become very large. In order to make the semantic relations between documents appear and to reduce computational complexity, the high dimensionality of the word-document matrix is reduced by means of singular value decomposition (SVD). A key factor in using SVD is in deciding how many dimensions to retain when approximating the original term-document matrix. When SVD is used to represent the data in a reduced dimensional space it puts an emphasis on strong relationships between terms and documents and it discards noise. It depends on the number of dimensions retained how well the original data is represented. If represented by too few dimensions, important relations are discarded. If represented by too many dimensions, the data will contain noise. After applying SVD, the data contains the *latent* semantic structure of the documents used as input. Each document represented in the data can be found by its unique vector in the matrix. This allows querying the semantic space for documents semantically related to the query document by translating the query document into a vector. This vector is then compared to the document vectors in the semantic space by taking the cosine of the angle between the vectors. A cosine between vectors of 1 (the vectors overlap) indicates a 100% semantic similarity (texts having the same meaning), while a cosine value of 0 (the vectors are at an angle of 90 degrees) indicates a 0% semantic similarity (texts having no meaningful relation to each other). In order to capture as many co-occurrences of words as possible, an LSA vector space should ideally be constructed from (very) large document collections. However, satisfying results can also be achieved by using smaller-scale document collections as the basis for semantic analysis, providing the corpus is highly sanitised (Van Bruggen, Sloep, Van Rosmalen, Brouns, Vogten, Koper & Tattersall, 2004; Jorge-Botana, Leon, Olmos & Escudero, 2010).

The TMG settings we used for the construction of the LSA space are depicted in the figures A4 and A5.



Figures A4 and A5: *the settings used in TMG to construct the LSA space.*

TMG reported the following corpus metrics: Number of documents = 2257, Number of terms = 19744, Average number of terms per document = 150.63, Average number of indexing terms per document = 86.53, Sparsity = 0.32%. The TMG application removed 216 stop words and 37 terms with a length above the term-length threshold.

Siks Dissertation Series

The complete list of SIKS dissertations (from 1998 onward) is available at http://www.siks.nl/dissertations.php

2009-01
Rasa Jurgelenaite (RUN)
Symmetric Causal Independence Models

2009-02
Willem Robert van Hage (VU)
Evaluating Ontology-Alignment Techniques

2009-03
Hans Stol (UvT)
A Framework for Evidence-based Policy Making Using IT

2009-04
Josephine Nabukenya (RUN)
Improving the Quality of Organisational Policy Making using Collaboration Engineering

2009-05
Sietse Overbeek (RUN)
Bridging Supply and Demand for Knowledge Intensive Tasks - Based on Knowledge, Cognition, and Quality

2009-06
Muhammad Subianto (UU)
Understanding Classification

2009-07
Ronald Poppe (UT)
Discriminative Vision-Based Recovery and Recognition of Human Motion

2009-08
Volker Nannen (VU)
Evolutionary Agent-Based Policy Analysis in Dynamic Environments
Feature Extraction from Visual Data

2009-18
Fabian Groffen (CWI)
Armada, An Evolving Database System

2009-19
Valentin Robu (CWI)
Modeling Preferences, Strategic Reasoning and Collaboration in Agent-Mediated Electronic Markets

2009-20
Bob van der Vecht (UU)
Adjustable Autonomy: Controling Influences on Decision Making

2009-21
Stijn Vanderlooy (UM)
Ranking and Reliable Classification

2009-22
Pavel Serdyukov (UT)
Search For Expertise: Going beyond direct evidence

2009-23
Peter Hofgesang (VU)
Modelling Web Usage in a Changing Environment

2009-24
Annerieke Heuvelink (VUA)
Cognitive Models for Training Simulations

2009-25
Alex van Ballegooij (CWI)
"RAM: Array Database Management through Relational Mapping"

2009-26
Fernando Koch (UU)
An Agent-Based Model for the Development of Intelligent Mobile Services

2009-27
Christian Glahn (OU)
Contextual Support of social Engagement and Reflection on the Web

2009-28
Sander Evers (UT)
Sensor Data Management with Probabilistic Models

2009-29
Stanislav Pokraev (UT)
Model-Driven Semantic Integration of Service-Oriented Applications

2009-30
Marcin Zukowski (CWI)
Balancing vectorized query execution with bandwidth-optimized storage

2009-31
Sofiya Katrenko (UVA)
A Closer Look at Learning Relations from Text

2009-32
Rik Farenhorst (VU) and Remco de Boer (VU)
Architectural Knowledge Management: Supporting Architects and Auditors

2009-33
Khiet Truong (UT)
How Does Real Affect Affect Affect Recognition In Speech?

2009-34
Inge van de Weerd (UU)
Advancing in Software Product Management: An Incremental Method Engineering Approach

2009-35
Wouter Koelewijn (UL)
Privacy en Politiegegevens; Over geautomatiseerde normatieve informatieuit-wisseling

2009-36
Marco Kalz (OUN)
Placement Support for Learners in Learning Networks

2009-37
Hendrik Drachsler (OUN)
Navigation Support for Learners in Informal Learning Networks

2009-38
Riina Vuorikari (OU)
Tags and self-organisation: a metadata ecology for learning resources in a multi-lingual context

2009-39
Christian Stahl (TUE, Humboldt-Universitaet zu Berlin)
Service Substitution -- A Behavioral Approach Based on Petri Nets

2009-40
Stephan Raaijmakers (UvT)
Multinomial Language Learning: Investigations into the Geometry of Language

2009-41
Igor Berezhnyy (UvT)
Digital Analysis of Paintings

2009-42
Toine Bogers (UvT)
Recommender Systems for Social Bookmarking

2009-43
Virginia Nunes Leal Franqueira (UT)
Finding Multi-step Attacks in Computer Networks using Heuristic Search and Mobile Ambients

2009-44
Roberto Santana Tapia (UT)
Assessing Business-IT Alignment in Networked Organizations

2009-45
Jilles Vreeken (UU)
Making Pattern Mining Useful

2009-46
Loredana Afanasiev (UvA)
Querying XML: Benchmarks and Recursion

2010-01
Matthijs van Leeuwen (UU)
Patterns that Matter

2010-02
Ingo Wassink (UT)
Work flows in Life Science

2010-03
Joost Geurts (CWI)
A Document Engineering Model and Processing Framework for Multimedia doc-uments

2010-04
Olga Kulyk (UT)
Do You Know What I Know? Situational Awareness of Co-located Teams in Mul-tidisplay Environments

2010-05
Claudia Hauff (UT)
Predicting the Effectiveness of Queries and Retrieval Systems

2010-06
Sander Bakkes (UvT)
Rapid Adaptation of Video Game AI

2010-07
Wim Fikkert (UT)
Gesture interaction at a Distance

2010-08
Krzysztof Siewicz (UL)
Towards an Improved Regulatory Framework of Free Software. Protecting user freedoms in a world of software communities and eGovernments

2010-09
Hugo Kielman (UL)
A Politiele gegevensverwerking en Privacy, Naar een effectieve waarborging

2010-10
Rebecca Ong (UL)
Mobile Communication and Protection of Children

2010-11
Adriaan Ter Mors (TUD)
The world according to MARP: Multi-Agent Route Planning

2010-12
Susan van den Braak (UU)
Sensemaking software for crime analysis

2010-13
Gianluigi Folino (RUN)
High Performance Data Mining using Bio-inspired techniques

2010-14
Sander van Splunter (VU)
Automated Web Service Reconfiguration

2010-15
Lianne Bodenstaff (UT)
Managing Dependency Relations in Inter-Organizational Models

2010-16
Sicco Verwer (TUD)
Efficient Identification of Timed Automata, theory and practice

2010-17
Spyros Kotoulas (VU)
Scalable Discovery of Networked Resources: Algorithms, Infrastructure, Applications

2010-18
Charlotte Gerritsen (VU)
Caught in the Act: Investigating Crime by Agent-Based Simulation

2010-19
Henriette Cramer (UvA)
People's Responses to Autonomous and Adaptive Systems

2010-20
Ivo Swartjes (UT)
Whose Story Is It Anyway? How Improv Informs Agency and Authorship of Emergent Narrative

2010-21
Harold van Heerde (UT)
Privacy-aware data management by means of data degradation

2010-22
Michiel Hildebrand (CWI)
End-user Support for Access to Heterogeneous Linked Data

2010-23
Bas Steunebrink (UU)
The Logical Structure of Emotions

2010-24
Dmytro Tykhonov
Designing Generic and Efficient Negotiation Strategies

2010-25
Zulfiqar Ali Memon (VU)
Modelling Human-Awareness for Ambient Agents: A Human Mindreading Perspective

2010-26
Ying Zhang (CWI)
XRPC: Efficient Distributed Query Processing on Heterogeneous XQuery Engines

2010-27
Marten Voulon (UL)
Automatisch contracteren

2010-28
Arne Koopman (UU)
Characteristic Relational Patterns

2010-29
Stratos Idreos(CWI)
Database Cracking: Towards Auto-tuning Database Kernels

2010-30
Marieke van Erp (UvT)
Accessing Natural History - Discoveries in data cleaning, structuring, and retrieval

2010-31
Victor de Boer (UVA)
Ontology Enrichment from Heterogeneous Sources on the Web

2010-32
Marcel Hiel (UvT)
An Adaptive Service Oriented Architecture: Automatically solving Interoperability Problems

2010-33
Robin Aly (UT)
Modeling Representation Uncertainty in Concept-Based Multimedia Retrieval

2010-34
Teduh Dirgahayu (UT)
Interaction Design in Service Compositions

2010-35
Dolf Trieschnigg (UT)
Proof of Concept: Concept-based Biomedical Information Retrieval

2010-36
Jose Janssen (OU)
Paving the Way for Lifelong Learning; Facilitating competence development through a learning path specification

2010-37
Niels Lohmann (TUE)
Correctness of services and their composition

2010-38
Dirk Fahland (TUE)
From Scenarios to components

2010-39
Ghazanfar Farooq Siddiqui (VU)
Integrative modeling of emotions in virtual agents

2010-40
Mark van Assem (VU)
Converting and Integrating Vocabularies for the Semantic Web

2010-41
Guillaume Chaslot (UM)
Monte-Carlo Tree Search

2010-42
Sybren de Kinderen (VU)
Needs-driven service bundling in a multi-supplier setting - the computational e3-service approach

2010-43
Peter van Kranenburg (UU)
A Computational Approach to Content-Based Retrieval of Folk Song Melodies

2010-44
Pieter Bellekens (TUE)
An Approach towards Context-sensitive and User-adapted Access to Heterogeneous Data Sources, Illustrated in the Television Domain

2010-45
Vasilios Andrikopoulos (UvT)
A theory and model for the evolution of software services

2010-46
Vincent Pijpers (VU)
e3alignment: Exploring Inter-Organizational Business-ICT Alignment

2010-47
Chen Li (UT)
Mining Process Model Variants: Challenges, Techniques, Examples

2010-48
Withdrawn

2010-49
Jahn-Takeshi Saito (UM)
Solving difficult game positions

2010-50
Bouke Huurnink (UVA)
Search in Audiovisual Broadcast Archives

2010-51
Alia Khairia Amin (CWI)
Understanding and supporting information seeking tasks in multiple sources

2010-52
Peter-Paul van Maanen (VU)
Adaptive Support for Human-Computer Teams: Exploring the Use of Cognitive Models of Trust and Attention

2010-53
Edgar Meij (UVA)
Combining Concepts and Language Models for Information Access

2011-01
Botond Cseke (RUN)
Variational Algorithms for Bayesian Inference in Latent Gaussian Models

2011-02
Nick Tinnemeier(UU)
Organizing Agent Organizations. Syntax and Operational Semantics of an Organization-Oriented Programming Language

2011-03
Jan Martijn van der Werf (TUE)
Compositional Design and Verification of Component-Based Information Systems

2011-04
Hado van Hasselt (UU)
Insights in Reinforcement Learning; Formal analysis and empirical evaluation of temporal-difference learning algorithms

2011-05
Base van der Raadt (VU)
Enterprise Architecture Coming of Age - Increasing the Performance of an Emerging Discipline.

2011-06
Yiwen Wang (TUE)
Semantically-Enhanced Recommendations in Cultural Heritage

2011-07
Yujia Cao (UT)
Multimodal Information Presentation for High Load Human Computer Interaction

2011-08
Nieske Vergunst (UU)
BDI-based Generation of Robust Task-Oriented Dialogues

2011-09
Tim de Jong (OU)
Contextualised Mobile Media for Learning

2011-10
Bart Bogaert (UvT)
Cloud Content Contention

2011-11
Dhaval Vyas (UT)
Designing for Awareness: An Experience-focused HCI Perspective

2011-12
Carmen Bratosin (TUE)
Grid Architecture for Distributed Process Mining

2011-13
Xiaoyu Mao (UvT)
Airport under Control. Multiagent Scheduling for Airport Ground Handling

2011-14
Milan Lovric (EUR)
Behavioral Finance and Agent-Based Artificial Markets

2011-15
Marijn Koolen (UvA)
The Meaning of Structure: the Value of Link Evidence for Information Retrieval

2011-16
Maarten Schadd (UM)
Selective Search in Games of Different Complexity

2011-17
Jiyin He (UVA)
Exploring Topic Structure: Coherence, Diversity and Relatedness

2011-18
Mark Ponsen (UM)
Strategic Decision-Making in complex games

2011-19
Ellen Rusman (OU)
The Mind' s Eye on Personal Profiles

2011-20
Qing Gu (VU)
Guiding service-oriented software engineering - A view-based approach

2011-21
Linda Terlouw (TUD)
Modularization and Specification of Service-Oriented Systems

2011-22
Junte Zhang (UVA)
System Evaluation of Archival Description and Access

2011-23
Wouter Weerkamp (UVA)
Finding People and their Utterances in Social Media

2011-24
Herwin van Welbergen (UT)
Behavior Generation for Interpersonal Coordination with Virtual Humans On Specifying, Scheduling and Realizing Multimodal Virtual Human Behavior

2011-25
Syed Waqar ul Qounain Jaffry (VU))
Analysis and Validation of Models for Trust Dynamics

2011-26
Matthijs Aart Pontier (VU)
Virtual Agents for Human Communication - Emotion Regulation and Involve- ment-Distance Trade-Offs in Embodied Conversational Agents and Robots

2011-27
Aniel Bhulai (VU)
Dynamic website optimization through autonomous management of design pat- terns

2011-28
Rianne Kaptein (UVA)
Effective Focused Retrieval by Exploiting Query Context and Document Structure

2011-29
Faisal Kamiran (TUE)
Discrimination-aware Classification

2011-30
Egon van den Broek (UT)
Affective Signal Processing (ASP): Unraveling the mystery of emotions

2011-31
Ludo Waltman (EUR)
Computational and Game-Theoretic Approaches for Modeling Bounded Rati oality

2011-32
Nees-Jan van Eck (EUR)
Methodological Advances in Bibliometric Mapping of Science

2011-33
Tom van der Weide (UU)
Arguing to Motivate Decisions

2011-34
Paolo Turrini (UU)
Strategic Reasoning in Interdependence: Logical and Game-theoretical Investigations

2011-35
Maaike Harbers (UU)
Explaining Agent Behavior in Virtual Training

2011-36
Erik van der Spek (UU)
Experiments in serious game design: a cognitive approach

2011-37
Adriana Burlutiu (RUN)
Machine Learning for Pairwise Data, Applications for Preference Learning and Supervised Network Inference

2011-38
Nyree Lemmens (UM)
Bee-inspired Distributed Optimization

2011-39
Joost Westra (UU)
Organizing Adaptation using Agents in Serious Games

2011-40
Viktor Clerc (VU)
Architectural Knowledge Management in Global Software Development

2011-41
Luan Ibraimi (UT)
Cryptographically Enforced Distributed Data Access Control

2011-42
Michal Sindlar (UU)
Explaining Behavior through Mental State Attribution

2011-43
Henk van der Schuur (UU)
Process Improvement through Software Operation Knowledge

2011-44
Boris Reuderink (UT)
Robust Brain-Computer Interfaces

2011-45
Herman Stehouwer (UvT)
Statistical Language Models for Alternative Sequence Selection

2011-46
Beibei Hu (TUD)
Towards Contextualized Information Delivery: A Rule-based Architecture for the Domain of Mobile Police Work

2011-47
Azizi Bin Ab Aziz(VU)
Exploring Computational Models for Intelligent Support of Persons with Depression

2011-48
Mark Ter Maat (UT)
Response Selection and Turn-taking for a Sensitive Artificial Listening Agent

2011-49
Andreea Niculescu (UT)
Conversational interfaces for task-oriented spoken dialogues: design aspects influencing interaction quality

2012-01
Terry Kakeeto (UvT)
Relationship Marketing for SMEs in Uganda

2012-02
Muhammad Umair(VU)
Adaptivity, emotion, and Rationality in Human and Ambient Agent Models

2012-03
Adam Vanya (VU)
Supporting Architecture Evolution by Mining Software Repositories

2012-04
Jurriaan Souer (UU)
Development of Content Management System-based Web Applications

2012-05
Marijn Plomp (UU)
Maturing Interorganisational Information Systems

2012-06
Wolfgang Reinhardt (OU)
Awareness Support for Knowledge Workers in Research Networks

2012-07
Rianne van Lambalgen (VU)
When the Going Gets Tough: Exploring Agent-based Models of Human Performance under Demanding Conditions

2012-08
Gerben de Vries (UVA)
Kernel Methods for Vessel Trajectories

2012-09
Ricardo Neisse (UT)
Trust and Privacy Management Support for Context-Aware Service Platforms

2012-10
David Smits (TUE)
Towards a Generic Distributed Adaptive Hypermedia Environment

2012-11
J.C.B. Rantham Prabhakara (TUE)
Process Mining in the Large: Preprocessing, Discovery, and Diagnostics

2012-12
Kees van der Sluijs (TUE)
Model Driven Design and Data Integration in Semantic Web Information Systems

2012-13
Suleman Shahid (UvT)
Fun and Face: Exploring non-verbal expressions of emotion during playful interactions

2012-14
Evgeny Knutov (TUE)
Generic Adaptation Framework for Unifying Adaptive Web-based Systems

2012-15
Natalie van der Wal (VU)
Social Agents. Agent-Based Modelling of Integrated Internal and Social Dynamics of Cognitive and Affective Processes.

2012-16
Fiemke Both (VU)
Helping people by understanding them - Ambient Agents supporting task execution and depression treatment

2012-17
Amal Elgammal (UvT)
Towards a Comprehensive Framework for Business Process Compliance

2012-18
Eltjo Poort (VU)
Improving Solution Architecting Practices

2012-19
Helen Schonenberg (TUE)
What's Next? Operational Support for Business Process Execution

2012-20
Ali Bahramisharif (RUN)
Covert Visual Spatial Attention, a Robust Paradigm for Brain-Computer Interfacing

2012-21
Roberto Cornacchia (TUD)
Querying Sparse Matrices for Information Retrieval

2012-22
Thijs Vis (UvT)
Intelligence, politie en veiligheidsdienst: verenigbare grootheden?

2012-23
Christian Muehl (UT)
Toward Affective Brain-Computer Interfaces: Exploring the Neurophysiology of Affect during Human Media Interaction

2012-24
Laurens van der Werff (UT)
Evaluation of Noisy Transcripts for Spoken Document Retrieval

2012-25
Silja Eckartz (UT)
Managing the Business Case Development in Inter-Organizational IT Projects: A Methodology and its Application

2012-26
Emile de Maat (UVA)
Making Sense of Legal Text

2012-27
Hayrettin Gurkok (UT)
Mind the Sheep! User Experience Evaluation & Brain-Computer Interface Games

2012-28
Nancy Pascall (UvT)
Engendering Technology Empowering Women

2012-29
Almer Tigelaar (UT)
Peer-to-Peer Information Retrieval

2012-30
Alina Pommeranz (TUD)
Designing Human-Centered Systems for Reflective Decision Making

2012-31
Emily Bagarukayo (RUN)
A Learning by Construction Approach for Higher Order Cognitive Skills Improvement, Building Capacity and Infrastructure

2012-32
Wietske Visser (TUD)
Qualitative multi-criteria preference representation and reasoning

2012-33
Rory Sie (OUN)
Coalitions in Cooperation Networks (COCOON)

2012-34
Pavol Jancura (RUN)
Evolutionary analysis in PPI networks and applications

2012-35
Evert Haasdijk (VU)
Never Too Old To Learn -- On-line Evolution of Controllers in Swarm- and Modular Robotics

2012-36
Denis Ssebugwawo (RUN)
Analysis and Evaluation of Collaborative Modeling Processes

2012-37
Agnes Nakakawa (RUN)
A Collaboration Process for Enterprise Architecture Creation

2012-38
Selmar Smit (VU)
Parameter Tuning and Scientific Testing in Evolutionary Algorithms

2012-39
Hassan Fatemi (UT)
Risk-aware design of value and coordination networks

2012-40
Agus Gunawan (UvT)
Information Access for SMEs in Indonesia

2012-41
Sebastian Kelle (OU)
Game Design Patterns for Learning

2012-42
Dominique Verpoorten (OU)
Reflection Amplifiers in self-regulated Learning

2012-43
Withdrawn

2012-44
Anna Tordai (VU)
On Combining Alignment Techniques

2012-45
Benedikt Kratz (UvT)
A Model and Language for Business-aware Transactions

2012-46
Simon Carter (UVA)
Exploration and Exploitation of Multilingual Data for Statistical Machine Translation

2012-47
Manos Tsagkias (UVA)
Mining Social Media: Tracking Content and Predicting Behavior

2012-48
Jorn Bakker (TUE)
Handling Abrupt Changes in Evolving Time-series Data

2012-49
Michael Kaisers (UM)
Learning against Learning - Evolutionary dynamics of reinforcement learning algorithms in strategic interactions

2012-50
Steven van Kervel (TUD)
Ontologogy driven Enterprise Information Systems Engineering

2012-51
Jeroen de Jong (TUD)
Heuristics in Dynamic Scheduling; a practical framework with a case study in elevator dispatching

2013-01
Viorel Milea (EUR)
News Analytics for Financial Decision Support

2013-02
Erietta Liarou (CWI)
MonetDB/DataCell: Leveraging the Column-store Database Technology for Efficient and Scalable Stream Processing

2013-03
Szymon Klarman (VU)
Reasoning with Contexts in Description Logics

2013-04
Chetan Yadati (TUD)
Coordinating autonomous planning and scheduling

2013-05
Dulce Pumareja (UT)
Groupware Requirements Evolutions Patterns

2013-06
Romulo Goncalves (CWI)
The Data Cyclotron: Juggling Data and Queries for a Data Warehouse Audience

2013-07
Giel van Lankveld (UvT)
Quantifying Individual Player Differences

2013-08
Robbert-Jan Merk (VU)
Making enemies: cognitive modeling for opponent agents in fighter pilot simulators

2013-09
Fabio Gori (RUN)
Metagenomic Data Analysis: Computational Methods and Applications

2013-10
Jeewanie Jayasinghe Arachchige(UvT)
A Unified Modeling Framework for Service Design.

2013-11
Evangelos Pournaras(TUD)
Multi-level Reconfigurable Self-organization in Overlay Services

2013-12
Marian Razavian(VU)
Knowledge-driven Migration to Services

2013-13
Mohammad Safiri(UT)
Service Tailoring: User-centric creation of integrated IT-based homecare ser- vices to support independent living of elderly

2013-14
Jafar Tanha (UVA)
Ensemble Approaches to Semi-Supervised Learning

2013-15
Daniel Hennes (UM)
Multiagent Learning - Dynamic Games and Applications

2013-16
Eric Kok (UU)
Exploring the practical benefits of argumentation in multi-agent deliberation

2013-17
Koen Kok (VU)
The PowerMatcher: Smart Coordination for the Smart Electricity Grid

2013-18
Jeroen Janssens (UvT)
Outlier Selection and One-Class Classification

2013-19
Renze Steenhuizen (TUD)
Coordinated Multi-Agent Planning and Scheduling

2013-20
Katja Hofmann (UvA)
Fast and Reliable Online Learning to Rank for Information Retrieval

2013-21
Sander Wubben (UvT)
Text-to-text generation by monolingual machine translation

2013-22
Tom Claassen (RUN)
Causal Discovery and Logic

2013-23
Patricio de Alencar Silva(UvT)
Value Activity Monitoring

2013-24
Haitham Bou Ammar (UM)
Automated Transfer in Reinforcement Learning

2013-25
Agnieszka Anna Latoszek-Berendsen (UM)
Intention-based Decision Support. A new way of representing and implementing clinical guidelines in a Decision Support System

2013-26
Alireza Zarghami (UT)
Architectural Support for Dynamic Homecare Service Provisioning

2013-27
Mohammad Huq (UT)
Inference-based Framework Managing Data Provenance

2013-28
Frans van der Sluis (UT)
When Complexity becomes Interesting: An Inquiry into the Information eXperience

2013-29
Iwan de Kok (UT)
Listening Heads

2013-30
Joyce Nakatumba (TUE)
Resource-Aware Business Process Management: Analysis and Support

2013-31
Dinh Khoa Nguyen (UvT)
Blueprint Model and Language for Engineering Cloud Applications

2013-32
Kamakshi Rajagopal (OUN)
Networking For Learning; The role of Networking in a Lifelong Learner's Professional Development

2013-33
Qi Gao (TUD)
User Modeling and Personalization in the Microblogging Sphere

2013-34
Kien Tjin-Kam-Jet (UT)
Distributed Deep Web Search

2013-35
Abdallah El Ali (UvA)
Minimal Mobile Human Computer Interaction

2013-36
Than Lam Hoang (TUe)
Pattern Mining in Data Streams

2013-37
Dirk Boerner (OUN)
Ambient Learning Displays

2013-38
Eelco den Heijer (VU)
Autonomous Evolutionary Art

2013-39
Joop de Jong (TUD)
A Method for Enterprise Ontology based Design of Enterprise Information Systems

2013-40
Pim Nijssen (UM)
Monte-Carlo Tree Search for Multi-Player Games

2013-41
Jochem Liem (UVA)
Supporting the Conceptual Modelling of Dynamic Systems: A Knowledge Engineering Perspective on Qualitative Reasoning

2013-42
Léon Planken (TUD)
Algorithms for Simple Temporal Reasoning

2013-43
Marc Bron (UVA)
Exploration and Contextualization through Interaction and Concepts

2014-01
Nicola Barile (UU)
Studies in Learning Monotone Models from Data

2014-02
Fiona Tuliyano (RUN)
Combining System Dynamics with a Domain Modeling Method

2014-03
Sergio Raul Duarte Torres (UT)
Information Retrieval for Children: Search Behavior and Solutions

2014-04
Hanna Jochmann-Mannak (UT)
Websites for children: search strategies and interface design - Three studies on children's search performance and evaluation

2014-05
Jurriaan van Reijsen (UU)
Knowledge Perspectives on Advancing Dynamic Capability

2014-06
Damian Tamburri (VU)
Supporting Networked Software Development

2014-07
Arya Adriansyah (TUE)
Aligning Observed and Modeled Behavior

2014-08
Samur Araujo (TUD)
Data Integration over Distributed and Heterogeneous Data Endpoints

2014-09
Philip Jackson (UvT)
Toward Human-Level Artificial Intelligence: Representation and Computation of Meaning in Natural Language

2014-10
Ivan Salvador Razo Zapata (VU)
Service Value Networks

2014-11
Janneke van der Zwaan (TUD)
An Empathic Virtual Buddy for Social Support

2014-12
Willem van Willigen (VU)
Look Ma, No Hands: Aspects of Autonomous Vehicle Control

2014-13
Arlette van Wissen (VU)
Agent-Based Support for Behavior Change: Models and Applications in Health and Safety Domains

2014-14
Yangyang Shi (TUD)
Language Models With Meta-information

2014-15
Natalya Mogles (VU)
Agent-Based Analysis and Support of Human Functioning in Complex Socio-Technical Systems: Applications in Safety and Healthcare

2014-16
Krystyna Milian (VU)
Supporting trial recruitment and design by automatically interpreting eligibility criteria

2014-17
Kathrin Dentler (VU)
Computing healthcare quality indicators automatically: Secondary Use of Patient Data and Semantic Interoperability

2014-18
Mattijs Ghijsen (VU)
Methods and Models for the Design and Study of Dynamic Agent Organizations

2014-19
Vincius Ramos (TUE)
Adaptive Hypermedia Courses: Qualitative and Quantitative Evaluation and Tool Support

2014-20
Mena Habib (UT)
Named Entity Extraction and Disambiguation for Informal Text: The Missing Link

2014-21
Kassidy Clark (TUD)
Negotiation and Monitoring in Open Environments

2014-22
Marieke Peeters (UU)
Personalized Educational Games - Developing agent-supported scenario-based training

2014-23
Eleftherios Sidirourgos (UvA/CWI)
Space Efficient Indexes for the Big Data Era

2014-24
Davide Ceolin (VU)
Trusting Semi-structured Web Data

2014-25
Martijn Lappenschaar (RUN)
New network models for the analysis of disease interaction

2014-26
Tim Baarslag (TUD)
What to Bid and When to Stop

2014-27
Rui Jorge Almeida (EUR)
Conditional Density Models Integrating Fuzzy and Probabilistic Representations of Uncertainty

2014-28
Anna Chmielowiec (VU)
Decentralized k-Clique Matching

2014-29
Jaap Kabbedijk (UU)
Variability in Multi-Tenant Enterprise Software

2014-30
Peter de Cock (UvT)
Anticipating Criminal Behaviour

2014-31
Leo van Moergestel (UU)
Agent Technology in Agile Multiparallel Manufacturing and Product Support

2014-32
Naser Ayat (UvA)
On Entity Resolution in Probabilistic Data

2014-33
Tesfa Tegegne (RUN)
Service Discovery in eHealth

2014-34
Christina Manteli (VU)
The Effect of Governance in Global Software Development: Analyzing Transactive Memory Systems.

2014-35
Joost van Ooijen (UU)
Cognitive Agents in Virtual Worlds: A Middleware Design Approach

2014-36
Joos Buijs (TUE)
Flexible Evolutionary Algorithms for Mining Structured Process Models

2014-37
Maral Dadvar (UT)
Experts and Machines United Against Cyberbullying

2014-38
Danny Plass-Oude Bos (UT)
Making brain-computer interfaces better: improving usability through post-processing.

2014-39
Jasmina Maric (UvT)
Web Communities, Immigration, and Social Capital

2014-40
Walter Omona (RUN)
A Framework for Knowledge Management Using ICT in Higher Education

2014-41
Frederic Hogenboom (EUR)
Automated Detection of Financial Events in News Text

2014-42
Carsten Eijckhof (CWI/TUD)
Contextual Multidimensional Relevance Models

2014-43
Kevin Vlaanderen (UU)
Supporting Process Improvement using Method Increments

2014-44
Paulien Meesters (UvT)
Intelligent Blauw: Intelligence-gestuurde politiezorg in gebiedsgebonden eenheden.

2014-45
Birgit Schmitz (OUN)
Mobile Games for Learning: A Pattern-Based Approach

2014-46
Ke Tao (TUD)
Social Web Data Analytics: Relevance, Redundancy, Diversity

2014-47
Shangsong Liang (UVA)
Fusion and Diversification in Information Retrieval

2015-01
Niels Netten (UvA)
Machine Learning for Relevance of Information in Crisis Response

2015-02
Faiza Bukhsh (UvT)
Smart auditing: Innovative Compliance Checking in Customs Controls

2015-03
Twan van Laarhoven (RUN)
Machine learning for network data

2015-04
Howard Spoelstra (OU)
Collaborations in Open Learning Environments

2015-05
Christoph Busch (UT)
Cryptographically Enforced Search Pattern Hiding

2015-06
Farideh Heidari (TUD)
Business Process Quality Computation - Computing Non-Functional Requirements to Improve Business Processes

2015-07
Maria-Hendrike Peetz (UvA)
Time-Aware Online Reputation Analysis

2015-08
Jie Jiang (TUD)
Organizational Compliance: An agent-based model for designing and evaluating organizational interactions

2015-09
Randy Klaassen(UT)
HCI Perspectives on Behavior Change Support Systems

2015-10
Henry Hermans (OU)
OpenU: design of an integrated system to support lifelong learning

2015-11
Yongming Luo(TUE)
Designing algorithms for big graph datasets: A study of computing bisimulation and joins

Open Universiteit