OPEN LEARNER MODELS FOR SELF-REGULATED LEARNING: EXPLORING THE EFFECTS OF SOCIAL COMPARISON AND GRANULARITY

by

Julio Daniel Guerra Hollstein

Bachelor in Computer Sciences, Universidad Austral de Chile

Submitted to the Graduate Faculty of

the School of Information Sciences, University of Pittsburgh in

partial fulfillment

of the requirements for the degree of

Doctor of Philosophy

University of Pittsburgh

2018

UNIVERSITY OF PITTSBURGH

SCHOOL OF INFORMATION SCIENCES

This dissertation was presented

by

Julio Daniel Guerra Hollstein

It was defended on

October 24th 2017

and approved by

Dr. Peter Brusilovsky, School of Computing and Information, University of Pittsburgh

Dr. Rosta Farzan, School of Computing and Information, University of Pittsburgh

Dr. Yu-Ru Lin, School of Computing and Information, University of Pittsburgh

Dr. Christian Schunn, Learning Research and Development Center, University of Pittsburgh

Dissertation Director: Dr. Peter Brusilovsky, School of Computing and Information, University of Pittsburgh

OPEN LEARNER MODELS FOR SELF-REGULATED LEARNING: EXPLORING THE EFFECTS OF SOCIAL COMPARISON AND GRANULARITY

Julio Daniel Guerra Hollstein, PhD

University of Pittsburgh, 2018

Open Learner Models (OLM) show the learner the internal model that the computer-based adaptive or tutoring system maintains. In the context of Self-Regulated Learning, where the learner is able to make decisions about what to learn and how to learn, OLM bring a wide variety of supporting features, ranging from metacognitive support, to navigational support, to engagement with the learning content. In prior work using OLM which featured social comparison features (OSLM), I have discovered interesting effects from these systems, regarding engagement with the system, encompassing considerable variations across different studies.

My thesis deepens the understanding of OLM and OSLM by a series of studies in which I evaluate different versions of Mastery Grids, incorporating features that were designed to match different motivational profiles, which are grounded in theories of Self-Regulated Learning and Learning Motivation. A large classroom study with more than 300 active students was conducted to deepen the exploration of the social comparison features in terms of engagement and navigation within the system. The results of this study confirmed the positive effects of the social comparison features and also brought insights into why certain students are influenced, based on their motivational orientations and prior-knowledge. A second large classroom study expanded the exploration by deploying the Rich-OLM, an extension of Mastery Grids featuring coarse- and fine-grained information about the learner model, which was designed to help students navigate the content contained in the system.

Results showed that students exposed to the fine-grained components took comparatively less time navigating the interface with higher rates of attempting content that they had opened. Results also raised concerns about increasing the complexity of the interface by integrating fine-grained visualization and social comparison features.

My work contributes to the understanding of the effects of Open Learner Models and additional features that provide social comparison and detailed information. It also contributes bringing learning motivation aspects into the understanding of Open Learner Models. Learning motivation in the context of self-regulated learning, provides a valuable theoretical basis to study how different students react and use learning tools.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

PREFACE

Completing my PhD has been a wonderful experience for me because of all the amazing people I have gotten to know, work with and love. Many colleagues and friends have accompanied and helped me in this quest, not merely by collaborating with me on these studies, completing development and research tasks and discussions, giving academic writing and presentation feedback; but also with their soul, encouragement, and smiles. I first want to thank all of these people, especially to my finacée, Yun Huang, the shinning light in all my endeavours.

I feel especially thankful for the support, encouragement and help of my advisor, Peter Brusilovsky, and the priceless role of all those on my committee: Rosta Farzan, YuRu Lin, and Chris Schunn. Their guidance has been of extreme importance not just to complete this thesis, but also in the whole learning process of becoming a PhD.

Last but not least, I am grateful to my family –my daughter, parents, brothers, and sister– for their unconditional support during all of the 5 years that I have lived so far away.

1.0 INTRODUCTION

1.1 MOTIVATION

Open Learner Models (OLM), also called Open Student Models (OSM), are learning tools that present the usually hidden internal model built by the adaptive or tutoring system to the learner. In the context of Self-Regulated Learning [Zimmerman, 1990b], where the student is able to make decisions about what to learn and how to learn, Open Learner Models bring a wide variety of supporting features, which are beautifully summarized by [Bull and Kay, 2010]: *"improving learner model accuracy by allowing the learner to make contributions to their learner model; promoting learner reflection through confronting students with representations of their understanding; facilitating planning and / or monitoring of learning; facilitating collaboration amongst learners; facilitating competition amongst learners; supporting navigation; the right of access to information stored about oneself; learner control over and responsibility for their learning; trust in the learner model content; formative assessment; summative assessment."*

A wide variety of OLM exists and different OLMs have been used in learning contexts, from simple *skillometers* [Duan et al., 2010], to more complex representations such as concept maps [Maries and Kumar, 2008]. Figure 1 shows examples of these types of OLM. Variations of OLM explore other features, too, such as editable, persuadable, and negotiable OLM [Mabbott and Bull, 2006], where the learner has an active role in providing different levels of direct feedback to her learner model; or OLM where the models are shared with others peers, teachers, and even parents. Our own OLM *Mastery Grids* [Loboda et al., 2014] exploits this *social* feature by showing to the learner an aggregated OLM of the rest of the class. This is what we have called the *Open Social Student Model* (OSLM). Mastery Grids also includes

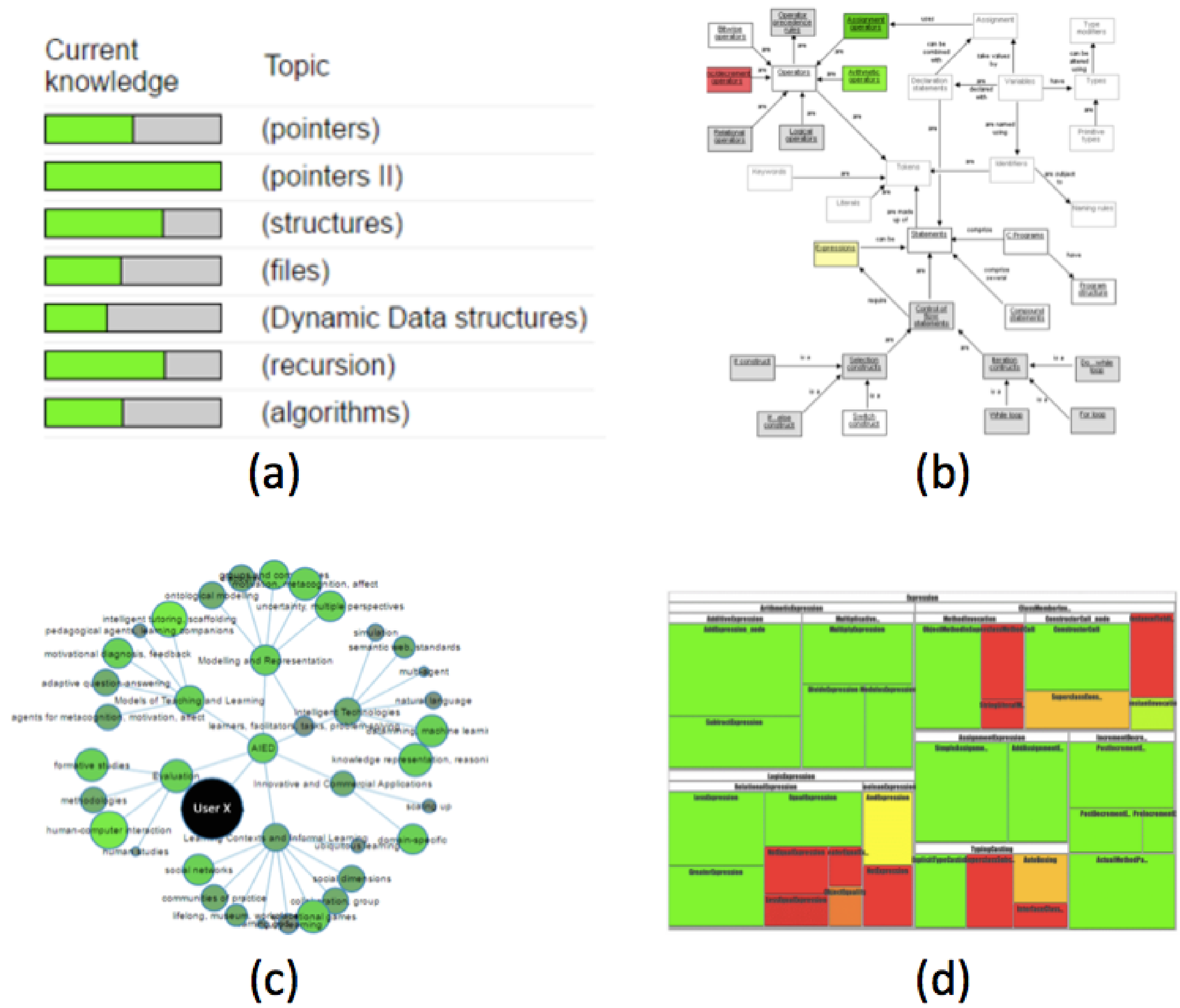


Figure 1: Examples of Open Learner Models of different complexity and showing different types of structural information. (a) Skillometers [Bull and Mabbott, 2006], (b) Concept Map [Mabbott and Bull, 2004], (c) Network [Bull et al., 2015], (d) Treemap [Brusilovsky et al., 2013].

another important feature: the tool is used to navigate through the learning content and is not just a static, visual representation of the learner model. Mastery Grids is explained in detail in chapter 3.

Through a series of classroom studies in which we have evaluated the effects of OLM and the social comparison features, we have seen benefits in how students engage with the system. We have found that while the social comparison features generally explain an increase in activity within the system (i.e., more learning content is completed), this positive effect varies considerably across studies. Also, the social comparison features have been shown to interact with multiple factors, such as gender [Brusilovsky et al., 2016] and motivation [Guerra et al., 2016]. However, past studies don't provide sufficient analyses, nor enough statistical power to support a solid explanation of the positive engagement phenomenon. More research that looks deeply at the effects of OLM, and particularly OSLM, is needed.

A deeper exploration of OLM is also needed to address another issue. While past research studies have evaluated different OLM, ranging from coarse-grained representations to more complex and structured visualizations, no research has looked at the combined or contrasted effects of different levels of granularity, i.e., different levels of detail being shown. I think that this exploration is necessary because the level of granularity of the information shown conveys a trade-off: a complex OLM may provide wider support, for example, by helping the learner navigate the system; but also, the accompanying higher level of details may result in an interface that is too complex to understand and use, which may overwhelm the learner.

My work is motivated by the need to better understand the effects of OLM. This understanding is important because it can guide the development of better tools, better personalization and adaptive mechanisms, and better use of such tools in supporting the learning experience. Specifically, I focused my thesis in two ways that extend prior OLM studies. First, I aim to dig deeper into the effects of OLM and the social comparison features by extending analyses incorporating factors that, from a theoretical perspective, could explain these effects. Secondly, I aim to explore the granularity feature by combining coarse-grained and fine-grained representations, and study the effects of these features in supporting learner navigation within the system. These two aspects of my work are explained in detail in the next sections.

1.2 EXPLORING OLM AND OSLM WITH INDIVIDUAL DIFFERENCES

In the context of Self-Regulated Learning (SRL), where the learner makes active decisions about her own learning process, the level of engagement and interaction with the learning system is strongly dependent on the students themselves. Because of this, understanding the effects of OLM better requires looking at factors that differentiate learners, particularly individual differences that may influence self-regulation. Individual differences range over a plethora of cognitive, personality, and demographic factors. To narrow down this space, I turned to the literature of Self-Regulated Learning (SRL) and learning theories, placing special attention on the factors that are theoretically related to exploration features, i.e., the social comparison features.

Learning Theories connect engagement in SRL with different aspects, including metacognitive skills [Bandura, 1986], learner beliefs [Dweck, 2000], and goal orientation [Wolters et al., 1996], among other things. In the definition of SRL, Zimmerman provides a common ground in terms of three aspects: metacognitive, motivational, and behavioral [Zimmerman, 1990a]. While the metacognitive dimension has been explored in the context of OLM [Bull and Kay, 2013], and the behavioral dimension is in the realm of observation (observation of engagement with SRL opportunities), the motivational aspect is a key individual difference that has not been studied in the context of Open Learner Models.

Another relevant individual factor is the level of the learner's knowledge or prior skills. For example, [Mitrovic and Martin, 2007] found that OLM produced significant positive differences between pretest and posttest only for "less able" students. Having prior knowledge may have a strong impact in deciding what to engage with in the learning content. The Expectancy-Value Theory of learning motivation [Wigfield and Eccles, 2000] helps to relate prior knowledge and engagement: the strength of learner skills influences the *subjective task values*, including the attained *cost* of performing the activity, and the *expectancy* in terms of potential benefits, thus influencing the decision of whether or not to do the activity.

These individual differences represent a key aspect to the contributions of my thesis, extending the exploration of OLM and OSLM interfaces and providing a foundation to guide the research work, the analyses and the interpretation of results.

1.3 GRANULARITY: EXTENDING NAVIGATIONAL SUPPORT

Mastery Grids provides topic-based navigational support that was designed based on past experience in coarse-grained navigational interfaces in learning systems. The value of this approach is demonstrated by its ability to guide students to the most appropriate topics, improve learning outcomes, and increase their engagement [Sosnovsky and Brusilovsky, 2015]. However, topic-level visualization has limitations. Topics aggregate information, hiding the learner's progress knowledge of more detailed components of the learner model, such as specific concepts. The learner may not be aware of knowledge "holes" in topics in which she could have a high overall progress. Also, coarse-grained visualization (topic-based) does not provide enough details to help the learner to choose activities within a topic, failing to provide useful content navigation support. I think that including detailed information in the representation of the learner model (LM) could substantially improve the navigational support of the system. The reasoning behind this belief has roots in the foundations of the information science and information visualization fields. On the one hand, detailed information can allow students to make decisions about what content to target by providing traces that improve support for *information foraging* [Pirolli and Card, 1999]. On the other hand, the learner model might make more sense if it is shown by means of *external anchoring* that detailed LM represent when it is visualized [Liu and Stasko, 2010]. These ideas can be summarized by stating that detailed information in the open learner model could improve the usefulness and the experience in the system by helping the learner to find useful content and to make sense of the information of her own learner model.

However, while detailed information may provide support for better guidance and self-reflection, it also increases the complexity of the tool and thus the cognitive effort necessary to understand and make sense of the detailed information shown. The problem is that this additional complexity could diminish the interest and comfort of the learners. For example, [Duan et al., 2010] found that simple indicators like *skillometers* are preferred by students over more elaborate and detailed representations of the Learner Model. The information overload in abstract visual representations is a foundational problem addressed by the Information Visualization field. In this regard, Shneiderman proposed a framework to address this problem, depicted masterfully in his famous information-seeking mantra: *Overview first, zoom and filter, then details-on-demand* [Shneiderman, 1996]. While in Mastery Grids OLM, the coarse-grained topic-based visualization accounts for the *overview*, the fine-grained information that could account for the *details-on-demand* is missing.

Following these ideas, my work embraces the task of displaying levels of granularity within the OLM that can support better navigation through content, but at the same time does not overload the learner. To this end, my work includes interviewing students and performing controlled studies to guide the design of a fine-grained visualization that balances support and complexity.

1.4 OVERVIEW AND RESEARCH QUESTIONS

There are many aspects, factors and variables involved in my thesis work. On the one hand, the OLM interface includes the features of social comparison and granularity that are the focus of my evaluation. These features may produce several effects related to the use of this system: engagement, navigation, the support of metacognitive outcomes, and even (indirectly) learning. On the other hand, the effects of these features are studied in conjunction with factors, individual differences, that are likely to influence the results, such as motivation, orientation, and prior knowledge.

Figure 2 is an attempt to summarize all the interwoven elements in this thesis. Elements marked with the letters (A), (B), (C) represent relationships that will be explored in this work and which generate research questions. Dotted lines represent other relationships that are not explored in this thesis, such as how the metacognitive support associated with the OLM affects learning, or what are the effects of individual differences in the learning experiences that occur outside the use of the practice system. How practicing content activities affects learning (dotted line labeled as “practice” in the Figure 2) is not a central aspect of this work because it is outside of the focus of my thesis to evaluate the quality of the learning material. Moreover, there is another reason why I do not focus on its effect on learning: the goal of the Mastery Grids system is to complement formal coursework, and as such, this system is not the only and probably not the main source of learning material, since students learn from different sources. However, the learning effect is included in the analyses to confirm the general beneficial impact of the learning system. The relationships and effects investigated are described below.

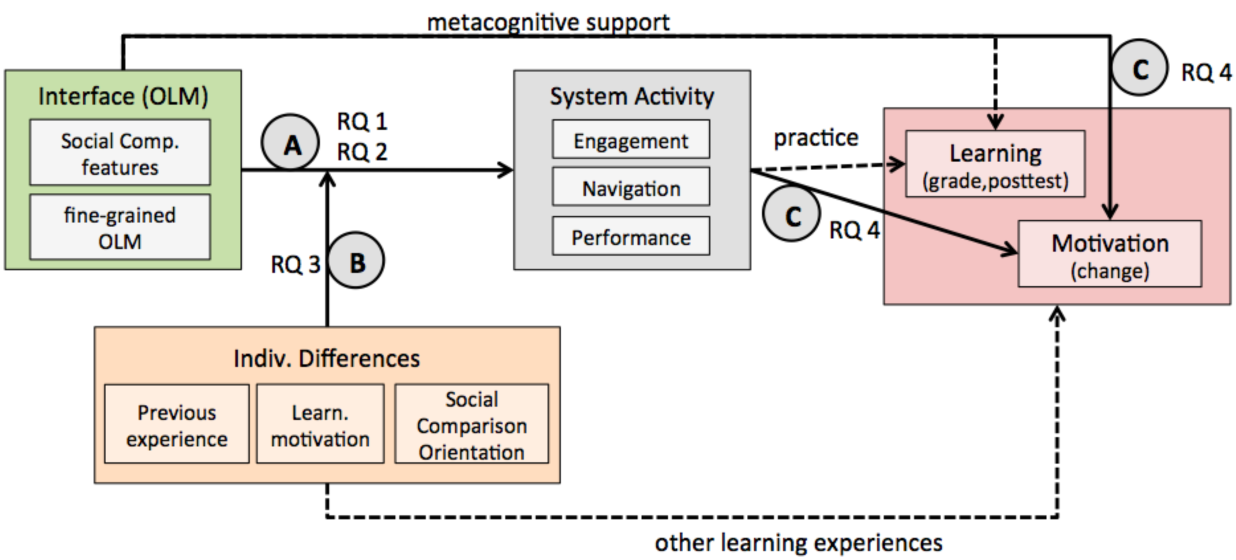


Figure 2: Research questions, factors and effects explored in this thesis.

(A) The effect of the interface on the use of the learning system

I am interested in looking at the effect that different features in the OLM have on how and how much the learner uses the system. I call the total combination of these the *system activity* and range:

- the *engagement* with the practice content, which refers to the amount of learning activities (examples, problems, etc) viewed and solved, and the pattern of engagement during the term (e.g., how regularly the system is used in the term);
- the *navigation* refers to how students navigate the system and that can be expressed in multiple indicators such as patterns of navigation in different types of content, rates

of times spent in navigating and viewing content, rates of content selected vs. content completed, etc.; and

- the *performance* in self-assessment content (questions, problems) such as success-rates.

The system activity variables of engagement, navigation and performance are introduced and described in Section 5.3, chapter 5.

Two research questions (RQ) are stated, according to the two different features explored of the OLM interfaces. The first RQ looks into the effects of using social comparison features in a OLM.

RQ 1 What are the effects of an OLM with social comparison features (or OSLM) compared to an individual-view OLM on *system activity*?

The second RQ looks into the effects of adding fine-grained information to the OLM:

RQ 2 What are the effects of including fine-grained OLM and OSLM on *system activity*?

(B) The role of Individual Differences

As explained before, in the context of Self-Regulated Learning, individual differences gain special relevance, when analyzing the effects of tools such as Open Learner Models. Individual differences range over a number of different factors that describe learners. In my work, I focus on relevant individual differences in the context of Self-Regulated Learning such as *previous knowledge* and *learning motivation*. The research question is stated as follows:

RQ 3 How do individual differences influence the effects of the OLM on system activity?

(C) The effects on Learning Motivation

Learning motivation literature states that motivational factors are not static and can change as a result of positive or negative learning experiences [Grant and Dweck, 2003], thus Learning Motivation is both an influencing factor and an outcome of the learning experience. As outcome, learning motivation may be influenced by all the factors that influence the learning experience: the individual differences, specifically the initial level of motivation; the exposure to the interface, which provides metacognitive support; the interaction with content material and all the other learning activities outside the use of this system. In this thesis, I focus on how system activity and different interfaces (OLM, OSLM, fine-grained components) affect motivation, thus research question 4 is located in two places in Figure 2.

RQ 4 How does the use of a learning system featuring OLM, OSLM and fine-grained elements affect motivation?

1.5 THESIS ORGANIZATION

This thesis is organized as follows. Next is chapter 2, which presents related work from seminal and varied research on Open Learner Models, as well as work that relates OLM to theories of Self-Regulated Learning and Learning Motivation. The goal is to help narrow the focus of OLM and provide grounds for studying OLM from the perspective of Self-Regulated Learning.

Chapter 3 presents Mastery Grids, our Open Learner Model and multi-type smart content architecture. A general view of the system, its visual interface, and the learning material that is contained in it is followed by a description of its technical architecture that allows independent smart content to be integrated on two levels: as accessed through the Mastery Grids, and when monitored in terms of knowledge progress.

Chapter 4 presents previous studies we have conducted using Mastery Grids and our findings on the effects that the Open Learner Models and the Open Social Learner Model produce in terms of engagement, navigation, and performance with the practice system. Three studies, previously published in relevant conferences, are summarized. Findings of these prior studies serve as a starting point to draw hypotheses and extend an exploration of the system.

Chapter 5 refines the research framework of this thesis, complementing research questions with expectations that connect prior research with theoretical foundations. This chapter also offers an overview of the studies depicted in the following chapters, along with a description and grounding of the individual difference measures and variables that are used across all the studies, including log variables, performance tests, learning motivation questionnaires, and social comparison orientation questionnaires. I have set this information aside to avoid repeating common details in later chapters.

Chapter 6 presents a semester-long large classroom study with a between-subjects design in which Mastery Grids is provided in two versions, with and without social comparison features. This study was performed in a large python programming course and focuses on exploring the *social* dimension of a coarse-grained OLM. The large size of the study allows it to include individual differences in the analyses, with reasonable statistical power.

Chapters 7 and 8 present the design, construction, and initial evaluation of the Rich-OLM through two controlled user studies. The Rich-OLM is an extension of Mastery Grids which includes both coarse-grained and fine-grained visual and interactive features. The process followed for designing this Rich-OLM is contained in chapter 7. It includes interviewing students and then performing a laboratory study, in order to choose a visual representation of the fine-grained OLM. Chapter 8 describes a second controlled laboratory study, in which three variations of the Rich-OLM are compared.

In chapter 9, I present another semester-long classroom study performed in a large python programming course. This study compares three variants of the Rich-OLM, thus exploring the fine-grained OLM and contributing to answer the research questions.

Chapter 10 offers a set of analyses across the studies of chapter 6 and chapter 9. Although the two studies presented in chapter 6 and 9, respectively, are similar, there were changes in the course content, grading process, and deployment of the study. These differences prevent me from doing a straightforward cross-studies analysis. However, some relative comparisons still accomplished the goal of complementing my previous observations and findings, related to the research questions.

Finally, chapter 11 summarizes conclusions, discussions, and limitations of the work.

2.0 BACKGROUND AND RELATED WORK

2.1 OPEN LEARNER MODELS

In traditional adaptive and personalized computed-based learning environments a User Model captures individual aspects, preferences, and learning progress of the student, allowing the system to perform adaptation and personalization tasks [Brusilovsky and Millán, 2007]. Open Learner Models (OLM), also called Open Student Models, provide the learner with some sort of representation of this internal model aiming at promoting reflection and encouraging self-regulated processes. According to [Bull and Kay, 2007], OLM can support a variety of aspects: "improving learner model accuracy by allowing the learner to make contributions to their learner model; promoting learner reflection through confronting students with representations of their understanding; facilitating planning and / or monitoring of learning; facilitating collaboration amongst learners; facilitating competition amongst learners; supporting navigation; the right of access to information stored about oneself; learner control over and responsibility for their learning; trust in the learner model content; formative assessment; summative assessment."

Different types of OLM have been explored, and [Bull and Kay, 2010] offered a review. The most common OLM is related to the representation of knowledge or learning progress of the learner. Researchers have explored different representations ranging from overall knowledge *skillmeters* (also called *skillometers*) [Mitrovic and Martin, 2007], to detailed knowledge elements [Kay and Lum, 2005], and structured representations such as *treemaps* [Brusilovsky et al., 2011] and *concept-maps* [Rueda et al., 2003, Pérez-Marín et al., 2007, Kumar and Maries, 2007]. Different representations of OLM are shown in the previous chapter in Figure 1. Different representations of the OLM at different levels of complexity, or using different visualization approaches may serve different purposes, such as providing an overview of progress, showing conceptual relationships, or highlighting misconceptions that the learner may have [Bull and Kay, 2016].

Presenting visualizations of the learning related information is an idea that is not exclusive of the OLM area. Learning analytics has gained attention in recent years [Verbert et al., 2014]. While learning analytics exploits diverse data of learning records, benefits from the big data phenomenon, and focuses extensively on providing the learning data to the institution [Siemens and Long, 2011], Open Learner Models use information that is generated by an *intelligent* system capable of making inference of the learner competencies [Bull et al., 2015].However, the distinction between OLM and Learning Analytics is blurry. In fact, as expressed by Bull and Kay [Bull et al., 2015], "Open learner model visualisations could be seen as a specific type of learning analytics, in that the visualisation is of the learner model." There is no doubt that these two areas present opportunities for synergy. For example, OLM could provide knowledge estimations to feed learning analytics dashboards, and OLM representation could be improved using visualization techniques and approaches explored in the learning analytics field.

An important concern related with OLM is how to represent the information of the User Model, which in some cases can be fairly complex, in an understandable manner [Bull, 2012, Law et al., 2015]. While some studies have found that simple indicators like *skillometers* are preferred by students [Duan et al., 2010], other studies have found support for more complex representations such as concept-maps [Maries and Kumar, 2008]. Moreover, it has been proposed to offer multiple OLM views, from simple to detailed to structured views, giving options satisfying different students' preferences [Bull et al., 2010, Duan et al., 2010, Conejo et al., 2011]. For example, Flexi-OLM offers the learner visualizations of prerequisite-based concept-maps, hierarchical representation of concept details, and hierarchical representation of the course organization, among others [Mabbott and Bull, 2006]. Our previous work on a questionnaire study of a wide variety of visualizations from different systems found that students expected structured visualizations such as Prerequisites and Hierarchical Tree (from [Mabbott and Bull, 2006]) to best support the task of identifying what to work on next [Bull et al., 2016]. However, it was unclear why students might prefer these representations over other structured views such as concept maps. Other work has taken the issue of complexity by extending the OLM with more elaborate features such as indicators of effort, progress or working style which offer pre-digested interpretative meaning [Papanikolaou, 2015]. Since OLMs show information that is based on estimations made by the system (knowledge), it necessarily conveys levels of uncertainty, which can be addressed using techniques borrowed from the information visualization field [Epp and Bull, 2015].

Beyond the role of visualizing the learner model, OLM can also incorporate different levels of interactivity. One approach is related to make the OLM into a navigational tool [Papanikolaou et al., 2003, Long and Aleven, 2013a, Hsiao et al., 2013], which is closely related to the area of adaptive navigation support [Brusilovsky, 2003], because the user model is used to generate indicators that are included in the interface to support guidance [Brusilovsky et al., 2004b, Hsiao et al., 2010]. A different approach of interactivity deals with OLM that is editable by the learner [Kerly and Bull, 2008, Mabbott and Bull, 2006]. Moreover, some systems implement OLM that are entirely constructed by the learner [Mabbott et al., 2007, Cimolino et al., 2004].

Open Learner Models can also be *opened* to others. For example, OLM can show peer models to the learner, or letting the teacher inspect the models of the students [Bull and McKay, 2004, Rueda et al., 2003, Pérez-Marín et al., 2007]. The review of [Bull and Kay, 2010] distinguished different approaches that incorporated this *social* dimension into the OLM. There is work inclined to construct group models, where group interactions are visualized to support collaboration and assessment of the collaborative work [Kay et al., 2006, Upton and Kay, 2009, Bull and Vatrapu, 2011].

Other approaches have explored awareness, social navigational support, and social comparison effects as a result of showing the models of other learners individually or aggregated [Brusilovsky et al., 2004a, Linton and Schaefer, 2000, Shi et al., 2014, Hsiao et al., 2013, Hsiao and Brusilovsky, 2012, Brusilovsky et al., 2015], which has been called Open Social Learner Model (OSLM), or Open Social Student Model (OSSM). The idea behind this approach to OSLM is that exposing the model of others will produce a competitive effect that has shown positive effects in encouraging participation in online communities [Vassileva and Sun, 2007], or stimulating activities in learning environments [Burguillo, 2010]. Consistently, OSLMs have demonstrated that they could boost system engagement and affect navigational patterns. For example, the work of [Hsiao et al., 2013, Hsiao and Brusilovsky, 2012, Falakmasir et al., 2012] in different studies consistently found that by showing the models of other learners, students covered more topics in the system, reached higher success rates in self-assessment problems, and that *strong* students went ahead in the course topics guiding *weak* students who followed later. Our later work confirmed these findings and added other components to the analyses, revealing different effects. For example, in [Brusilovsky et al., 2015], we showed how the experimental group, which was exposed to social comparison visualizations, presented higher rates of system usage, higher learning effectiveness ([Paas and Van Merriënboer, 1993]), and interaction effects of gender. Recent work has also shown how the social comparison features accounted for better completion rates in MOOCs [Davis et al., 2017]. Other recent work has also investigated OSLM from the broader perspective of Learning Analytics. For example, the recent work of Shi and Cristea [Shi and Cristea, 2016] incorporated visual indicators of different learning related information such as learning paths and learner contributions, into a multifaceted OSLM.

While we have repeatedly demonstrated positive uses of OSLM in classroom studies, our past work explored a relatively simple visualization of the learner progress using a coarse-grained representation based on topics. My thesis focuses on taking this exploration further, and studying the effects of OSLM combining coarse-grained and fine-grained representations.

In general OLMs are evaluated in terms of engagement, guidance, metacognition, and satisfaction, i.e., the extent to which an OLM engages students to use the learning system [Brusilovsky and Sosnovsky, 2005, Brusilovsky et al., 2015, Hsiao and Brusilovsky, 2012], guides students to better content [Brusilovsky et al., 2004b, Loboda et al., 2014, Hsiao et al., 2010, Mitrovic and Martin, 2007], facilitates awareness and reflection about knowledge [Bull and Kay, 2013, Lazarinis and Retalis, 2007, Bull et al., 2003, Dimitrova et al., 2001], and the extent to which learners find it useful [Mabbott and Bull, 2004, Mazzola and Mazza, 2010] or desirable [Bull, 2004]. The impact on learning outcomes or learning performance is limited or indirect, because an OLM is a tool that supports learning metacognition [Bull and Kay, 2013], but is not the content material or the tutoring tool itself. However, some researchers have encountered positive learning effects of using OLMs. For example, [Kumar and Maries, 2007] used a concept map representation and found evidence that students might learn concepts from OLMs which were not covered by the learning tutor. The work of [Mitrovic and Martin, 2002, Mitrovic and Martin, 2007] found that a simple representation of an OLM had a positive impact on weak students' performance measured by post-test. They also found that strong students (more-able students) showed higher self-assessment skills when using OLMs which translated into better selection of problems to work with.

Other more recent work has considered the evaluation of OLMs from the perspective of Self-Regulated Learning (SRL), and supported the claim that OLM can enhance SRL processes of self-assessment, planning and motivation [Long and Aleven, 2017, Law et al., 2017]. However, the incorporation of factors such as motivational traits is rare in the literature of digital learning systems. In this context, my thesis work contributes to the literature of OLM with the exploration of the role of Learning Motivation together with other factors such as prior knowledge.

2.2 SELF-REGULATED LEARNING AND LEARNING MOTIVATION

Self-Regulated Learning and Learning Motivation theories are relevant to my work because they offer theories and frameworks which serve the understanding of the learning experience phenomenon, particularly when the learner is exposed to learning opportunities that require self-regulation (e.g., when a learning system is offered in a non mandatory way.) Moreover, this background becomes more relevant if the OLM related tools are specially designed to support the self-regulation process.

Self-Regulated Learning (SRL) is a positive and desirable condition that defines a learner as an active participant who monitors and applies strategies to control her own learning process cognitively, meta-cognitively, and emotionally [Zimmerman, 1990b]. Zimmerman summarizes three dimensions in which SRL has been studied and considered: (i) the dual focus in self-regulation process and strategies targeting those processes, (ii) the key role of continuing feedback enabling SRL to happen, and (iii) the interdependence between motivation and self-regulating processes [Zimmerman, 1990b]. SRL and motivation are interdependently related. For example, the social cognitive view of SRL focuses on self-efficacy, a measure of self-regulation, which is considered to be an important force behind motivation [Bandura, 1986]. Other authors have confirmed the positive relation between self-efficacy and learning performance [Zimmerman, 1990b] and demonstrated its relations to other motivational elements such as goal-setting [Schunk, 1990].

Learning motivation is framed from different perspectives and with different emphases, and includes general intrinsic motivation such as “fascination” [Moore et al., 2011], self-beliefs [Dweck, 2000], self-efficacy [Zimmerman, 2000], competency-beliefs [Moore et al., 2011], values, and goal-orientation [Elliott and Dweck, 1988], among others. Theoretical frameworks have been put forward to articulate these factors and relate them to learning performance, or more generally, to the learning experience. One such framework is the Achievement-Goal Orientation Framework [Elliot and Murayama, 2008]. This framework proposes that the factors that influence motivation, e.g., beliefs, values, intrinsic motivation, induce the learner to embrace different goal orientations when facing a learning activity. The goal orientation could be defined as *Mastery* goal orientation or *Performance* goal orientation, and has a “valence” that could be *approach* or *avoidance*. Then accordingly, four different motivation orientation exist: Mastery-Approach oriented students pursue learning, while Performance-Approach oriented students want to demonstrate mastery and they are usually more sensitive to comparison and scores. Mastery-Avoidance students avoid achieving less than what they think they can achieve, and Performance-Avoidance students avoid to perform worse than others or receive lowest scores [Elliot and Murayama, 2008]. Since the goal orientations “encapsulate” diverse motivational factors that internally explain them, the framework is especially relevant to my work: it allows me to focus on the effect of motivation at different OLM interfaces, rather than elaborate on the internal interplay of the motivational factors, which is out of the reach of my work. Moreover, the framework allows to make direct associations between the motivation orientations and the system that is the subject of study in this thesis. For example, the social comparison features of the system are expected to generate a competitive effect that will be stronger on students that are highly *Performance* oriented. In my own preliminary work I showed evidence supporting this: the level of engagement with Open Social Learner Model was positively correlated to changes in motivation factors such as Performance-Approach, while this correlation didn't seem to hold for students engaged in OLMs without social comparison features [Guerra et al., 2016].

Researchers have also studied the factors that can foster different achievement-goal orientations, suggesting that Mastery oriented environmental factors, such as an environment supporting autonomous work, can foster the adoption of Mastery goals [Ciani et al., 2010], while Performance oriented elements can account for the adoption of Performance goal [OḰeefe et al., 2013]. Research has also established relationships between the different goal orientations. For example, although Mastery orientation and Performance orientation seem to represent opposite values, they can coexist [Ames, 1992]. A student can present high levels of performance and mastery orientation goals at the same time. These last elements of the achievement-goal orientation are important for my work because they support the idea that a learning system with performance and mastery oriented features could affect these motivational orientations of the students.

2.3 SOCIAL COMPARISON ORIENTATION

An important aspect explored in my work relates to providing the learner model to other learners, which is called Open Social Learner Model (OSLM). The main idea of OSLM is that the learner can compare her achievements to the achievements of other learners individually or in a group. Social comparison is a well-studied area in psychology. The core of this idea is that by being able to compare to others, a person may adopt different behaviors and set different thresholds for evaluating her opinions and abilities [Festinger, 1954], and that this effect is stronger when comparison is made with similar or known people [Cialdini et al., 1999]. The importance of social comparison in social sciences is considerable and [Buunk and Gibbons, 2007] states that social comparison "has developed from a focused theoretical statement on the use of others for self-evaluation into a lively, varied, and complex area of research encompassing many different paradigms, approaches, and applications." Researchers have used the ideas of social comparison in different areas including the virtual world. For example, social comparison has been applied successfully to increase participation in online communities [Harper et al., 2010]. The study of social comparison in educational settings is also important. While researchers have put social comparison in the center of the idea of Social Learning Environments [Vassileva, 2008], the effects of manipulating social comparison have shown to be beneficial in some settings [Huguet et al., 2001], and detrimental in others, for example, when students compete instead of cooperate [Buchs and Butera, 2009]. These contradictory findings raise interesting questions about how to effectively use social comparison in learning systems such as the one featuring OLM in this thesis.

Learning Motivation theories also connect to social comparison. For example, achievement-goal researchers explained that Performance oriented learners are prone to compare to others [Elliot and Murayama, 2008], and suggested that the positive or negative effect of performing upward or downward comparisons was mediated by the goal orientation of the learner [Grant and Dweck, 2003].

3.0 MASTERY GRIDS OSLM

The Mastery Grids system is an attempt to design an intelligent interface for accessing learning content that provides support for Self-Regulated Learning (SRL) and allows learners to monitor their course progress. At its core, it follows earlier work that integrated content navigation with OLM-based knowledge progress visualization [Hsiao et al., 2013]. To complement the benefits of OLM, Mastery Grids (MG) also engage the power of the Open Social Student Model by incorporating visualization based on the models of other students. The MG interface presented below adds several features to its first version presented in [Loboda et al., 2014].

A basic version of the interface is shown in Figure 3. The interface organizes the course as a sequence of cells representing the topics of the course, in this case of a Java programming course with topics such as *Variables*, *Primitive Data Types*, *Constants*, etc. Each topic cell can be clicked allowing the learner to access content pieces or *activities*. Each activity is also represented by a slightly smaller cell that can be clicked to display the content activity on the screen. When the learner completes an activity, its corresponding cell is painted green and contributes to darken the cell of its topic. In this way, darker topic cells mean the learner has more activities completed on that topic. Mastery Grids can be configured to use different colors to represent the progress. In the first reported work [Loboda et al., 2014] we used shades of purple and in a more recent work, and the work reported in this thesis, we used shades of green.

The interface has been designed to allow social comparison features which basic version is shown in Figure 4. The grid now has several rows. There are three rows in the main grid.

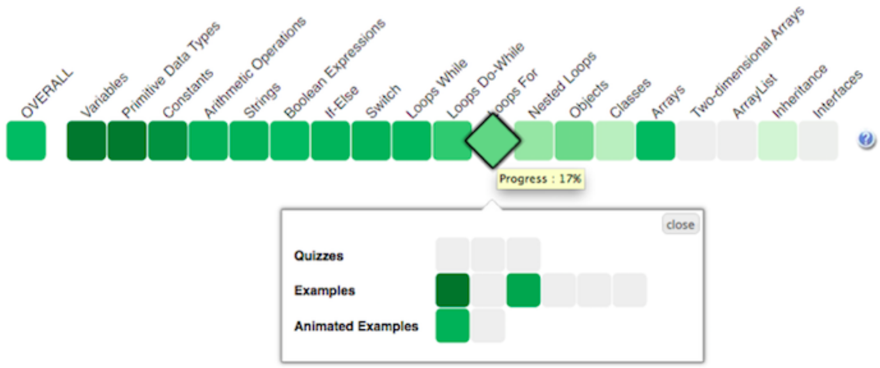


Figure 3: Basic Mastery Grids interface.

The first row (*Me*) represents the knowledge progress of the learner and is the same row of the basic interface shown before. The third row represents the progress of the rest of the class, which we label *Group* aggregating the progress of the other learners who have logged into the system at least once. In this row, we use different color shades, and we have chosen blue. The second row (*Me vs group*) shows a comparison between the learner and the group and its cells become green if the learner has a higher knowledge progress in the topic, blue if the group has higher progress, or remains gray if the learner and group have the same level of progress. The intensity of the color represents the intensity of the difference.

Below the main grid with three rows there is a grid with a set of thinner rows representing the progress of all peers individually and ordered top to bottom according to the level of progress (higher at the top). Here each peer is represented also with shades of blue and the learner with shades of green. Neither names, nor any identifier for the learner is shown, and only the row corresponding to the learner (in green) is labeled as "*Me* ->" and showing the position number of the learner in the ranking list. To speed up the interface loading, the ranked list of peers is only shown when clicking on the button "Load the rest of the learners," which is located below the 3rd grid and does not appear in Figure 4.

By clicking on any topic cell in the interface, the user can access the practice content of this topic, shown as activity cells organized in rows by content type (number 5 in Figure 6). By clicking on an activity cell, the content is loaded in an overlaid window. Figure 5 shows

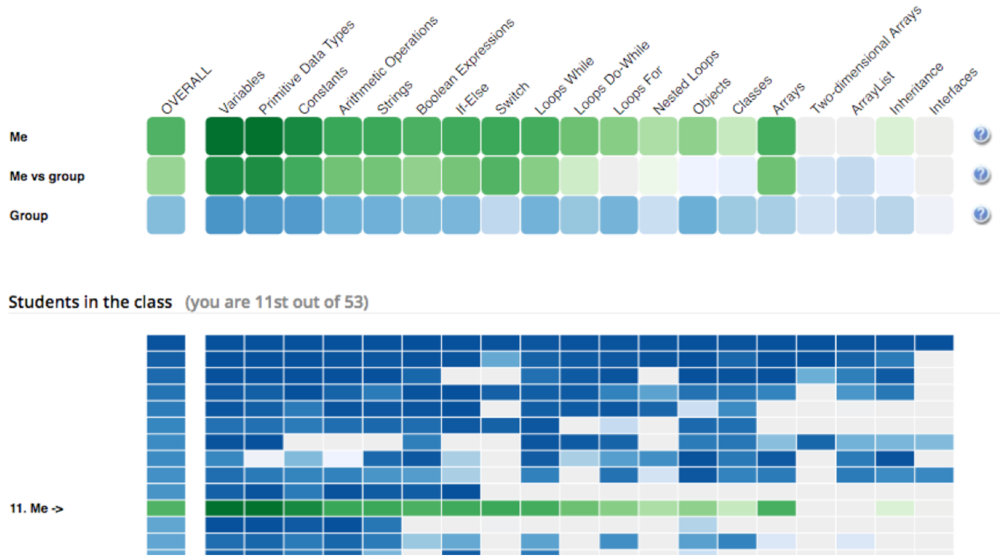


Figure 4: Mastery Grids with social comparison features.

the overlay window when the learner has opened an activity. The interface is shaded out to stress the overlay window.

Since the system can include access to activities of different type (Figure 3 show three types of content *Quizzes*, *Examples*, and *Animated Examples*), the interface can be configured to display more details and levels of aggregation. The full interface of the Mastery Grids system (Figure 6) follows the same idea than the simpler version of the interface, and now each row is "opened" into a grid showing different aggregations of progress information. This means that each row shown in Figure 4 is represented as a grid in the full interface shown in figure 6. In all grids, columns represent topics and rows represent different types of content (such as problems, examples, or animations) maintaining consistency with the simpler version of the interface. The first grid (1 in the Figure 6) shows an extended OLM that visualizes the learner's own progress over several kinds of content, and where the first row (*OVERALL*) represent the aggregated information. The third grid (number 2 in Figure 6) represents the average progress of the *reference group* using a varying density of blue color. A combo box in the menu bar allows the student to use the whole class, or just the top students, as a reference group (number 7 in Figure 6). Second grid (number 3 in Figure

Figure 5: Overlay window showing an activity opened in Mastery grids.

6) is the comparison grid showing the difference between the learner and the group. When the group has a higher progress than the learner on a specific kind of content in a specific topic, the corresponding cell in the second grid becomes blue. Otherwise, it becomes green.

In the bottom part (number 4 in Figure 6), a progress grid for each of the students in the group is shown (with the top progressing students shown first). As mentioned before, the list does not show the names of the students. To be consistent with the colors used in the first grid, each peer grid is represented in shades of blue and the learner is represented in shades of green, which also facilitates locating the learner in the list. Figure 6 shows the learner in the 3rd position of the list.

To allow the exploration of a broader design space, different interface components can be loaded with different combinations. A selector widget in the menu bar allows students to select among different progress visualizations for the different content types (number 8 in Figure 6). The user can choose a *full* view in which each grid has separate rows for each content type (as shown in Figure 6), and can also select to display averages by the type of content (for example, showing only progress in the examples), or an *overall* view where all the three first grids are collapsed in one grid with one row for the learner progress, one row for the comparison, and one row for the group progress, as shown in Figure 4. The *overall* view mode is set as the default view. In addition, all comparison features can also be completely hidden by clicking the button "Individual" (number 6 in Figure 6), which leaves only the personal part of the interface visible, as shown in Figure 3. The Mastery Grid interface can be configured to hide or show the menu controls (numbers 6, 7, and 8 in Figure 6), as well as to enable or disable the social comparison features for a specific group or for individual users. For example, this allows us to show social comparison features only to some students, or to enable all features for the instructor.

3.1 SMART CONTENT IN MASTERY GRIDS

Mastery Grids integrates different types of content activities which are online "smart" content from different content providers. "Smart" content [Brusilovsky et al., 2014b] interac-

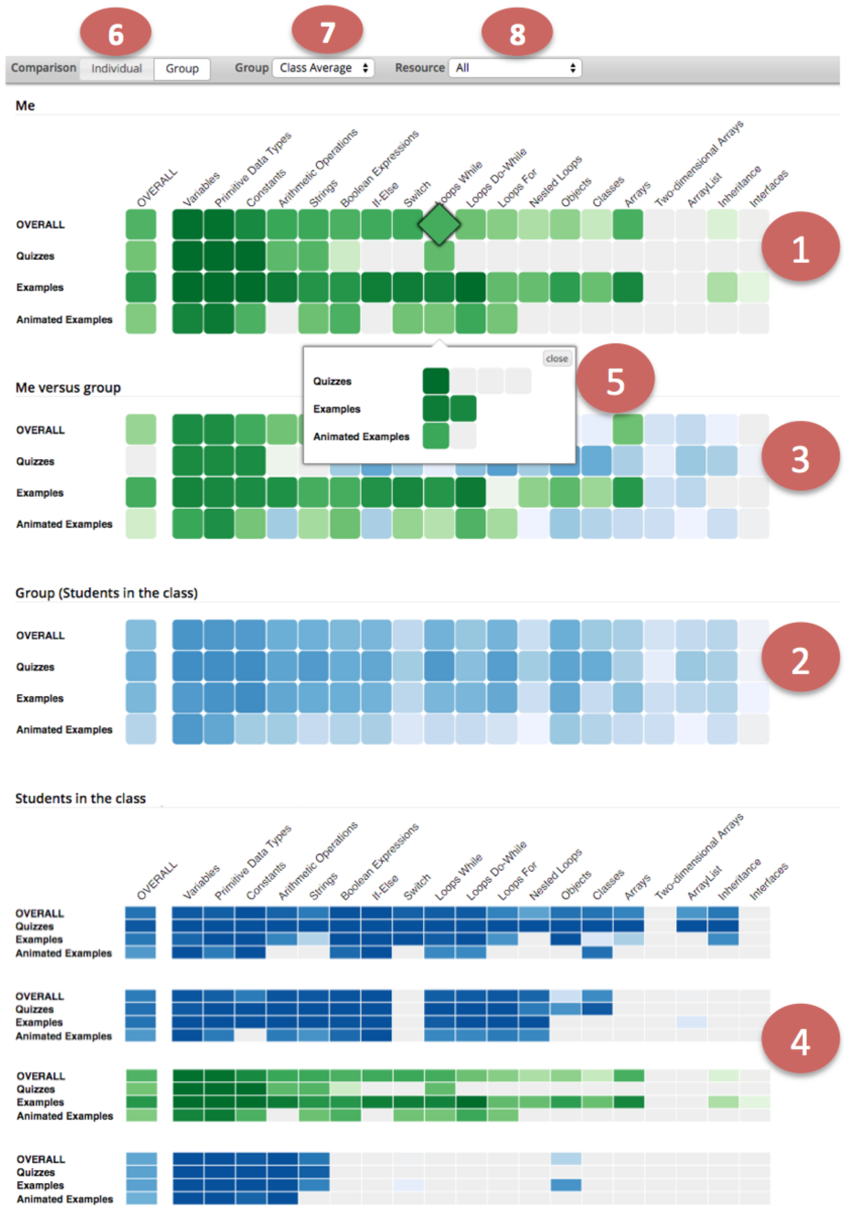


Figure 6: The full Mastery Grids interface. A menu bar contains controls to change the view of the group or the details shown. Circled numbers have been added in the image to support explanations.

tively engages students, provides mechanisms to store and retrieve student activity data, and ultimately, incorporates feedback mechanisms. Different type of activities is included in different domain courses implemented in Mastery Grids. We have implemented courses for Java, Python and SQL programming. In Java and Python programming course we have used the content of the type programming problems or parameterized problems (also called questions or quizzes) [Hsiao et al., 2010], annotated examples [Brusilovsky and Yudelson, 2008], and program animations (or animated examples) [Sirkiä and Sorva, 2015]. In Python programming course we also had used Parsons problems [Parsons and Haden, 2006]. In SQL programming course we have used SQL problems [Brusilovsky et al., 2010] and annotated examples [Brusilovsky and Yudelson, 2008]. Each of the content types is shown in the following.

3.1.1 Annotated Examples

Annotated examples provide interactively explorable text explanations of code examples. Figure 7 illustrates an annotated example in the topic "Logical Operators" in a Python programming course. A green bullet next to a line indicates that an explanation is available for that line. Once the student clicks on the line or the green bullet next to it, the explanation opens up below the line. Each explanation emphasizes important concepts in the line or the result of the line being executed.

Annotated examples are delivered by a system called WebEx [Brusilovsky and Yudelson, 2008]. All interactions of students with these examples are reported to the user modeling server. The reported data includes information about each example's lines that the student has viewed, along with the timestamp of those activities. We used this data in our analysis to evaluate the use of examples and their impact on student performance. Currently, we have developed annotated examples for Java, Python and SQL programming courses.

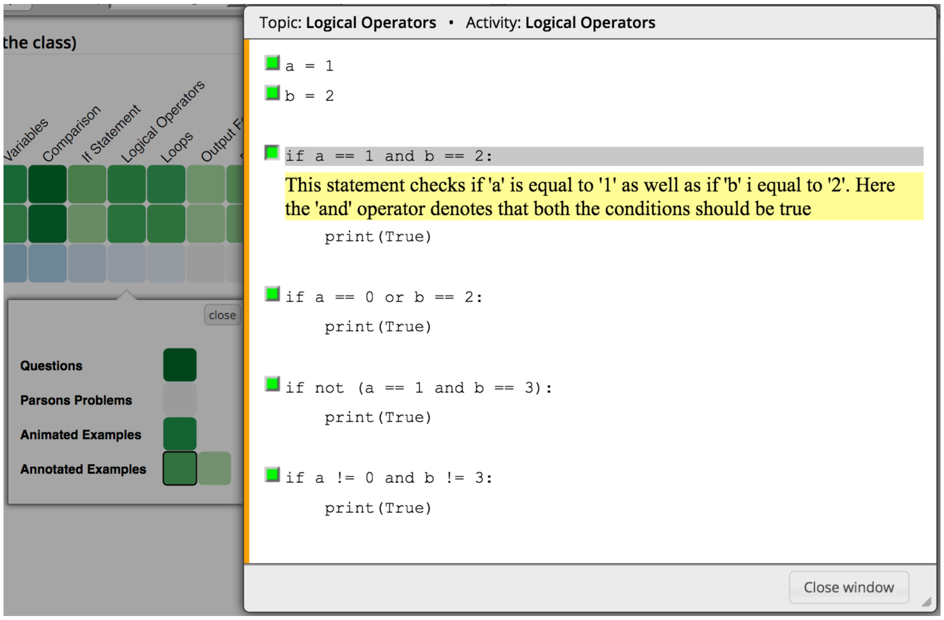


Figure 7: An instance of an annotated example loaded from Mastery Grids. Here, the student has clicked on the third line and an explanation is shown below the line that demonstrates the result of executing this line in the example program.

3.1.2 Animated Examples

Animated examples (Figure 8) provide an expression-level visualization of the code execution. The aim of these examples is to visually demonstrate how various programming constructs are executed by showing how each execution step changes the program state. These examples are implemented with the Jsvee library [Sirkiä, 2016] and are delivered using the Acos content server which is located in Finland.

Animated examples can visualize arithmetic operations, assignment statements, conditional statements, different kind of loops, functions with parameters and return values, lists and indexing, classes and instances, and references. Currently, we are integrating animated examples for Java and Python programming course.

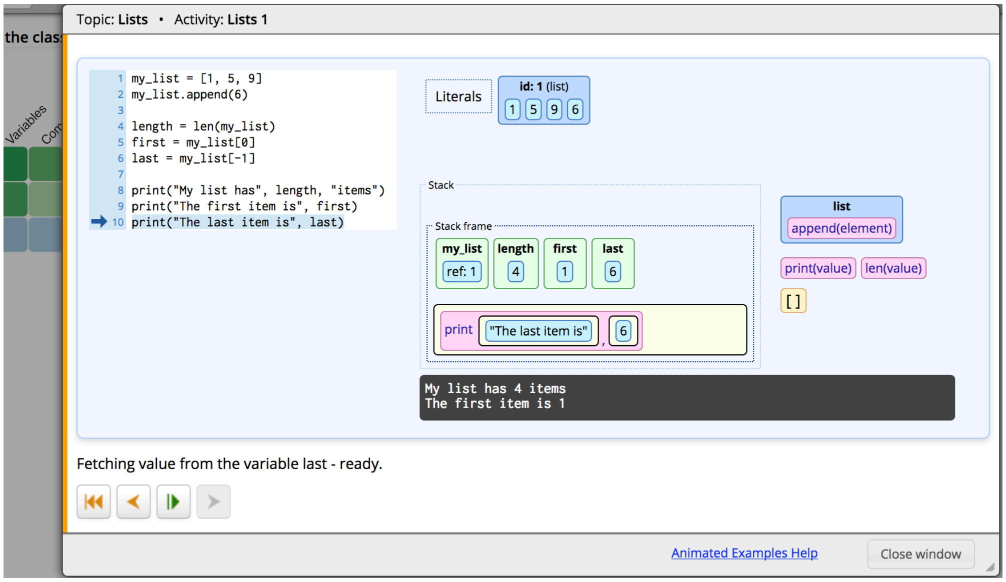


Figure 8: An animated example in the Master Grids system. The right panel shows the state of the stack frame and the output printed in the console when the program execution reaches the last line of the example.

3.1.3 Parameterized Problems

Semantic problems are parameterized exercises that test student comprehension of program execution by asking them about the output of the given program or the final value of a specific variable after the program is executed. For python domain, these problems are generated by the QuizPET system (Quizzes for Python Educational Testing), which is a re-design of QuizJET, an earlier Java-based system [Hsiao et al., 2010]. Since these exercises are parameterized, students can practice the same problem several times, each time with randomly selected values for the problem's parameter.

Figure 9 shows an instance of a parameterized problem for the "If Statement" topic in python. The student writes his/her answer in the text box area below the problem. Once the student's answer is submitted, QuizPET evaluates it and presents feedback to the student, along with the correct answer. Figure 10 shows the feedback presented to the student when the answer is evaluated as correct. The student can repeat the same problem with different parameter values by clicking on the "Try Again" button.

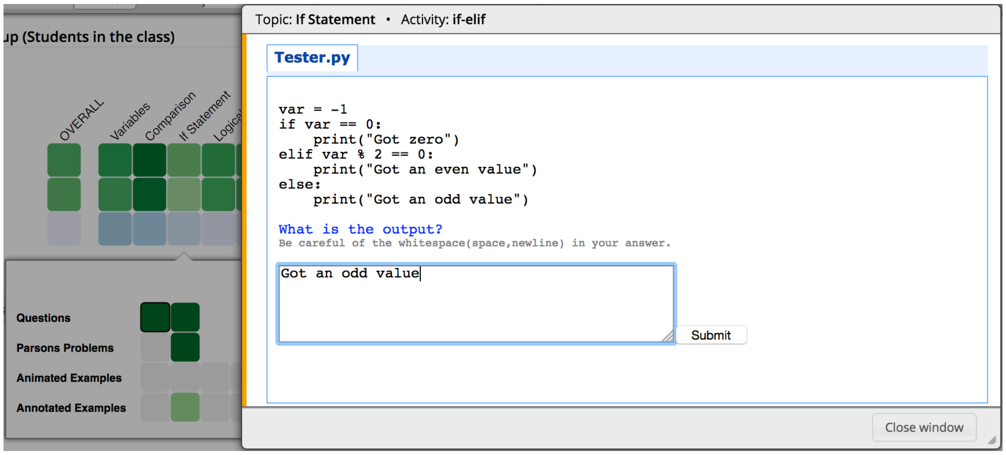


Figure 9: An instance of a parameterized problem for python in the Mastery Grids system.

In the domain of SQL programming parameterized problems are served by the SQL-Knot system [Brusilovsky et al., 2010]. An instance of this type of problems is shown in Figure 11.

3.1.4 Parsons Problems

Parsons problems are code construction exercises in which students do not need to type code. The original idea presented by Parsons and Haden [Parsons and Haden, 2006] describes the exercises so that there is a limited number of code fragments available in a random order. To solve the exercise, the student must construct the program described by putting the fragments in the correct order. Figure 12 shows an instance of Parsons problems in the Mastery Grids system.

Parsons problems are implemented with a JavaScript Js-parsons library provided by Ihantola and Karavirta [Ihantola and Karavirta, 2011] and delivered by the *Acos* server. For Python exercises, the library requires correct indentation. Therefore, the fragments must

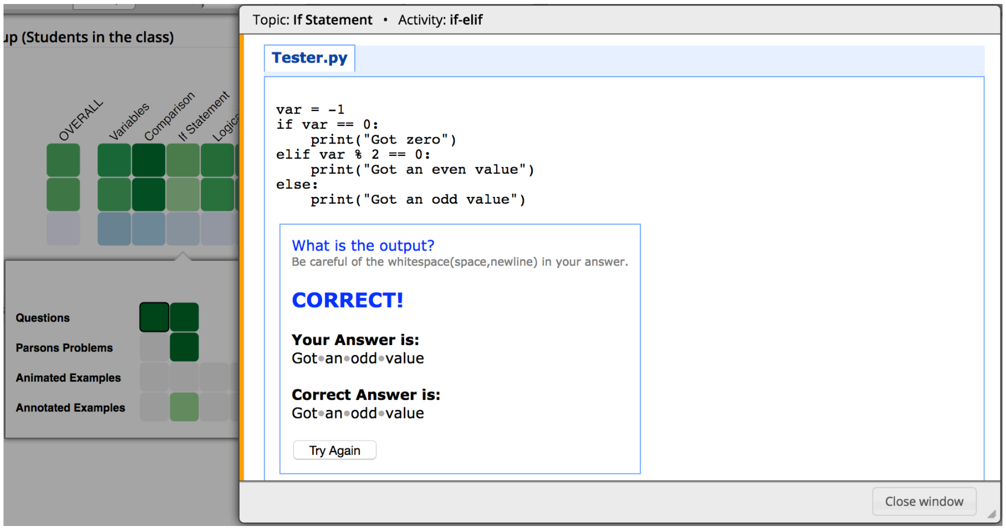


Figure 10: Feedback shown to the student after the system evaluates the submitted answer.

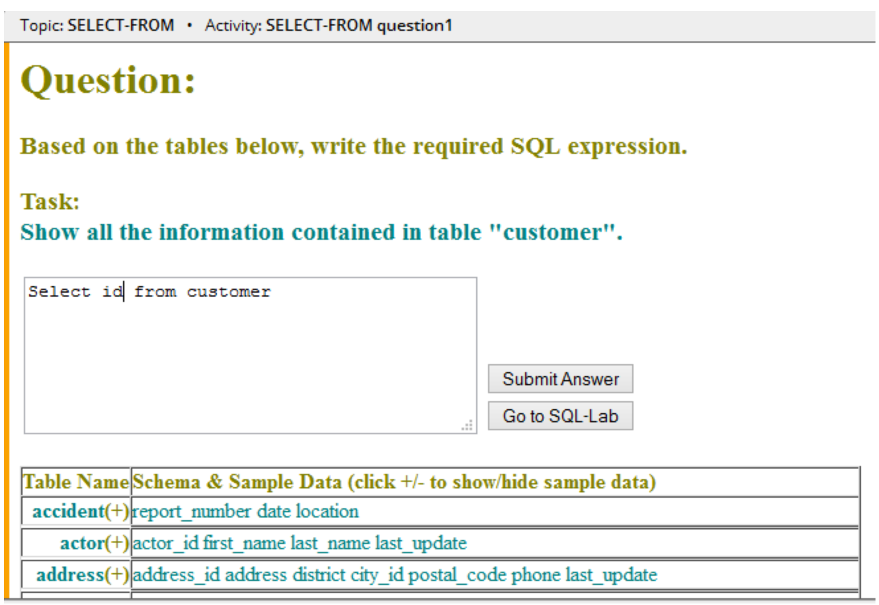


Figure 11: A parameterized problem for SQL programming served by the system SQLKnot.

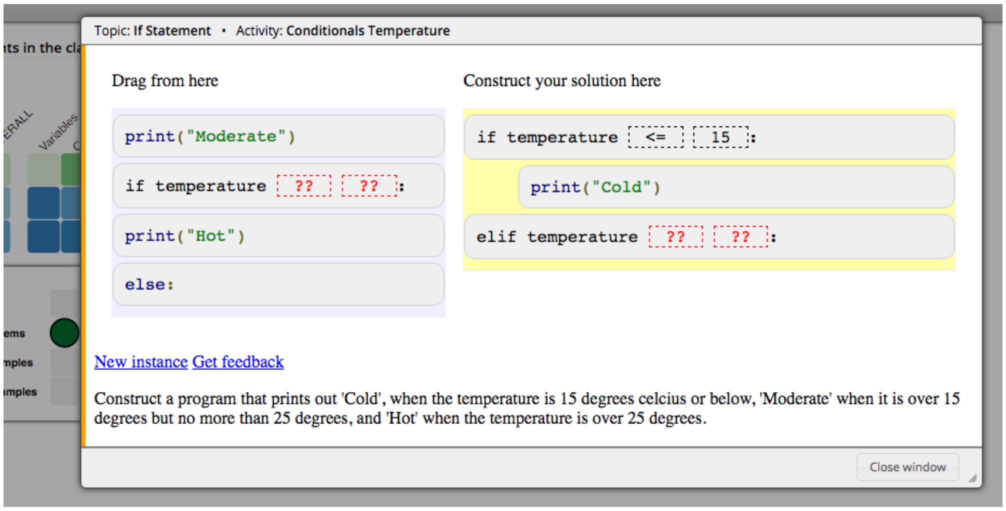


Figure 12: An instance of a Parson problem in the Mastery Grids system. The student assembles the solution to the question (written at the bottom) in the right side.

not only be in the correct order, but must also be indented correctly. The Js-parsons library also supports distractors; i.e., when not all the given fragments may be necessary for the solution. The fragments may also contain toggleable elements, which are shown as gaps. For these fragments, the student must select the correct operator to fill the gap (see the segmented squares with question marks '??' in Figure 12). In addition to providing feedback based on the positions of the fragments, Js-parsons exercises provide unit tests that can run the solution and check the results against the test cases.

3.2 SYSTEM ARCHITECTURE

This section explains the back side of Mastery Grids: its underlying architecture that makes the integration of several types of smart content possible. Mastery Grids are an attempt to implement the vision of the ACM ITiCSE working group on the use of smart content in computer science education [Brusilovsky et al., 2014a]. It brings together several types of smart learning content that are independent of the host system, fully reusable and hosted by different physical servers that are, in fact, located in different countries. For example, Animated Examples and Parsons problems are hosted on the *Acos* server[1] located in Helsinki, Finland. Parameterized problems and Annotated Examples are served by specialized *QuizPET* and *WebEx* content servers, respectively, that are located in Pittsburgh, USA. In this context, the Mastery Grids interface works as a *aggregator* that contains links to the content that can belong to different content servers or different applications, and transparently delivers the selected content to the students. The students might not be aware of which external system actually provides each type of content, what they see is a holistic system with the Mastery Grids interface and diverse learning content.

The ability to provide such transparent access to multiple kinds of reusable content while supporting data collection and personalization is supported by the Mastery Grids infrastructure. This infrastructure is an extension of the *ADAPT2* infrastructure[2], which extends the early KnowledgeTree framework [Brusilovsky, 2004]. The Mastery Grids infrastructure includes several types of components that inter-operate by using standardized communication protocols which are summarized in Figure 13. The main components are smart content providers such as several content servers, the Mastery Grids interface with its back-end services called *Aggregate*, and student modeling servers, such as the CUMULATE server [Yudelson et al., 2007].

[1] http://acos.cs.hut.fi/
[2] http://adapt2.sis.pitt.edu/wiki/ADAPT2

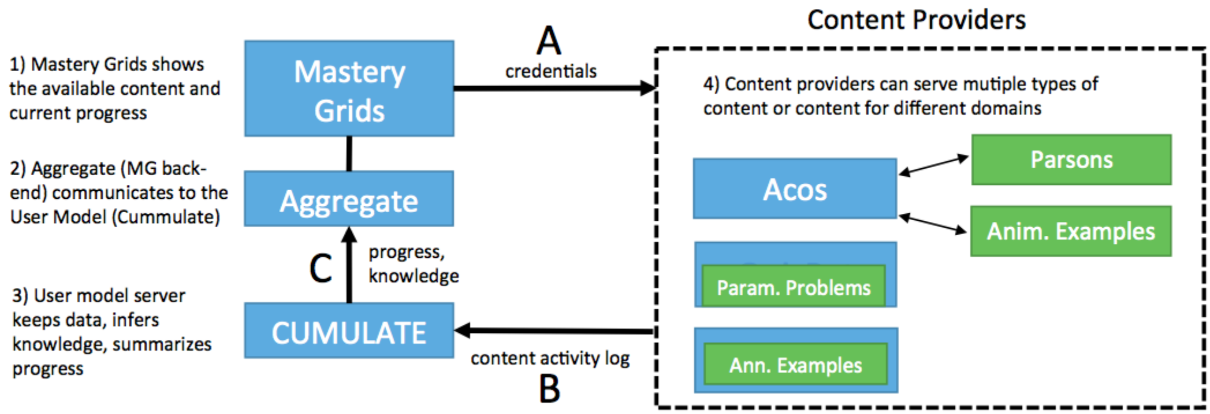


Figure 13: An overview of the different components in the Mastery Grids System and the communication that occurs between them. The arrows indicate the direction of the communication. The explanations give an overview of how information is transferred when a student visits Mastery Grids and uses the content provided by an external server. Blue boxes represent different content servers, and green boxes represent different content types. Note that QuizPET and WebEx content types are shown inside a small blue box, which indicates that there is a dedicated server that hosts only this type of content.

Three communication protocols support the smooth cooperation of these independent components within the system. The first *content invocation* protocol (the arrow labeled with letter A in the Figure 13) defines how a learning content item could be invoked from a specific server by a portal (i.e., from the Mastery Grids interface). The protocol is implemented as an HTTP GET request to the content provider, which identifies the requested activity and also passes the user's credentials: user identifier, session identifier, and group (class) identifier. The content is loaded into a *iframe* and there is no further direct communication between the content interface and the Mastery Grids interface. This first protocol imposes a requirement on the content provider: single content items should load independently into a *iframe* through a unique URL.

The second *event report* protocol (the letter B in Figure 13), also known as the CUMULATE protocol[3], defines how learning data are reported and logged to the student modeling

[3] http://adapt2.sis.pitt.edu/wiki/CUMULATE_protocol

server. Interactive content generates *learning events* based on user actions. For example, each time a student moves to the new line in an animated example, it will send an event to its content server (for example Acos in the Figure), which will deliver the event to the student modeling server CUMULATE using the learner modeling protocol. Parameterized problems and annotated examples send learning data as a flow of events directly to the student modeling server. CUMULATE uses the flow of learning events to calculate the knowledge progress. Since each type of content may require a different approach to compute the progress that a student has made on it, the student model needs to know how learning events can be processed to estimate knowledge progress. A set of services in the CUMULATE user model has been developed to provide such computations for all types of content accessible through Mastery Grids. For example, to mark a parameterized problem as completed, the user model checks if there is at least one attempt answered correctly by the learner; to compute the progress of an animated example, the student model computes the ratio of the different lines that have been seen by the learner and all the lines in the animation.

The third *knowledge query* communication protocol (the letter C in Figure 13) defines how Mastery Grids, from its back-end Aggregate can request information about student and group knowledge progress from the student modeling server. This communication channel is important to support personalization, learning analytics, and knowledge visualization. In the context of Mastery Grids, this information is used to present the comparative knowledge visualization that is shown in Figure 6. Aggregate takes the progress knowledge information reported from the user model and *aggregates* to the topic level to be shown in the Mastery Grids interface.

With this data flow design, all the components have their own tasks that make them highly reusable. For example, the main task for smart content providers is to deliver smart content activities and maintain student interactions with them. The content does not have to worry about authentication, storing grades, or logging interaction data because there is a predefined interface of how to communicate with the other parts of the system. It is also easy to add new types of content just by implementing the same interfaces that are used by the other content types. As a result, the architecture is fully open. New content servers could be easily added to offer other kinds of smart content. Different portals could be designed to maintain different types of interfaces with students, such as a more traditional "folder" interface of learning management systems, or electronic books such as Open DSA [Shaffer, 2016].

The presence of standard communication protocols also simplifies the integration of Mastery Grids with other learning systems. For example, in the context of the studies presented later in this thesis (see chapters 6 and 9), Aalto University students accessed the Mastery Grids system through a URL link, which authenticates the student to Mastery Grids (using an account mapping for anonymization) and loads it in a separate window.

3.3 ACTIVITY LOG

As explained before, Mastery Grids system is supported by a software platform of smart content providers and user modeling services that logs and process the activity within the system. All the activities of the students with the content is saved including attempts to parameterized questions and parsons problems, and interaction with examples and animated examples. Additionally, the Mastery Grids interface tracks the cells clicks, for example when the user clicks o open an activity; clicks on the on the menu options; mouseover cells; and scroll. The detailed activity data tracked from the interface and the activity logged in the user model (from content servers) can be combined allowing us to inspect detailed sequence logs of each learner interaction with the system. The detailed log allows, for example, to post-process the data to compute time spent in each action by subtracting time stamp from the time stamp of the next action logged.

A series of different activity variables can be computed from the activity log. In chapter 5 I described activity variables that I computed from this combined log data and that I later used in the analyses of classroom studies presented in chapters 6 and 9.

4.0 PREVIOUS STUDIES AND FINDINGS

Mastery Grids has been used in several classroom studies that are reported in different articles. In this chapter, I summarized the findings of three articles reporting studies in which I was involved and in which the general goal has been to evaluate the coarse-grained OLM of Mastery Grids and its social comparison features, in terms of how they engage and support navigation through the content of the system. Altogether, the studies and their findings represent a starting point of my thesis and set the motivation and directions to explore further work that is later presented in this document.

4.1 INITIAL MASTERY GRIDS STUDIES

The first version of Mastery Grids was deployed and evaluated in semester-long classroom studies in Fall term 2013 involving a course of Java Programming and 2 courses of Database Administration Systems, which took place in the School of Information Sciences at the University of Pittsburgh. The main goal of these studies was to verify if the tool has the positive impact on engagement and navigation that motivated its development, and to collect feedback that allows us to improve the system.

In the Java course, Mastery Grids supported all the content of the course, and in the Database courses, Mastery Grids supported the programming section of the course with SQL content. The system and the studies are reported in [Loboda et al., 2014]. In the Java domain, Mastery Grids were deployed organizing 75 WebEx examples (annotated examples) and 94 QuizJet questions (parameterized problems), while in the SQL courses there were a total of 64 WebEx examples and 46 SQL-Knot questions. In both versions, the content was organized into 19 topics.

In these studies, students were given two alternative ways of accessing content. One way was through the Mastery Grids interface. The alternative was a simple two-level hierarchy of HTTP links which from now on I refer to as *Links* interface or simply Links. The first level links listed topics and the second level links listed activities (i.e., questions and examples). We offered two alternative forms of access because we looked for contrasting the benefit of the Mastery Grids interface and not the quality of the content contained in it. Both access tools were introduced to students in the second week of classes in the Java course and on the fourth week of classes in both database courses (when SQL programming was introduced in the course according to the course syllabus). Students were informed that the use of these tools was non-mandatory and that there was no penalty for not using them. To engage students, the Java course instructor offered extra points (5 out of 100) towards class participation in solving at least 15 questions using either Links or Mastery Grids. In the database course, a similar amount of extra points was offered. In all courses and at the end of the term, students were asked to fill a questionnaire about usability and usefulness of the system.

Findings from these studies show several effects of the Mastery Grids interface in navigation through the content and engagement in the content activity. We observed that students who used Mastery Grids had a higher ratio of questions answered correctly than those who used *Links* interface only. It is possible then that the visualization guided students to the questions which were more suited to their level of understanding of the material. It was also observed that the visualization, directed students to new material at rates higher than the alternative Links interface which is consistent with previous work in the context of adaptive explanatory visualization [Loboda and Brusilovsky, 2008]. We hypothesized that students advancing faster may be the result of the visualization attempting to stay in sync with students progress and thus being able to direct them to new content more quickly.

When analyzing the relation of the activity in the system and the learning (as measured by the course grade) we observed that it was the total amount of activity, considering content activities and interface interaction, the measure that has some predicting power on grade. However, we pointed out that more studies were needed with specific planned intervention to support this kind of claim. More precisely, the current study cannot adjudicate causality, i.e., if it is indeed the case that using the visualization more helped students with getting a better grade or if instead, students which ended up getting a better grade were also the ones more likely to be engaged with supplementary educational tools. This is a common observation in studies of this nature and it is a reason to extend the studies to consider other factors that can influence the usage of the system and can mediate or moderate the effects in learning, such as motivational traits.

Finally, student feedback analysis demonstrated that students assessed the usefulness and usability of the system quite positively. At the same time, some features of the system were regarded as less positive than others which were important information to improve the system.

4.2 THE VALUE OF SOCIAL FEATURES: STUDY IN DATABASE COURSE

We performed a semester-long classroom study in a Database Administration Systems course offered in the School of Information Sciences at the University of Pittsburgh during the Fall 2015 term. The study has several purposes. First, while the previous studies looked at the general effects of Mastery Grids interface and contrast it against a non-visual interface, we now focus on evaluating the effects of the OLM with and without social comparison features. The set up of the course, split into two similar sections, allowed us to design a "clean" study in which each section was exposed to a different version of Mastery Grids (with and without social comparison orientation). Also, this study allowed us to evaluate some of the changes that we did in Mastery Grids considering the feedback received in the previous studies. Details of this study are in the article [Brusilovsky et al., 2016].

Since the class cohort was separated in two sections taught by the same instructor, we deployed a version without social comparison features, that we called *OSM* (Open Student Model) in one section (see Figure 14), and a version with the social comparison features named *OSSM* (Open Social Student Model) in the other section of the course (see Figure

15). Both versions, provided access to the same educational content, which includes parameterized SQL problems provided by SQLKnot system and annotated examples from the system WebEx. In the OSSM version of the interface, the group information was based on the progress of this group alone (it did not include the data of the students using the OSM version). Pretest and posttest were also collected, and the final grade in the course was also available for analyses. Also, the Iowa-Netherlands Comparison Orientation Measure (INCOM) developed by [Gibbons and Buunk, 1999] was administered to measure social comparison Orientation. This questionnaire is also used in later studies reported in this thesis and is described and replicated in the APPENDIX C.

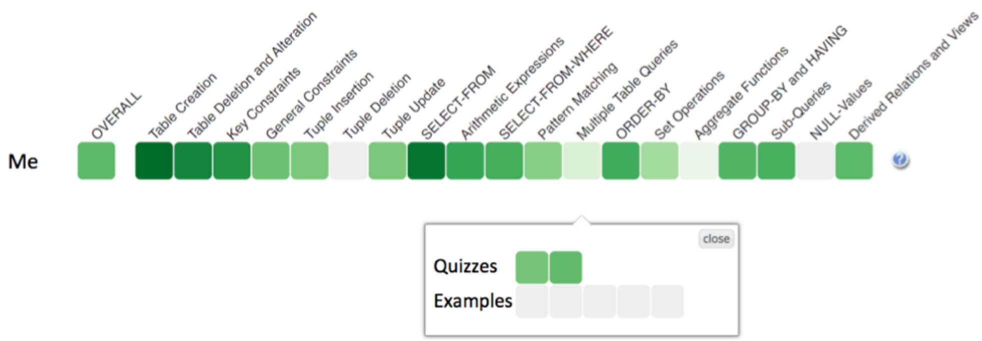


Figure 14: Mastery Grids without social comparison (*OSM*) for SQL programming course.

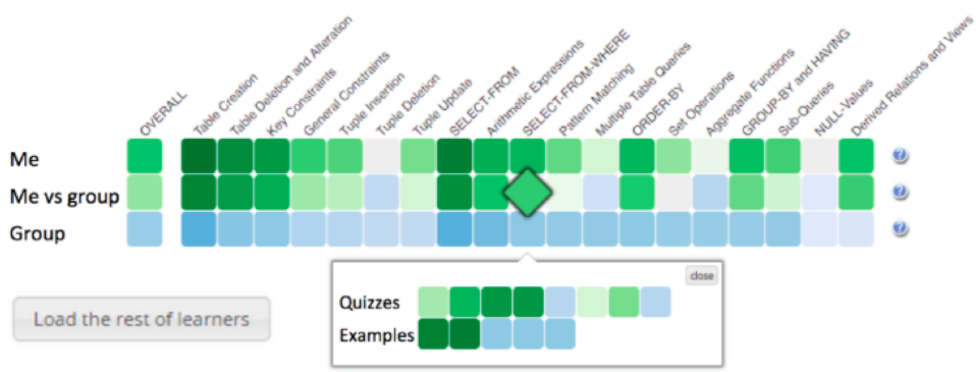


Figure 15: Mastery Grids with social comparison (*OSSM*) for SQL programming course.

The total number of students in the two sections of the course was 103, however, 14 students never logged into the system and were excluded from the analyses. Of the remaining 89 students, 47 (52.8%) worked with the OSM and 42 (47.2%) worked with the OSSM interface. Most of the participants (77%) were graduate students in the Information Science program.

The results of the study demonstrated a strong impact of the social features in students engagement and retention with the system, and on performance with assessment activities. Regarding retention, we compared the percent of students who engaged with OSM and OSSM at six different levels (0 or more activities completed, 10 or more, 20 or more, etc.). In Figure 16(a) we compare the percentage of students who logged in at least once, and continue doing activity in the system. A difference emerged between the groups early and then persisted. For OSSM, almost 70% of the students decided to explore the system further attempting at least one question. In contrast, for OSM, less than 30% of them did so. At the level of 30+ questions that we could consider as a serious engagement with the system, the OSSM group still retained more than 50% of its original users while OSM engagement was below 20%. Figure 16(b) provides an alternative look at the student engagement treating the number of students who attempted at least one problem as 100% in each group. Still, we see that OSM group is losing students at a higher rate than the OSSM group, even with this adjustment. These observations demonstrate that the OSSM interface was much more successful than the OSM interface in engaging and retaining students.

Regarding system usage, the results demonstrated a remarkably higher level of activity in the OSSM group, with significant differences in all system activity variables compared. These variables include a number of sessions, topics covered, raw count of problems attempts and example viewed and activity in the interface, like topic cell clicks, or time in the interface. Table 1 shows a means of activity in these different variables and the result of Mann Whitney U test.

The results indicated that students who used the OSSM interface were significantly more engaged with the system. The difference is not only significant, but shows double, triple, or even larger increases in student activity. The number of attempted problems more than tripled and the number of problems solved correctly quadrupled in the OSSM group. OSSM

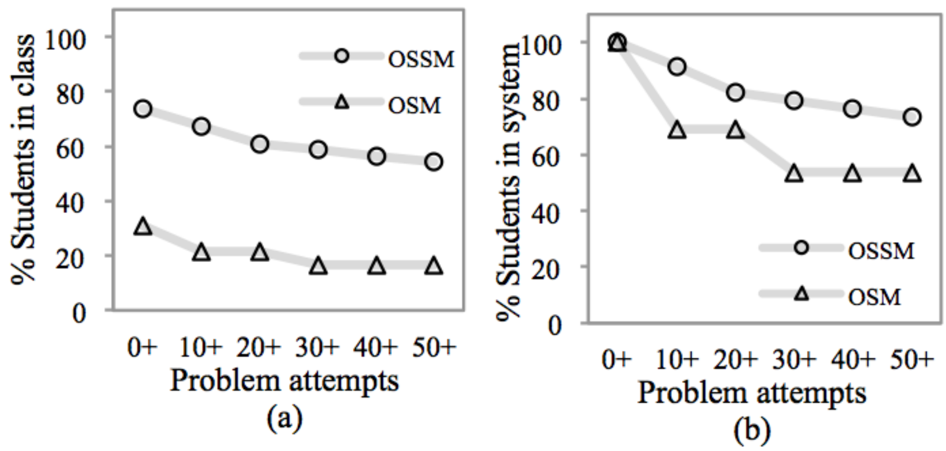


Figure 16: Students according to number of problem attempts in the OSM and OSSM groups: (a) as percent of students who ever logged in; (b) as percent of students who attempted at least one problem.

students viewed twice as many examples and example lines and covered three times as many topics. The OSSM group also worked more extensively with the Mastery Grids interface, and overall spent almost twice as much time in the system.

We also observed that times in activity in OSSM group were significantly lower than in the other group. Considering that these students also did more activity, then we claim that students who used the OSSM interface worked more efficiently. We believe that this is a result of the social navigation support provided by the OSSM interface guiding students to the right content at the right time. We can't rule out another possible reason students may rush to move ahead of their classmates in the OSSM group where class progress was visible. In this rush, they may skim examples too fast to understand them. It is harder, however, to argue that OSSM students rushed through all content. Their work on questions was as thorough as the work of OSM group: no significant difference for the success rate (percentage of correct attempts) was found (median OSM =61%; median OSSM =64%). We complemented these analyses by comparing Instructional Effectiveness between groups [Paas and Van Merriënboer, 1993]. This measure includes the correct attempts to assessment items (problems) and the time invested. According to results of Mann Whitney U test (U=116.000,

Table 1: System usage in OSM and OSSM groups. Significance is marked: * ($p < 0.05$), and ** ($p < 0.01$).

Variable	OSM Mean	OSSM Mean	Mann Whitney U
Sessions	3.93	6.26	685.50*
Topics coverage	19.00%	56.40%	567.50**
Total attempts to problems	25.86	97.62	548.50**
Correct attempts to problems	14.62	60.28	548.00**
Distinct problems attempted	7.71	23.51	549.00**
Distinct problems attempted correctly	7.52	23.11	545.00**
Distinct examples viewed	18.19	38.55	611.50**
Views to example lines	91.6	209.4	609.00**
MG loads	5.05	9.83	618.50**
MG clicks on topic cells	24.17	61.36	638.50**
MG click on content cells	46.17	119.19	577.50**
MG difficulty feedback answers	6.83	14.68	599.50**
Total time in the system	5145.34	9276.58	667.00**
Time in problems	911.86	2727.38	582.00**
Time in MG (navigation)	2260.1	4085.31	625.00**

p=0.045), instructional effectiveness scores of students who studied with the OSSM interface were significantly higher (N=32, mean=0.22) than the scores of students who studied with the OSM interface (N=12, mean=0.03).

Regarding the influence in learning, there was no significant difference between the groups in normalized learning gain (ngain= (posttest-pretest)/ (maxscore-pretest)) when we looked at all students who used the system. However, when we split students into weak and strong students according to their pretest score and selected only students who at least did 5 activities within the system, we found differences. The mean learning gain was higher for both weak and strong students in the OSSM group compared to the OSM group and the difference was significant for weak students (according to the results of independent samples t-test (t=-2.22; p=.033). More advanced analyses using regression showed that the number of problems is a significant predictor of the final course grade, with a β of 0.09, which indicated that attempting 100 problems will increase the final grade by 9 (final grade goes from 0 to 100). Putting these result together: in both groups, more attempts on problems were associated with gaining a better grade in the final exam, and in OSSM group, students do more work, including more problem-solving.

Regarding differences in gender, the analyses found significant interactions between the effects of gender and interface type (OSM, OSSM) on almost every system usage parameter. The nature of the effect is explained: while the presence of social comparison features in OSSM positively affected usage for both genders, male students were significantly more affected by social comparison. As the data show, female students in the OSM group used the system more than males in almost every aspect. However, in the OSSM group the situation is completely reversed: male students demonstrated much higher system usage in every aspect. We also saw that male students were significantly more interested to compare themselves with others as they used the comparison features more. This finding is consistent with several previous studies showing that females are often more reluctant to compete than males [Niederle and Vesterlund, 2011].

Subjective evaluation through a usability and usefulness questionnaire was applied to 81 students (42 in OSSM group, 39 in OSM group) and showed a positive opinion towards the system that was stronger in the OSSM group. Interestingly, results also indicated that while students in the OSSM group used the system much more than the students in OSM group, OSSM students were also more eager to attribute it to the ability to their own progress. To examine the impact of in-system experience, we clustered students into usage groups, low (N=26) and high (N=27) from the standardized values of the system usage variables. We expected that students who used the system more would evaluate it higher, as it frequently happens with complicated systems, but we did not find any significant difference here. We hypothesized that the system was sufficiently simple and usable to be sufficiently mastered even by the low group.

As mentioned before, in the study we also collected the social comparison orientation of the students measured using a questionnaire [Gibbons and Buunk, 1999]. Interestingly, we did not find any effect or relationship of this measure in system usage, nor with performance. However, we found that high social comparison orientation students were more positive when evaluating the social comparison features in the subjective evaluation of the system.

4.3 MASTERY GRIDS WITH SOCIAL COMPARISON: A STUDY IN A JAVA COURSE

This work, described in details in [Guerra et al., 2016], evaluates the use of Mastery Grids with and without social comparison orientation in a Java programming course. There were two main reasons behind this work. First, to evaluate the power of the Open Social Learner Model in another domain (a previous study comparing Mastery Grids OLM and OSLM was in a Database course with SQL content). Second, the studies reported here are the first contextualizing Mastery Grids as a Self-Regulated Learning tool, and including in elements of Learning Motivation in the evaluation. Specifically, these studies used the Achievement-Goal Orientation framework to measure Learning Motivation, setting up a path that is further explored in this thesis.

Following the results obtained in our previous study in which we found a strong positive effect of the social comparison features in Mastery Grids, (see Section 4.2), we now analyze a similar setup in two semester-long classroom studies in an introductory Object-Oriented Java programming class during 2014-2015. Classes were taught by the same instructors and had the same setup on the two semesters. Students were offered Mastery Grids as a voluntary practice system. Half of the students were exposed to a version with social comparison (*OSSM* group) and the other half, without (*OSM* group). Both studies collected pretest and posttest, and to characterize he Learning Motivation, we used the Achievement-Goal Orientation questionnaire [Elliot and Murayama, 2008], which I repeatedly use in the studies contained in this thesis in chapters 6 and 9, and which is explained in details in Section 5.3.3 and the APPENDIX B.

Regarding the overall effect of Mastery Grids, the results of the studies were compared to a previous classroom study in which the same content activities for Java programming were offered in the same course without the Mastery Grids. This baseline study was called *Portal* stressing that the content was offered in a portal fashion with a set of links to the activities rather than in a visual OLM interface. Results are replicated in Table 2 and show a significant positive effect of Mastery Grids over the baseline (here p-values represent the significant differences between the values of the columns at the left with the column *Portal*). The OSM/OSSM interface made the Mastery Grids system arguably more addictive than the basic portal: the average number of sessions and examples viewed were significantly higher in all conditions of the Mastery Grids system (MG, OSM, and OSSM). Progress tracking also allowed students to better distribute their efforts: on average, when using Mastery Grids, students explored and solved more distinct problems. This difference becomes significant for the OSSM group, where they accessed about 1.6 times more distinct problems than in the *Portal*. This indicates that the navigation support available in Mastery Grids decreases the students tendency of staying with the same content (for example, repeating problems they have already mastered), and as a result, students moved on to new problems more quickly. This data correlates (but not significantly) with a slightly lower *success rate* in the Mastery Grids system. Our data shows that in the absence of mastery indicators and navigation support offering guidance across course topics, students tended to over-stay within the topics, repeating the same problems even after solving them correctly, which resulted in a larger fraction of successful attempts on the same problems.

These observations indicate that the Mastery Grids system is more beneficial than a traditional portal, in terms of student engagement and effort allocation.

We then analyzed the difference between the OSM and the OSSM groups. Because in both studies the social comparison features were introduced a few weeks after the system (OSM version) was introduced in the beginning of the term, then we labeled the activities of the students as *Part 1* and *Part 2* to refer to the periods before and after the social features were introduced in the OSSM group. Results showed that while there were no significant

Table 2: The Mean±SD of system usage statistics: comparison between a portal of course materials and the Mastery Grids system across all groups (MG), the OSM group, and the OSSM group. Significant level: ***: <.001; **: <.01; *: <.05; .: <.1

Parameters	Portal	MG	p-value	OSM	p-value	OSSM	p-value
Logged-in students	17	89	-	43	-	34	-
Active students	14 (82%)	80 (90%)	-	40 (93%)	-	30 (88%)	-
Sessions	2.71±1.49	7.54±6.05	***	7.45±5.76	***	9.37±6.49	***
Distinct topics	9.21±4.85	9.4±5.6		9.3±5.71		11.47±4.75	
Problem attempts	72.36±67.25	78.88±62.18		76.45±54.33		100.83±69.51	.
Distinct problems	32.79±21.67	43.74±28.26		43.6±28.25		53.2±25.86	*
Distinct problems solved	28.43±19.2	41.92±28.13		41.88±28.7		50.77±25.57	**
Success rate	.707±.147	.648±.144		.639±.147		.627±.127	.
Repeats per problem	2.52±1.81	1.8±0.77		1.83±0.84		1.85±0.74	
Success per problem	1.8±1.36	1.11±0.35		1.11±0.33		1.12±0.39	.
Failure per problem	0.72±0.61	0.69±0.56		0.72±0.67		0.73±0.45	
Examples viewed	13.27±9	32.52±26.23	*	31.02±26.67	*	41.57±24.67	***

differences overall, the groups had different patterns of engagement and system usage from Part 1 to Part 2. Using repeated measures Anova on the amount of activity from part 1 to part 2, we found that students in the OSSM group increased their amount of activity per session, while OSM students decreased it, $F(1,55) = 4.972$, $p = 0.03$, $\eta_p^2 = .083$. This effect can be seen in Figure 17. Another significant interaction was found for the factors *Time*, *Gender* and *Social* group on the number of examples displayed, $F(1,45) = 6.467$, $p = .014$, $\eta_p^2 = 0.126$. Female students in the OSSM group tended to increase the number of examples displayed from Part 1 to Part 2, while male students decreased the number of examples displayed, and both female and male students decreased the number of examples displayed in the OSM group (Figure 18).

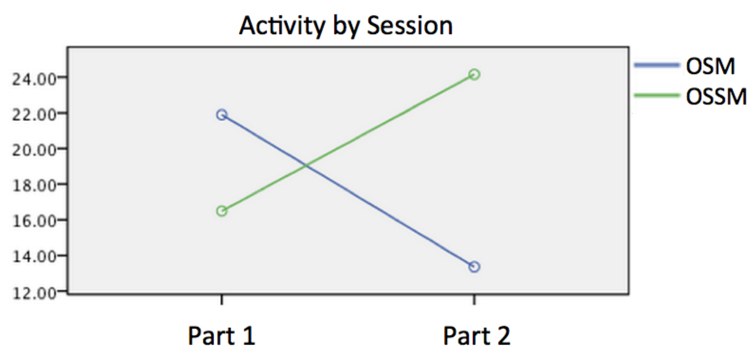


Figure 17: Interaction between Time (Part 1, Part 2) and *Social* factor (OSM/OSSM).

Similarly, we found differences in the change of Instructional Effectiveness [Paas and Van Merriënboer, 1993] between the groups from Part 1 to Part 2. A repeated-measure analysis of variance with both groups (OSM/OSSM) and gender as factors showed the main effect of time (Part 1, Part 2) is significant ($F(1,40) = 27.02$, $p < .001$). The within-subject test indicates that the interaction of time and group is also significant ($F(1,40) = 4.72$, $p = .036$), in Part 2 the effectiveness scores of the OSSM group ($M = 0.18$, $SE = .426$) were higher than in the OSM group ($M = -2.81$, $SE = .389$). Also, the interaction of gender and group was marginally significant ($F(1,40) = 3:59$, $p = .065$), male students in the

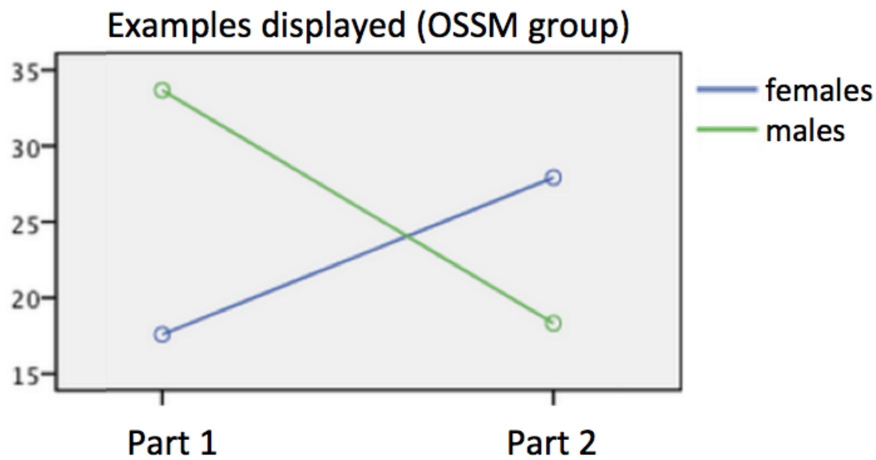


Figure 18: Examples displayed by female and male students in the OSSM during Part 1 and Part 2.

OSSM group had higher effectiveness scores ($M = 0.12$, $SE = 0.40$) than male students in the OSM group ($M = -0.48$, $SE = 0.37$) during Part 2. In general, we observed a tendency to decrease the effectiveness scores from Part 1 to Part 2, except for male students in the OSSM group.

To explore the way students navigated through the system in both groups, we computed a *ratio of non-sequential navigation* as the ratio between the count of times the learner goes from one activity to another activity in a different topic that is not the next topic (*jump-forward* and *jump-backward*), by the total times she transitioned from one activity to another. Analyses showed that the non-sequential patterns increased more in the OSM group than in the OSSM group from Part 1 to Part 2, i.e., OSSM became more sequential. This could be due to the social nature of the OSSM that makes students more conservative in their navigation – they tend to sequentially follow their peers rather than browsing the content space by their own, which is often a non-sequential process. More interestingly, there was a positive association between non-sequential navigation patterns and learning gain (form pretest to posttest, normalized): those who had a higher proportion of non-sequential patterns gained more knowledge. Although the two groups (OSM and OSSM) were not different in terms of the learning gain, this suggests that students in the social group might gain more knowledge if other adaptive features are added to the social interface, such as individual or personalized guidance. We concluded that future studies should be conducted to investigate this hypothesis.

As explained before, Motivation was measured by the Achievement-Goal questionnaire. The questionnaire provides 4 factors: Mastery-Approach, Mastery-Avoidance, Performance-Approach, and Performance-Avoidance. We applied the questionnaire three times: at the beginning of the term, at the middle point (before midterm), and at the end of the term. Results of Repeated measures Anova analyses showed that while all motivational factors changed from the initial to the final measure (all decreasing), a significant interaction existed for the Performance-Approach orientation and *group* factor (OSM, OSSM), $F(1, 50) = 7.506$, $p = .009$, $\eta_p^2 = .131$. Students in the OSSM group showed a flatter slope of the Performance-Approach level (decreased less) than students of the OSM group (Figure 19). These results suggest either that students who did not decrease their Performance-Approach orientation are becoming engaged by the social comparison features, or that social comparison features are influencing students positively in their Performance orientation. Both of these explanations have support in the achievement-goal literature, and further research is needed to establish a causal relationship. It is interesting to highlight that the *Social* factor presented no interaction effect, nor a main effect on the change of other Achievement-Goal factors like Mastery-Approach orientation. Even when the social comparison features might foster performance orientation, they are not negatively influencing the mastery orientation.

We did not find relationships between the motivation factors and the instructional effectiveness score. However we found a relationship between motivation and the ratio of non-sequential navigation. A significant negative correlation between the proportion of non-sequential patterns and the Mastery-Approach orientation score was found, $\rho = -.378$, $p = .043$, $N = 29$. This suggests that highly motivated students are more sequential in their patterns of navigation. A significant negative correlation was also found in the difference of the proportion of the non-sequential patterns (Part 2 - Part 1) and Mastery-Approach level, $\rho = -.429$, $p = .02$, $N = 29$. When looking at the OSM and OSSM groups separately, the

Figure 19: Different decrease of Performance-Approach orientation in OSM and OSSM groups.

negative correlation between the difference of non-sequentiality and the Mastery-Approach orientation is stronger in the OSSM group $\rho = -.62$, $p = .018$, $N = 14$, and is not significant in the OSM group. These results suggest that more motivated students become more sequential in their patterns of navigation after being exposed to social features.

Subjective evaluation through a questionnaire which included questions of usability and usefulness of the system was collected. In general, the evaluation of the OSM interface (the ability to monitor your own progress) is positive in terms of both usability and usefulness. Students also agreed that Mastery Grids motivated them to work on problems and examples. When crossing this subjective answers with the achievement-goal factors we found that high Mastery-Approach students were more positive towards the usefulness of the system. In the OSSM group alone, high Mastery-Approach students value the interface more and think higher of the usability of the system than low Mastery-Approach students.

4.4 CONCLUSIONS OF THIS CHAPTER

Through several classroom studies, we have seen several positive effects of Mastery Grids OLM and the social comparison features on engagement with practice content, performance in self-assessment content activities, and navigation through the system. We have seen consistently that the visual interface showing the knowledge progress of the learner (OLM) makes her do more activity and move forward in the content quicker than other interface without OLM features. The social features (OSLM or OSSM) enhance these effects: produce more activity, students move quickly, and sometimes become more sequential in their navigation.

An interesting observation across studies is that the level of impact of the social features greatly varies: increase of activity was strong and clear in the database course study, but it was not as clear in the Java studies. The effects on performance with self-assessment content, as measured by success rates and instructional effectiveness also showed differences across studies. We have seen that social comparison features have shown increase and decrease in performance. An explanation is that social comparison features may produce effects on navigation that counter each other in terms of performance. In one hand, as seen in the studies, social comparison make students more sequential in their navigation, which means that they complete more content in order and without jumping ahead to more complex activities. As a result, this sequential navigational trend may generate higher success rates. On the other hand, social features have demonstrated to encourage students to move quickly to the next content activities, avoiding to overstay in the same activities they already solved, thus decreasing their success rates. More research is needed to understand better these phenomena.

I think that other factors, such as cultural background, education, motivation, etc. might explain the differences observed across studies and could bring ground to better understand the potential impact of OLM and OSLM. An example of this is provided in the last studies reported, where I included the Achievement-Goal motivational orientation in the analyses and I observed a relationship between the change of Performance Orientation and the engagement with the learning content in the social comparison group. However, until now, the studies have been conducted in courses with small or medium size cohorts which allow studying the overall effects, but fail to provide enough statistical power to *dig deeper* including more elaborate analyses with other factors. A central contribution of my Thesis work is to evaluate Mastery Grids OSLM in bigger classroom studies, allowing including in the analyses other factors such as motivational traits, and their combined impact on the system usage.

Although Mastery Grids has demonstrated positive effects on engaging, we think that its role as a content navigational support tool is limited because of its coarse-grained approach. Currently, the system only shows the knowledge progress on the topics and the completion of the content activities within, but does not provide much information to help the learner to choose which activities are more suitable for her to do. Other related works have exploited content recommendations approaches to address this issue [Hosseini et al., 2015a]. From the Open Learner Model perspective, we believe that the navigational support of the system could be improved if we include detailed information about the content of each topic and how is the learner doing with it. This represents a strong motivational element of my Thesis, and I explore this issue adding fine-grained OLM features to Mastery Grids in later chapters.

5.0 RESEARCH FRAMEWORK, OVERALL DESIGN AND STUDIES

This chapter describes how I refined the research framework that has guided the work of my thesis. On the one hand, the chapter describes how I developed hypotheses connecting the goals and research questions with expectations set by prior knowledge and theoretical foundations. On the other hand, it presents an overall view of the studies conducted in the chapters following it, and how they contributed to the choice of research questions considered in this thesis work. Also, chapters 6 and 9 report on two similar classroom studies which share many measures and variables, such as pretest, posttest, motivation questionnaire, and system activity log variables. These are described in detail here, to avoid repeating this information later.

5.1 RESEARCH FRAMEWORK

Research questions, first stated in the Introductory chapter and repeated here, were phrased somewhat vaguely, to reflect the exploratory character of my work. However, both the findings of previous studies and theoretical foundations related to this work have set expectations regarding the effects to be observed. For example, prior findings show a consistent increase in the amount of activity in the system (amount of problems and examples completed, called *system activity*) when the social comparison features are activated in the Mastery Grids interface. These expectations influenced me to set some of the hypotheses in each of the aspects to be explored, which are identified in each of the research questions.

RQ 1 What are the effects of an OLM with social comparison features (or OSLM) on *system* *activity* compared to an individual-view OLM?

As mentioned before, prior studies have shown an increased amount of system activity in the group that is exposed to the social comparison features in Mastery Grids [Brusilovsky et al., 2016]. The effect, although of very different magnitude across studies, has been consistent. Thus, I state the following hypothesis:

H1 *Students exposed to an OLM with social comparison features increase the level of activity in the system.*

Social comparison features have also shown other effects, such as sequential *navigation* [Hosseini et al., 2015b], and contradictory effects in *performance* in self-assessment content. Even though I will look at the effects on navigation and performance, I do not state specific hypotheses regarding them.

RQ 2 What are the effects of using a fine-grained OLM on *system activity*?

The main goal of adding the fine-grained feature in the OLM is to support students when they are navigating the system. Fine-grained information, in the form of detailed information about the learner model, could provide the student with an additional alternative to explore within the content of the system and, at the same time, support her in making decisions about which content to target. Assuming that both general goals (exploration and searching specific content) are targeted by students in their free usage of the practice system, I would expect that navigation becomes more efficient, meaning that students have to spend less effort (time and number of actions) searching for content or finding *interesting* content. Considering that the students are free to engage as they wish with the system, efforts can only be measured in relative terms. This means, for example, that a measure of the time spent in navigating the OLM interface should be considered relative to the total amount of time in the system, or that the number of actions are reported as proportional to the total number of content activities attempted or completed. I'll return to these measures of navigation efficiency in the Section 5.3. The following hypothesis is stated:

H2 *Fine-Grained OLM helps students to navigate the content of the system more efficiently.*

RQ 3 How do individual differences influence system activity within an OLM? This research question refer to individual differences that I have narrowed down to three factors: prior knowledge, learning motivation, and social comparison orientation. Thus, three sub-research questions are stated accordingly:

RQ 3.1 How does prior knowledge influence system activity within an OLM?

RQ 3.2 How does learning motivation influence system activity within an OLM?

RQ 3.3 How does social comparison orientation influence system activity within an OLM?

I expect prior knowledge to be an important factor related to the engagement in the practice system. However, the expected effects might neutralize each other. On the one hand, having prior knowledge of programming makes it easier for students to understand and complete content activities within the system. On the other hand, the practice system may be seen as more valuable by students with little or no experience, who realize that they need more practice. The research question remains exploratory in its nature and no hypotheses are stated to indicate the valence of expected effects. However, I will focus my research on the relationships between prior knowledge and the two OLM features explored: social comparison orientation and fine-grained OLM. This means that I am interested in the aggregated effects of 1) prior knowledge and the presence of social comparison features, and 2) prior knowledge and fine-grained OLM features.

The theoretical background related to learning motivation, specifically related to the Achievement-Goal Orientation framework, supports certain expectations about how different interface features explored in this work will influence students with different motivational profiles. *Performance orientation* is defined as the motivational goal in which the learner pays more attention to scores and ranks and become specially sensitive to social comparison [Elliot and Murayama, 2008, Grant and Dweck, 2003]. Thus, I expect that:

H3 *Social comparison features will increase the engagement of students who are highly performance oriented.*

On the other hand, *mastery* oriented students set their goals toward learning (*I want to learn as much as I can*) and tend to engage in metacognitive tasks that allow them to make sense of their learning process [Grant and Dweck, 2003]. For these students, the details offered by the fine-grained OLM may gain relevance by facilitating the visual projection of the internal metacognitive model of understanding of the content being learned into the *external anchoring* that the fine-grained OLM conveys [Liu and Stasko, 2010]. Thus, I expect that students with a higher mastery orientation will get more value from a fine-grained interface, which will translate into more activity in the interface (although not necessarily more practice content completed) compared to other less mastery oriented students.

H4 *Students with a higher Mastery orientation will use the fine-grained components more.*

Finally, the Social Comparison Orientation factor is a subjective measure of the extent to which a person tends to compare to others [Gibbons and Buunk, 1999]. It is natural to expect that this factor will affect engagement in a system featuring social comparison. Thus, I expect that:

H5 *The effects of social comparison features of the system will be stronger for students with higher Social Comparison Orientation.*

RQ 4 How does the use of a learning system featuring OLM, OSLM and fine-grained elements affect motivation?

The last research question focuses on the potential effects that system interfaces features have on motivation and is grounded on the fact that motivations can change. Specifically, literature states that achievement-goals are not fixed orientations, and that they can change as a result of learning experiences that favor certain orientations. For example, in a context in which scores and ranks are stressed, students may become performance oriented [Grant and Dweck, 2003, OḰeefe et al., 2013]. Along this line, my previous work has shown evidence that the social comparison features included in the Mastery Grids are related to maintaining performance orientation and thus will not decrease the mastery orientation (as opposed to finding a decrease in this motivational factor in the group without social features). Following these results, I expect:

H6 *The active use of OLM with social comparison features will increase the Performance orientation of the students.*

Homologous to this, and also based on the relationship established between the Mastery orientation and fine-grained features of the OLM, I expect that this goal orientation could be affected by exposure to a detailed OLM:

H7 *The active use of an OLM with fine-grained features will increase the Mastery orientation of the students.*

5.2 CLASSROOM STUDIES AND CONTROLLED STUDIES

An overall view of the studies contained in this thesis are shown in Figure 20. The left side of the figure summarizes the four variations of the Mastery Grids interface to be explored in this work:

- coarse-grained Mastery Grids, with and without social comparison features (OLM, OSLM), and
- coarse- + fine-grained Mastery Grids, with and without social comparison features.

Note that granularity is explored by adding the fine-grained component to the coarse-grained component, in an interface that I call Rich-OLM.

On the right side of Figure 20, the studies conducted are summarized. The exploration of social comparison features (OLM, OSLM) in the coarse-grained Mastery Grids is performed in chapter 6, with a classroom study. This study focuses on research question 1, *What are the effects of an OLM with social comparison features (or OSLM) on system activity*

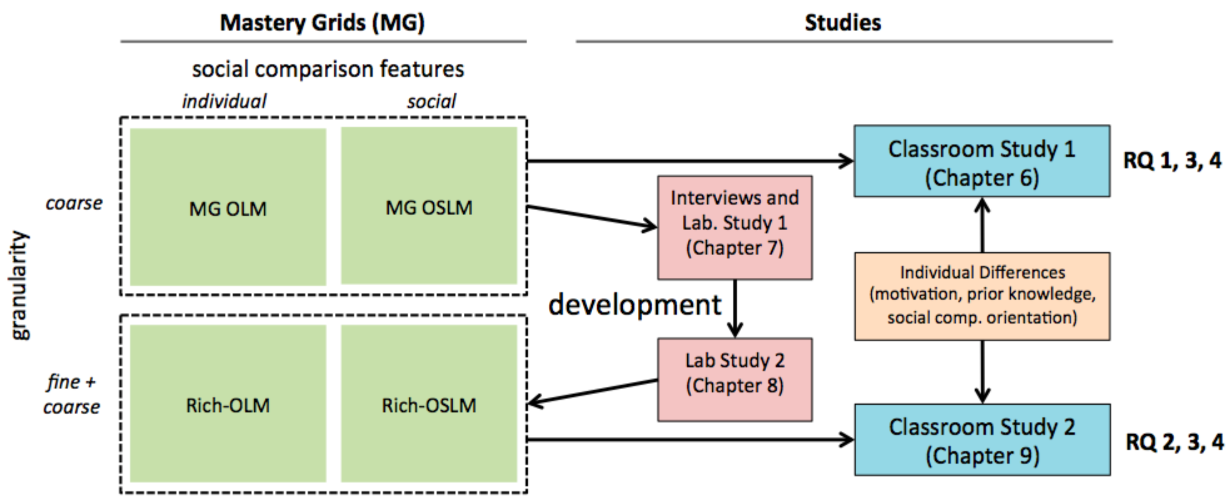


Figure 20: Diagram summarizing the studies of the chapters 6, 7, 8 and 9 and their contribution to Research Questions.

compared to an individual-view OLM?. Individual difference factors are added in order to answer research question 3, *How do individual differences influence system activity within an OLM?*, and research question 4 *How does the use of a learning system featuring OLM, OSLM and fine-grained elements affect motivation?*.

Chapters 7 and 8 report on the work and laboratory studies performed to support the design and development of the Rich-OLM. The exploration of the Rich-OLM, i.e., OLM and OSLM with an additional fine-grained component, is performed in chapter 9, in a second classroom study. Thus, this study focuses on research question 2 *What are the effects of fine-grained OLM on system activity?*, and adds individual difference factors that contribute to research questions 3 and 4.

5.3 COMMON MEASURES AND VARIABLES

5.3.1 Prior knowledge and learning

Prior knowledge and learning are measured using a *pretest* and *posttest*. Both tests consisted in the same set of 10 python programming small problems and are reproduced in APPENDIX A. The problems cover the concepts included by the content activities contained in Mastery Grids. The score of the tests is expressed as a number between 0 and 1. Also, normalized learning gain is reported in the studies. Normalized learning gain balances for differences on posttest and pretest depending on the pretest level. Equation 5.1 shows this measure. In the equation, $MaxScore_{posttest}$ is always 1.

$$LearnGain = \frac{Score_{posttest} - Score_{pretest}}{MaxScore_{posttest} - Score_{pretest}} \tag{5.1}$$

5.3.2 System activity: engagement, navigation and performance

I call *system activity* to a set of variables that involve different aspects of the interaction between the learner and the system, and that are extracted from the logs of both the Mastery Grids system and the user model. The variables include measures of *engagement* (amount of activity), indicators of patterns of *navigation* through the system, and *performance* measures on the self-assessment content activities such as questions and problems.

5.3.2.1 *Engagement* variables

Completion of activity (*mg_completion*) This measure is computed by dividing the number of distinct content activities attempted by the student by the number of different activities that exist in the course. The percentage of completion is computed considering distinct examples and animated examples viewed at least once, and the distinct self-assessment content activities that has been attempted successfully at least once. Repeated activity does not sum for the completion measure. In the Mastery Grids course for Java programming there are 254 activities (102 parameterized questions, 102 examples and 50 animated examples), and in the Python course there are 161 activities (37 parameterized questions, 32 parsons problems, 39 examples and 53 animated examples).

Attempts to questions and Parson problems (*n_questions*, *n_parsons*) These variables correspond to the raw count of attempts made by the student in parameterized questions and Parsons problems and do not consider the correctness of the attempts, nor do them discard repetitions. In case of Parsons problems, it does not include the movement of lines within the problem.

Views of examples and animated examples (*n_examples*, *n_ae*) These measures count the number of distinct examples and animated examples viewed by the student. The number of lines viewed in each example or animated example is not considered.

Regularity of activity in the term (*term_regularity*) To build a measure of how regular were the students in using the system through the term, I subdivide the term in N bins of 2 weeks each. Then I compute the proportion of activity done by each student in each of the bins and then compare this vector to the vector of perfect regularity, in which each bin has $1/N$ of the activity. To compare the vectors I use cosine similarity which is a popular measure to compute the similarity of two vectors by measuring the angle between them in a N-dimensional space.

5.3.2.2 *Navigation* variables

Probability of attempt (*prob_attempt*) Opening a content activity by clicking in the corresponding cell does not imply that the student attempted or even viewed the content. Students can click in activities and close the overlay window without doing it. To capture this phenomena I count the number of times an activity cell is clicked followed by a record of an attempt to this activity. The probability of attempting (or viewing) and activity that has been opened is then computed as shown in Equation (5.2).

$$prob\_attempt = \frac{\text{count act open and attempted}}{\text{count act opened}} \tag{5.2}$$

Ratio of time spent in the interface (*ratio_gui*) This variable represents the amount of time spent in navigating through or inspecting the interface relative to the total amount of time spent in the system. The measure is computed as follows. In the first pre-processing step applied to the data, the time of each action is computed by substracting the date and time of the action from the date and time of the next action. Since the system tracks every action in the interface and every content activity submission, these computed times are considered reliable enough. Then the amount of time in the interface is simply obtained by summing the times of all actions that are interface actions (clicking cells, mouseovers, etc). Extreme long times (greater than 30 minutes) were discarded. The ratio of time in the interface divides the total time spent in interface actions by the total time spent in the system (sum of all action times).

5.3.2.3 *Performance* in self-assessment variables

Instructional effectiveness (*eff_questions*, *eff_parsons*) Instructional effectiveness is a measure balancing the success on self-assessment activities and the time spent to reach the success. The measure is described by [Paas and Van Merriënboer, 1993]. To compute the effectiveness of parameterized questions, first the Z-score of the number of distinct solved items (Z_solved_q) and the Z-score of the total time spent in parameterized questions are computed (Z_time_q). Z-scores are computed subtracting the group mean and dividing by the standard deviation. Same is done for the number of distinct solved Parsons problems and the time spent in Parsons. Then the instructional effectiveness scores are computed as shown in Equation (5.3)

$$\begin{aligned} eff\_questions &= \frac{Z\_solved\_q - Z\_time\_q}{\sqrt{2}} \\ eff\_parsons &= \frac{Z\_solved\_p - Z\_time\_p}{\sqrt{2}} \end{aligned} \tag{5.3}$$

Success Rates (*sr_questions*, *sr_parsons*) Success rates are computed by dividing the number of correct attempts by the total number of attempts to assessment items. The success rate is computed for parameterized questions and for Parsons problems.

5.3.3 Learning Motivation

To measure learning motivation we used the Motivational Questionnaire (APPENDIX B) which join together sets of questions of two instruments: the Learning Activation questionnaire and the Achievement-Goal Orientation questionnaire. The Learning Activation questionnaire was developed to measure learning motivation in STEM activities and includes four motivational factors: Fascination, Competency Beliefs, Values, and Scientific Sense-making [Moore et al., 2011]. From this questionnaire I kept a core set of questions for the factors *Fascination* (4 questions), *Competency Beliefs* (5 questions), and *Values* (5 questions). I did not include questions about *Scientific Sense-making* because this factor corresponds to domain-specific skills. Since the original questions were designed for the subject of *science*, I modified these questions, maintaining the phrasing but changing the subject to *computer programming*. Items of the *Fascination* factor measure the extent to which the student like programming ("In general, I find programming. . . " with options "very boring","boring","interesting","very interesting"). *Competency Beliefs* questions ask students if they think they can deal positively with the subject ("I can figure out how to finish a programming class project at home", with answers in a 5-point scale from "I'm sure I CAN'T do it" to "I'm sure I CAN do it"). *Values* questions measure to which extent students think the programming subject is important for their lives and professional development ("I think programming will be useful for me in the future", with options "NO", "no", "yes", and "YES").

The Achievement-Goal questionnaire is a 12-question survey that measures Goal-Orientation, a fundamental motivational factor in Self-Regulated learning experiences [Elliot and Murayama, 2008]. Goal-Orientation is conformed of 4 factors that are not exclusive: *Mastery Approach* orientation is related to the motivation of mastering the learning content ("My goals is to learn as much as possible"); *Mastery Avoidance* is related to the avoidance of failing to learn ("My aim is to avoid learning less than I possibly could"); *Performance Approach* relates to motivation to perform, score, or doing better than others ("My goal is to perform better than the other students"); and *Performance Avoidance* is the orientation to be motivated to avoid failing, scoring under the minimum, or doing worst than others

("My aim is to avoid doing worse than others"). Each factor has 3 questions. Questions are measured in a 7-point scale with extremes labeled as "Not at all true of me" and "Very true of me", and a middle point "Unsure".

In the studies reported in the chapters 6 and 9, I mainly focus on exploring the role of Achievement-Goal factors and the Competency Beliefs, and opt to set aside the other two Learning Activation factors Fascination and Values. The reason of this is that the constructs of the Achievement-Goal Orientation framework and the Learning Activation factors are related and the nature of this relation positions the Achievement-Goal orientations as *closer* factors to explain the engagement in the use of a learning system. More precisely and from a theoretical perspective, there are motivational factors such as intrinsic fascination, domain specific values, and self-beliefs in abilities and skills which determine the goal that internally a student sets when facing a learning opportunity, and which has been framed by the Achievement-Goal theory in four orientations Mastery Approach, Performance Approach, Mastery Avoidance and Performance Avoidance [Elliot and Murayama, 2008, Grant and Dweck, 2003]. Additionally, I put special attention in Competency Beliefs because these represent a measure of prior knowledge, which can be contrasted to the pretest, which is an objective measure of prior knowledge.

Then, why to include Fascination and Values in the measurement? One reason is to validate the measurements of Competency Beliefs using factor analyses. This is because the Principal Component Analyses (PCA) applied to the whole instrument (the qhole questionnaire) should result in three components aligned to the three theoretical factors. Additionally, if results of PCA confirm what other researchers, the creators of the instrument, have found, then it will give validity to the measure. For example, researchers have found that Competency Beliefs share a component with Fascination. Another reason, perhaps more important, to measure Fascination and Values is that while these factors are *behind* the achievement goals in the theoretical structure of motivation that explains activity within the learning system, they are not necessarily distal when analyzed as outcomes of the process, i.e., when I look at the change of motivation (research question 4). Thus I will explore the change of Fascination and Values over the term, and not just Achievement-Goals and Competency Beliefs.

5.3.4 Social Comparison Orientation

To measure Social Comparison Orientation I use the INCOM questionnaire [Gibbons and Buunk, 1999], reproduced in APPENDIX C. Social Comparison Orientation is measured with 11 statements in a 5-point likert scale (*Strongly Disagree*, *Disagree*, *Neither Disagree nor Agree*, *Agree*, *Strongly Agree*). Example of the items are: "I often compare myself with others with respect to what I have accomplished in life", "If I want to learn more about something, I try to find out what others think about it", and "I always pay a lot of attention to how I do things compared with how others do things".

5.4 EDUCATIONAL CONTEXT OF THE CLASSROOM STUDIES

The classroom studies that I present in this thesis (chapters 6 and 9) were deployed in a particular educational context of a well known University in Finland. This section summarizes aspects of the educational context that are relevant to later ponder the findings.

Classroom studies were deployed in the course "*CSE-A1111 Basic Course in Programming Y1*" in Aalto University in Finland. Aalto University is considered one of the best technical universities in Finland, and the best choice for students of engineering. The course covers introductory level programming and receives students from different engineering programs, particularly of the School of Electrical Engineering and School of Engineering which includes the Departments of Civil Engineering, Mechanical Engineering and Built Environment.

In Aalto, students are set with recommended plan to take courses, but this plan is not mandatory, and there is no punishment for students that drop courses (at any time) or students who deviate from the course plan. Prerequisites between courses exist, but are rarely considered. In particular for the course in which the studies where deployed, this has no special pre-requisites and also, even when is a required course, it does not delay students if they want to drop it or take it later. All students are supported by the state scholarship (to cover living expenses; there is no tuition fees in Finland) which requires a minimum of 45

ETCS credits each year. This represents 3/4 of the recommended plan of courses (60 ETCS a year). Also, all courses, even if they are outside of the program curricula, count for this requirement.

The course that I focus in has 5 ETCS which correspond of approximately 133 hours in the term. The grade of the course is mainly computed from mandatory exercises, voluntary practice, and an exam. Voluntary practice is also considered for the final grade, but in a has a very small impact. Mandatory exercises are given in rounds during the term, and the requirement is that the learner obtain no less than 50% of the points in each round. The last round, corresponding to object-oriented exercises, does not have this constraint. All of this means that students are not required to solve all mandatory exercises to pass the course, and there is some flexibility in the assignments.

This configuration of (flexible) requirements is in line with the general educational culture and *mood* of Finnish students, which is summarized by the instructor of the course: *"Some students in our university are ambitious, but most are not. They think that if they apply for a job in industry, it does not matter, which grades they have and whether they have used a couple of extra years to have the degree. The most important thing (according to their opinion) is that they have the degree and what kind of work experience (from summer jobs and part time jobs during university years) they have. I suppose that their opinion is a little exaggerated, but I have heard that quite often the employer in industry does not even look at the grades."*

These considerations are of importance because they may have a strong influence in the motivational traits of the students, specially because the learning system explored in this thesis, Mastery Grids, is offered as a voluntary and complementary practice system in top of the regular (and flexible) course work that include other content material, including the mandatory exercise rounds.

6.0 CLASSROOM STUDY 1: SOCIAL COMPARISON FEATURES ON A COARSE-GRAINED OLM

In this chapter I explore the role of social comparison features in the coarse-grained Mastery Grids OLM. This chapter contributes to research question 1 *What are the effects of an OLM with social comparison features (or OSLM) on system activity compared to an individual-view OLM?*, research question 3 *How do individual differences influence system activity within an OLM?*, and research question 4 *How does the use of a learning system featuring OLM, OSLM and fine-grained elements affect motivation?*.

6.1 MOTIVATION

Previous studies conducted using Mastery Grids have already looked at the effects of a social comparison and have found that these features generally explain positive engagement and navigation within the practice system. Studies in a database course, where Mastery Grids was loaded with contents covering SQL programming, found a remarkable increase of activity in the group exposed to social comparison features [Brusilovsky et al., 2015]. Another study conducted in a Java course, found milder positive effects in the group with social comparison [Guerra et al., 2016]. Regarding navigation, previous studies have found that both OSM and OSLM improve the efficiency of navigation, i.e., making students move forward faster, and that OSLM encourages sequential navigational patterns.

I believe that social comparison features may have different effects on different students. Students can have different tendency to compare to others, and could have different motivational orientations when use the system. Students also have varying levels of competency or skills (based on previous knowledge and practice) and may also have different confidence in their skills. All of these factors may contribute to how they engage and navigate differently within the system.

To better study the effects of social comparison, including individual differences such as the ones described above, and to confirm or complement previous findings, I conducted a semester-long classroom study offering Mastery Grids in the course "*CSE-A1111 Basic Course in Programming Y1*" in Aalto University in Finland. This course covers procedural Python programming and includes students from a variety of non-Computer Science programs. They typically take it in their first fall semester of their Bachelor studies. The course enrollment is traditionally large (around 700 students) which meant it would be a good opportunity to study individual differences, such as motivational traits, with reasonable statistical power. Also, this course brought the opportunity to use Mastery Grids with a different audience than in previous studies, thus contributing toward generalizing the findings. The study was conducted with the support and help of the course instructor and a research assistant, who helped in technical issues regarding the deployment of the tools.

Students received the course grade based on mandatory exercises, voluntary practice content, and an exam. The exam contributed 50% of the final grade. The *exercise grade* contributed to the other 50%. It covered the mandatory content, divided into 9 rounds of exercises, where most of the grade (about 92%) was determined by small programming tasks. Voluntary practice, which was measured by the use of Mastery Grids, contributed with a 3% bonus on the exercise grade, which represented a bonus of 1.5% to the final course grade.

6.2 STUDY DESIGN

A version of Mastery Grids was prepared with Python content on 14 topics: Variables, Comparison, If Statement, Logical Operators, Loops, Output Formatting, Function, Lists, Strings, Dictionary, Values and References, Exceptions, File Handling, Classes and Objects. Four types of content were included: 37 parameterized problems, 32 Parsons problems, 39 animated examples, and 59 annotated examples.

The use of Mastery Grids was non-mandatory (optional) and complementary to the mandatory content exercises, which were accessed by students through a Learning Management System (LMS). To access Mastery Grids, a link was available for the students after logging into the LMS with their accounts. The link to Mastery Grids was personalized and mapped the student account in the mandatory exercise system to the Mastery Grids account. This approach had two benefits: students did not have to use a different account to access Mastery Grids, and all registered activity did not contain personal information. This mapping was performed within the exercise system and was implemented and managed by employees of Aalto University. An incentive of 3% of extra credit was added to the exercise grade (which contributed 50% of the course grade) was given to those who solved at least 15 problems in Mastery Grids. We offered such a bonus to encourage students to try the system.

6.2.1 Treatment groups

Regarding the version of Mastery Grids used, students were randomly assigned into 2 groups. We called these "treatment groups" to distinguish them from other groupings in further analyses. The first group, *Individual*, accessed a version of Mastery Grids where all social comparison features were deactivated, as shown in Figure 21. The *Social* group accessed a version of Mastery Grids with the social comparison features enabled, as shown in Figure 22. These features included the topic cell row of the aggregated progress of the group (blue row in Figure 22), the comparison row (middle row in Figure 22), and the peer comparison list that appears when the button *Loads the rest of learners* is clicked and which shows the position of the learner within her group, according to current progress of the learner through the course content. A more detailed explanation of these features can be found in chapter 3.

Furthermore, groups were subdivided into 8 subgroups (4 subgroups in each treatment group). This splitting had a technical reason: the social comparison features require the loading of a considerably large amount of data to present the Open Learner Model visualization of peer ranking, which can slow down the initial loading of the system when there are groups of several hundred students. Given that we expected between 600 and 700 students in the course, by dividing them into 8 subgroups, the system had to deal with less than 100 students each time the system was loaded.

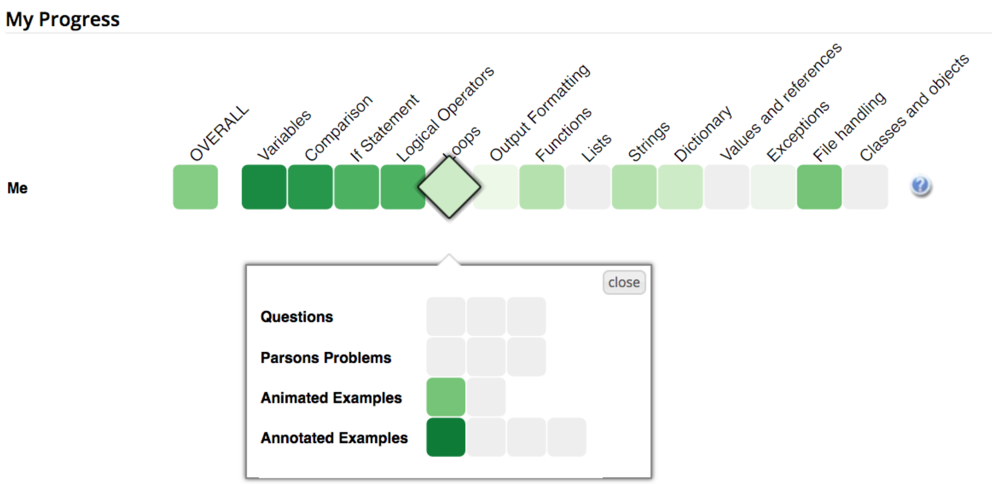


Figure 21: Mastery Grids version for Python programming, with minimal features, also called individual view (all social comparison features have been disabled).

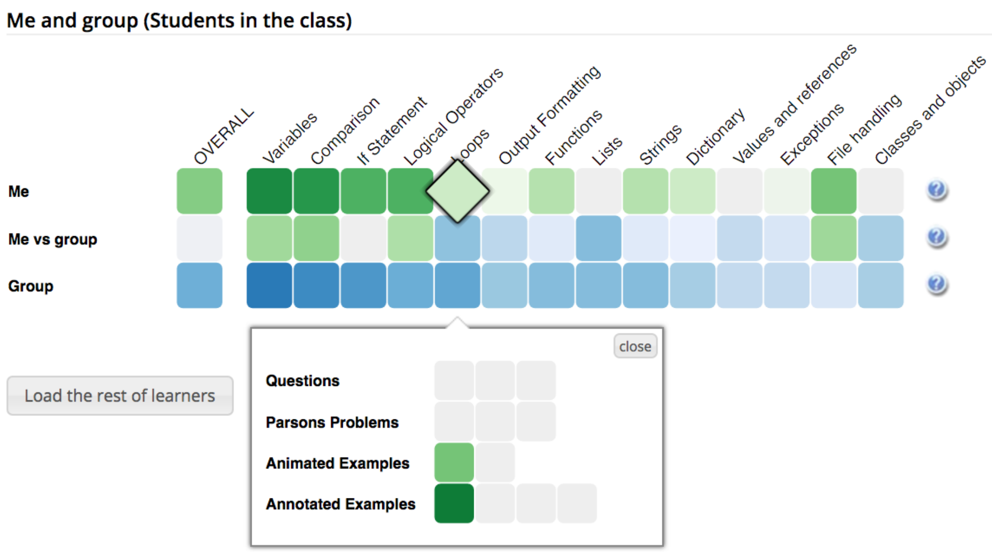


Figure 22: Mastery Grids version for Python programming, with social comparison features enabled.

6.2.2 Data collection

Pretest and posttest data were collected at the beginning and end of the term, respectively. Mastery Grids was enabled during the whole term, from the date that the pretest was taken, to the date of the posttest. Both the pretest and posttest were created as an online survey using the Qualtrics system (provided by the Katz School of Business at the University of Pittsburgh) and both included the same 10 questions about Python programming (see APPENDIX A).

Similar to the pre- and posttest, the Motivation Questionnaire (see APPENDIX B) was implemented as an online survey using Qualtrics and was given at the beginning and end of the term. In the questionnaire given at the end of the term, we also included the social comparison set of questions (see APPENDIX C).

System activity was measured with several variables related to the completion of activity, activity in different types of content, regularity of use, and performance on self-assessment content such as questions and Parsons problems. Details of these measures are described in section 5.3.2 of the previous chapter.

6.2.3 Approach followed in the analyses

The exploratory character of the research questions made a gradual approach suitable for these analyses. First, I present statistics about the data collected and perform reliability and factor analyses of the questionnaire answers.

Then, knowledge and learning (pretest and posttest) differences are contrasted between the treatment groups. Initially, I checked to see if both groups had differences in the pretest. Then an overall effect, using the posttest, was verified with the intention of confirming the positive effect of using the system in learning. Although this aspect is not a goal of this thesis work, it is an important element that needed to be checked.

After the knowledge differences analyses, I looked closer at the effect of prior knowledge, as objectively measured by the pretest and subjectively measured by Competency Beliefs, and its relationship to system activity. The analyses contributed to research question 1 *What are the effects of an OLM with social comparison features (or OSLM) on system* *activity compared to an individual-view OLM?* and research question 3 *How do individual differences influence system activity within an OLM?*. Then, other individual differences were included to contribute more insight into research question 3. At first, I looked at the Social Comparison Orientation, followed by Learning Motivation comprising the Achievement-Goal orientations. In these analyses, the exploration included two sets of regression models on system activity variables. The first set of analyses subdivided students by individual differences (e.g., low/high motivation) and built regression models in each group to see the *local* effects of treatment. The second set of regression analyses used all of the students and added interaction terms (e.g., treatment X motivation) to formalize the observations made in the first set of regressions.

Finally, research question 4 *How does the use of a learning system featuring OLM, OSLM and fine-grained elements affect motivation?* was addressed by analyzing the change of motivation during the term and its relationship with system activity and the treatment groups. Regressions on the motivation orientations measured at the end of the term were built to include the motivation orientations measured at the initial point in the term and other factors, such as the treatment group (*Individual*, *Social*) and pretest data.

6.3 DATA OVERVIEW AND PRE-PROCESSING

6.3.1 Data collected

A total of 697 students were assigned to Mastery Grids accounts. This represented the number of students who initially enrolled in the course. However, only 553 students (79%) finished the course by taking the final exam. In general, there were more males than females and the proportion reached 77% of males among students who provided gender information (N=636). Among students who finished the course, 324 students did at least some activity in Mastery Grids (active students). A relatively large proportion of students completed the pretest, posttest and motivation questionnaires, as shown in Table 3.

The average scores of performance variables among students who finished the course are shown in Table 4, which includes normalized learning gain. We observed similar performance values in both treatment groups.

Table 3: Number of students who completed the course (take exam), answer questionnaire at the initial (i) and final (f) term points, including the Social Comparison Orientation questionnaire, and number of students who did activity in Mastery Grids (active).

	Take exam	pre+post	Motiv initial (i)	Motiv final (f)	Motiv (i+f)	SC survey	Active	Active & exam
All	553	422	636	451	424	451	350	324
Individual	279	216	314	225	214	225	184	173
Social	274	206	322	226	210	226	166	151

Table 4: Summary statistics of performance measures

		All	Individual	Social
pretest	Mean	0.222	0.238	0.206
	SD	0.208	0.218	0.197
	SE	0.009	0.013	0.012
posttest	Mean	0.602	0.620	0.583
	SD	0.262	0.263	0.261
	SE	0.013	0.018	0.018
Norm. learning gain	Mean	0.466	0.469	0.463
	SD	0.427	0.458	0.395
	SE	0.021	0.032	0.028

In general, the levels of activity were considerable, but lower than other previous studies in which we have used Mastery Grids [Brusilovsky et al., 2015, Guerra et al., 2016, Loboda et al., 2014]. Table 5 shows the number of activities performed by each group. Table 6 shows some of the system activity variables (described in chapter 5) averaged across *active* students who completed the course (N=324, 59%). In general, in further analyses we considered only students who completed the course, because students who dropped out might have had very different reasons to disengage.

Table 5: Raw count of activity performed in the system

	Individual	Social	Total
Attempts to Questions	3814	4003	7817
Attempts to Parsons	4290	4575	8865
Examples viewed	3028	3062	6090
Animated Examples viewed	2155	2312	4467
Total Activity	**13287**	**13952**	**27239**

Table 6: General statistics of engagement variables

	Individual group			Social group		
	Mean	SD	SE	Mean	SD	SE
mg_completion	0.30	0.28	0.02	0.34	0.28	0.02
n_questions	21.39	20.32	1.54	24.11	21.07	1.72
n_parsons	24.16	32.66	2.48	28.09	35.35	2.88
n_examples	16.88	16.97	1.29	18.79	16.46	1.34
n_ae	11.98	12.05	0.92	14.26	12.26	1.00
sr_questions	0.45	0.21	0.02	0.48	0.18	0.02
sr_parsons	0.61	0.27	0.02	0.57	0.24	0.02

Distributions of the completion of activities for each of the groups are shown in a histogram in Figure 23. Clearly, these distributions are not close to normality. In the *Individual* treatment group, the distribution of completion of activity shows a non-linear decrease from a 0% completion to around 70% completion. Then there is a small spike for a few students with a very high level of completion. The *Social* group presents a different distribution of completion with the summit being at around 20% of completion instead of at 0%. The positive shift observed in the *Social* group suggests that effects on (*engagement*) may not be noticeable in overall activity, but rather within only certain sections of the activity distribution.

Figure 24 shows attempts at activity during the term. The y-axis represents the position at which the activity performed is located with respect to the order of the activities in Mastery Grids. A higher point means an activity is in a more advanced topic. All activity of all students in each treatment group is used in this chart. Vertical spikes of activity show patterns of single (or only a few) student(s) who completed many activities within one session within the system. Note the visible spike of activity at the end of the term in the *Individual* group. This activity corresponds to the last minute participation of 4 students within the

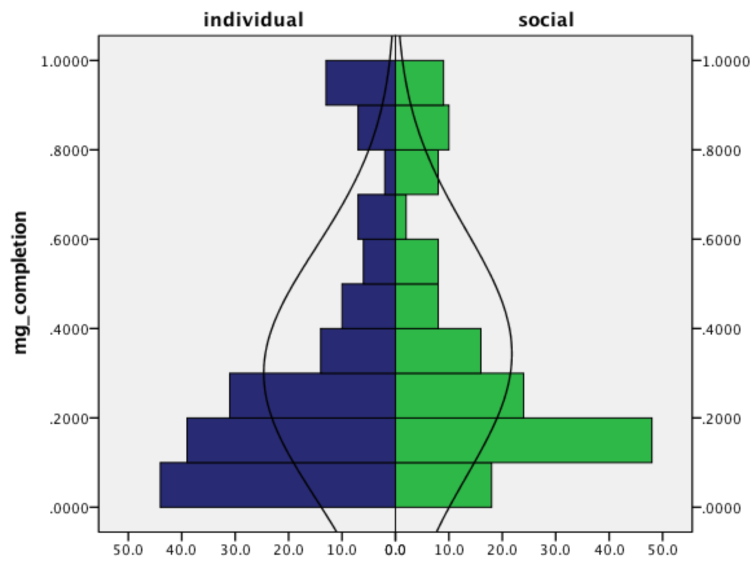


Figure 23: Histograms of mg_completion in both treatment groups

Individual group and does not represent a relevant bias. Figure 25 presents another view of the same data as a density plot, which visualizes the differences between the groups better. Note the moderate increases of activity in the *Social* group contrasted to the "valleys" in the *Individual* group. Also, note the higher concentration of activity near the end (before the exam), which is even higher in the *Individual* group. These observations suggest that differences in *engagement* may exist in the patterns of system activity during the term. Analyses to test these differences are performed later in this chapter.

6.3.2 Questionnaire Reliability and Factor Analyses

Before summarizing and using the measures of Learning Motivation and Social Comparison Orientation, I performed a reliability analysis to verify that the answers of all items within each factor measured were coherent.

The Social Comparison Orientation was measured using the INCOM instrument ([Gib-

Figure 24: Attempts during the terms by treatment group. Y-Axis position represents the position in the course order (higher are in later topics). Last spike in individual group is due to 4 students who did 2425 actions after the exam.

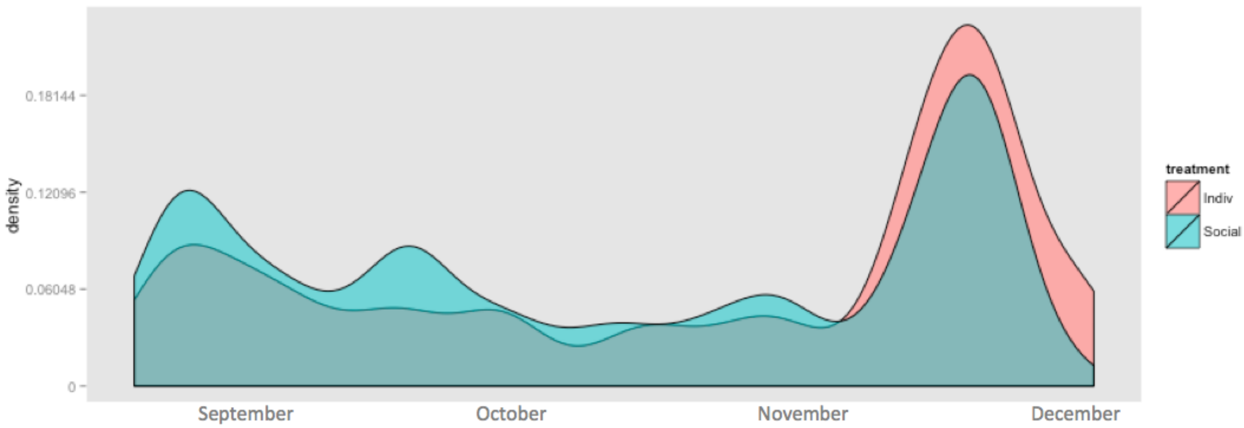


Figure 25: Density of system activity during the term by treatment group.

bons and Buunk, 1999], and see APPENDIX C), which has 11 statements (9 positive, 2 negative, in which the score was reversed) about the inclination to compare oneself to others. Reliability reaches a Cronbach's Alpha of .799 which is acceptable. Factor analyses failed to find the two theoretical orientations described in the literature as *ability* and *opinion* [Gibbons and Buunk, 1999]. Table 7 shows the *loadings* of the factor analyses. The column *Orientation* corresponds to the theoretical subfactor that was defined by [Gibbons and Buunk, 1999]. As can be seen in the table, data extracted factors are not clearly aligned to these theoretical factors. With this in mind, I opted to compute a unique scoring of Social Comparison Orientation by averaging the scores of all 11 statements (items 6 and 10 were reversed before computing the SC score).

Regarding the Learning Motivation questionnaire, reliability analyses showed good scores for all factors except Mastery Avoidance (MAv) which had a reliability score (Cronbach's Alpha) below 0.7. Table 8 shows the scores of all measures, including the Learning Activation factors: Fascination (F), Competency Beliefs (CB), Values (V); and the Achievement-Goal factors: Mastery Approach (MAp), Mastery Avoidance (MAv), Performance Approach (PAp), and Performance Avoidance (PAv).

A Principal Component Analysis (PCA) was performed with an extraction based on Eigenvalue and Varimax rotation with Kaiser Normalization to corroborate that the different groups of questions within the instrument were measuring different factors. Regarding the section of Learning Activation, PCA extracted 3 factors which explained the 61.9% and 66.2% variance in the initial and final questionnaires, respectively. The *loadings* (associations between each question and the latent extracted factors) that can be seen in Table 9 matched the designed factors Fascination, Competency Beliefs and Values. This means that, according to the answers, we can distinguish the 3 theoretically defined motivational factors. As explained before, we will only include the Competency Beliefs factor in further analyses. However, the factor analysis is important because it corroborates that the construct is distinguishable from other motivational traits, such as Fascination and Values.

In the Achievement-Goal Orientation section, the PCA extracted only 3 factors, because Performance Approach and Performance Avoidance loaded together. This suggests that students of this cohort did not distinguish between Performance-Approach and Performance-Avoidance items. Note also that consistently, in both initial and final measures, the first item of the Mastery Avoidance factor loaded more strongly when loaded with the component with the Mastery Approach items. Factors extracted explained the 72.2% and the 73.4% variance in the initial and final questionnaires, respectively. Table 10 shows the results of the PCA analysis. Loadings lower than 0.3 have been removed to facilitate the interpretation of the table.

Table 7: Results of the Factor Analyses on Social Comparison Orientation Questionnaire. Rotated matrices show the loadings greater than .3.

		Component		
Question	Orientation	1	2	3
1	ability	0.673		
2	opinions	0.572		
3	ability	0.733		
4	ability	0.58		0.408
5	opinions	0.424	0.579	
7	ability	0.473	0.338	
8	opinions		0.745	
9	opinions		0.822	
11	ability	0.673		
6 (R)	ability	-0.626		0.447
10 (R)	opinions			0.823

Table 8: Reliability analyses of the Motivation questionnaire taken at the beginning of the term (initial) and at the end of the term (final).

	Cronbach's Alpha						
	F	CB	V	MAp	MAv	PAp	PAv
Initial	.805	.840	.820	.744	.652	.900	.901
Final	.868	.827	.842	.810	.682	.905	.886

Table 9: Results of the Factor Analyses on the Learning Activation section of the Motivation Questionnaire. Rotated matrices show the *loadings* greater than 0.3.

	Component				Component		
	1	2	3		1	2	3
F1i		0.739		F1f	0.718	0.403	
F2i		0.738		F2f	0.744	0.396	
F3i		0.743		F3f	0.711		
F4i		0.719		F4f	0.76		
CB1i	0.766			CB1f	0.392	0.695	
CB2i	0.818			CB2f		0.779	
CB3i	0.832			CB3f		0.781	
CB4i	0.806			CB4f		0.757	
CB5i	0.547			CB5f		0.586	
V1i		0.424	0.595	V1f	0.7		0.308
V3i			0.744	V3f			0.833
V4i		0.429	0.696	V4f	0.579		0.602
V5i		0.408	0.705	V5f	0.67		0.496
V6i			0.782	V6f			0.81

Table 10: Results of the Factor Analyses on the Achievement-Goal section of the Motivation Questionnaire. Rotated matrices show the *loadings* greater than 0.3.

	Component				Component		
	1	2	3		1	2	3
MAp1i		0.776		MAp1f		0.765	
MAp2i		0.724		MAp2f		0.788	
MAp3i		0.821		MAp3f		0.826	
MAv1i		0.65		MAv1f		0.734	
MAv2i			0.874	MAv2f			0.864
MAv3i			0.827	MAv3f		0.312	0.764
PAp1i	0.831			PAp1f	0.823	0.312	
PAp2i	0.879			PAp2f	0.86		
PAp3i	0.87			PAp3f	0.875		
PAv1i	0.835			PAv1f	0.841		
PAv2i	0.844			PAv2f	0.814		0.346
PAv3i	0.872			PAv3f	0.854		

Based on these results we decided to discard Mastery Avoidance and Performance Avoidance from further analyses.

6.3.3 Statistics of questionnaire

Results of the Learning Motivation and Social Comparison Orientation questionnaires are shown in Table 11. Scores have been computed by averaging items of the questionnaire and by moving them to the range of 0 - 1, using a simple linear transformation. Histograms of the motivation factors are shown for the initial measures in Figures 26,27,28. A histogram of the Social Comparison Orientation is shown in Figure 29.

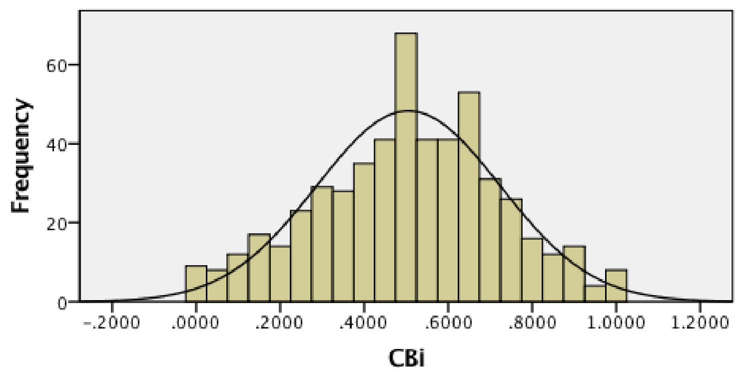


Figure 26: Histogram of Competency Beliefs measured at the beginning of the term.

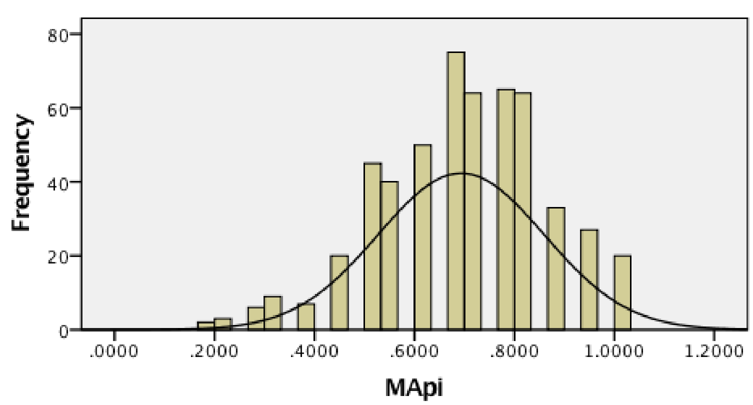


Figure 27: Histogram of Mastery Approach measured at the beginning of the term.

Table 11: Basic statistics of motivational factors.

		Initial	Final
Competency	Mean	0.505	0.656
Beliefs	SD	0.219	0.197
	SE	0.010	0.009
Mastery	Mean	0.693	0.641
Approach	SD	0.167	0.200
	SE	0.007	0.010
Performance	Mean	0.582	0.568
Approach	SD	0.229	0.226
	SE	0.010	0.011
Social	Mean	-	0.589
Comparison	SD	-	0.128
Orientation	SE	-	0.006

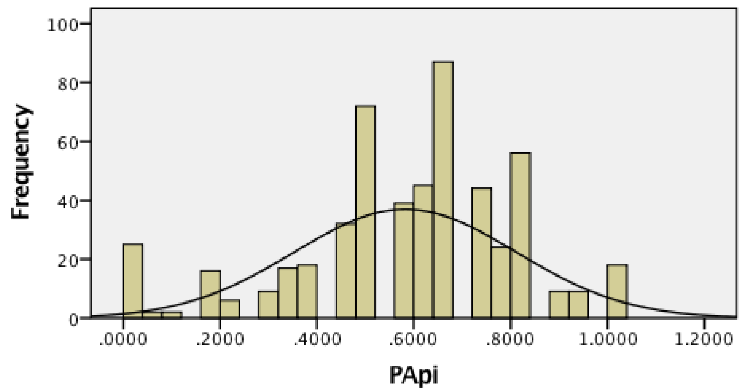


Figure 28: Histogram of Performance Approach measured at the beginning of the term.

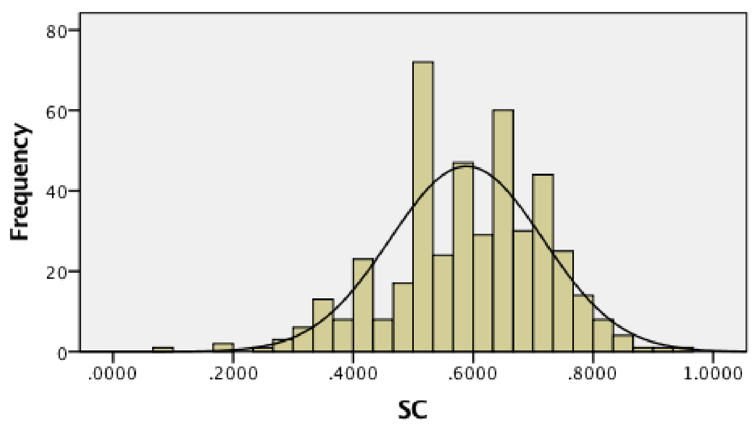


Figure 29: Histogram of Social Comparison Orientation.

The Social Comparison Orientation scores showed a relatively small Standard Deviation, which means its power of discrimination is at risk. However, these values are similar to the statistics obtained by [Gibbons and Buunk, 1999], which after being scaled to the range 0-1, are Mean=.61, SD=.154 and Mean=.65, SD=.145 in the studies reported.

6.4 PRIOR AND POST KNOWLEDGE DIFFERENCES

I started by looking at differences in pretest and posttest scores between the two treatment groups, *Individual* and *Social*. These analyses have a double purpose. On the one hand, they serve as a validation that the treatment groups are no different in their pre-condition (pretest). On the other hand, differences in the posttest might reveal positive (or negative) effects from using the system. Even when I did not expect to see great differences, it was necessary to perform such verifications.

As shown in Table 4, the pretest, posttest and learning gain (normalized) scores were similar in both *Individual* and *Social* groups. Some differences were observed in the pretest and posttest, where the *Individual* group had slightly higher scores (pretest=.238, posttest=.620) than the *Social* group (pretest=.206, posttest=.583). However, considering that those students who finished the course (took the exam) and did at least one activity in Mastery Grids (active students), the non-parametric Mann-Whitney did not find any significant difference between treatment groups, whether on the pretest, posttest, or in a learning gain. Table 12 shows the results of these tests.

Table 12: Non-parametric test on performance measures between treatment groups.

	Pretest	Posttest	Learning Gain
Mann-Whitney U	11567	8716	8215
Z	-1.079	-0.859	-0.336
Asymp. Sig. (2-tailed)	0.28	0.39	0.737

Since the posttest was not independent of the pretest variable, a better analysis of differences on this dependent variable (DV) between treatment groups can be done using regression models that consider pretest. I performed a linear regression on the posttest with the pretest and treatment group as predictors. The treatment group variable was set as a dummy variable *social*, taking the value 1 for the *Social* group and 0 for the *Individual* group [1]. The first model was run incorporating only the pretest as a predictor and results showed it was a significant model ($R2 = .144$, $p < .001$, $B_{pretest} = .421$). A second model added the factor *social* and results showed no effect for this variable to explain the variation of the DV posttest ($R2change = .001$, $B_{social} = -.011$, $p = .690$).

These small or non-existent differences are expected. Giving that groups were split randomly, we expected no differences on the pretest. Regarding the posttest, no observed difference is not necessarily a negative result. Mastery Grids was used as a complementary practice system, and students already had a mandatory exercise system as their primary source of practice. Also, I understand that the potential effect is due to practice with the system, and not directly because of the interface design. In other words, if the social comparison features have an effect, it will be noticed in activity within the system (engagement, navigation), and then indirectly this will translate to an increase or lack of increase in learning.

With this in mind, I next looked into the relationship between the level of engagement, measured as the completion of activities (*mg_completion*), and the *posttest*. We used linear regression with *posttest* as a dependent variable, and the *pretest* and *mg_completion* as independent variables. Once again, I built two models. The first model incorporated the *pretest*, and the second model added the variable *mg_completion*. The results, detailed in Table 13, show a significant prediction model where both predictors are significant. The pretest is the stronger predictor ($\beta = .386$), followed by mg_completion ($\beta = .185$). Completing the content in Mastery Grids predicts almost 20% of an increase in the posttest, after controlling for the pretest.

Table 13 also shows the results of similar regressions run separately on both groups. It is interesting to notice how the effects are stronger in the *Social* group. To test the combined contribution of pretest and social features of the interface, I performed a regression analysis incorporating the interaction *pretest*social*. The analysis built 3 consecutive models. In

[1]Note that the word *social* is capitalized to refer to the treatment group, and not capitalized to refer to the dummy factor used in regression analyses

Table 13: Summary of regression on Posttest with predictors Pretest and mg_completion.

	Model 1 R^2	Model 2 R^2	pretest β	mg_completion β
Overall	.144	.188	.386	.185
Indiv	.113	.152	.359	.172
Social	.181	.232	.407	.204

all models, the dependent variable was *posttest*. Model 1 included only *pretest* as predictor. Model 2 added the dummy variable *social*. Model 3 added the interaction term *pretest*social*. The results are shown in Table 14. Model 1, shows the clear contribution of the pretest, which decreases marginally when Models 2 and 3 add the factor *social* and the interaction term. Neither Model 2 nor Model 3 improved prediction of posttest scores. Extending this analyses by adding *mg_completion* showed a significant contribution by this variable (β = .172, p = .011) but no significant contribution from the interaction term *mg_completion*social* (β = .032, p = .745).

Table 14: Regression analysis on posttest. Interaction *pretest*social* is not significant predictor of posttest.

	R^2	$Sig.FChange$	$B_{pretest}$	B_{social}	$B_{pretest*social}$
Model 1	.144	.000	.421 (.000)	-	-
Model 2	.144	.69	.420 (.000)	-.011 (.690)	-
Model 3	.146	.406	.371 (.000)	-.36 (.382)	.106 (.406)

Overall, results showed that there were no important differences between the treatment groups on either the pretest or posttest. Also, the level of activity in the system (mg_completion) contributed to explain the final performance, taking into consideration the strong predictive power of the pretest, and that the interface (OSM, OSLM) does not change this relation. These results supported the idea that increasing activity within the practice system is beneficial.

6.5 THE IMPACT OF PRIOR KNOWLEDGE IN STUDENT ENGAGEMENT

In the previous section, I showed how completion of activities has a positive effect on performance, as measured by the posttest, even after considering the strong predictive power of the pretest. In this section, I turn to explore the role of prior experience, as objectively measured with the *pretest* and subjectively measured with the *Competency Beliefs* factor. Although the effects of pretest and other individual differences on system activity, i.e., *engagement*, *navigation* and *performance* are later explored in this chapter (targeting Research Question 3), I focus first on the role of prior experience as related to engagement in the treatment groups. As it will be shown in this section, prior experience, measured by pretest, has a strong influence on the usage of the system and cannot be set aside when exploring other factors. This section contributes to research question 3.1: *How does prior knowledge influence system activity within an OLM?*.

Simple correlations between *pretest* and *mg_completion* show a significant positive relation in the *Social* group ($Pearson$ = .257, p = .002), but not in the *Individual* group ($Pearson = .030, p = .695$). Interestingly, *Competency Beliefs* do not seem to have the same strong relationship with *mg_completion* that pretest does, neither for the initial measure (Person = .093, p=.262), nor for the final measure (Person = .126, p=.152).

To confirm these observations, I ran a multiple regression on *mg_completion* with the predictors *pretest* and *Competency Beliefs*, separated by treatment group. I used *stepwise* regression to compare the variables. The analysis was performed using *Competency Beliefs*, measured separately at the initial and final points (CBi, CBf). Results are shown in Table 15. Only the model built for the *Social* group was significant ($R^2 = .064$, $p = .009$), and only *pretest* was a significant predictor ($\beta = .321$, $p = .004$). A scatterplot of the *mg_completion* and pretest scores (Figure 30, left side) shows a positive correlation between engagement and pretest –both are increasing in the *Social* group. Contrasting with this, a similar scatterplot with Competency Beliefs (CB) in Figure 30, center and right side, shows a decreasing slope (negative correlation) within the *Social* group line.

Table 15: Regressions on *mg_completion* with predictors pretest and Competency Beliefs measured at initial and final points (CBi and CBf).

	Model	Included predictor	Excluded predictors	
	R^2 (*sig F change*)	$B_{pretest}$ (p)	B_{CBi} (p)	B_{CBf} (p)
Individual	-	-	-	-
Social	.077 (.002)	.339 (.002)	.006 (.951)	.023 (.809)

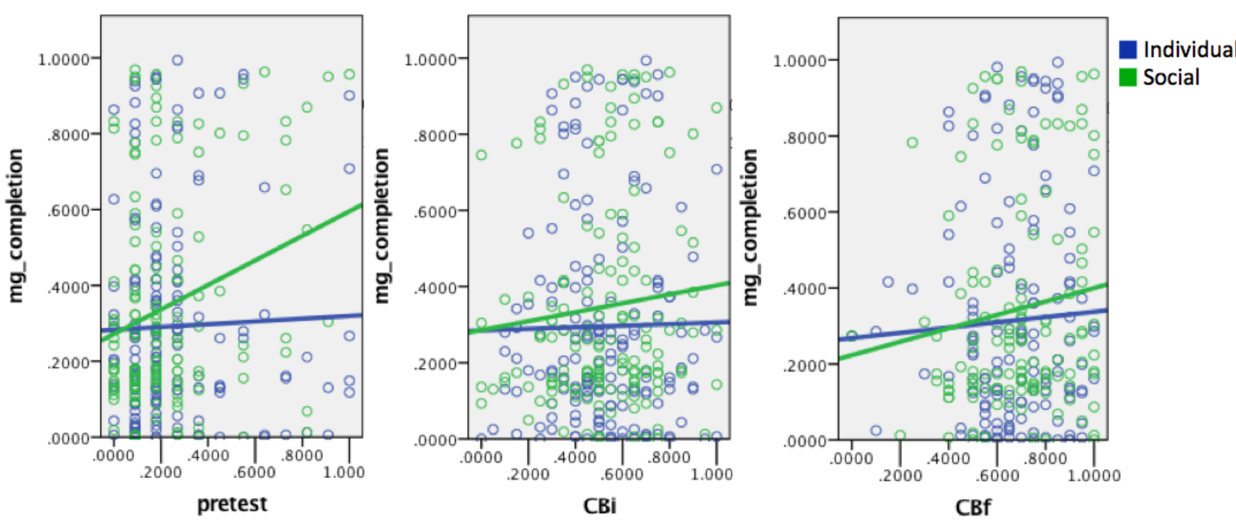


Figure 30: Scatterplots of mg_completion and pretest, and mg_completion and Competency Belief at initial (CBi) and final points (CBf).

More differences were apparent between the *Individual* and *Social* groups, based on the relationship between the *pretest* and system activity in each group, as shown in Figure 31. Here, students have been classified in 3 bins depending on their pretest score and the percentiles 33.3 and 66.7: Low (pretest $\leq .09$), Medium ($.09 <$ pretest $\leq .18$), and High (pretest $> .18$). Error bars in the figure represent 2 SE (standard errors of the mean). A general trend of higher activity was observed in the *Social* group for all levels of pretest, which does not show an interaction effect. In a second attempt, I divided students into 4 levels of pretest using the percentiles 25, 50, and 75: Low (pretest $\leq .09$), Medium Low ($.09 <$ pretest $\leq .18$), Medium High ($.18 <$ pretest $\leq .27$), and High (pretest $> .27$). Note that because of the discrete nature of the pretest score (it contains only 10 questions graded correct/incorrect, which makes the score a discrete scale), this grouping maintained the Low and Medium group and subdivided the High group. The Figure 39 presents this division and the level of engagement in each group, showing the most important result in the very high group: the greatest difference between *Individual* and *Social* is in the higher pretest group.

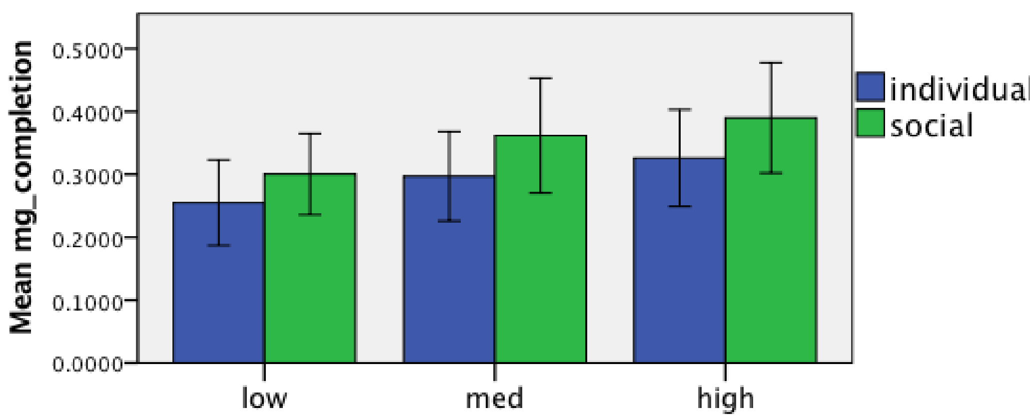


Figure 31: Mean mg_completion across different levels of pretest and between *Individual* and *Social* group.

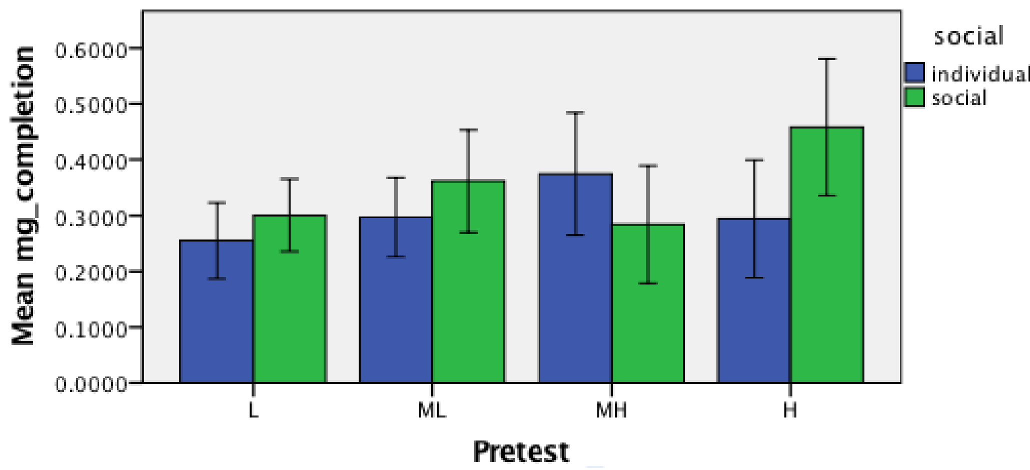


Figure 32: Mean mg_completion across different levels of pretest and between *Individual* and *Social* group.

Overall, I observed that since *pretest* and not *Competency Beliefs* has a positive correlation with the activity in the system, the results suggest that it is what students actually know and not what they believe they know, that explains their increasing engagement with the system. The results also confirm similar observations made by other studies in which previous experience and not competency beliefs explain students' persistence in activities in a MOOC [Higashi et al., 2017].

Another finding of these analyses is that the role of previous experience is stronger when the social features are present in the system. In other words, it seems that the social features produce the highest engagement in students with higher prior knowledge. This selective effect of the social comparison features upon pretest values may be due to several reasons. One reason is that higher pretest students probably find it easier to complete activities in the practice system since they are already familiar with some of the contents), thus they might start using the system earlier and with less difficulty than their lower pretest peers do. As a result, when these early-to-engage students notice that they are progressing ahead of the rest of the class, if they are exposed to the social comparison features (in the *Social* group), they may realize that they have gained a higher status and want to keep it. This hypothesis is explored in the next section. Another plausible reason, not exclusive to the first, is that higher pretest students also have different levels of motivation, which can encourage them more when the social features are present. This idea is explored in later sections, when I add motivational factors into the analyses.

Interestingly, the selective effect of *social* and pretest is a finding that has helped to explain the contrasting effects of social features on performance (success rates), which was hypothesized in previous studies: we observed that social comparison pressures the students to move forward to new content faster (decreasing their success rates), but on the other hand, social comparison features also tends to engage higher pretest students, who have higher success rates to begin with.

6.6 HIGH PRETEST STUDENTS ARE EARLY STUDENTS

In previous sections, I observed a positive effect on engagement in the *Social* group conditioned to the level of pretest. One possible explanation is that since they were stronger students to begin with (high pretest), they are better able to start and advance through the activities in the system with less effort, so that the students in the *Social* group may tend to continue activities because they see that they are ahead of the rest of the group, especially if they start using the system early in the term. To explore this hypothesis, I looked at the students who started the system earlier in the term and noted any engagement differences between the treatment groups. Activity throughout the term was split into 7 bins of 2 weeks each. Then, the number of activities was counted for each student in each bin. Students who had activity in the first 3 bins were marked as *early* students. Students who started only after the 3rd bin (7th week) were labeled as *late*. Table 16 shows the number of *early* and *late* students in both groups and the mean of completion of activity and pretest. It is clear and not surprising that *early* students have greater levels of completion. It is interesting to note that *early* students also have higher pretest scores than *late* students, although there is no real difference between the average pretest scores of *early* students in the *Individual* and *Social* group. This is evidence that having a higher pretest score influences students to

Table 16: Number and means of *mg_completion* and pretest of students in the *Social* and *Individual* groups who entered the system early and late.

		students (N)	mg_completion (mean)	pretest (mean)
Early	Individual	85	0.336	0.267
Early	Social	79	0.390	0.274
Late	Individual	78	0.291	0.225
Late	Social	69	0.299	0.179

start the system earlier. The key evidence from the table is that among *early* students, the *Social* group has more activity (M=.390) than the *Individual* group (M=.336), although the pretest is not really different. The same effect is not observed for *late* students in the *Social* group. This evidence supports the hypothesis that starting early in the system is the result of having both a higher level pretest and the presence of social comparison features.

To test the strength of this effect, I built regression models on *mg_completion* within the *Social* group, with the predictors *pretest*, *early* (dummy variable with value 1 for students having early activity), and interaction term *pretest*early*. Regression models were built using a stepwise forward and backward method. Results consistently found that the strongest predictor (and the only one entering or remaining in the model) is the interaction *pretest*early* ($\beta = .390$, $p < .001$). The nature of this interaction can be seen in Figure 33.

6.7 EFFECTS ON SYSTEM ACTIVITY

Now I will present analyses that show the effects of social comparison features on system activity. As explained before, system activity includes variables that measure *engagement* with completing content activities, *navigational* patterns through the system, and *performance*

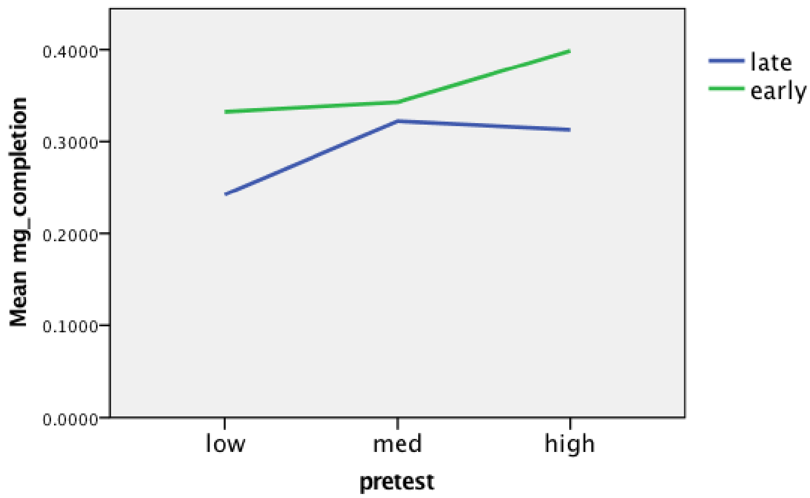


Figure 33: Mean mg_completion across levels of pretest and between *early* and *late* students.

in the self-assessment content items. Several variables were then analyzed and a description of each of these variables can be found in chapter 5, Section 5.3.2. Table 17 presents the mean and standard deviation of the system activity variables for the *Individual* and *Social* groups. The following analyses targeted **RQ 1**: *What are the effects of an OLM with social comparison features (or OSLM) on system activity compared to an individual-view OLM?*, and focuses in testing the hypothesis **H1** *Students exposed to an OLM with social comparison features increase the level of activity in the system.*

In general, the *Social* group showed a consistent but small positive difference in the engagement variables (completion, number of attempts), and a shift of sign for the effectiveness scores (negative for *Individual* and positive for *Social*. Moreover, a density plot of activity, shown in Figure 25, reproduced here as Figure 34, suggests that differences might be in engagement throughout the term. Also, the distributions of levels of activity, shown in Figure 23 and also reproduced here as Figure 35, suggest that differences between treatment groups might be more subtle, affecting a specific region of the engagement distribution. The following analyses seek to formalize and complement these observations.

Table 17: System activity by treatment.

	Individual		Social	
	Mean	SD	Mean	SD
mg_completion	0.300	0.280	0.344	0.277
n_questions	21.393	20.315	24.113	21.075
n_parsons	24.156	32.664	28.093	35.348
n_examples	16.879	16.966	18.795	16.465
n_ae	11.983	12.054	14.265	12.257
term_regularity	0.436	0.083	0.432	0.084
eff_questions	-0.025	0.449	0.028	0.386
eff_parsons	-0.058	0.906	0.066	0.538
sr_questions	0.448	0.207	0.483	0.183
sr_parsons	0.615	0.273	0.567	0.242
prob_attempt	0.420	0.244	0.408	0.206
ratio_gui	0.862	0.273	0.912	0.211

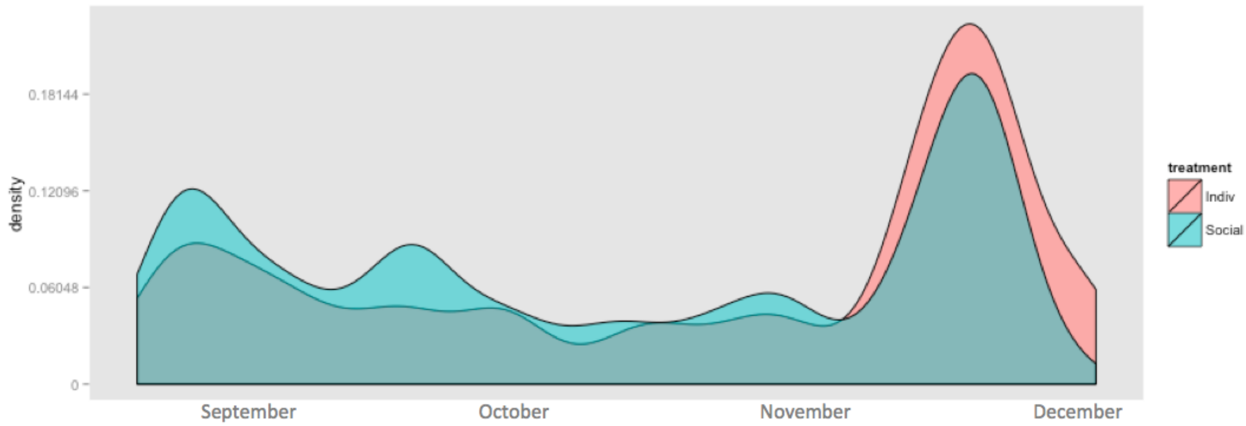


Figure 34: Density of system activity during the term by treatment group.

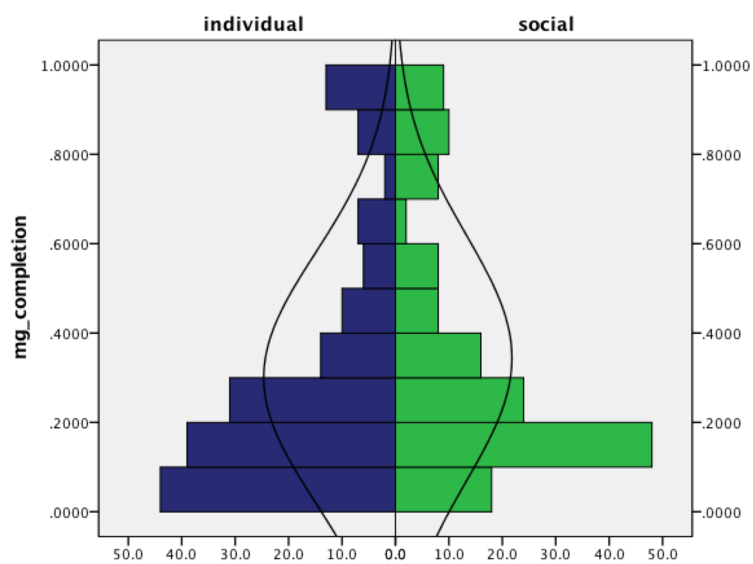


Figure 35: Histograms of mg_completion in both treatment groups.

6.7.1 Regression analyses

The analyses were performed by building linear regression models on each of the system activity variables, treating them as dependent variables with the predictors being *pretest* score, the variable *social* (dummy representing the treatment group), and the interaction term *pretest*social*. Since an external incentive was offered to perform at least 15 problems in the system, I repeated these analyses for students who went beyond this incentive threshold and accumulated more than 15 activities. Table 18 shows the β values for the predictors and the significance is marked in each of the cells. The analyses of these results are broken into effects on *engagement*, *navigation* and *performance*.

6.7.2 Effects on *engagement*

The first 6 rows of Table 18 show the *engagement* variables. When looking at the results of the analyses that include all of the students (the first 3 columns of the results in Table 18, we can notice a positive effect of the interaction *pretest*social* on the completion of activities, except for the parsons problems. Significant positive interaction terms mean that students in the *Social* group do more activity, but this is also affected by their pretest level. This confirms the selective effect of the social features in regards to the pretest level, which was observed in the previous sections. Interestingly, the majority of the effects observed for the interaction term become weaker and significance disappears when only the students who did more than 15 activities are considered (see the incentive threshold, the last 3 columns in the Table 18), although the coefficients show the same relationship (coefficient signs).

The results also show that the social comparison features have a positive effect on the regularity of system activity during the term (*term_regularity*). The overall effect can be seen in Figure 34 where a density plot shows deeper *valleys* of activity in the *individual* group during the term. Regression models show that this regularity effect, although reduced after the incentive threshold, is still visible for the more engaged students. Regressions also showed that the effect is conditioned by the pretest. This observation complements previous analyses where I found that the selective effect of pretest and social comparison features is explained as a result of the early activity of high pretest students. Putting the observations together, high pretest students enter the system early, and in the *Social* group, they tend to maintain their advantage status, thus they must also keep active during the term.

A general view of the regularity in the engagement through the term can be tested by comparing the levels of activity before and after the exam preparation time. Figure 34

Table 18: Results of regressions on engagement variables (rows) in two cases: all the students (columns 2-4), and for students who has engaged beyond the 15 activities incentive threshold (columns 5-7). Values are raw coefficients. Significance is marked with the cell background color and with symbols '.' (.1-.05), '*' (.05-.01), '**' (.01-.001), '***' (<.001).

	Active students			More than 15 act.		
	pretest	social	pre*social	pretest	social	pre*social
mg_completion	.036	-.011	.29*	.11	-.03	.21
n_questions	-4.70	-1.16	19.28 .	-5.48	-.87	11.37
n_parsons	5.67	3.37	9.71	9.16	4.79	-.27
n_examples	1.64	-1.39	16.72*	4.48	-3.76	16.09
n_ae	1.06	.22	10.72 .	5.37	-.99	8.36
term_regularity	-.03	-.03*	.12**	-.03	-.03	.11 .
eff_questions	.37**	.013	.26	.62*	.03	.21
eff_parsons	.06	.01	.59	.09	.07	.51
sr_questions	.19**	.06 .	-.08	.18**	.02	-.03
sr_parsons	.21*	-.02	-.10	.20 .	.05	-.16
prob_attempt	.006	-.012	.009	-.014	.042	-.110
ratio_gui	-.19*	-.02	.28*	-.17 .	-.04	.27 *

shows a spike at the end of the term that corresponds to activity just before the final exam. The pattern of activity between treatment groups seems to shift in this figure, from more activity in the *Social* group in the early to the middle of the term, to more activity in the *individual* group just before the exam. To test this difference, I focused on the amount of activity that occurs before the spike of exam preparation. The bar chart on Figure 36 shows a considerable difference between treatment groups on amount of activity (counted as the number of attempts to do content activities) that occurs before the spike of the exam (1 week before the exam). A non-parametric Mann-Whitney Test showed a significant difference, Mann-Whitney $U = 5493.5$, $p = .012$. Among students who were active in the practice system before the exam, students in the *Social* group practiced with the content more.

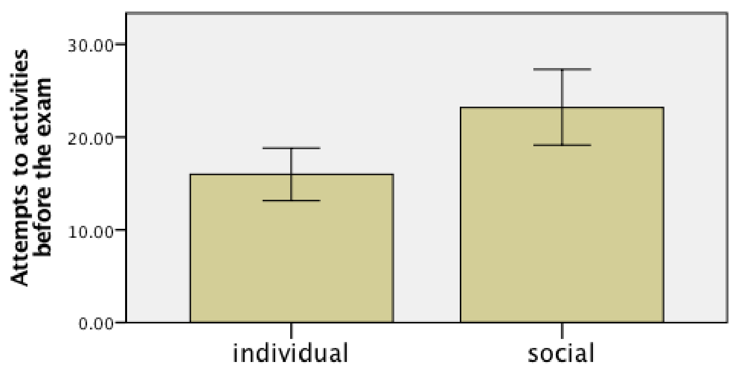


Figure 36: Average number of attempts to content activities from the beginning of the term until 1 week before the exam. Error bars represent two standard error of the mean.

Another view of general engagement across treatment groups can be seen in the distribution of engagement levels. Distributions of the completion of the activities for each of the groups are shown in a histogram in Figure 35. A difference can be seen in the lower level of activity, which is consistent with the results of Table 18, showing little or no noticeable effect beyond the 15-activity engagement threshold. The partial difference in the lower level of activity is explained in Figure 37, where the number of students of different levels of engagement relative to the number of active students in each group, are plotted for each treatment group. Differences between levels of engagement are very small and only noticeable between the 15+ and 30+ sections (having more than 15 and 30 activities, respectively). This figure contrasts with Figure 16 (b), presented in Section 4.2, which showed remarkable differences in how *Social* maintained higher levels of engagement in previous studies [Brusilovsky et al., 2016].

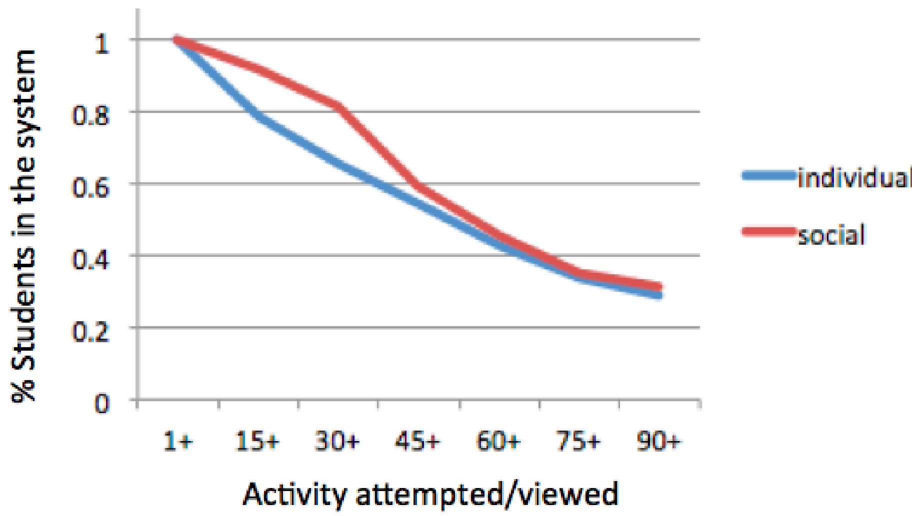


Figure 37: Number of students by level of activity (number of attempts to content items.)

6.7.3 Effects on *performance*

Regarding performance in self-assessment items (rows 7-10 in Table 18), results show, as expected, that pretest is a clear positive predictor of success rates and effectiveness scores, which confirms the intuitive idea that higher pretest students have higher success rates. Putting this finding together with the previously observed selective effect of the social comparison features and pretest (previous sections) brings a more solid explanation of why performance seems not to be affected by the social comparison features. While *social* makes students move forward more quickly, thus lowering the success rates because they avoid to repeat attempts to known content; on the other hand, *social* also contributes more through the engagement of higher pretest students, who also tend to have higher success rates. The marginally significant positive effect of *social* in the success rate of questions (*sr_questions*) indicates that is not just the pretest which explains the higher success rates, but also that *social* may have a positive effect by itself. We have hypothesized in the past that this might be due to the sequential navigational patterns that the *social* features induce, making students advance progressively through the content.

6.7.4 Effects on *navigation*

As observed before, *social* has a positive effect in success rates that might be explained by the navigational support that the social comparison features convey and the recently observed positive interaction of *pretest*social*. While *social* engages the higher pretest students more, who tend to succeed more easily, *social* also makes students move forward quickly. Because success rates are computed as the rate of successful attempts divided by the total number of attempts, they do not tell if students are repeating successful attempts or moving forward faster. To check this, I computed a "strict" success rate by dividing the distinct questions solved (no repetition) by the number of attempts to questions. Regressions run on this variable produced almost the same results as the effects observed for *sr_questions* (marginally positive significant effect of *social*), which strengthens the observation that in this study, *social* has a positive effect on moving students forward.

Another effect on navigation was observed. The interaction term *pretest*social* is significant and positive for the ratio of time spent interacting with the interface (*ratio_gui*), suggesting that for high pretest students, the social features make them spend more time in navigating the interface relative to the total time spent in the system. However, the probability of attempting activities that are open is not different between the treatment groups, suggesting that the extra time spent in the interface is not due to a misguided effect. At the same time, students in the *Social* group accomplish more activities, which supports the idea that extra navigation in this group translates to more activity.

Altogether, these results show that when students are exposed to social comparison features they do more, confirming hypotheses 1: *Students exposed to an OLM with social comparison features increase the level of activity in the system.* Also, students exposed to social comparison features in the OLM spent relatively more time interacting with the interface (navigating), and thus they have better success rates on questions, even if they tend to move forward quickly and avoid overstaying in already solved activities. However, a concern exists: since students in the *Social* group who “jump” into the system at some point in the term will observe that others have completed previous topics (already covered topics), it is possible that they feel motivated to complete past activities, which means doing *easy* activities just to “get the cells green” in the interface. If this is true, we should be able to observe more activity for the *Social* group in the “lower” topics when compared to the group without social features. To observe this phenomena, Figure 38 shows the density of activity performed in each treatment group, where the x-axis represent the position of the activity in the course. Position of the course indicates where the activity is located in the order of all activities organized in Mastery Grids. Activity on the left is related to activity in the first topics. Clearly, from the figure, both groups present the same pattern of activity, meaning that there is no trend in the *Social* group to complete early topic activities, at least judging by overall activity.

Another concern exists regarding possible outliers. Distributions of the completion of activities presented earlier in this chapter (Figure 25) show a small number of students who completed all the content. To discard the distortion of these outliers in the previous analyses, I repeated them discarding the students with more than 90% of completion ($N_{Individual} = 13$, $N_{Social} = 9$). However, the same pattern of results is observed as in Table 18, indicating that these high level activity students do not introduce much distortion into the analyses.

6.8 THE ROLE OF SOCIAL COMPARISON ORIENTATION

So far, analyses have shown the important role of previous experience in regards to system activity when the interface contains social comparison features. I now explore another

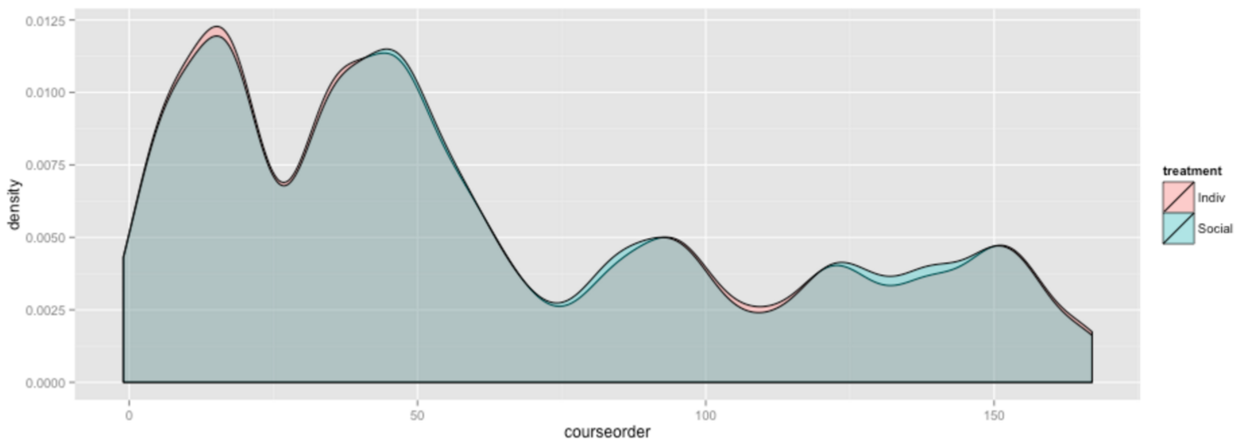


Figure 38: Density chart of the activity of each treatment group. X-Axis represent the order in the course of the activities (right side are activities in advanced topics). Blue shade correspond to the group with social features enabled (*Social*) and red shade is the group with the *Individual* view (*indiv*).

variable which is theoretically relevant: the Social Comparison Orientation of the students, targeting research question 3.3 *How does Social Comparison Orientation influence system activity within an OLM?* It is natural to assume that students who declare they have a tendency to compare themselves to others may be more sensitive to the social comparison features in Mastery Grids.

To analyze the role of the Social Comparison Orientation (SCO), I split the students along the median of Social Comparison Orientation scores in low and high groups. For simplicity, well call these the *Low SCO* group and the *High SCO* group. The mean and standard deviation of the system activity variables in the *low* and *high SCO* groups are shown in Table 19. I noticed a consistent lower level of activity in the *high SCO* group.

To analyze the relationship of the *SCO* to the interface features, I built regressions models in each *SCO* group for each engagement variable, with predictors *pretest*, the dummy variable *social*, and the interaction term *pretest*social*. Table 20 shows the β coefficient of each predictor in these regression models. Significance is marked with symbols '.' ($p < .1$),

Table 19: System activity in low and high Social Comparison Orientation groups.

	Low SCO		High SCO	
	Mean	SD	Mean	SD
mg_completion	0.354	0.320	0.302	0.243
n_questions	25.807	23.916	20.244	17.366
n_parsons	31.829	43.131	21.970	24.563
n_examples	18.971	18.672	17.393	15.486
n_ae	14.457	13.611	12.400	11.056
term_regularity	0.429	0.075	0.443	0.091
eff_questions	-0.004	0.413	0.020	0.463
eff_parsons	0.045	0.773	-0.040	0.833
sr_questions	0.444	0.203	0.493	0.184
sr_parsons	0.570	0.264	0.615	0.258
prob_attempt	0.418	0.226	0.407	0.226
ratio_gui	0.890	0.246	0.882	0.244

'*' ($p < .05$), '**' ($p < .01$) and '***' ($p < .001$). A second round of analyses was performed on the whole group (without splitting by SCO), by adding the predictors *SCO* (raw score) and the interaction *SCO*social*. These analyses were performed to confirm the significance of observations, as shown in Table 20, in which I include the significant interaction values in the last column of the table.

Table 20 shows only a few effects of the Social Comparison Orientation in the system variables, mainly in the *engagement* aspect (rows 1-6 in the table) and in the *high SCO* group. An interesting observation is that, although not particularly significant, a trend is observed regarding the role of pretest: pretest presents consistently higher coefficients in the *low SCO* group. Worth noting is the positive effect of the interaction between *pretest*social* in the *low SCO* group and effectiveness on the parsons problems (*eff_parsons*), which suggests a counter-intuitive phenomena: a higher the pretest in the *Social* group results in being more efficient in solving parsons problems for students who don't have the tendency to compare themselves to others. Being more efficient means that they required a smaller number of attempts and spent less time, on average, to solve the parsons problems. It does not mean that they solved more problems. It might happen that these students actually did less problems of this type, as the row *n_parsons* seems to show ($\beta_{pre*social}$ = -20.748), although this effect is not significant. Another key observation along this line of thinking is that the only significance obtained in the second round of analyses for the interaction *SCO*social* is for the variable *sr_parsons*. This interaction complements the earlier observation and shows that the SCO has a negative effect on the success rate of parsons problems in the *Social* group (i.e., higher success rates are obtained by low SCO students).

Another observation from the table is in regards to the regularity measure (*term_regularity*). Here the effect of *social* and the interaction *pre*social* is concentrated in the *High SCO* group. Although the effect of *social* is negative, it is also weak compared to the positive effect of the interaction. This points out that the *social* features have a positive effect on making students more regular in their use of the system along the semester, but that this effect is correlated to higher levels of pretest and works better for students who tend to compare themselves to others (*high SCO*).

Although this last result makes sense in terms of the theoretical positive relationship

Table 20: Coefficient values of regressions on engagement variables for students with Low and High Social Comparison Orientation. Significance is marked with the cell background color and with symbols '.' (.1-.05), '*' (.05-.01), '**' (.01-.001), '***' (<.001).

	Low SCO			High SCO			Interaction
	pretest	social	pre*social	pretest	social	pre*social	SCO*social
mg_completion	.168	.028	.196	.015	-.069	.289	
n_questions	6.891	4.051	5.932	-6.468	**-7.989 .**	**23.707 .**	
n_parsons	39.868	8.676	-20.748	7.926	2.367	1.951	
n_examples	3.728	-0.276	16.619	1.371	-3.921	16.124	
n_ae	4.658	2.001	8.706	.235	-2.992	10.777	
term_regularity	-.050	-.002	.086	-.043	**-.044 .**	**.207 ****	
eff_questions	.158	.02	.537	.38	-.022	.247	
eff_parsons	-.711	-.114	**1.503 ***	.241	.07	.363	
sr_questions	.276	.072	-.185	.099	-.004	.086	
sr_parsons	.389	.022	-.228	.130	-.047	-.076	**-.532 ***
prob_attempt	-.107	-.046	.174	-.060	.001	.011	
ratio_gui	.083	.016	.025	**-.180 .**	.014	.271	

between SCO and the presence of social comparison features, the findings are surprising because I expected the Social Comparison Orientation to have a stronger discrimination power on engagement when *social* features are present. One observation is that the deviation of the SCO scores is relatively low (M=0.589, SD=0.128, see section 6.3.3) which means that it may not be possible to discriminate between the different orientations. To address this, I performed a second subdivision on SCO, splitting it into 3 groups of equal size, and then discarded the middle group from the analysis. However, the same patterns of results were observed.

6.9 THE ROLE OF LEARNING MOTIVATION

Continuing to explore individual differences in system activity, I now look at motivational traits, targeting research question 3.2 *How does learning motivation influence system activity within an OLM?* My interest is in looking at how students with different Achievement-Goals orientations engage with the system when the social features are present and not present. To analyze this, I first split the students into motivational groups, classifying them in *low* or *high* groups, based on the central value. Looking at the distribution of the motivational variables (see Figures 27 and 28) and to generate balanced groups, I decided to use the median. Table 21 shows means and standard deviations of the system activity variables in the low and high groups for both splits. The Mastery Approach is labeled as *MAp*, and the Performance Approach is labeled as *PAp*. Notice the higher engagement rates for the high Mastery Approach group. In general, high Mastery oriented students have higher rates of completing the system.

To explore the relationship between motivation and the interface features (*social*), regression models were built for each of the motivational groups, where the engagement variables were predicted by *pretest*, the dummy variable *social* and the interaction term *pretest*social*. Considering the reliability and factor analyses performed earlier, I decided to keep only the *Mastery Approach* and *Performance Approach* factors. Table 22 and Table 23 show the results of these regressions, concerning the Mastery Approach and Performance Approach orientations, respectively.

6.9.1 Mastery Approach Orientation

Regarding *engagement* levels, there is an observable effect in the *High MAp* group, with a marginally significant positive effect of the interaction *pretest*social* on the general completion of activities (*mg_completion*) and the number of questions attempted (*n_questions*). However, the engagement through the term (*term_regularity*) shows a significant effect of the *social* and the interaction *pretest*social* in only the *low MAp* group. It seems that only the higher pretest, low motivated students in the *Social* group became more regular.

A general positive effect of pretest is also observed in *High MAp* for the effectiveness and success rate measures of questions and is marginally significant for the success rate of parsons. The effect of the *pretest* on the effectiveness of the parsons appears in the interaction term *pretest*social*, and it is only marginally significant.

Regarding *navigation* patterns, a negative effect of the pretest on the ratio of time spent in the interface (*ratio_gui*) is compensated by the positive effect of the interaction *pretest*social*, suggesting that while high Mastery Approach students spend little time in the interface, since their pretest is higher; in contrast, they also tend to spend more time in the interface if they are in the Social group, which makes sense considering that the social features add information to the system interface.

A second round of analyses were performed for the whole group (no splitting by MAp) and by adding the MAp raw score and the interaction *MAp*social* to verify the significance of the differences observed. However, none of the regressions on engagement measures showed any significant effect of the interaction term. This means that regardless of local effects observed in low and high Mastery Approach, this factor does not significantly cause differences in system activity between the *Individual* and *Social* groups.

6.9.2 Performance Approach Orientation

Similar to the Mastery Approach, Table 23 shows the results of regressions on system activity variables that were performed separately for low and high Performance Approach students

Table 21: System activity in low and high Mastery Approach (MAp) and Performance Approach (PAp) groups.

	Low MAp		High MAp		Low PAp		High PAp	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
mg_completion	0.292	0.272	0.356	0.278	0.322	0.278	0.318	0.275
n_questions	21.073	20.799	24.794	20.294	24.174	22.190	21.231	18.961
n_parsons	23.670	33.445	28.199	31.941	27.445	36.321	23.863	29.059
n_examples	16.056	16.741	20.066	16.266	17.645	17.053	17.925	16.264
n_ae	12.313	11.863	14.022	12.433	13.387	12.237	12.725	12.040
term_regularity	0.428	0.084	0.442	0.082	0.435	0.082	0.433	0.085
eff_questions	-0.042	0.416	0.056	0.418	-0.073	0.446	0.071	0.380
eff_parsons	-0.082	0.876	0.097	0.564	-0.020	0.823	0.010	0.698
sr_questions	0.432	0.210	0.502	0.173	0.441	0.195	0.484	0.198
sr_parsons	0.590	0.257	0.591	0.262	0.536	0.243	0.640	0.263
prob_attempt	0.416	0.240	0.409	0.208	0.402	0.224	0.425	0.229
ratio_gui	0.879	0.251	0.899	0.236	0.880	0.247	0.894	0.243

Table 22: Regressions on engagement with the system with predictors pretest, social and interaction term pretest*social, for low and high Mastery Approach oriented students. Significance is marked with the cell background color and with symbols '.' (.1-.05), '*' (.05-.01), '**' (.01-.001), '***' (<.001).

	Low Mastery Approach			High Mastery Approach		
	pretest	social	pretest*social	pretest	social	pretest*social
mg_completion	.098	.008	.117	.054	-.008	**.345 .**
n_questions	6.196	1.714	.342	-6.936	-2.156	**27.501 .**
n_parsons	25.599	8.396	-22.566	1.598	1.439	23.947
n_examples	3.339	-.817	10.681	2.903	-.835	17.618
n_ae	-1.030	.401	7.756	4.010	.458	12.106
term_regularity	-.060	**-.045 ***	**.157 ***	-.031	-.015	.095
eff_questions	.263	-.021	.355	**.453 ***	.061	.134
eff_parsons	-.459	.017	.776	.225	-.071	**.728 .**
sr_questions	.155	.050	-.105	**.178 ***	.063	-.029
sr_parsons	.259	-.034	-.052	**.232 .**	.006	-.218
prob_attempt	.332	.017	-.244	-.057	-.001	.008
ratio_gui	-.040	-.005	.124	**-.306 ****	-.008	**.344 ***

Table 23: Regressions on engagement with the system with predictors pretest, social and interaction term pretest*social, for low and high Performance Approach oriented students. Significance is marked with the cell background color and with symbols '.' (.1-.05), '*' (.05-.01), '**' (.01-.001), '***' (<.001).

	Low Performance Approach			High Performance Approach			Interaction
	pretest	social	pre*social	pretest	social	pre*social	PAp*social
mg_completion	.000	-.022	.190	.179	.048	.187	**.351 ***
n_questions	-2.811	1.337	8.731	1.823	-.788	18.261	**25.960 ***
n_parsons	10.984	8.493	-14.061	13.553	2.843	12.159	**31.988 .**
n_examples	-3.755	-3.391	14.708	11.430	3.287	8.449	**21.866 ***
n_ae	-2.739	-.139	8.425	8.087	2.481	6.217	**12.130 .**
term_regularity	-.040	**-.036 .**	.045	-.016	-.013	**.124 ***	
eff_questions	.075	-.04	.067	**.583 *****	.111	.126	
eff_parsons	-.173	-.042	.733	.317	.110	.343	
sr_questions	.138	.024	-.136	**.233 ****	**.106 ***	-.116	**.203 ***
sr_parsons	.259	-.071	-.008	.160	.005	-.172	
prob_attempt	.168	.022	-.053	-.065	-.065	.060	
ratio_gui	-.225	-.045	.308	**-.223 ***	.006	**.270 .**	

(*low PAp*, *high PAp*). The last column in the table shows a significant coefficient of the interaction term *PAp*social* from a second round of regression analyses where the whole cohort is considered (no splitting). As we can observe in the table, although there is no significance in completion of activity, whether they were *low PAp* , or high PAp students, the interaction term (last column on the table) shows a consistent positive significant effect on the amount of activity (but only marginally for the parsons and animated examples): *social* has a positive effect on performing more activities, but is constrained by the level of Performance Approach orientation. This effect is expected because of the performance orientation nature of the social comparison features. This effect is visualized in Figure 39. Here, students have been classified into 4 bins, depending on their Performance Approach (PApi) score and the percentile 25, 50, and 75: Low (PApi $\leq .5$), Medium Low ($.5 <$ PApi $\leq .61$), Medium High ($.61 <$ PApi $\leq .72$), and High (PApi $> .72$). Error bars in the figure represent 2 SE (standard errors of the mean). Note that the *Low* (L) group has students with the Performance Approach below the middle point (.5). A trend can be observed of higher activity in the *Social* group compared to the individual group for all levels above the Low Performance Approach group. In the individual group, the levels of activity are not much different across levels of Performance orientation. On the contrary, in the *Social* group, levels of activity increase for higher performance oriented students. Although differences are not very high, the chart shows that it is not that low Performance Approach students do less in the *Individual* group, but that higher performance oriented students do more in *Social* group.

Regularity of system activity during the term also showed an interesting effect when looking at the *low* and *high PAp* groups: while social negatively contributes to regularity for *low PAp* students, *high PAp* students become more regular when exposed to social features, depending on their pretest level.

Other effects are shown in the Table 23, mainly in the *high PAp* group. Regarding *performance*, a strong positive effect of pretest is observed in the effectiveness score and the success rate of questions (*eff_questions* and *sr_questions*), and this last measure also shows a significant interaction with *PAp*social*, which suggests that social features contribute to better success rates, depending on the level of Performance Orientation of the students. A

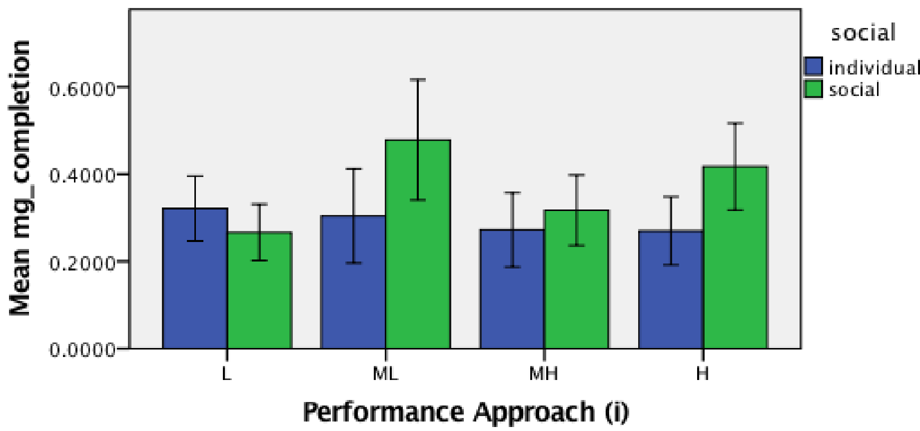


Figure 39: Mean mg_completion across different levels of Performance Approach (initial measure) and between *Individual* and *Social* group.

concern exists with this observation: it might be that students with a high performance orientation who are exposed to the social comparison features tend to do easy activities just to complete their model, compared with the rest of the class. However, a density chart of activity separated for *low* and *high PAp* groups in Figure 40 shows that in the *high PAp* group (bottom chart), the *Social* group actually does more advanced activity than the *Individual* group.

Regarding *navigation* patterns, the ratio of time spent in the interface also show a different effect between Performance Approach groups, with *social* being a positive predictor conditioned by the level of pretest (*pre*social*).

6.9.3 Do motivation orientations explain the selective effect of pretest in *Social* group?

I observed in previous sections a positive effect on engagement within the *Social* group, as conditioned by the level of pretest. I also observed that this effect is reasonably explained by the fact that high pretest students start using the system earlier. However, it is possible that the motivations of these students also contributes to explaining the effect, at least in part. To test this, I performed two series of regressions in the *Social* group alone, with the system

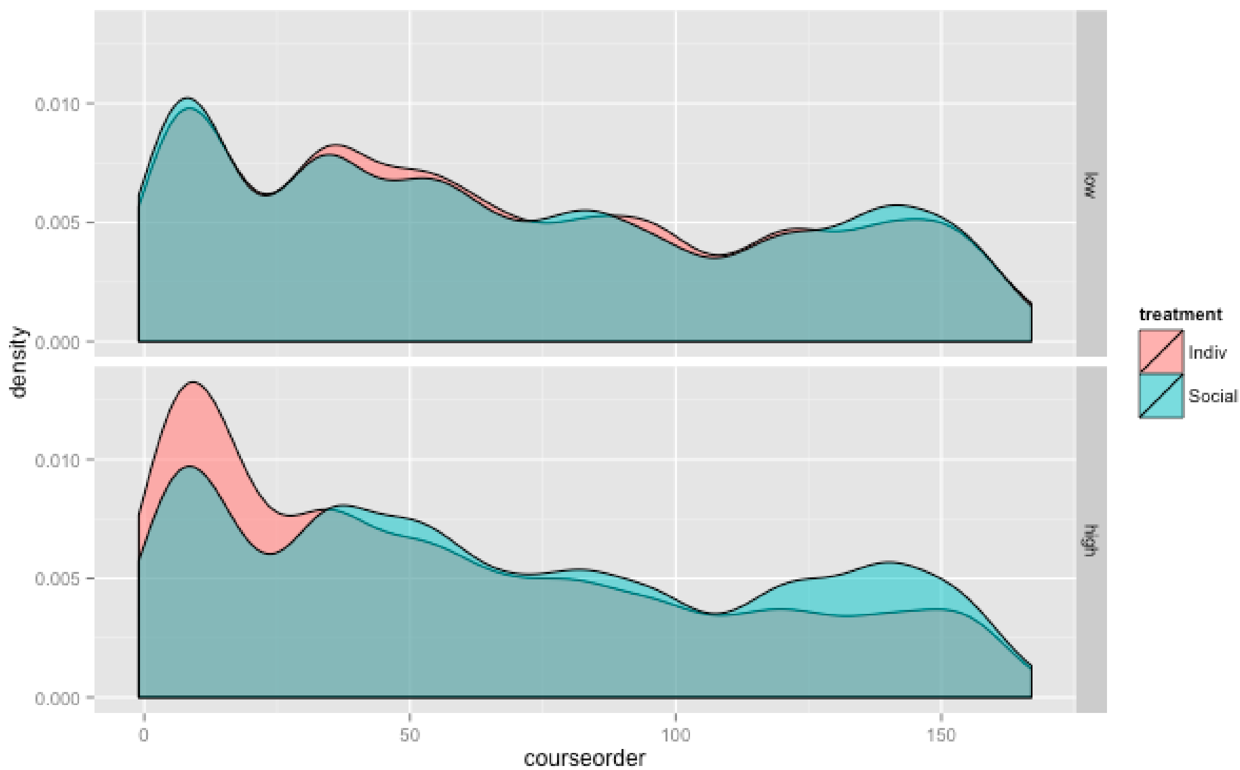


Figure 40: Density of system activity by PAp high/low groups and Social/Individual groups. The x-axis correspond to the position of the content in the course. To the left is activity that is in early topics, to the right is the activity on more advanced topics.

activity variables as dependent variables. In the first series, the predictors were *pretest*, *Mastery Approach* (MAp score, initial measure) and the interaction term *pretest*MAp*. In the second series of regressions, the predictors were *pretest*, *Performance Approach* (PAp score, initial measure) and the interaction *pretest*PAp*. None of the regressions analyses that included Mastery Orientation found a significant contribution from the interaction term on any engagement variable. However, significant contributions from the interaction term were found in the second series of regressions. The interaction *pretest*PAp* was significant and positive for the number of examples viewed ($\beta_{pre*PAp} = 59.999, p < .05$), marginally significant for the number of animated examples viewed ($\beta_{pre*PAp} = 42.008, p < .1$), and significant for the effectiveness score of questions ($\beta_{pre*PAp} = 1.553, p < .05$). This last effect complements the previous finding in which we observed a strong positive effect of pretest on effectiveness of questions in the High Performance Approach group in Table 23.

Overall, the results show a partial influence of the Performance Approach orientation on the selective effect of pretest in the *Social* group. Although the effect of Performance orientation and pretest is not strong enough to be observable in the overall completion of activities, it is observable for the raw number of activities in examples. The interpretation is that the role of the pretest in the *Social* group is also determined by the Performance orientation level, which is theoretically related to the presence of social comparison features.

6.10 THE CHANGE OF MOTIVATION

Now, I turn my attention to motivation as a dependent variable to address research question 4 *How does the use of a learning system featuring OLM, OSLM and fine-grained elements affect motivation?* In this section I focus in the relationships between using an OLM and OSLM and change in motivation. Motivation is known to change [Elliot and Murayama, 2008, Moore et al., 2011], and it is expected that motivation will evolve as the semester progresses. Recall that motivational factors were measured at the beginning and at the end of the term. This does not allow me to see the whole pattern of motivation evolution, but allows me to look for differences in the overall change in motivation. As happens with learning outcomes, I understand that there could be many factors which influence a change in motivation including the formal course experience and other content resources and activities performed by students. Thus, I don't expect to see a clear, nor strong influence by our system, which was by voluntary access and complements the other mandatory exercise system, in changing motivation. However, I am interested in seeing the relationship between engagement with the system and patterns of motivational change. The results presented below do not aim to establish causal relationships, but associations contributing to enrich the understanding of how the system is used. A note about the motivational measures which were considered in the following analyses: while in the previous sections I argue that the Achievement-Goal factors were relevant to conduct the analyses (in those analyses motivation was an independent variable and achievement goal, i.e., Mastery and Performance orientation were theoretically closer to explaining engagement with the system); while in this section I include all the motivational factors measured, because I am interested in seeing any change in motivational orientations, no matter what their theoretical relationships or structure are. These measures are: Fascination (F), Competency Beliefs (CB), Values (V), Mastery Approach (MAp), and Performance Approach (PAp). To simplify notation, the different measures receive the suffix 'i' or 'f' to refer to the initial or final measure, respectively. For example, *CBf* stands for Competency Beliefs measured at the end of the term.

Although analyses cover all motivational factors, one explicit hypothesis were stated in chapter 5 regarding change in motivation and the use of the system with social comparison features: **H6** *The active use of OLM with social comparison features will increase the performance orientation of the students.*

A series of paired sample t-tests, whose results are presented in Table 24, show that differences in Fascination, Competency Beliefs and Mastery Approach Orientation are significant, while Values and Performance Approach Orientation does not vary enough to be significant. Its interesting that while Fascination and Competency Beliefs increased, Mastery Approach decreased. Note also that even in cases of significant difference, the mean difference is not very big. Fascination has a mean difference of close to 2%, Mastery Approach about 5%. And Competency Beliefs showed the greatest difference, with a mean of 17%. These analyses included all students: those who used and those who did not used the system. The pattern

Table 24: Paired Samples t-tests for motivation measured at the beginning (i) and at the end (f) of the term.

DV	Mean	SE	t	p
Fi - Ff	-.018	.008	-2.312	.021
CBi - CBf	-.165	.011	-15.497	<.001
Vi - Vf	.003	.008	.427	.670
MApi - MApf	.049	.009	5.388	<.001
PApi - PApf	.012	.010	1.157	.248

of results is the same when repeating these analyses, whether considering only those who used the system or only those who did not use the system with only one difference: among the students who did not use the system, Fascination does not change (t=.696, p=.487).

To explore the relationship between change in motivation and the use of the system with and without social comparison features, I performed a series of regression analyses. A regression model was built for each motivational factor measured at the end of the term (e.g., *Ff*). The predictors include the motivational factor at the beginning of the term (e.g., *Fi*), the dummy variable *social*, the overall amount of activity performed in the system measured with the variable *mg_completion*, and the interaction term *mg_completion*social*. Taking into consideration the results shown before, I only included Fascination, Competency Beliefs and Mastery Approach in these regressions. Additionally, I repeated the regression analyses, filtering out all students with less than 15 activities in the system. The results are shown in Table 25.

Results of the regression analyses show no effect associate with the use of the system, nor any effect due to interactions with the interface features (*social*), thus hypothesis **H6** cannot be confirmed.

Overall, while the results do not provide evidence of correlation, the lack of significance

Table 25: Coefficients (β) of regressions on motivational factors at the end of the term with predictors motivation at the beginning of the term (Xi column), social, mg_completion (mg) and interaction mg_completion*social (mg*social). Significance is marked with symbols '.' (.1-.05), '*' (.05-.01), '**' (.01-.001), '***' (<.001). The left side (columns 2-5) shows coefficients on regressions performed for all students with at least 1 activity performed in the system. Right side (columns 6-9) shows results of regressions performed only for students who has more than 15 activities in the system.

	All students with activity				Students with more than 15 activities			
	Xi	social	mg	mg*social	Xi	social	mg	mg*social
Ff	**.705** ***	.010	.057	.007	**.633** ***	.077	.103	-.098
CBf	**.410** ***	.012	.029	.017	**.499** ***	.061	.039	-.064
MApf	**.577** ***	-.007	.028	.054	**.591** ***	.033	.019	-.012

is not necessarily bad, as it provides consistent evidence that the use of this system, particularly, the use of this system with comparison features, does not harm the motivation of the students: social comparison features do not make students become more Performance oriented, do not make them become less Mastery Oriented, nor damage their Fascination. These results contrast with previous findings in which I observed a significantly smaller decrease in the Performance Orientation in students exposed to social comparison features, in previous studies in a Java programming course (see [Guerra et al., 2016], or Section 4.3 in this thesis). One possible reason may be due to cultural differences in the populations involved in these studies. While the currently analyzed study was performed in a University in Finland, the previous studies were conducted in a University in the United States. One cross-study observation was that the levels of Performance Orientation (in fact the level in all Achievement-Goal factors) were much lower in the Finland study than in the previous Java studies. Exploring cultural differences and motivation was not a target of this Thesis and remains an open question.

6.11 CONCLUSIONS OF THIS CHAPTER

Results of the analyses performed in this chapter confirm previous findings on the effects of social comparison features in an OLM, and provide several other observations regarding the roles of prior knowledge and learning motivation. Overall, the study confirms the benefit of having practice content. Analyses showed that completing the content in Mastery Grids predicts almost a 20% increase in the posttest, after controlling for the pretest. Treatment groups were similar in terms of prior knowledge and prior motivational orientations, and were similar in the final performance (posttest). Although the general effects of the social comparison features on system activity are smaller than what we have observed in the past (see [Brusilovsky et al., 2016]), these are still noticeable.

Regarding research question 1 *What are the effects of an OLM with social comparison features (or OSLM) on system activity compared to an individual-view OLM?*, analyses showed that OSLM affects *engagement*, *performance* with self-assessment content items and *navigation*. OSLM makes students do more in the system confirming H1 (*Students exposed to an OLM with social comparison features increase the level of activity in the system*). Students also become more efficient in self-assessment items, and navigate better through the system interface.

Regarding research question 3.1 *How does prior knowledge influence system activity within an OLM?*, an important finding of this study is the relationship between the presence of social comparison features and prior knowledge. First, I notice that prior knowledge has a positive strong correlation with the level of activity in the system, only in the group that is exposed to the social comparison features. Second, that this relationship exists for the objective measure of prior knowledge, i.e., *pretest*, but it does not exist for the subjective measure *Competency Beliefs*, a phenomena which has been observed in other related work ([Higashi et al., 2017]). Third, that this interaction between pretest and the presence of social comparison features exists on the higher pretest levels. Fourth, to better explain this effect, analyses of system engagement throughout the term showed that high pretest students are also more likely to engage early in the term, thus gaining an advanced status, which is displayed by social comparison features, encouraging them to want to keep this status by continuing to interact with the system.

These observations, regarding the positive effect of social comparison features in engaging high pretest students, also serve to explain previous findings regarding performance with self-assessment activities, i.e., success rates. While previous work consistently found that social comparison features make students move forward and not overstay in already known content, we expected it to also show lower success rates. Since this did not happen, I hypothesize that *social* features may also have a positive effect on navigation, which translates to higher success rates. Then, *social* produced neutralizing effects, both lowering and raising performance. In this study, I found that another reason behind the raise of success rate is that the social comparison features tend to engage higher prior-knowledge students, given that higher prior-knowledge students have higher success rates.

Other individual differences were also analyzed. On the one hand, the Social Comparison Orientation scale failed to show effects (RQ 3.3 *How does Social Comparison Orientation influence system activity within an OLM?*) rejecting hypothesis **H5** *The effects of social comparison features of the system will be stronger for students with higher Social Comparison Orientation.*

On the other hand, Learning Motivation, measured by the Performance Approach and Mastery Approach orientations showed interactions with the social comparison component of the system (RQ 3.2: *How does learning motivation influence system activity within an OLM?*). The greatest effect is observed in the relationship of Performance Approach orientation and the *social* features on engagement with system activity. Students with Performance Orientation above the middle point (positive opinion towards these items in the questionnaire) are more sensitive to the OSLM (the *Social* group). In other words, the benefits of the social comparison features depend on how high the Performance Orientation of the students is. This observation confirms hypothesis **H3**: *Social comparison features increase the engagement of students who are highly performance oriented.* which reflects what is expressed by the Learning Motivation literature: high Performance Oriented students tend to compare themselves to others. Mastery Grids has successfully translated this orientation into the benefit of practicing with the system more.

Regarding research question 4 *How does the use of a learning system featuring OLM,* *OSLM and fine-grained elements affect motivation?*, no effect of OSLM was observed in relation to change in motivation in any of the factors measures: Fascination, Competency Beliefs, Values, Mastery and Performance Approach orientation. Hypothesis **H6** is not confirmed: *The active use of OLM with social comparison features will increase the performance orientation of the students.* However, I judge the results as not conclusive enough to reject that hypothesis: a closer look at change in motivation is necessary to make stronger claims about the potential effect of a system such as an OSLM, for example, by studying the motivation variation in shorter time spans, or by doing a more controlled study where more qualitative observations can be made.

7.0 DESIGNING A FINE-GRAINED OLM

This chapter devotes to the work conducted to develop Rich-OLM, an extension of Mastery Grids incorporating a Fine-Grained view of the learner model. The potential benefits of showing learners with a more detailed information of their learner model include the ability to identify "holes" of knowledge, which are not visible in the coarse-grained visualization of Mastery Grids; a better understanding of the domain, as detailed views add information about underlying relations between domain concepts and content; and better guidance or support to choose content to practice or improve learning.

As the amount of information displayed increases, it also increases the complexity of the interface and risks to produce information overload. Because the system visualizes details, it could become hard to understand, overwhelming the learner. Moreover, complexity is a special concern in this scenario, because the end-user is not an "expert" per se, and may not be willing to spend the required effort. This is why it is extremely important to design such complex visualizations carefully, balancing complexity and potential support. From the perspective of the Information Visualization field, complexity caused by information overload is a foundational problem of the field. This is the main motivation of the famous Information Seeking Mantra: *overview first, zoom and filter, then details-on-demand* [Shneiderman, 1996]. Following Shneiderman's principle, the design of the Rich-OLM start with the idea of incorporating together the coarse-grained (overview) and the fine-grained (details) representation together. I approached the development of the Rich-OLM consulting students from the very beginning, and continued a development process conducting two controlled studies.

7.1 THE UNDERLYING LEARNER MODEL

Before describing details of the design of the Rich-OLM, I describe the information that the Learner Model of Mastery Grids provides, which is the starting point to analyze options of visualization: what information does the Learner Model manage?

Mastery Grids is built on top of a user modeling and personalization framework [Brusilovsky et al., 2005] that includes a two-level domain model, a learner model, and a content model. Since we use the same framework to implement the fine-grained visualization, I introduce the most essential components of this framework below.

- The *Fine-Grained Domain Model* is composed of a set of Knowledge Components (KCs) that represent elementary units of knowledge such as skills or concepts. For example, the Java domain, uses 114 KCs from an ontology developed by our group. Examples of KCs are *int data type*, *addition*, *variable initialization*, *String concatenation*, *for loop*, *constructor*, and *inherited method*. The concepts and associated topics of the two domains used in this thesis (Java and Python programming) are in Appendix D.
- The *Coarse-Grained Domain Model* is composed of a list of topics that represent relatively large fragments of domain knowledge. While the KC-level model is defined by the structure of the domain, the list of topics reflects a pedagogical approach to teach the domain. Our infrastructure allows different instructors to introduce their preferred sequence of topics for the domain. Structurally, each topic could be mapped into a subset of KCs. Taken together, topics and KCs define a two-level hierarchical domain model.
- *Activity-KC mapping* is used to connect learning activities (examples, problems, animations) to a set of KCs so that students can practice the activities addressing the KCs. This mapping can be established manually or automatically. For the Java domain, this mapping is done automatically by the content parser presented in [Hosseini and Brusilovsky, 2013], with optional expert refinement (see [Huang et al., 2016]). In this domain, content activities have between 2 and 70 associated KCs.
- *Activity-Topic mapping* associates each course activity with one of the course topics. This mapping, which essentially defines the structure of a course, is usually done manually by course instructors who adopt a specific sequence of topics. In both domains used

in this thesis, Java and Python programming, the structure of activities and topics was assembled with the help of instructors of the programming courses. The sequence of topics and activities associated in each of them is carefully decided to ensure that an activity only contains KCs of the topic in which it belongs or from topics covered previously.

- The *Learner Model* represents an estimation of learner knowledge for each component of the domain model. The sources for this knowledge estimation are activity traces produced by the learner's work with different learning activities. The Learner Model uses these activity traces and the mapping between activities and domain model components (topics or KCs) to update the learner's knowledge level for each topic or concept related to the activity performed. For example, when the learner solves a problem that contains the KC *for-loop*, the Learner Model will consider this as evidence of knowing the KC and will update its estimation. Note that the knowledge level for domain topics visualized by the MG interface can be modelled independently or calculated as an aggregation of knowledge of concepts included in the topic. In past studies of the MG interface, we explored both approaches. Details of the current Learner Model implementation can be found in [Huang et al., , Huang et al., 2016].

Figure 41 shows the relations between topics, activities and concepts (KC) of the Domain and Content models. It is a partial view. The Learner Model is represented by the color of the concept nodes in a gray scale (darker color corresponds to higher estimated knowledge of the KC). Each topic is associated with a set of concepts. Activities (content) in the bottom of the figure have many concepts associated which could belong to different topics.

7.2 INTEREST FOR DETAILED LEARNER MODEL

Before designing a visualization of the fine-grained OLM, I asked myself to which extent all of the levels of this information could be helpful to students, and whether this detailed information could support the different ways the students use the system. I understand that students may use the system with different goals. To better understand these issues, I talked

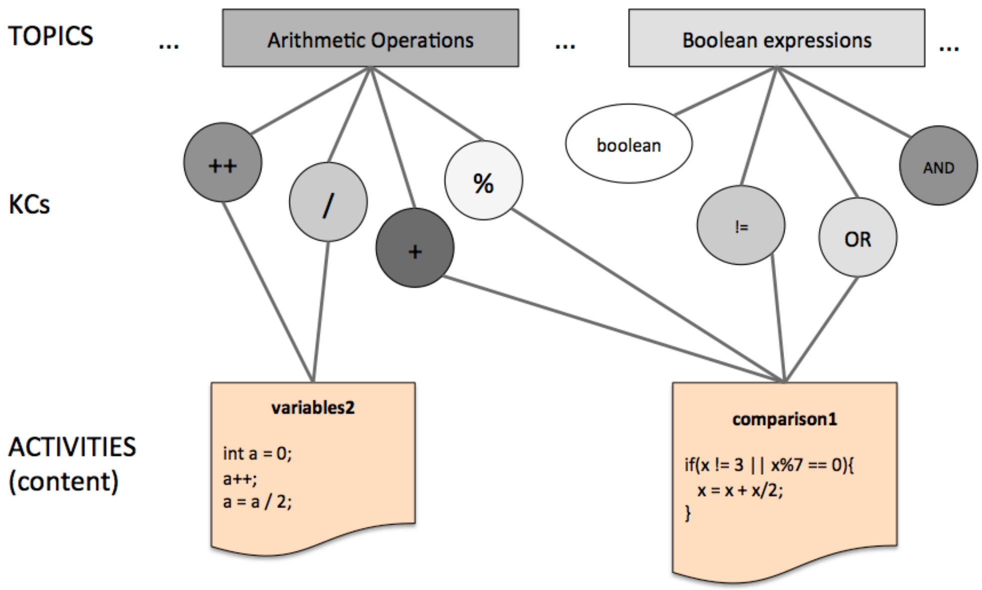


Figure 41: Partial representation of Domain, Content and Learner Models. Learner model is represented by the darkness of the KCs and topics.

to students. I performed semi-structured interviews with 9 students who were familiar with the coarse-grained Mastery Grids. Participants were undergraduate and graduate students of the Information Science School at the University of Pittsburgh, and received a compensation of US$20, and signed a Consent Form.

The interview was structured in two parts. In the first part the conversation focused on understanding to which extent the participant used Mastery Grids, to which extent she understood what it shows, and the goals the learner had when using it. Guiding questions in this part of the interview were: *What do you think is or are the goals of this system?*, *How do you think this system helps students?*, *What do you think is the ideal way to use it?*. There were some guiding questions about the presence of social features in the system too: *What did you make of the progress of others in the system?*, *Did you consider the progress of others when using MG?*, *Did you feel lagging behind or getting ahead?*. The conversation was then conducted towards a scenario in which the participant has to prepare for a quiz, and questions aimed to clarify how helpful Mastery Grids could be for this goal. Interviews were performed individually and in two opportunities, in groups of two students together. From the first part I highlighted several ideas that were expressed by the participants.

- The general perception of the system was positive, some participants even expressed that it had a clearer structure of content than the book.
- Students used the system differently. Some liked to go sequentially and do everything, some just wanted to verify if they knew all what is relevant in the topic. Some participants mentioned the idea of having a "super" quiz in each topic that summarizes all the content of the topic.
- Regarding social comparison features, opinions were positive ("encourages the competitive spirit", "useful to quantify / want to catch up"), or indifferent ("I don't care"), and only one participant expressed that it could be discouraging if you are lagging behind. Interestingly, one participant gave a different interpretation of the progress of others: "I think [the darker cells on the others' rows] means that people are struggling with it", thus attaining a higher level of difficulty to the material that showed more aggregated activity.

The second part of the interview centered around the idea of presenting more details in the visualization, i.e., the fine-grained space. Participants were first introduced to the concept space, and the topic-KC, activity-KC relations were explained and examples were presented. Then the conversation was guided by questions such as *Do you think the information of others will be helpful for you? How?*, *Do you think this information will be helpful for others? How?*. Special attention was paid to representing all the different information associated: the concepts in each topic, the relations of concepts and activities, the level of knowledge and progress the learner will see in each concept, and the possibility of seeing this fine-grained information of the rest of the class. Participants were instructed that the Learner Model could basically estimates the level of knowledge in each concept (KC), but also the amount of effort spent from the amount of work (amount of activities done) associated with each concept.

Form the second part, I summarize the following ideas.

- I corroborated the idea that fine-grained information about students and peers' progress and knowledge if of value for students, although different levels of such information might not be of interest to some of them. For example students did not make clear distinctions between progress as completion of the content, and progress as the amount of knowledge gained. Also, some students expressed no interest in social comparison features, especially at the fine-grained level.
- We also confirmed that, although adding more details is generally considered useful, a clear concern arose about complexity. As more information is added, the OLM could become more complicated to understand and interpret.
- It was recommended to maintain the topic visualization because it provides the context. It was easier for students to navigate the content through ordered topics. The fine-grained view has to be linked and complement the coarse-grained view.
- It was recommended to represent the links between topics and concepts, because it was useful to know "what is inside" the topic.
- It was recommended to limit the information provided for each concept and I choose to represent "progress of knowledge" as the estimation of knowledge provided by the LM, and I discard content completion information (at least at the level of concepts).

These ideas serve as a basis to guide he development of the fine-grained visualization. They provide a first level understanding of the potential value of showing fine-grained information and offer directions of which information to show. However it does not tell much about how to visualize the information. I advance in this issue in the next section.

7.3 STUDY 1: COMPARING DESIGN OPTIONS FOR A FINE-GRAINED OLM

7.3.1 Motivation and set up

To have a better idea of how to visualize the KC space and how much information is needed we designed a controlled user study, that I call *Study 1*. With the help and suggestions of my advisor and other professors (see Acknowledgements), I designed five different visualizations with different levels of information about the concept space and its relationships. All visualizations included the topic level visualization (Mastery Grids). We excluded the social comparison features in order to focus on the complexity issues of the fine-grained level. These designs, together with a control version (Mastery Grids alone) are presented in Figure 42. Visualization options varied in terms of the amount of information displayed (showing KCs only within the topic, showing all KCs at the same time, or showing connections between KCs), and the visual element representing each KC (bars or circles). Knowledge in each KC is represented with shades of green as in Mastery Grids, and in the case of using bars to represent KCs, we represent such information with both color and size. This decision was motivated to avoid possible biases caused by the use or non-use of color. The different visualizations were inspired by visual representations previously used in OLM such as *skillometers*, which are the most common visualizations (e.g., [Bull and Mabbott, 2006, Corbett and Bhatnagar, 1997, Long and Aleven, 2013b, Mitrovic and Martin, 2007, Weber and Brusilovsky, 2001]), *bar charts* or *histograms* (e.g., [Mazzola and Mazza, 2010, Shi and Cristea, 2016]), and *concept-maps* (e.g., [Duan et al., 2010, Mabbott and Bull, 2006, Pérez-Marín et al., 2007]).

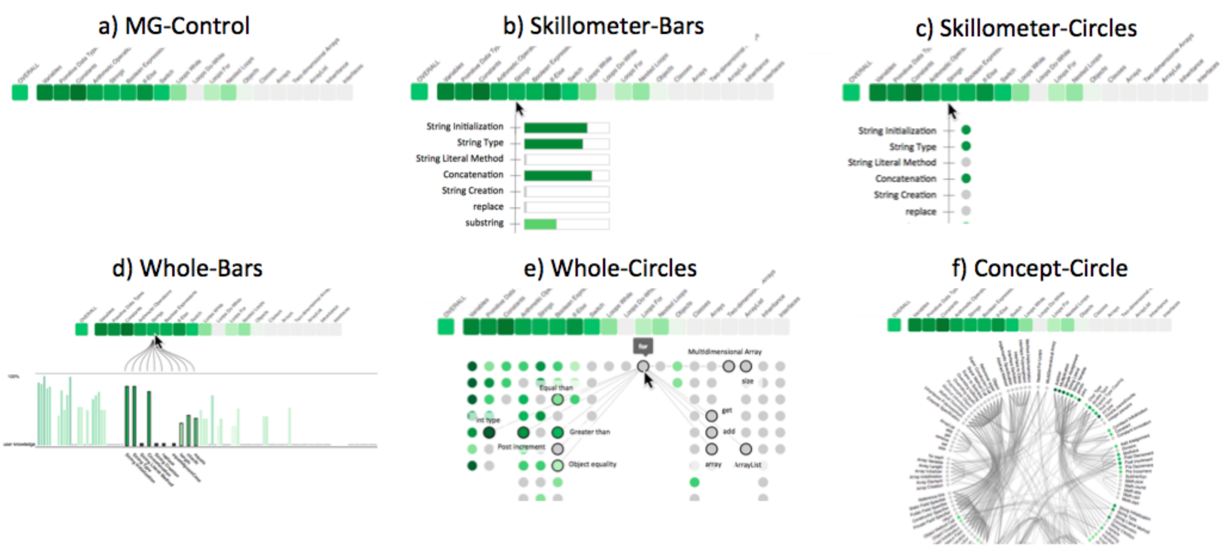


Figure 42: The 6 visualization variations evaluated in Study 1.

Although the prototypes were presented as paper mockups to subjects, we described them as functional prototypes with some interactivity features (e.g., how they react when a concept is mouseovered).

- *Skillometer-Bars*: They show the list of KC associated with a specific topic when you mouse over the topic. Each KC is represented with its name and a bar indicating the estimated knowledge.
- *Skillometer-Circles*: They are similar to Skillometer-Bars, but KCs are represented with colored circles here.
- *Whole-Bars*: They show all KCs in the course (114 in the Java course) with bar chart parallel to the coarse-grained visualization. The idea is that when topics are pointed to, the related concepts are highlighted.
- *Whole-Circles*: This visualization also shows the whole space of KC ot once. KCs are positioned under the topic to which they belong and are represented with colored cir-

cles. When you mouse over a concept, the name is shown and the connections to other concepts are also shown with the names of the related concepts. These connections are Skill-Combinations [Huang et al., 2016] and represent pairs of concepts that should be practiced together.

- *Concept-Circle*: This is another view of the whole space where names and connections are shown all at the same time. KC are represented with small colored circles. Mousing over a KC will highlight its connected KCs. Mousing over a topic will highlight the group of related concepts in the circle.

Subjects were first offered a presentation with explanations of the Learner Model, including all the information described above, and a description of each of the visualizations shown in Figure 42. We provided several mockups for each of them to describe interactivity. Clarifications were provided when needed. To ensure that subjects could give valuable feedback, we required that all had previous experience using Mastery Grids in a course.

Then subjects received a survey with three parts, each setting a different context or *scenario* in order to collect a broader subjective evaluation. Part 1 set a general scenario. Part 2 set the scenario of preparing for a hypothetical quiz on a specific topic. Part 3 set the scenario of preparing for a midterm exam that covers a number of topics. In each part of the survey, questions were repeated and phrased to match the specific scenario. The questions covered different aspects (the examples in parenthesis are the questions phrased for Part 1): *preparation checking* ("The visualization helps me to check whether I am doing well enough in the course"), *knowledge reflection* ("The visualization makes me think about my knowledge in the course"), *strength and weaknesses identification* (2 questions: "The visualization helps me to identify the strengths (weaknesses) in my knowledge of the course content"), *motivation to explore* ("The visualization motivates me to look for further material to learn more about the course content"), *easy understand* ("The visualization is easy to understand"), and *topic awareness* ("The visualization helps me to have a better idea of the content involved in each of the topics of the course"). Each part of the survey was presented as a matrix, with the rows containing the questions and the columns containing the 6 visualizations to facilitate comparative answers. In Part 2 and 3, where the overall stated goal is to prepare for a quiz or midterm exam, we included two additional items: *plan* *next* ("The visualization helps me to plan what to do next in order to prepare for the quiz"), and *quantify work* ("The visualization helps me to quantify how much work I should do to prepare for the quiz").

At the end of the session subjects were asked to indicate the best and the worst visualization, and to provide an explanation of their choices.

7.3.2 Study 1 results

Forty two subjects completed the study. The subjects were Information Sciences Master students and Computer Science undergraduate students at the University of Pittsburgh. Each received US $20 for participating and signed an informed consent. Multilevel linear regression analysis was performed for each of the aspects measured (dependent variables). Random effect of subject in the repeated measures was specified, and models were built using Maximum Likelihood method. I used R and the function *lme* to run these analyses (see chapter 13 in the book [Field, 2012] for a detailed explanation of how this is performed). For space constraints, and since I am not looking for detailed differences but want to inform design decisions, I report only general trends observed.

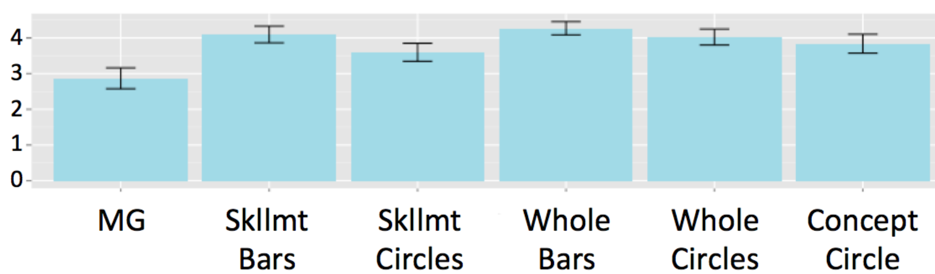


Figure 43: Overall perception of usefulness of the different visualizations for planning what to do next. Error bars represent 2 Standard Errors of the mean.

A first run of the analysis contrasted the perception of the visualizations, averaging the answers per visualization across questionnaire parts (scenarios) for each of the questions in

Table 26: Study 1, the visualizations that were most often chosen as the best or the worst.

Visualization	Best	Worst
Whole-Bars	14	1
Concept-Circle	14	13

the survey. I found a significant effect of the visualization on all the aspects measured. The patterns of preferences showed a preference for *Whole-Bars* and were similar across items of the survey with slight differences. Figure 43 shows, as an example, the average evaluation in helping to plan what to do next (*plan next*) across part 2 and 3 of the survey. Post-hoc comparisons were performed with Tukey contrast between the visualization options. Results showed a clear advantage of all visualizations over the control version (*MG-control*) and the *Skillometer-Circles* for all dependent variables, except for *easy understand*, where *MG-control* is, not surprisingly, generally better evaluated. While generally evaluated higher, *Whole-Bars* did not show significant differences to *Skillometer-Bars*. These two visualizations using bars were evaluated higher than visualizations using circles to represent KCs.

A second run of analysis was performed for each survey item separately in each of the scenarios. Results showed lower scores in the *quiz* scenario, especially for the aspects *strengths and weaknesses identification*, *knowledge reflexion*, *motivation to explore*, and *topic awareness*, which suggest that there is room to improve the system to support more specific tasks.

Interestingly, the overall preferences (best and worst) were divided between complex representations. Table 26 shows that while the same amount of participants choose *Whole-Bars* and *Concept-Circle*, this last visualization is chosen as the last preferred visualization because of "overwhelming" complexity.

From Study 1, I learned that students prefer bars to circles for representing their knowledge of concepts. They also think that bars are easier to understand. These findings are consistent with preferences for skillometers found in previous research [Duan et al., 2010], but also suggest that the preference might be due to the visual element used (the bar) and not necessarily the level of complexity offered (no difference between *Whole-Bars* and *Skillometer-Bars*). Visualizations with connections, which were evaluated as more complex, were not judged as more helpful in any of the aspects. However, preferences for *Concept-Circle* were extremely divided (best and worst). Multiple preferences have been recognized in the literature and addressed presenting alternative visualizations [Duan et al., 2010, Conejo et al., 2011]. We also learned that visualizations might bring different levels of support depending on the scenario. These scenarios involve different goals students have while using the system. A takeaway is that the current alternatives do not seem to support the *quiz* scenario well, and other features might be needed to improve this. The evidenced differences between scenarios also suggest that it is important for evaluations to specify well defined tasks. Although evaluation for *Whole-Bars* and *Skillometer-Bars* are similar in the questionnaire, subjects stated that for tasks like preparing for a midterm, they would prefer to use a visualization that shows the whole concept space. This was a strong reason to select a visualization that includes both global and local context. We conclude that the sweet spot is the *Whole-Bars* visualization, though there is an interesting research idea in exploring *Concept-Circle* as an alternative visualization.

7.4 THE RICH-OLM

Attending to the results of Study 1, I implemented a Rich-OLM based on the *Whole-Bars* prototype. It shows the topics with their progress and all the concepts of the course in parallel. The basic interface of the Rich-OLM is shown in Figure 44 for a course of Java programming. The same interface is shown in Figure 45 with the comparison features now enabled. Comparison features have been added to both the coarse-grained elements (topics) and the fine-grained elements, the last are represented in the form of blue bars opposite to the green bars of the learner's knowledge progress.

When a topic is moused over, related concepts are highlighted (the rest are shaded) as shown in Figure 46. When the learner clicks a topic, the activities contained are shown, and

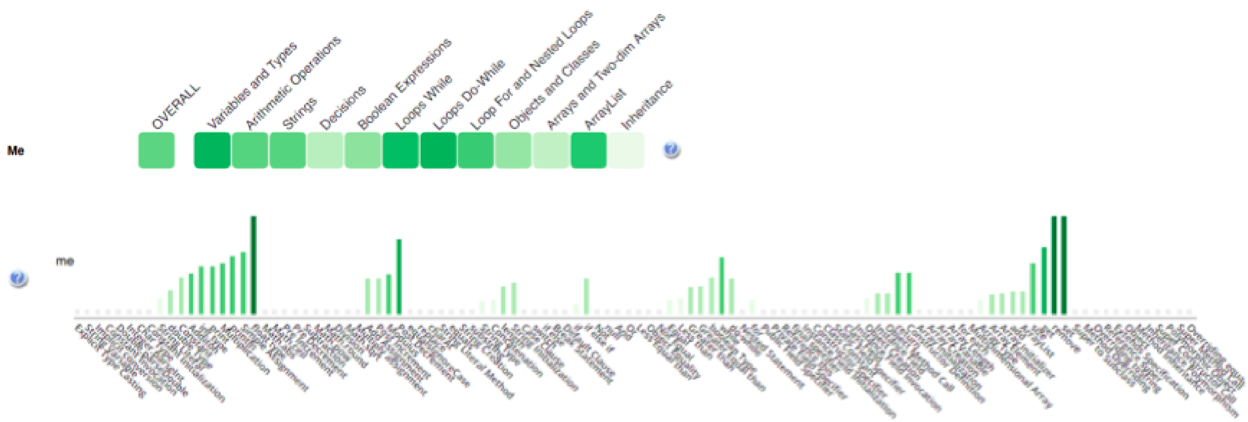


Figure 44: Rich-OLM interface without social comparison features.

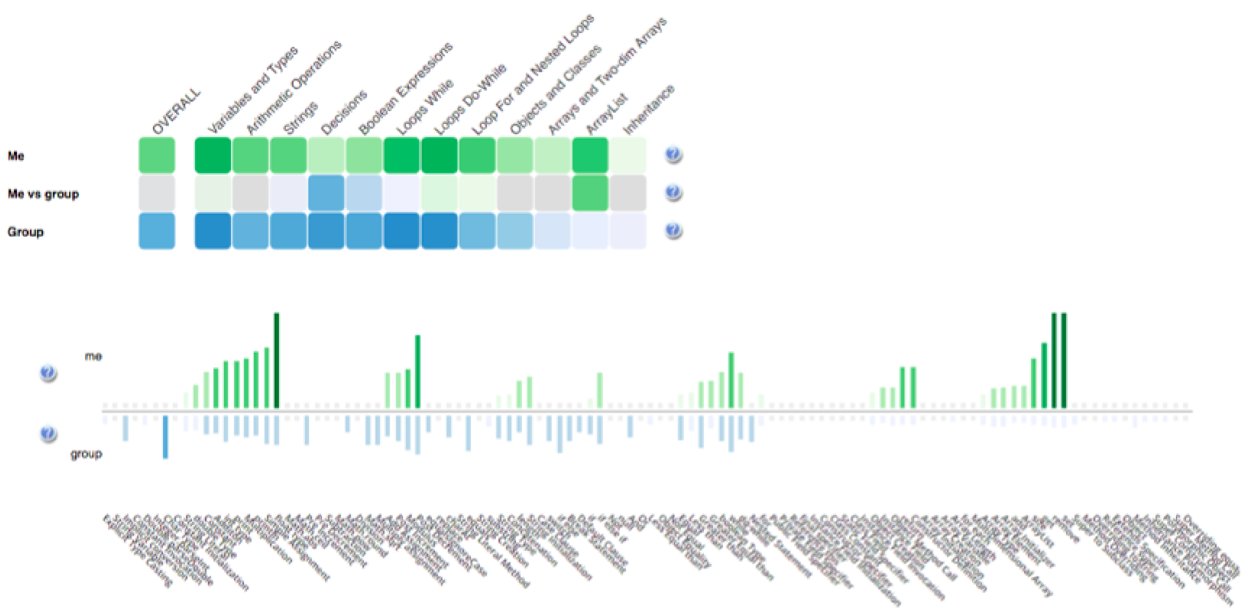


Figure 45: Rich-OLM interface with social comparison features.

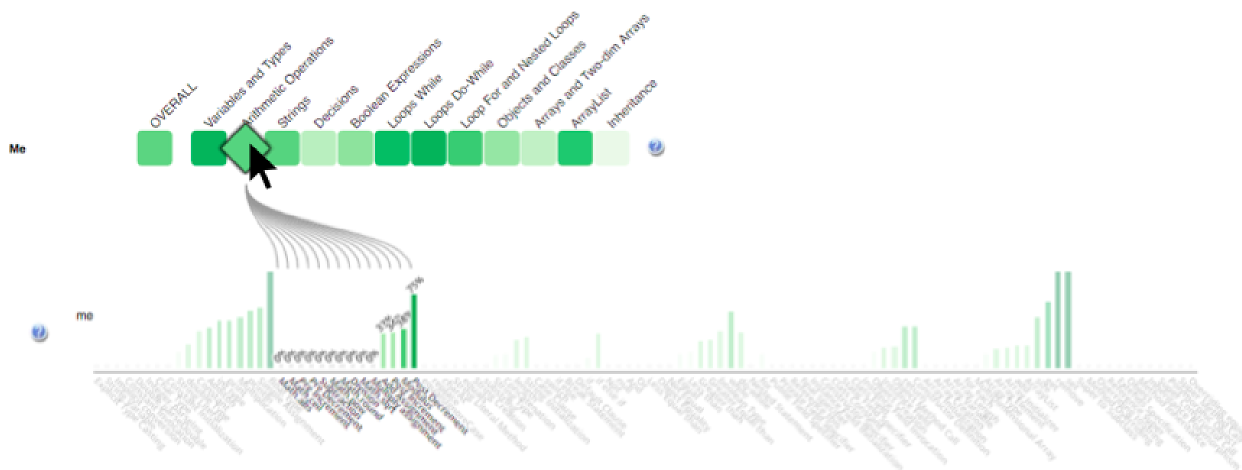


Figure 46: Rich-OLM interface with a selected topic.

the concepts related to this topic are highlighted, with their names at the bottom of each bar. Figure 47 shows a screenshot when entering the topic *Strings*, and Figure 48 shows a similar screenshot in another course assembled in the Rich-OLM with the social comparison features enabled. When the learner *mouse overs* a cell corresponding to an activity inside a topic, its related concepts will be highlighted in the bar-chart, as it can be seen in Figure 49.

Attending to the concern of complexity, expressed by participants in the study 1 and previous interviews, the Rich-OLM interface was further extended adding a visual aid to help learners to interpret the fine-grained information associated with the content activities within the system. This visual-aid should be able to express which activity is potentially more useful for a user that seeks learning. Gauges are popular to represent single values and at the same time to set meaningful boundaries, and have also been used in learning analytics visualizations [de la Fuente Valentín and Solans, 2014, Fulantelli et al., 2013, Khan and Pardo, 2016, Falakmasir et al., 2012]. We then designed the *learning gauge*, or simply, *Gauge*. The *Gauge* does not add extra information, as the social comparison feature does, but instead presents an interpretive view of the information shown in the concept bar chart:

Figure 47: Rich-OLM interface: entering a topic.

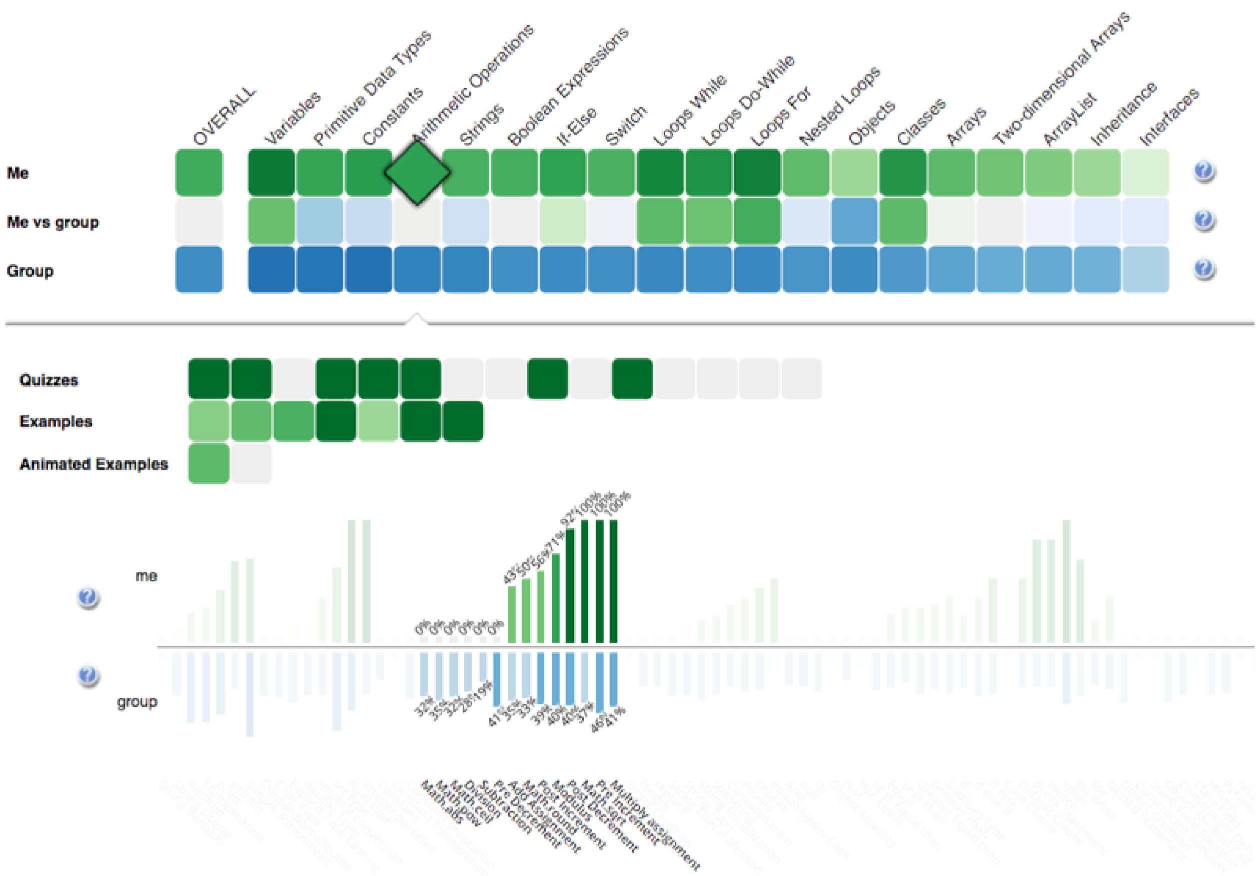


Figure 48: Full Rich-OLM interface. The user has entered a topic and the concept bar chart has faded with only related concepts highlighted.

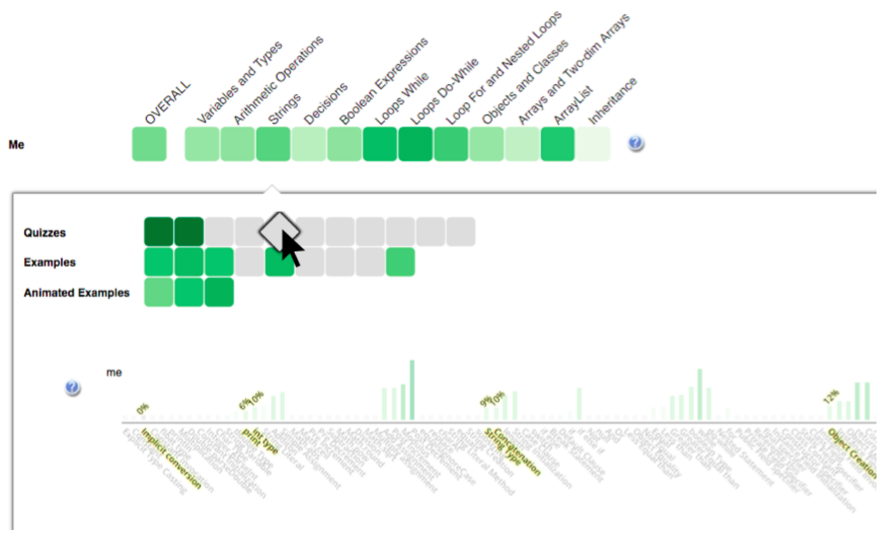


Figure 49: Rich-OLM interface: mousing over a content activity.

when the learner is inside a topic and mouse over a content activity, the gauge shows an estimation of the potential learning (which can also be considered as a measure of *difficulty*) by counting the number of related concepts (KCs) that are already known, familiar (or partially known) and not known (or new) to the learner based on predefined thresholds. This is shown in Equation 7.1.

$$learning_{estimated} = \frac{0.5 * kcs_{familiar} + kcs_{new}}{kcs_{known} + kcs_{familiar} + kcs_{new}} \tag{7.1}$$

The *learning gauge* is only shown inside a topic and when an activity is moused over. Figure 53 shows a screenshot of the individual Rich-OLM and the gauge in it, and Figure 51 shows the detail of the *learning gauge*. As mentioned before, this gauge aims to guide students to choose learning content to maximize learning, either by alerting the student of the content that does not provide new knowledge, as alerting the student of content that might be too difficult (to many new concepts). I materialize this expected guidance effect by complementing the hypothesis **H2** (Fine-Grained OLM helps students to navigate the content of the system more efficiently) with the sub-hypothesis **H2G**: *Fine-Grained OLM complemented with the Learning Gauge helps students to navigate the content of the system more efficiently.*

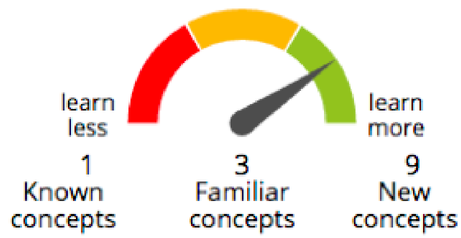


Figure 51: Details of the Gauge visual aid.

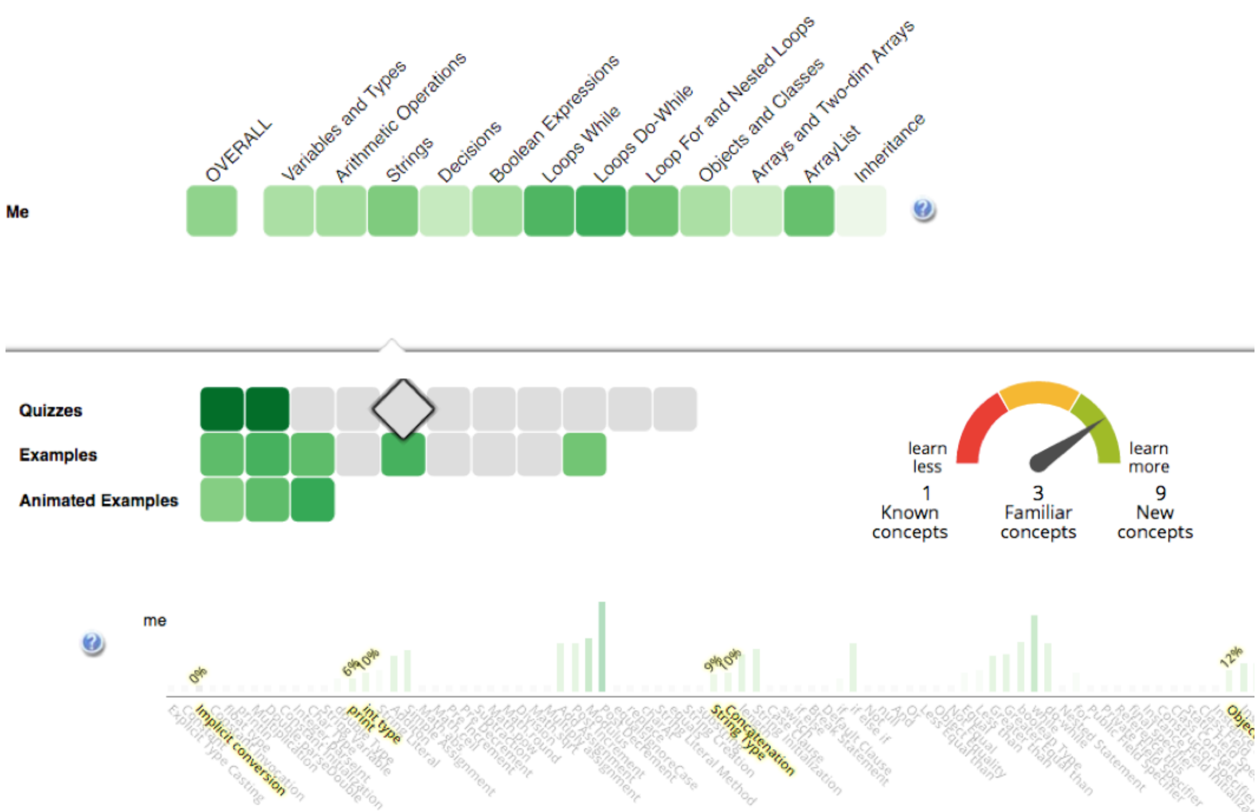


Figure 50: KCG (KCs+Gauge).

The visualization of the bar-chart was completely built using the javascript library *d3* (www.d3js.org) and was integrated into Mastery Grids. The set of services that supports

Mastery Grids was also modified to include the fine-grained information.

While evaluating Rich-OLM will require different aspects aligned to different purposes of use and self-regulated learning tasks, I prioritize here the evaluation of the support that the system brings when students are focusing on a specific topic and searching for the best activity to engage with. I then designed and performed a second controlled user study to inform this, which is described in the following chapter.

8.0 EVALUATION OF RICH-OLM

To evaluate the Rich-OLM, we designed a controlled user experiment contrasting different versions of the visualization for a specific task: find the piece of content that best helps the student to increase their level of mastery in a specific topic. This task is aligned with the main *navigational* goal of the fine-grained visualization, and seeks to find initial validation for the hypotheses **H2** *Fine-Grained OLM helps students to navigate the content of the system more efficiently*, and **H2G** *Fine-Grained OLM complemented with the Learning Gauge helps students to navigate the content of the system more efficiently.* The next sections present the details of this study, the different variations of the interface, and the results.

8.1 STUDY DESIGN

8.1.1 Visualizations and the system

The first version of the visualization is shown in Figure 52 and is simply called *KC*. It includes the basic features of the visualization of concepts (or KCs) without social comparison.

The second version called *KCG* (KC + Gauge) is shown in Figure 53. This version adds the *learning gauge* visual aid, specifically designed to direct the interpretation of the information displayed by the KC visualization towards a sense of the relevance of each of the activities within a topic.

The third version of the visualization in the study, *KCS* (KC + Social Comparison), is shown in Figure 54. This visualization provides all the information of the full Rich-OLM interface, including the social comparison features, but it does not include the *learning gauge*.

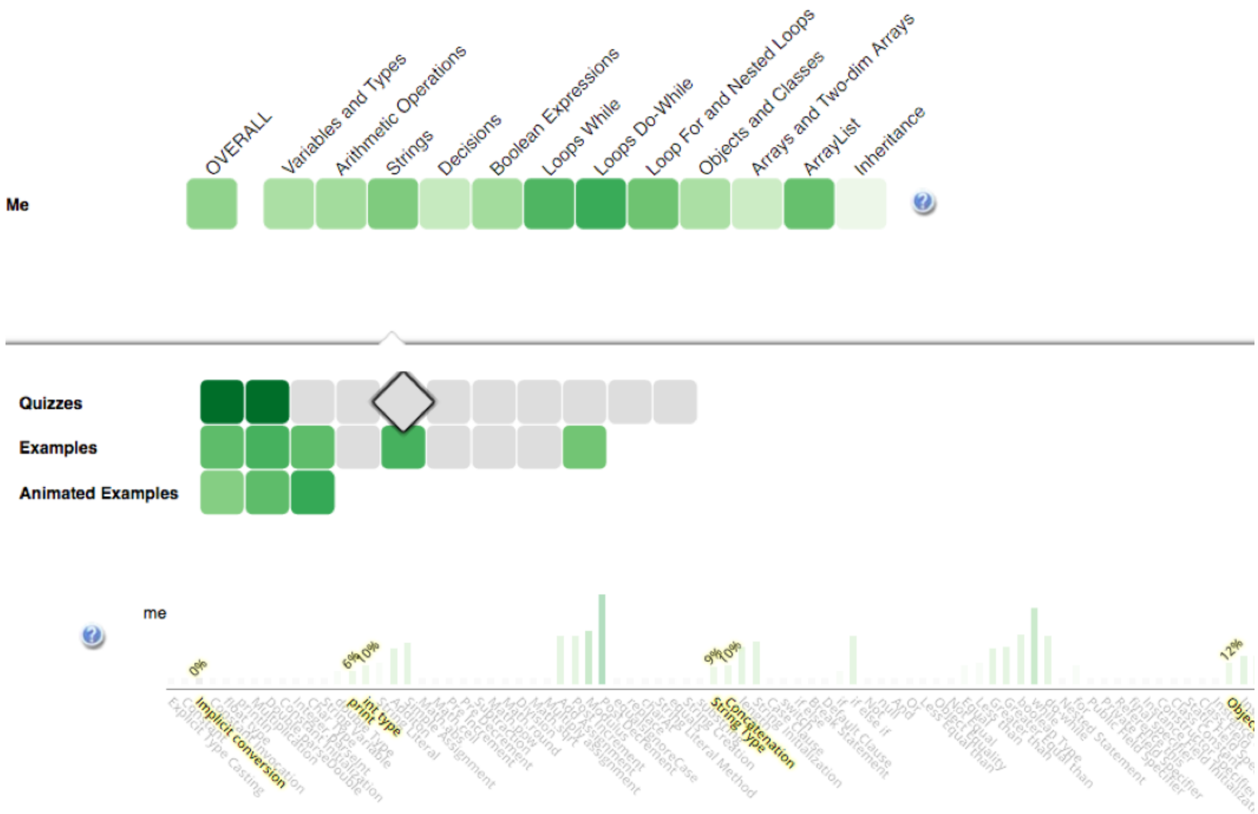


Figure 52: KC basic visualization.

Figure 53: KCG (KC+Gauge).

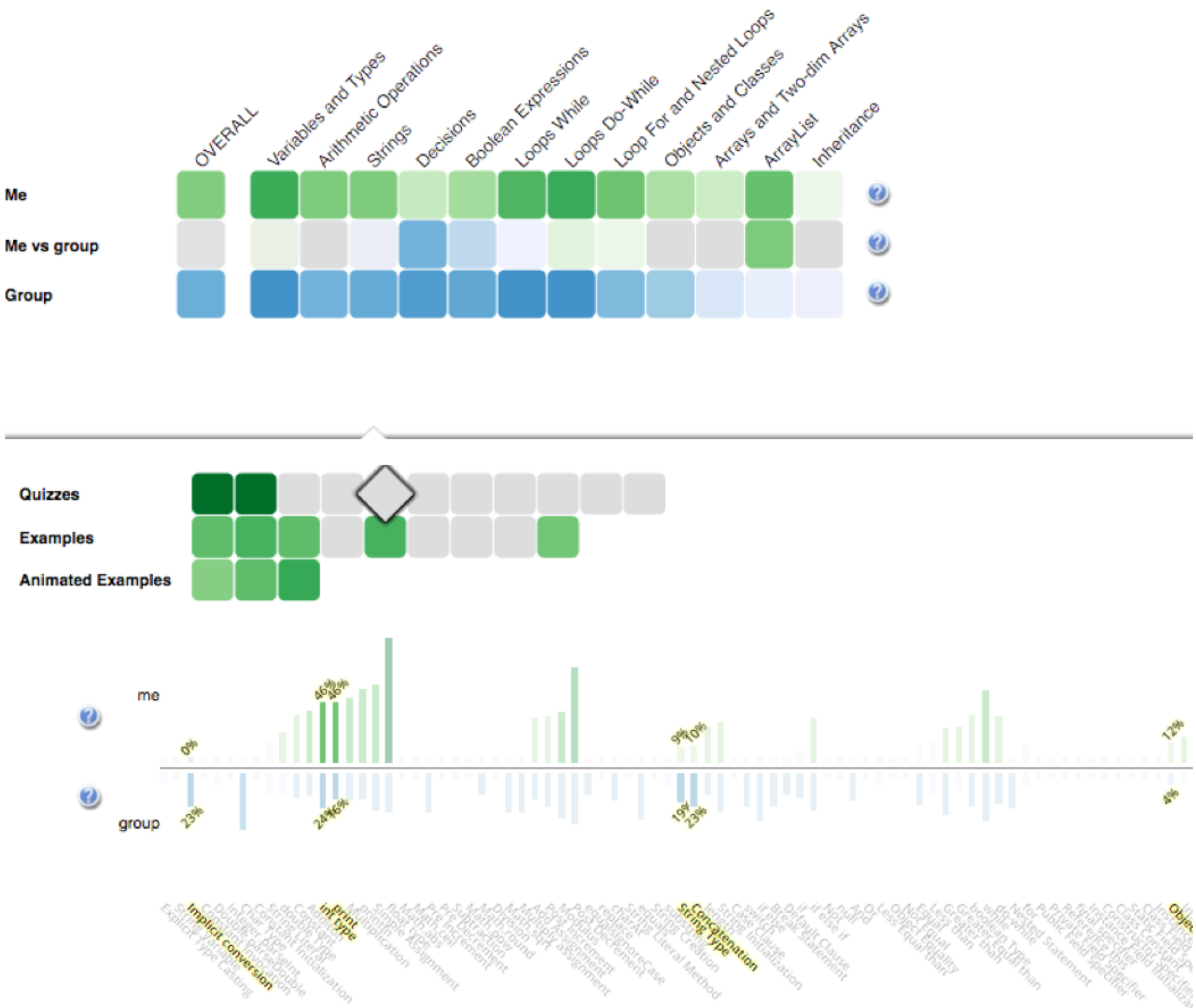


Figure 54: KCS (KC+Social Comparison).

I used a version of the *Java* course with 12 topics (Variables and Types, Arithmetic Operations, Strings, Decisions, etc). Each topic has between 13 and 29 content activities of different types including parameterized java problems, annotated examples, and animated examples (see Section 3.1 in chapter 3). Multiple topics allow me to ask subjects to repeat the task using different visualizations, implementing a within-subject design. To carry out Study 2, I developed a simple interface with which subjects can follow the steps of the study at their own pace.

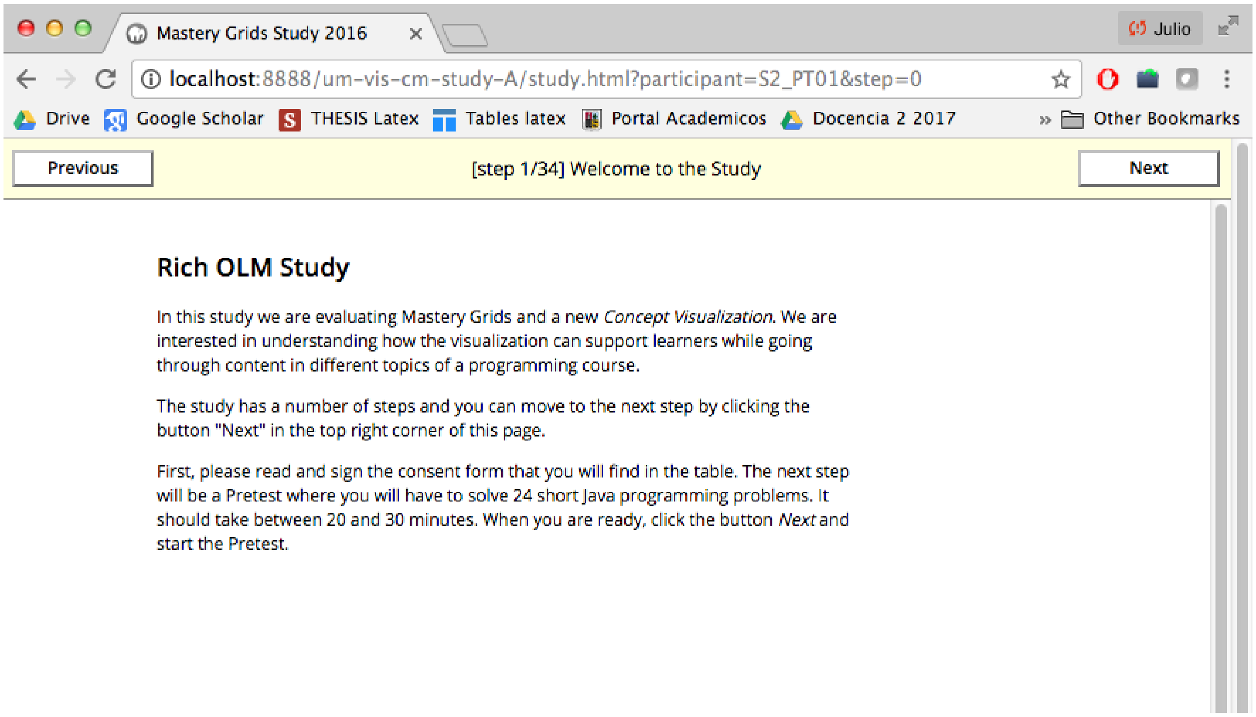


Figure 55: Study interface. Subject advanced through the steps with the button *Next* at the top right corner.

8.1.2 Pretest

Before starting the tasks, subjects completed a *pretest* consisting of 24 problems covering the 12 topics (2 problems from each topic). The goals were: (a) to have a measure of the prior knowledge of the subjects, that will be used in the analyses, and (b) to feed the Learner Model to be shown in tasks. The study interface with the pretest can be seen in Figure 56.

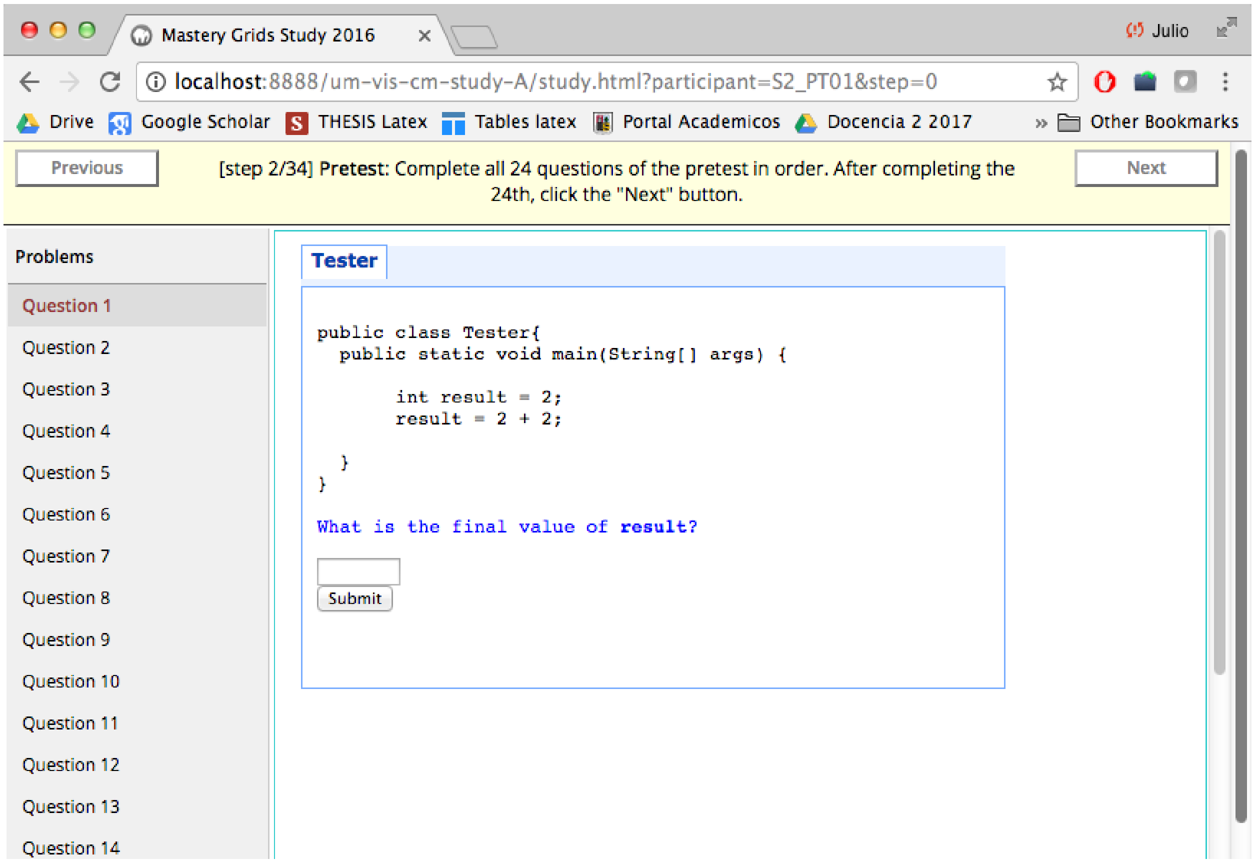


Figure 56: Pretest step in the study interface

8.1.3 Introduction and general Instructions

Following the pretest, subjects viewed a short video hosted in youtube[1] and embebed in the study interface. The video explains the basic KC visualization and its interactive features. Right after the video, the next step presented in Figure 57 presents the general instructions.

[1]https://youtu.be/lJZG4WEF4-8

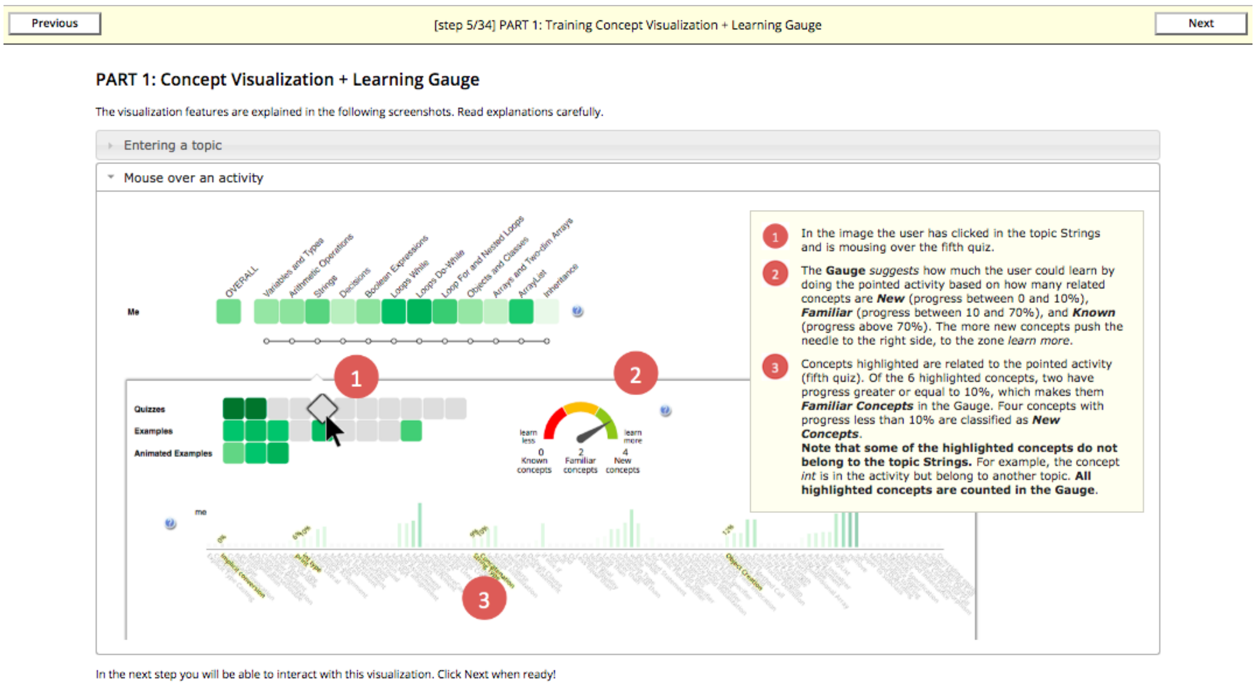
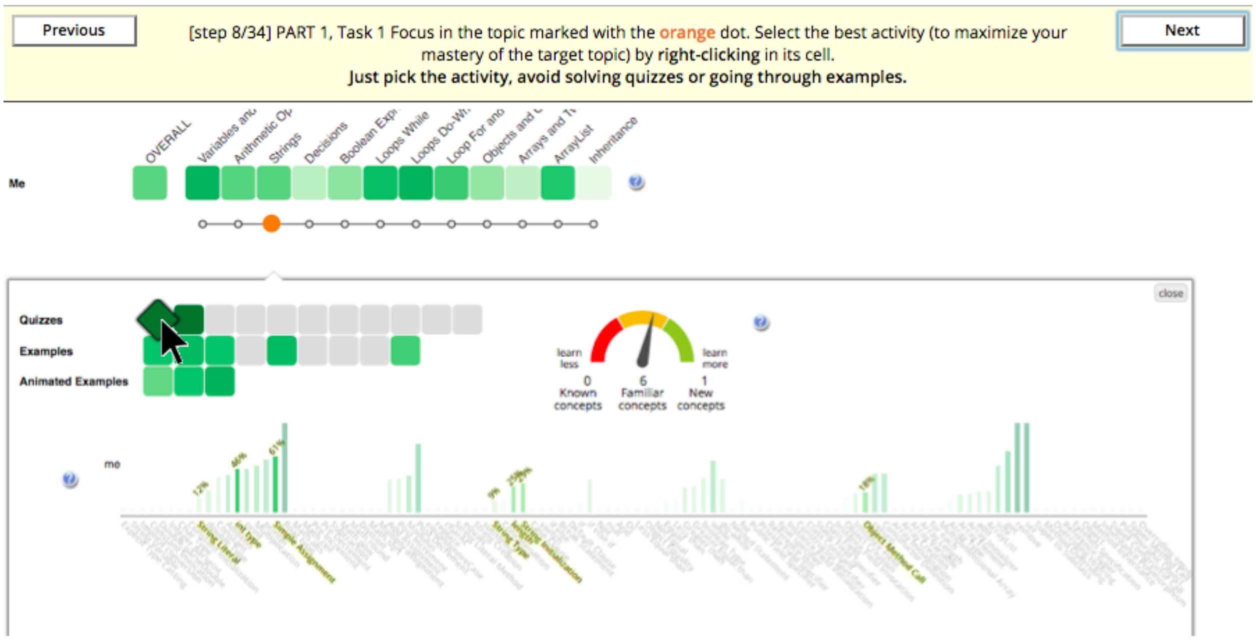


Figure 57: General instructions in the study interface

8.1.4 Tasks

Tasks were presented in groups of 4 for each visualization. Visualizations were introduced to subjects in different orders following a Latin-Square design. Each visualization was first presented with a short tutorial explaining its features, a training step where subjects were free to try the visualization, and an interactive self-assessment test to corroborate that subjects understood the features (if failed, subjects were asked to call the study coordinator for clarifications). Then the tasks for the visualization were introduced one by one, and each task involved one specific topic. The instructions were: "*Focus on the topic marked with the orange dot. Select the best activity (to maximize your mastery of the target topic) by right-clicking its cell. Just pick the activity, avoid solving quizzes or going through examples.*" Each topic is inspected only once (12 topics = 1 topic per task, 4 tasks for each visualization, 3 visualizations) and topics were assigned randomly to avoid bias due to the variability of the topics.



s

Figure 58: The study interface during a task

8.1.5 Task Survey

After every two tasks, the subjects were asked to fill out a *task survey* about their experience performing the previous two tasks. It covers the usefulness of the visualization, and its influence in making them reflect on their knowledge. Table 27 shows the items of the survey. Answers options are on a 7-point Likert scale (1:Strongly disagree - 7:Strongly agree). Some items were reversed (R). To facilitate the analyses in the next section, questions were given an identifier, which is shown in the first column of Table 27. Additionally, I included four questions from the *NASA-TLX*[2] survey (see Table 28). These questions are presented with sliders running from 0 to 1.

Finally, after the series of 12 tasks were completed, subjects were asked to fill out a *final survey* in which were asked to (1) rank the three interfaces according their own preference

[2]NASA Task Load Index: https://humansystems.arc.nasa.gov/groups/tlx/

Table 27: Usefulness and self-reflection *task survey*.

Item	Statement
confidence	I am confident that I selected a good activity for the tasks
usefulChoose	The visualization was useful to decide which activity to choose
ledUseless (R)	The visualization at times led me to less useful activities
findLearn	The visualization helped me to find activities where I think I can learn something new
thinkKnowledge	The visualization made me think about my own knowledge in programming concepts
notHelpful (R)	The visualization did not help me much while searching for a good activity for the target topic
avoidEasier	The visualization helped me to avoid choosing activities which I think are too easy for me
avoidHarder	The visualization helped me to avoid choosing activities which I think are too hard for me
criticalEfficacy	Without the visualization I will probably fail to select a good activity for the target topic
criticalEfficiency	Without the visualization I will probably spend more time selecting an activity for the target topic

and explain their ranking, and (2) rate the ease of understanding and ease of use of each visualization using a 7-point Likert scale (1:Extremely easy - 7: Extremely difficult).

Table 28: NASA-TLX survey.

Item	Statement
TLX1	Mental Demand: How mentally demanding was the task? (0:Very low - 1:Very high)
TLX4	Performance: How successful were you in accomplishing what you were asked to do? (0:Perfect - 1:Failure)
TLX5	Effort: How hard did you have to work to accomplish your level of performance? (0:Very low - 1:Very high)
TLX6	Frustration: How insecure, discouraged, irritated, stressed, and annoyed were you? (0:Very low - 1:Very high)

8.2 RESULTS

8.2.1 Data collected

Twenty nine subjects completed Study 2, with all of them completing all steps and surveys. However, some subjects did not explicitly select an activity at the end of the tasks: one subject missed the activity selection in all 12 tasks, two missed this in 10 tasks, four missed it in two tasks, and one subject missed it in 1 task. Analysis involving selected activities does not include these missing cases. Subjects spent roughly between half an hour and an hour and a half completing the *Study 2* (median = 40 minutes, mean = 50 minutes).

Table 29 shows the basic statistics for each of the questions in the task survey. Recall that responses were measured on a Likert scale from 1 to 7. NASA TLX survey items 1, 4, 5 and 6 were measured with a continuous scale from 0 to 1. The results of the pretest revealed that subjects had a relative high level of experience ($Median$ = .79). Only one subject scored less than 50%. I further classified the subjects into a *pretest group*: low or high. I grouped using the median as a compromise to avoid having very small groups in the statistical analyses.

Table 29: Statistics of task surveys.

Question	Mean	SD	Question	Mean	SD
confidence	5.97	0.98	avoidHarder	4.83	1.83
usefulChoose	5.80	1.05	criticalEfficacy	5.18	1.46
ledUselessR	3.53	1.81	criticalEfficiency	5.7	1.14
findLearn	6.13	0.90	TLX1	0.29	0.27
thinkKnowledge	6.14	0.92	TLX4	0.20	0.22
notHelpfulR	2.92	1.52	TLX5	0.27	0.24
avoidEasier	5.63	1.23	TLX6	0.15	0.18

8.2.2 Survey differences among visualizations

Averages of survey responses show a general tendency to evaluate the treatment KCG higher, although significant differences were not found. Since correlations were significant and high for many pairs of questions in the survey, and before advancing with more elaborated analyses, I performed a Factor Analysis using Varimax rotation. Three factors were discovered. The first factor groups together the items *confidence*, *usefulChoose*, *findLearn* and *avoidEasier*. Since *confidence* is conceptually a different aspect, I created the score *USEFUL* only averaging *usefulChoose*, *findLearn* and *avoidEasier*. The second factor discovered contains *criticalEfficacy*, *criticalEfficiency* and *thinkKnowledge*. Again, this last item is conceptually different, so I computed the score *CRITICAL* by averaging *criticalEfficacy* and *criticalEfficiency*. The third factor groups the reversed questions, *ledUselessand* and *notHelpful*, which I averaged in the score *UNHELPFUL*.

To uncover differences among treatments (visualizations), I performed repeated-measures ANOVA methods on the scores *USEFUL*, *CRITICAL* and *UNHELPFUL* by treatment. Pretest-group (high, low) was added as a between-subjects factor. A significant effect of treatment was found for the score *USEFUL*, $F(1.4, 37.7) = 3.961$, $p = .041$, partial $\eta^2 = .128$.

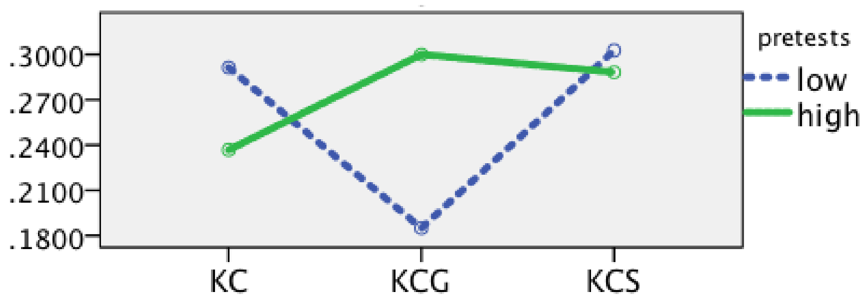


Figure 59: Interaction between treatment and pretest group for the measure of Effort (TLX5).

The sphericity assumption was violated in this analysis, so the Greenhouse-Geiser correction was applied. No significant difference was found for the other two variables *CRITICAL* and *UNHELPFUL*. Also, no significant effect of pretest group, nor interaction between pretest group and treatment were found. Simple contrast (comparing KCG against KC and KCS) showed a marginally significant difference between KCG and KCS, $F(1, 27) = 4.134$, $p = .052$, partial $\eta^2 = .133$, indicating higher evaluation of *Gauge*. However, more elaborated pairwise comparison using Bonferroni correction only found a marginal difference between treatments KC and KCS ($p = .074$). Subjects tended to judge the *USEFUL* of the visualization lower in KCS ($Mean = 5.604$) than in KC ($Mean = 5.953$) for the task defined in the study.

Similar analyses were run for TLX items (*mental demand*, *performance*, *effort* and *frustration*). No main effect of treatment was found for any of them, nor main effect of pretest groups. Nevertheless, a marginally significant interaction of the treatment and pretest groups was found for the perception of *effort* (TLX 5), $F(2, 54) = 2.936$, $p = .062$, partial $\eta^2 = .098$. Figure 59 shows this interaction: lower pretest group (which in fact represents subjects with a medium level of knowledge) expressed less *effort* when using the interface containing the *Gauge*. Similar patterns, despite not resulting in significant effects or interactions, were observed for the other TLX scores.

8.2.3 Behavior differences among treatments

Click activity collected while performing a task is summarized in the following variables:

- *countSelectActs*: number of activities selected in the task (subjects might have thought twice before going to the next task).
- *lastSelectedActDifficulty*: difficulty value of the last activity selected in the task which corresponds to the computed *estimated learning*, presented earlier in this chapter (see Equation 7.1 in Section 8.1.1) .
- *lastSelectedActRelativeRanking*: if all the activities that the user has moused over are ranked by their difficulty scores, this is the position of the last selected activity divided by the number of activities moused over. The value ranges between 0 and 1, 0 being the higher ranking.
- *countMouseoverActivities*, *timeMouseoverActivities*: number and sum of time spent in mouseover activities. I only counted mouseover actions that lasted for 1 second or more to reduce noise of involuntary actions.
- *countMouseoverConcepts*, *timeMouseoverConcepts*: number and sum of time spent in mouseover concepts (KCs). Similar to before, only mouseover actions of more than 1 second are counted.
- *countActivityOpened*: although I advised subjects not to open activities, in some situations they did so.

Table 30 reports mean and standard deviation (SD) of the variables computed. Note that subjects rarely moused over concepts. The difficulty of the last activity selected is close to the overall mean of difficulty ($Mean$ = .75, SD = .12). Very high correlations were found between *countMouseoverActivities* and *timeMouseoverActivities* (r = .89) and between *countMouseoverConcepts* and *timeMouseoverConcepts* (r = .84), thus I discarded the time variables and keep the counts in the following analyses.

To analyze differences of behaviors among treatments, I aggregated the log data variables grouping tasks within each treatment (4 tasks in each treatment) and performed repeated-measures ANOVA on log activity variables by treatment. Pretest group was added as a between subject factor. Subjects who did not select activities in tasks were removed from

Table 30: Log activity summary.

Variable	Mean	SD
countSelectActs	0.98	0.62
lastSelectedActDifficulty	0.75	0.13
lastSelectedActRelativeRanking	0.37	0.30
countMouseoverActivities	3.76	5.33
countMouseoverConcepts	0.67	1.63
timeMouseoverActivities	13.62	24.07
timeMouseoverConcepts	3.49	13.75
countActivityOpened	1.75	2.98

these analyses. The normality (Shapiro-Wilk) assumption holds only for the variable *lastSelectedActRelativeRanking*. Sphericity (Mauchly's test) holds for variables *lastSelectedActRelativeRanking* and *countMouseoverActivities*.

Results of the analysis found a significant effect of treatment only on *lastSelectedActRelativeRanking*, $F(2, 46) = 4.700$, $p = .014$, partial $\eta^2 = .170$. Pairwise comparisons with a Bonferroni correction showed a marginally significant difference between treatments KCG and KC ($p = .083$), and between KCG and KCS ($p = .053$). Subjects selected more difficult activities (relative to the difficulty of the activities inspected) when using KCG ($Mean = .299$, $SE = .038$), compared to when using KC ($Mean = .414$, $SE = .046$) or when using the KCS ($Mean = .410$,$SE = .033$). Figure 60 shows the pattern of this effect.

No significant interaction between treatment and pretest group was found for any of the log variables. However, a significant effect of pretest group was found for countMouseoverActivities, $F(1, 23) = 8.709$, $p = .007$, partial $\eta^2 = .275$, and countActivityOpened, $F(1, 23) = 6.477$, $p = .018$, partial $\eta^2 = .220$. High pretest subjects did fewer mouseoveryhI removed hyphen activities, but they opened activities more, regardless of the visualization.

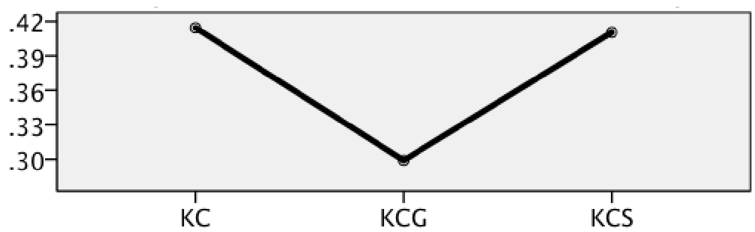


Figure 60: Ranking of (relative) difficulty of the selected activity. Lower value means higher ranking.

Table 31: Survey 2 summary. Count of rank preferences (rank 1 is top preference), and statistics on the ease of understanding and ease of use expressed by subjects.

	Ranking of Visualizations					Understand		Use	
	Rank1	Rank2	Rank3	Mean	SD	Mean	SD	Mean	SD
KC	0	13	15	2.54	0.51	2.48	1.43	2.38	1.42
KCG	20	6	2	1.36	0.62	2.21	1.50	1.90	1.23
KCS	8	9	11	2.11	0.83	2.52	1.50	2.48	1.43

8.2.4 Relations between survey and log variables

To better understand the subjective evaluation (survey), I now consider its relations to the log data (objective measures). Since log variables were collected by tasks and there was one survey for every two tasks, I aggregated log variables across tasks for each survey: Counting variables were added, whereas difficulty of the last activity selected and its ranking were averaged. Correlations (using Spearman) between task survey items and log variables revealed some interesting associations. In general, when subjects did more mouse over activities (which can be considered as more work) they lowered their perception of confidence in the task (*confidence* *countMouseoverActivities*, $r_s = -.222$, $p = .003$), they thought the system was less helpful to avoid harder activities (*avoidHarder* *countMouseoverActivities*, $r_s = -.281$, $p < .001$), but also declared lower frustration (*TLX6* *countMouseoverActivities*, $r_s = -.210$, $p = .006$). Variable *lastSelectedActDifficulty* was negatively correlated to both reversed measures *ledUseless* ($r_s = -.343$, $p < .001$) and *notHelpful* ($r_s = -.273$, $p = .001$), which suggests that positive perception of the support given by the system followed the selection of more difficult activities. Similar correlations were found for *countSelectedActs*, and this variable also shows a negative correlation with *frustration* ($r_s = -.239$, $p = .001$), which indicated less frustration when subjects did not complete the task in one shot. Finally, *countActivityOpened* was negatively correlated to *TLX4* (performance), which means that lower levels of failure were perceived after opening more activities.

8.2.5 Overall perception of the visualizations

At the end of the study session, subjects provided an overall evaluation of their experience. Table 31 summarizes the ranking that subjects gave to the three visualizations and the mean and standard deviation of the responses to questions about ease of understanding and ease of use. It can be seen the tendency of KCG to be considered as easier to understand and use, but differences were not significant. With reference to ranking, the KCG was considered the best by 20 subjects and the worst only for 2 subjects, with a Friedman test shows is a significant difference, $\chi^2(2) = 19.929$, $p < .001$. Free text explanations of the rankings were requested. Ten subjects explicitly referred to the advantages of using the

Gauge. For example, one subject said "the Gauge provides a summary/overview of the knowledge both the student have mastered and haven't learned, which saves a great bunch of time for comparing between different concepts and keeping a clear track of all processes". Five subjects expressed the value of social comparison features, for example "in the social comparison I have a direct and obvious guide as to where other skills are and therefore where my skills should probably be". Four subjects valued comparison as motivating: "comparison motivates us to perform better and improve our knowledge in the programming concepts". However, 7 subjects expressed a negative perception of these features: "I am not concerned about the progress of the class and how much I have completed when compared to them". Three subjects expressed concern about the gauge and how it works: "the gauge is somewhat distracting because some exercise covers concepts under other topics, and the number in the gauge always seduce me choose the one that can cover more new topics".

8.3 CONCLUSIONS OF THIS CHAPTER

This chapter presents a study in which I evaluated the Rich-OLM in the context of problem selection. To determine whether the Rich-OLM offers the right amount of information to support this task, I compared three versions of the Rich-OLM interface: a basic Rich-OLM, a version with a support tool to help the user in comprehending the OLM data (the gauge), and a version that offers additional information on the top of the basic version data by including social comparison in both topic and concept level. Evaluation also focused on a clearly defined task: to find activities to increase students' mastery of specific topics. This tasks was defined with the purpose of evaluating the potential support in content navigation that the interface features provide.

Results showed the positive effect of the gauge, especially in reducing the effort that less-prepared learners needed to complete the task, along with a very clear preference declared by subjects when comparing to other visualizations. These results help to confirm the idea that to allow effective navigation support while using a learning system, a fine-grained OLM can be enhanced with visual elements helping to interpret the data shown (which could in many cases be of high complexity) [Papanikolaou, 2015]. Thus, this study helps to confirm **H2G**: *Fine-Grained OLM complemented with the Learning Gauge helps students to navigate the content of the system more efficiently*. However, the study has only the purpose of validating the design. Extended environmental valid studies are needed to accept or reject this hypothesis. One such study is presented in the next chapter.

9.0 CLASSROOM STUDY 2: THE EFFECTS OF *RICH-OLM*

9.1 MOTIVATION

The main reason of adding the fine-grained component to Mastery Grids is to provide support for navigation, this is, to help students find useful content. In the previous chapter I evaluated the potential positive impact of the *Rich-OLM* and its fine-grained features in a *controlled* situation where the goal of using the system was fixed and defined as *finding the learning content that maximizes the real mastery of the participant*, i.e., a navigation task towards learning efficiency. However, the previous study has limitations, especially related to its environmental validity. We learned that the system is used in different ways, thus the task of *finding learning content to maximize mastery*, although a reasonable learning task, might not be what students necessarily set as a goal when using the system.

To extend the evaluation of the Rich-OLM, I performed a semester-long classroom study designed to contrast different version of the interface in a real learning environment. The study aims to answer research question 2 *What are the effects of fine-grained OLM on system activity?*, focusing on the effects on navigational support. Recall the hypotheses **H2** stated in chapter 5 *Fine-Grained OLM helps students to navigate the content of the system more efficiently*, and the later addition of **H2G** in chapter 7 *Fine-Grained OLM complemented with the Learning Gauge helps students to navigate the content of the system more efficiently.* The study also contributes research questions 3 and 4, which involve the exploration of individual differences and the change of motivation when the system includes a fine-grained OLM.

Similarly, than the study described in chapter 6, we offered Mastery Grids, now Rich-OLM, as a voluntary practice system in the course *CS-A1111, Basic Course in Programming Y1* in Aalto University during Fall term 2016. This course covers basic programming contents and uses Python as a programming language. As described in chapter 6, this course traditionally receives several hundreds of students from different programs different than Computer Science, and include bachelor students at the School of Engineering and at the School of Electrical Engineering. Since many aspects of the study presented in this chapter are similar to the previous study and for simplicity, I will now on refer to the previous study as the “2015 study” or as chapter 6 study. Again, the current study was made possible thanks to the fruitful collaboration with the course instructor and a researcher that helped to set up the system at Aalto University.

The structure of this chapter is similar than chapter 6. First, the design of the study is described, followed by data collection and pre-processing, including questionnaire reliability and factor analyses. Then the differences are analyzed in relation to prior and post knowledge (pretest, posttest) to verify possible distortions in the treatment groups. Then the analyses focus on the effects of treatment group features in system activity (*engagement*, *navigation* and *performance*) adding prior knowledge, social comparison orientation and learning motivation as factors. Finally, the analyses look into the change of motivation.

9.2 STUDY DESIGN

9.2.1 Course context

As in the previous 2015 study, the version of the system, now Rich-OLM includes Python content in 14 topics: Variables, Comparison, If Statement, Logical Operators, Loops, Output Formatting, Function, Lists, Strings, Dictionary, Values and References, Exception, File Handling, Classes and Objects. Four types of content were included: 37 parameterized problems, 32 parsons problems, 39 animated examples, and 59 annotated examples.

Access to Rich-OLM was provided with a personalized link to each student from the mandatory exercise platform, and in the same way than before, students did not have to log in Rich-OLM separately. Also, the same incentive was offered: 3% of extra credit on the *exercise grade* was given to whom at least solved 15 problems in Rich-OLM. *Exercise* *grade* contributes in 50% of the course grade, so the total impact of the extra credit offered for using Rich-OLM is about 1.5% of the course grade. We offered such bonus to encourage students to try the system.

Although we used the same content than in the study reported in chapter 6, there were some differences in the formal course structure and mandatory content between the terms that are worth to mention. The main difference was that in 2015, there were 18 lectures together covering all course topics, and the main online learning resource was a static PDF file [1] written by the lecturer. In 2016, the number of the lectures was reduced to 10, the lectures concentrated on basic concepts, and the students were asked to self-study the more advanced examples using an interactive e-book [2] which was developed from the previous year static resource. The text in both books was almost the same. However, the e-book included interactive animations and annotated code examples. Moreover, in 2016 there was also a direct link from the exercises to the corresponding e-book chapter, while in 2015 no such link existed. All of the interactive animation or examples contained in the e-textbook were different than the ones available in Rich-OLM. Since these course differences may account for performance and motivational differences between terms, I decided to analyze this study by itself, and not in conjunction with the previous Fall 2015 study. However, comparative analyses are presented in the next chapter.

9.2.2 Treatment groups

Students were randomly assigned into 3 "treatment groups" with different versions of Rich-OLM. The three versions offered were the same than in the laboratory study described in Section 8.1.1 in chapter 8 and were called *KC*, *KC+Gauge* (or *KCG*) and *KC+Social* (or simply *KCS*). As it can be seen in figures 61, 62 and 63, all versions include the topic based OLM and the fine-grained OLM. The variations among versions consider the presence of the *learning gauge*, and *the social comparison features*, both at the coarse- and fine-grained levels. Similarly, like I did in the 2015 study, groups were subdivided for technical reasons in 2 subgroups each (6 subgroups in total).

[1] http://www.cse.hut.fi/fi/opinnot/CSE-A1111/S2015/kalvot/opetusmoniste2015.pdf
[2] https://grader.cs.hut.fi/static/y1/

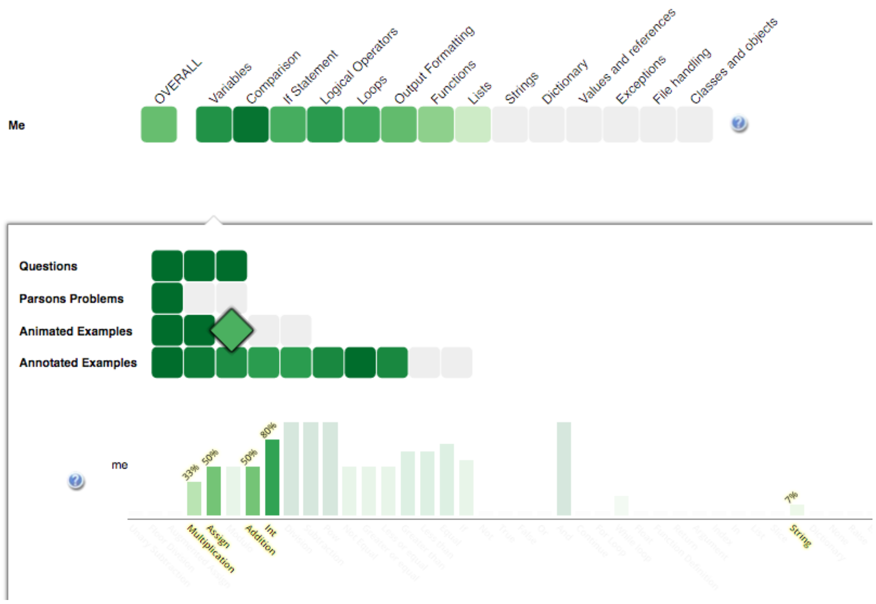


Figure 61: RichOLM interface without social comparison and without gauge. This interface was used by 1/3 of the students in the group called “KC”.

Figure 62: RichOLM interface with learning gauge. This interface was used by 1/3 of the students in the group called “KCG”.

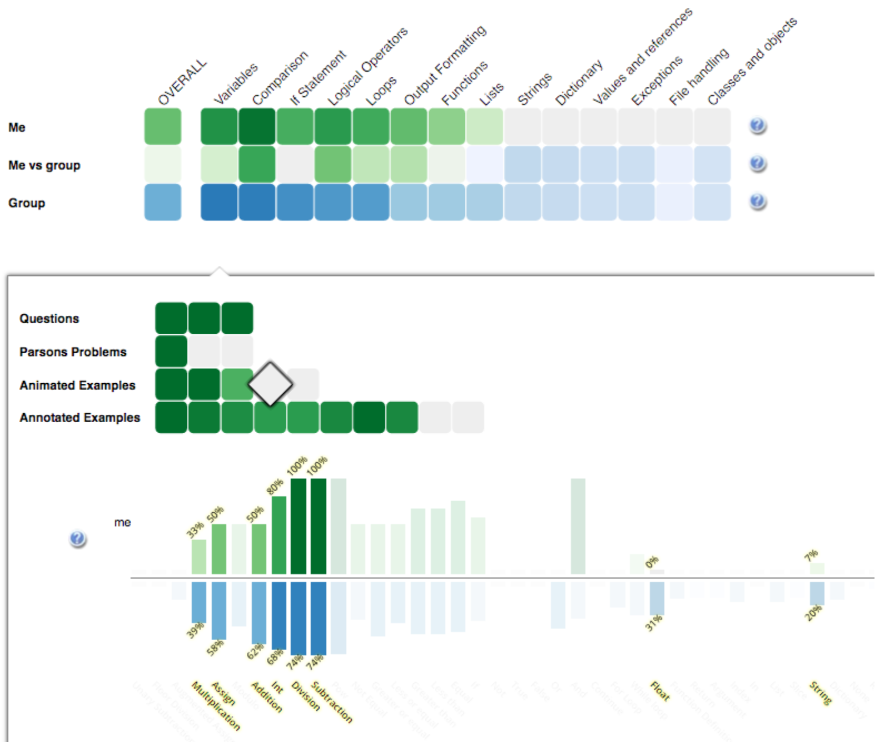


Figure 63: RichOLM interface with social comparison and without gauge. This interface was used by 1/3 of the students in the group called “KCS”.

The setup of the study with three experimental groups was intended for combined analyses with the previous study (chapter 6). The idea was to compare groups of the current study to groups of the previous study, and at the same time, being able to contrast different Rich-OLM configuration (*gauge*, *social*) with strong statistical power. Then, research question 2 *What are the effects of fine-grained OLM on system activity?* and its related hypothesis **H2** *Fine-Grained OLM helps students to navigate the content of the system more efficiently* could be fully answered in analyses involving control groups from the previous study. However as stated before in Section 9.2, there were some differences in the course implementations between the years that to some extent jeopardize the cross study comparisons. Because of this reason, I decide to concentrate in this chapter only the analyses within the 2016 study, and perform analyses cross studies in the next chapter.

9.2.3 Data collection

Data collection was similar than in the previous classroom study reported in chapter 6. Pretest and posttest were collected at the beginning and the end of the term, respectively. Both pretest and posttest were created as an online survey using Qualtrics system (provided by the Katz School of Business at the University of Pittsburgh) and both include the same 10 questions of python programming (see APPENDIX A). Similarly, like pre and posttest, the Motivation Questionnaire (see APPENDIX B) was implemented as an online survey using Qualtrics and applied at the beginning and at the end of the term. In the questionnaire applied at the end of the term, we also included the Social Comparison Orientation set of questions (see APPENDIX C).

Rich-OLM was not enabled immediately, but a week after the pretest and initially included only a basic individual view for all groups without fine-grained components, nor social comparison features, nor gauge. This initial version was the same than the *Individual* version of the previous study and can be seen in Figure 21. Two weeks later, the different treatment versions were introduced. The system was re-introduced during lectures and a link to a PDF containing a user manual of the system was included in the top right corner of the Rich-OLM interface. This manual was different for each treatment group and covers the specific features of each version of the interface. The late introduction of the features was due to technical problems in the development of the Rich-OLM and in the personalized link included in the mandatory exercise system. Since this delay and the re-introduction of the system will probably impact the patterns of system engagement, these represent another reason to avoid directly comparing groups in this study with the groups in the study reported in chapter 6.

System activity involves several variables concerning the completion of the activity, the activity in different types of content, the regularity of use, and the performance on self-assessment content such as the Questions and Parsons problems. These measures of system activity are described in chapter 5. Since the system features include now the concept visualization and the *learning gauge*, the system was also enabled to track activity on mouseover in the cells of the interface, and the value of the relative *difficulty* as computed for the learning gauge (see Section 8.1.1 in chapter 8), which is also available in the treatment groups where the gauge is disabled. Thus 2 new system activity variables are included in the analyses in this chapter and are described below.

Difficulty of open activities (*act_difficulty*) Difficulty is a relative measure of the effort that an activity could take to the student. To approach a measure of difficulty, I use the estimation of the level of knowledge that the user model engine maintains in each of the concepts related to the activity and then count how many of these concepts are "Known", "Familiar" or "New", depending if their estimated level of knowledge is above or below certain thresholds. Then, the difficulty is computed by weighting more the *new* concepts than the *familiar* and *known* concepts. The rationale behind this is that the more *new* concepts are in an activity, the more difficult it will be for the student. Details of the computations are shown in Section 8.1.1 of chapter 8 and in equation 7.1. The difficulty on open activities is considered an indicator of *engagement* and *navigation*, since it could reflect differences that can be attributed to the navigational support that the interface features provide.

Mouseover activity cells (*mouseover_act*) The Rich-OLM also implemented more logging capabilities, including recording all mouseover activity in the interface. From this data, I use the mouseover on cells that correspond to activities. Recall that mouseover on an active cell activates the concepts related to the concepts visualization (in all treatment groups) and shows the gauge (in *KCG* group). The, mouseover activity cells is an important measure that reflects pattern of *navigation*, and is a direct measure of how much the students use the new features. All mouseover actions that last for less than one second were discarded to avoid counting the involuntary or transitional mouseovers.

9.3 DATA OVERVIEW AND PRE-PROCESSING

9.3.1 Data collected

A total of 711 students were assigned to Rich-OLM accounts. This represents the number of students that initially enrolled in the course. However, 552 students (78%) finished the course taking the final exam. In general, there are more males than females and the proportion reaches 77% of males among students who provided this information (N=647). This is the same proportion than in the previous study reported in chapter 6. Among students who finished the course, 336 students did at least some activity within the system (active students). Nineteen students have activity in Rich-OLM, but dropped out from the course. A relatively large proportion of students completed pretest, posttest and motivation questionnaires, as shown in Table 32.

Table 32: Number of students who completed the course (take exam), answer questionnaire at the initial (i) and final (f) term points, including the Social Comparison Orientation questionnaire (SC), and number of students who did activity in Rich-OLM among who completed the course (active).

	Take exam	pre + post	Motiv (i)	Motiv (f)	Motiv (i+f)	SC survey	active
All	552	458	647	454	444	454	336
KC	176	146	210	146	141	146	104
KCG	189	160	219	155	152	155	111
KCS	187	152	218	153	151	153	121

Average of prior and post knowledge (pre and posttest) in the treatment groups, including only students who finished the course and had some activity in the system (*active*), are shown in Table 33. We observe similar performance values in all treatment groups, with a slightly smaller pretest and posttest in the *KCG* group. The differences on these measures among groups are analyzed later in this chapter. However, looking at the knowledge gained I discovered several cases with negative values. Figure 64 shows a scatterplot of pretest and posttest. A small noise was added to these measures in the Figure in order to visualize better the overlapping points. As can be seen in the figure, there are several points under the diagonal of 0 learning gain (22 cases). I further remove these cases of analyses involving posttest.

Table 33: Summary statistics of performance measures.

		All	KC (N=104)	KCG (N=111)	KCS (N=121)
pretest	Mean	0.202	0.213	0.182	0.213
	SD	0.208	0.223	0.182	0.219
	SE	0.009	0.017	0.013	0.016
posttest	Mean	0.663	0.688	0.645	0.660
	SD	0.241	0.230	0.251	0.241
	SE	0.011	0.019	0.020	0.020
lgain	Mean	0.526	0.543	0.511	0.528
	SD	0.679	0.789	0.732	0.480
	SE	0.032	0.066	0.058	0.039

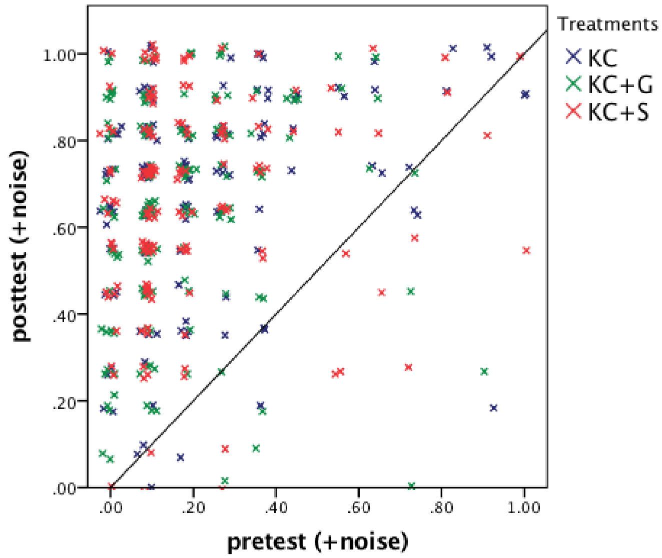


Figure 64: Pretest versus posttest.

Table 34: General statistics of engagement variables.

	KC			KCG			KCS		
	Mean	SD	SE	Mean	SD	SE	Mean	SD	SE
mg_completion	0.257	0.223	0.022	0.297	0.241	0.023	0.236	0.188	0.017
n_questions	26.644	21.607	2.119	27.622	21.552	2.046	21.934	16.185	1.471
n_parsons	29.510	35.590	3.490	31.640	33.301	3.161	23.719	22.701	2.064
n_examples	7.365	12.757	1.251	8.811	13.816	1.311	7.587	11.205	1.019
n_ae	9.269	11.187	1.097	12.739	11.904	1.130	8.736	9.204	0.837
sr_questions	0.621	0.188	0.019	0.606	0.165	0.016	0.647	0.173	0.016
sr_parsons	0.487	0.239	0.025	0.473	0.196	0.020	0.516	0.244	0.023

Levels of system activity in Rich-OLM are shown in Table 34. These values average across active students who completed the course. In general in further analyses we consider only students who completed the course because students who dropped might have had different reasons to disengage.

9.3.2 Questionnaire Reliability and Factor Analyses

Before summarizing and using the measures of motivation and social comparison orientation, I performed reliability analysis to verify that the answers of all items within each factor measured are consistent.

Social Comparison Orientation (*SCO*) was measured using the *INCOM* instrument ([Gibbons and Buunk, 1999], and see APPENDIX C), which have 11 statements (9 positive, 2 negative, which score was reversed) about the inclination to compare to others. Reliability reaches a Cronbach's Alpha of .750 which is acceptable. Factor analyses failed to find the two theoretical orientations described in the literature as *ability* and *opinion* [Gibbons and Buunk, 1999]. Table 35 shows the *loadings* of the factor analyses. The theoretical orientation assigned by the literature to each item is in the column "Orientation". As it can be seen in the table, data extracted factors are not clearly aligned to these theoretical factors. With this, I opted to compute a unique score of Social Comparison Orientation (*SCO* score) by averaging the scores of all 11 statements (items 6 and 10 were reversed before computing the SCO score).

Table 35: Results of the Factor Analyses on Social Comparison Orientation Questionnaire. Rotated matrices show the loadings greater than .3.

		Component		
Question	Orientation	1	2	3
1	ability	0.715		
2	opinions			0.464
3	ability	0.653		
4	ability	0.301		0.579
5	opinions		0.619	
7	ability	0.497	0.420	
8	opinions		0.806	
9	opinions		0.711	
11	ability	0.653		
6 (R)	ability	-0.736		
10 (R)	opinions			0.721

Regarding the motivation questionnaire, reliability analyses showed good scores for all factors. The lower value at both initial and final measure point was Mastery Avoidance (*MAv*) which has a reliability score (Cronbach's Alpha) below of .730 and .705, respectively. Table 36 shows the scores of all measures.

Principal Component Analysis (PCA) was performed with extraction based on Eigenvalue and Varimax rotation with Kaiser Normalization to corroborate that the different groups of questions within the instrument are measuring different factors. Regarding the section of Learning Activation, PCA extracted 3 factors which explained the 62.769% and the 65.869% of the variance in the initial and final questionnaires, respectively. The *loadings* (associations between each question and the latent extracted factors), that can be seen in Table 37, matched the designed factors *Fascination*, *Competency Beliefs*, and *Values*. This

Table 36: Reliability analyses of the Motivation questionnaire taken at the beginning of the term (initial) and at the end of the term (final).

	Cronbach's Alpha						
	F	CB	V	MAp	MAv	PAp	PAv
Initial	.831	.818	.830	.779	.730	.889	.884
Final	.879	.812	.848	.823	.705	.895	.845

means that according to the answers, we can distinguish the 3 theoretically defined motivational factors. As explained before in chapter 5, Section 5.3.3, I will only include the *Competency Beliefs* (*CB*) factor in further analyses. However, the factor analysis of the other motivational constructs is important because it verifies that the construct CB is different (enough) than other motivational traits such as Fascination and Values.

In the Achievement-Goal Orientation section, PCA also produced very similar results than in the previous study. Three factors were recognized having *Performance Approach* and *Performance Avoidance* loading together, as if students did not distinguish the Approach-Avoidance distinction of this dimension. Also similarly than in the 2015 study, the first item of the *Mastery Avoidance* factor loaded strongly within the *Mastery Approach* construct. Factors extracted explained the 71.9% and the 72.5% of the variance in the initial and final questionnaires, respectively. Table 38 shows the results of the PCA analysis. Loadings lower than 0.3 has been removed to facilitate the interpretation of the table.

Results of both reliability and factor analyses on the Learning Activation and Achievement-Goal questionnaires are very similar than the previous 2015 study, thus I make the same conclusions and decide to discard Mastery Avoidance and Performance Avoidance in the further analyses.

Table 37: Results of the Factor Analyses on the Learning Activation section of the Motivation Questionnaire. Rotated matrices show the *loadings* greater than 0.3.

	Component				Component		
	1	2	3		1	2	3
F1i		0.785		F1f	0.718	0.318	0.31
F2i		0.776		F2f	0.801		
F3i		0.819		F3f	0.766		
F4i		0.659		F4f	0.79		
CB1i	0.748			CB1f	0.443	0.635	
CB2i	0.817			CB2f		0.781	
CB3i	0.816			CB3f		0.757	
CB4i	0.8			CB4f		0.728	
CB5i	0.469			CB5f		0.604	
V1i		0.46	0.545	V1f	0.538		0.531
V3i			0.742	V3f			0.768
V4i		0.36	0.737	V4f	0.409		0.716
V5i		0.357	0.733	V5f	0.463		0.673
V6i			0.793	V6f			0.795

Table 38: Results of the Factor Analyses on the Achievement-Goal section of the Motivation Questionnaire. Rotated matrices show the loadings greater than 0.3.

	Component				Component		
	1	2	3		1	2	3
MAp1i		0.801		MAp1f		0.848	
MAp2i		0.761		MAp2f		0.78	
MAp3i		0.833		MAp3f		0.859	
MAv1i		0.587	0.389	MAv1f		0.66	
MAv2i			0.849	MAv2f			0.857
MAv3i		0.301	0.784	MAv3f			0.813
PAp1i	0.813			PAp1f	0.805		
PAp2i	0.837			PAp2f	0.872		
PAp3i	0.844			PAp3f	0.881		
PAv1i	0.802			PAv1f	0.782		
PAv2i	0.798		0.367	PAv2f	0.784		
PAv3i	0.842		0.308	PAv3f	0.847		

9.3.3 Statistics of questionnaire

The results of the Learning Motivation and the Social Comparison Orientation questionnaires are shown in Table 39. Scores have been computed by averaging items of the questionnaire and by moving them to the range of (0 ,1), both inclusive, doing a simple linear transformation. We have considered all responses to these questionnaires from students who finished the course. Histograms of the motivation factors are shown for the initial measures in Figures 65,66,67. Histogram of Social Comparison Orientation is shown in Figure 68. The segmented shape of the histograms reveal the low resolution of the scale. For example, Mastery and Performance Approach are each scored using 3 questions with alternatives between 1 and 7 points each, which means that the scale of each factor has a minimum of 3 and a maximum of 21 (before normalizing), which are 19 discrete positions in the continuous.

Table 39: Basic statistics of motivational factors.

		Initial	Final
Competency	Mean	0.477	0.682
Beliefs	SD	0.215	0.187
	SE	0.009	0.009
Mastery	Mean	0.700	0.675
Approach	SD	0.176	0.192
	SE	0.008	0.009
Performance	Mean	0.584	0.590
Approach	SD	0.224	0.220
	SE	0.010	0.011
Social Comparison	Mean		0.584
Orientation	SD		0.118
	SE		0.006

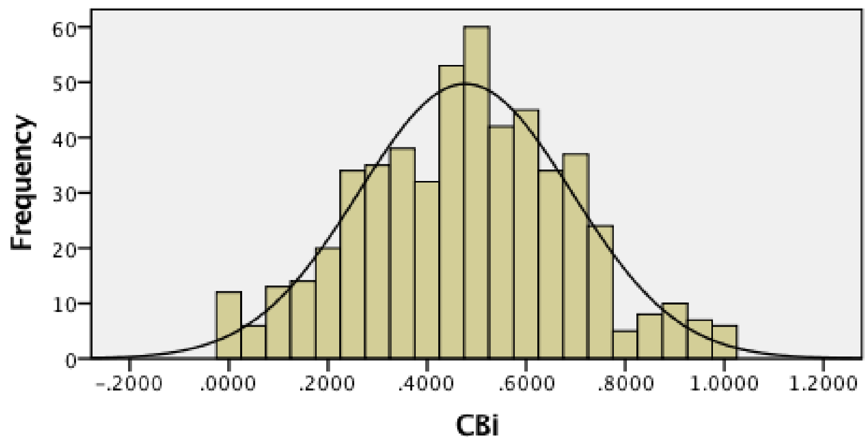


Figure 65: Histogram of Competency Beliefs measured at the beginning of the term.

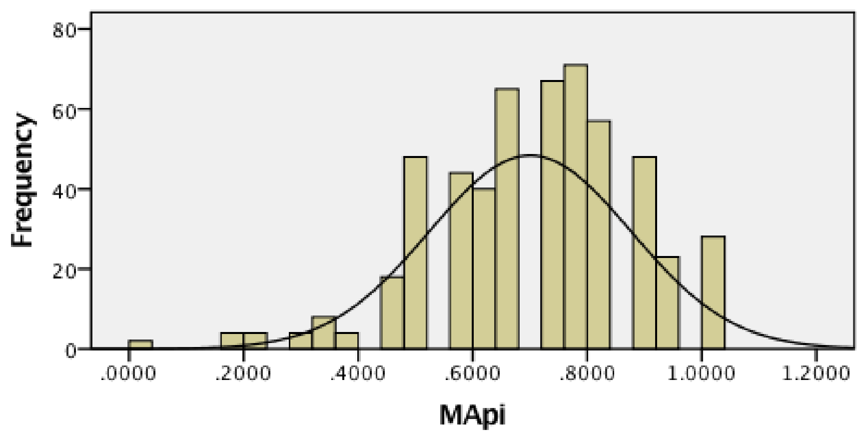


Figure 66: Histogram of Mastery Approach measured at the beginning of the term.

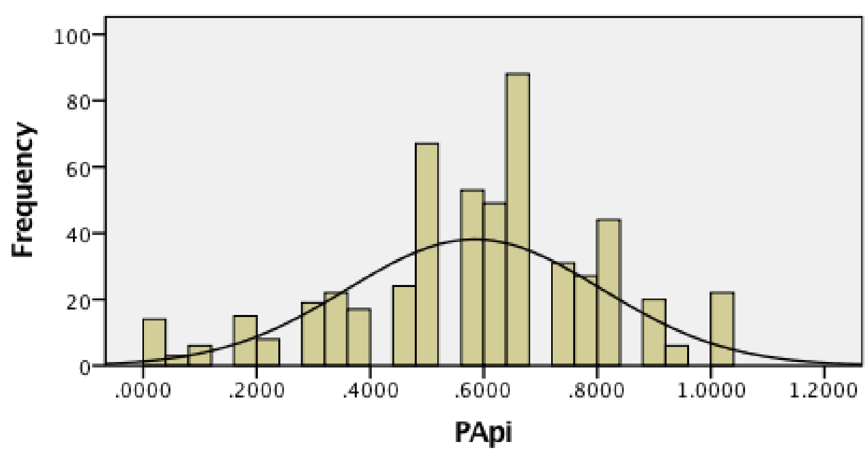


Figure 67: Histogram of Performance Approach measured at the beginning of the term.

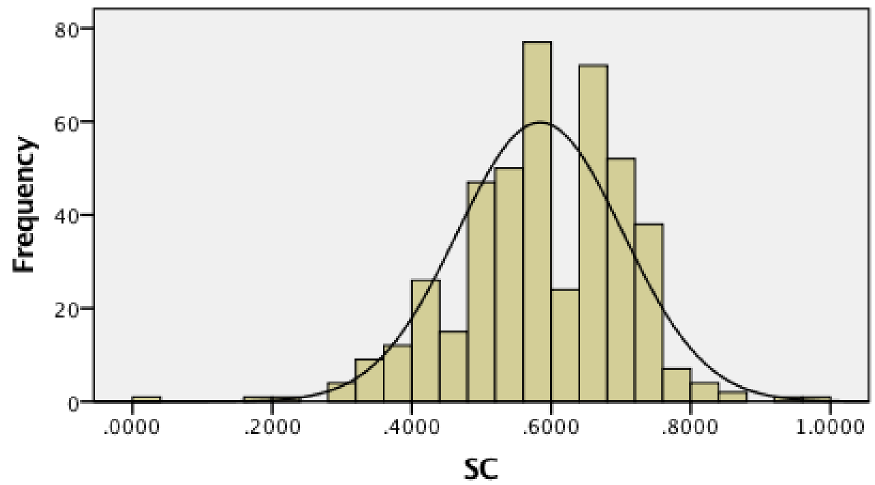


Figure 68: Histogram of Social Comparison Orientation.

As I noted in the previous similar study in chapter 6, Social Comparison Orientation presents a relatively low variance (low Standard Deviation, SD), which may impact in the *power of discrimination* of this measure in explaining the potential effects of the social comparison features of the system. This consideration is taken later when this factor is included in the analyses.

9.3.4 Initial motivation across groups

The mean and standard deviation of the motivational factors measured at the beginning of the term are shown in Table 40. This data consider all students who finished the course and took the questionnaires. Differences across groups are small and a non-parametric Kruskal-Wallis Test did not find any significant difference for any of these motivational measures (Table 41). These results confirm that groups are not significantly different in their motivational traits at the beginning of the term.

Table 40: Mean and Standard Deviation of motivational factors measured at the beginning of the term in each of the treatment groups.

	Fi		CBi		Vi		MApi		PApi	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD	Mean	SD
KC	0.576	0.159	0.500	0.216	0.720	0.160	0.699	0.151	0.581	0.235
KCG	0.572	0.181	0.467	0.211	0.691	0.167	0.701	0.190	0.593	0.221
KCS	0.587	0.175	0.465	0.217	0.700	0.164	0.701	0.185	0.577	0.217

Table 41: Kruskal-Wallis test on motivational factors measured at the beginning of the term across treatment groups (KC, KCG, KCS).

	Fi	CBi	Vi	MApi	PApi
Chi-Square	1.000	2.593	2.535	.425	.589
p-value	.607	.273	.282	.809	.745

9.4 PRIOR AND POST KNOWLEDGE DIFFERENCES

I start looking at prior and post knowledge levels between the treatment groups (*KC*, *KCG* and *KCS*). Prior and post knowledge is measured using pretest and posttest. The normalized learning gain is also reported. Also, the average of these values in the three treatment groups is reported in Table 33 (Section 9.5) and shown in Figure 69 presented here. Levels of pretest and posttest are very similar across groups.

Regarding pretest, differences observed in Table 33 reveal a slightly lower average score in the *KCG* group (M=.182) compared with the other two groups, *KC* (M=.213) and *KCS*

Table 42: Non-parametric test on performance measures between treatment groups.

	pretest	posttest	lgain
χ^2	0.330	1.176	0.593
df	2	2	2
Asymp. Sig.	0.848	0.555	0.743

(M=.213). Non-parametric Kruskal-Wallis test was performed among the *active* students (who at least did one activity in the practice system) and no significant differences were found for any of the knowledge levels (see Table 42).

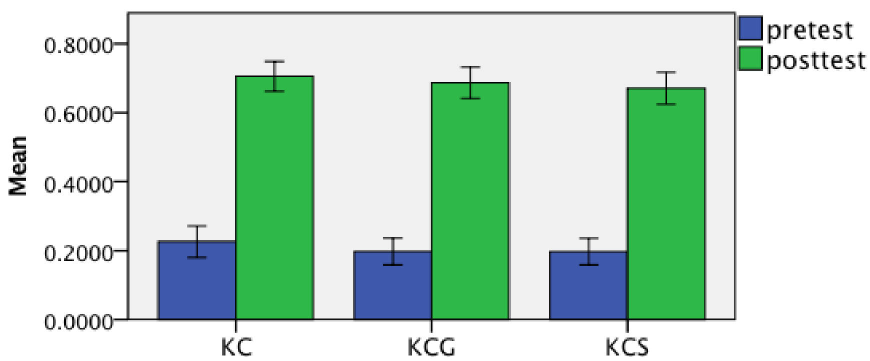


Figure 69: Average pretest and posttest among treatment groups.

An Anova test with Bonferroni correction performed on pretest among treatment groups confirmed the results, F(2,331)=.408, p=.665, partial η^2=.002, i.e., there was not a significant difference in pretest scores among the three treatments.

Learning gain also does not show a significant difference between the groups. To confirm this observation, I test for differences in posttest conditioned to pretest using regression analyses. An initial model using pretest (IV) to predict posttest (DV) is built, and then the factors *gauge* and *social* are added in a *stepwise* manner. Both of these factors are set as dummy variables taking the values 0 or 1. Results are reported in Table 43 and shows that only the pretest has a strong role predicting posttest. Neither social nor gauge resulted in a significant increase of the model, meaning that there is no impact of these features in the variability of posttest.

Table 43: Regression model built on posttest with predictors pretest, and dummy variables social and gauge. Significance '***' means p ¡ .001.

DV	$\beta_{pretest}$	β_{social}	β_{gauge}
posttest	.430 ***	.010	-.010

These results confirm that the non-existent differences neither in the pretest, nor in posttest, nor in the differences from pretest to posttest among the groups, as shown in Figure 69.

To completely discard the role of the interface features, in a second regression analysis I added the interaction terms *pretest*social* and *pretest*gauge*. Using *stepwise* regression, neither interaction term results in a significant improvement of the model, thus suggesting that pretest does not produce different effects in the treatment groups.

9.5 EFFECT OF SYSTEM PRACTICE ON POST-KNOWLEDGE

As I described before in chapter 6, the impact of the interface in learning is likely indirect, because students learn from multiples sources. The real effect of the interface is most likely to lay on the engagement and navigation with the practice system. To explore this, I

Table 44: Regression models on posttest including pretest, mg_completion and its interactions with *social* and *gauge*. Significance is marked with the cell background color and with symbols '.' (.1-.05), '*' (.05-.01), '**' (.01-.001), '***' (<.001).

Model	R^2	$\beta_{pretest}$	β_{mg}	$\beta_{mg*social}$	$\beta_{mg*gauge}$
1	.125	**.430** ***	-	-	-
2	.139	**.425** ***	**.109** *	-	-
3	.139	**.425** ***	**.109** *	-.047	.026

performed regression analyses on *posttest* as the dependent variable. Models are built consecutively adding predictors. The first model adds *pretest*. The second model adds the level of completion of activities in the system, measured by *mg_completion*. The third model adds the interactions *mg_completion*social* and *mg_completion*gauge* using a *stepwise* method. For these analyses, I discarded students that presented negative learning gains, as described before in section . With this approach, I could test the overall relevance of the completion of the activity and then test if this activity in the system had a different role across treatment groups.

Results reported in Table 44 show that after pretest, the level of completion in the system is a positive predictor of posttest, and that it is not conditioned to the interface features. These results confirmed the positive effect on posttest associated with doing activity in the system. I also notice that the β coefficient of the predictor *mg_completion* ($\beta = .109$) is lower than the value obtained in the previous study (chapter 6, Section 6.4, Table 13).

9.6 THE IMPACT OF PRIOR KNOWLEDGE IN STUDENT ENGAGEMENT

I now explored the role of pretest in the engagement with the practice system. In the previous study (chapter 6), we found that pretest was a significant predictor of the activity in the system only in the *social* group. In this study, we observe a similar trend that was, however, not as strong. First, Pearson correlation shows a significant relation between *pretest* and *mg_completion* in the *KCS* group ($Pearson = .248$, $p = .007$), but not significant in *KC* group ($Pearson = -.021$, $p = .834$), nor in *KCG* ($Pearson = .025$, $p = .791$). Also, similarly than in the previous classroom study, *Competency Beliefs* did not show any significant correlation with *mg_completion*.

To confirm these observations, I ran a multiple regression on *mg_completion* with predictors *pretest* and *Competency Beliefs*, separately in each treatment group. Results are shown in Table 45. None of the model resulted in significant prediction. The stronger effects observed are the predictive strength of *pretest* in the *KCS* group (as noted before), and the negative effect of *Competency Beliefs* in the *KC* group (marginal). Altogether, and differently than what I observed in the previous study (chapter 6), neither *pretest*, nor *Competency Beliefs* seemed to explain the variances on the usage of the system.

Another round of regression analyses was run where *mg_completion* was set as dependent variable and *pretest*, *social*, *gauge*, and interaction terms *social*pretest* and *gauge*pretest* were added as predictors. The results were mild, showing only a marginally significant effect of the interaction *pretest*social* ($\beta = .263$, $p = .065$), and a marginally significant negative effect of *social* ($\beta = -.073$, $p = .079$). The same regression using the *stepwise backward* method amplified these effects, but *forward* regression failed to show them. These results suggest that the selective effect of *pretest* in the *social* group exist but is weak, and that there is no selective effect of *pretest* regarding the *gauge* feature on the overall engagement with the practice system.

Table 45: Regressions on mg_completion with predictors pretest and competency beliefs (CB).

	R^2	$\beta_{pretest}$	β_{CB}
KC	.039	.075	**-.226 .**
KCG	.001	.025	.030
KCS	.019	.134	-.002

9.7 EFFECTS ON SYSTEM ACTIVITY

In the previous section, I had observed that the role of prior-knowledge is null in predicting the engagement, which is a counter-intuitive observation and contrast with previous findings. However, the positive relation between pretest and social comparison features is weak but still observable, confirming (although weakly) the previous findings in chapter 6. In this section I look closer to the effects of the different interface options (*social*,*gauge*) in several system activity variables beyond *mg_completion*, and which represents measures of different aspects of using the practice system: *engagement*, *navigation*, and *performance*.

9.7.1 Overall differences

Table 46 shows the mean and standard deviation of the system activity variables for each of the treatment groups. Notice higher levels of engagement (completion, attempts to questions, Parsons, etc) in the *KCG* group and lower levels of the same variables in the *KCS*. Further analyses using regressions will search for significance regarding these observations.

Figure 70 shows a density plot of the system activity (as the total amount of activity clicks in the practice content) over the term for each treatment groups. The figure allowed me to see potential differences in the pattern of activity during the term. Note that *KCS* group presents higher levels of activity at the beginning which sustain over a month, while levels of

Table 46: System activity by treatment.

	KC		KCG		KCS	
	Mean	SD	Mean	SD	Mean	SD
mg_completion	0.257	0.223	0.297	0.241	0.236	0.188
n_questions	26.644	21.607	27.622	21.552	21.934	16.185
n_parsons	29.510	35.590	31.640	33.301	23.719	22.701
n_examples	7.365	12.757	8.811	13.816	7.587	11.205
n_ae	9.269	11.187	12.739	11.904	8.736	9.204
term_regularity	0.467	0.083	0.482	0.092	0.459	0.080
eff_questions	-0.003	0.444	-0.032	0.392	0.032	0.401
eff_parsons	0.033	0.473	-0.037	0.508	0.005	0.477
sr_questions	0.621	0.188	0.606	0.165	0.647	0.173
sr_parsons	0.487	0.239	0.473	0.196	0.516	0.244
prob_attempt	0.584	0.240	0.522	0.219	0.555	0.216
ratio_gui	0.295	0.132	0.324	0.125	0.342	0.135
mouseover_act	13.817	13.967	18.252	18.114	11.463	14.109
act_difficulty	0.503	0.294	0.478	0.278	0.453	0.301

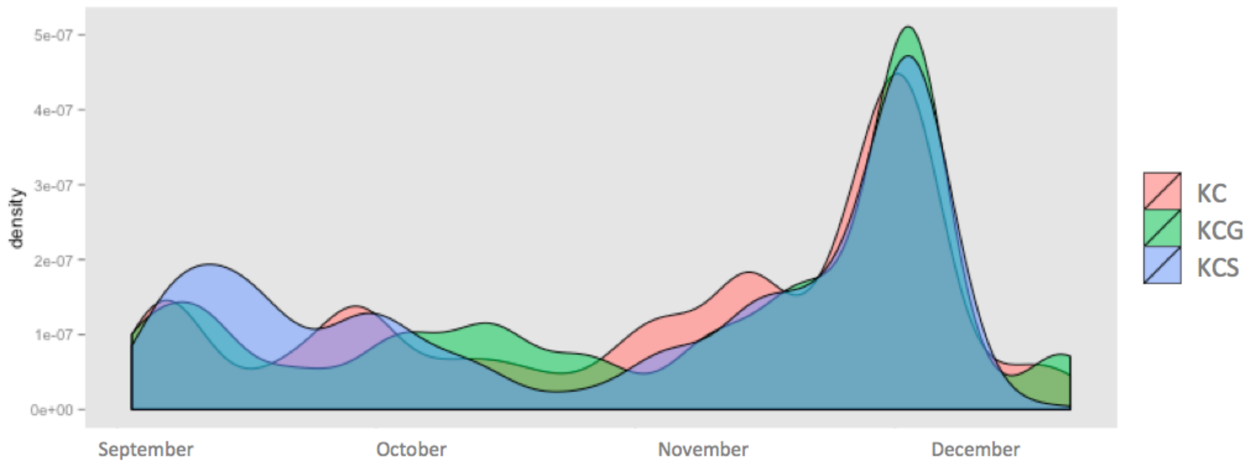


Figure 70: Density plot of system activity over the term.

activity of the *KCG* group increased towards the second month, a period that represents a "valley" on the activity of the other two groups (*KC*, *KCS*). These observations point to the idea that even when interface features might not have an impact on the overall engagement levels, they might impact in the way the students use the system during the term. However, regression analyses performed on the amount of activity in the first two weeks and in the first 4 weeks did not show any significant role of the interface features.

Figure 71 shows overall levels of engagement (mg_completion) across the treatment groups. A higher levels can be seen for the *KCG* group. In the next I explore this differences formally together with other variables that account for engagement and navigation in the system.

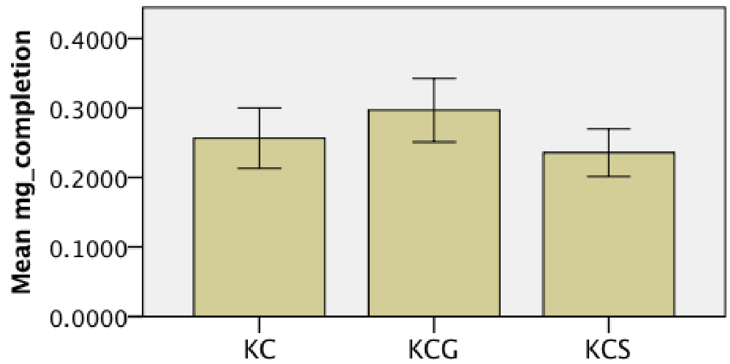


Figure 71: Mean completion of activity across groups. Error bars represent two standard errors of the mean.

9.7.2 Regression analyses

To explore the effects on system activity I performed linear regressions on each of the system activity variable as dependent variables with predictors *pretest*, *gauge*, and *social*, the later two predictors set as dummy variables indicating the presence of the *gauge* and the *social* features, respectively. Also the interaction terms *pretest*social* and *pretest*gauge* were added. Since an external incentive was offered to perform at least 15 problems in the system, I repeated these analyses for students who went beyond this incentive threshold (more than 15 activities). Table 47 shows the β coefficients resulting of these regression models with their significance. The results of the table are analyzed and commented from the perspective of *engagement*, *navigation* and *performance* with the system.

9.7.3 Effects on *engagement*

Looking at the engagement variables (rows 1-6 in Table 47), I notice only marginal effects that tend to disappear after the incentive threshold ($>$ 15 activities). I observe a consistent negative effect of the social features that seems to be compensated by the positive

Table 47: Results of regressions on system activity variables (rows) in two cases: all the students (columns 2-6), and for students who have engaged beyond the 15 activities incentive threshold (columns 7-11). Values are raw β coefficients. Significance is marked with the cell background color and with symbols '.' (.1-.05), '*' (.05-.01), '**' (.01-.001), '***' (<.001).

	All					>15 activities				
	pretest	social	gauge	pre*social	pre*gauge	pretest	social	gauge	pre*social	pre*gauge
mg_completion	-0.021	**-0.073 .**	0.029	**0.263 .**	0.055	0.023	-0.073	0.049	0.252	0.077
n_questions	-12.248	**-7.431 .**	-1.304	13.104	10.200	-11.439	**-7.27 .**	0.226	9.052	11.442
n_parsons	-17.274	-9.16	-2.606	15.768	22.131	-22.834	**-12.227 .**	-0.553	18.69	28.855
n_examples	-0.906	-3.018	0.157	**15.746 .**	6.462	1.179	-3.109	1.025	16.265	8.672
n_ae	-1.274	-2.908	**4.146 .**	**11.648 .**	-3.599	2.029	-2.636	4.283	11.894	-1.827
term_regularity	-0.006	-0.002	0.017	-0.029	-0.014	-0.019	0.005	0.027	-0.024	0.02
eff_questions	**0.341 .**	-0.019	0.021	0.305	-0.212	**0.507 ***	-0.004	0.076	0.275	-0.453
eff_parsons	0.304	-0.04	-0.073	0.091	0.053	0.341	-0.046	-0.1	0.067	0.129
sr_questions	**0.207 ****	0.023	0.006	0.03	-0.085	**0.171 ***	0.018	0.038	0.096	-0.193
sr_parsons	**0.303 ****	0.055	0.026	-0.096	-0.158	**0.337 ****	0.036	0.034	-0.073	-0.177
prob_attempt	-0.007	-0.004	**-0.099 ***	-0.115	.193	-0.042	-0.002	**-0.079 .**	-0.120	0.119
act_difficulty	-0.151	-0.013	-0.027	-0.235	-0.083	-0.133	0.027	-0.02	-0.139	-0.154
ratio_gui	-0.004	**0.044 .**	0.034	0.006	-0.026	-0.07	0.034	0.025	0.072	-0.006
act_mouseover	-6.995	**-6.578 ***	3.622	**20.705 ***	3.326	-6.963	**-6.953 .**	4.427	**22.185 .**	6.563

role of the interaction with pretest (*pretest*social*), as we can see for *mg_completion* and the activity on examples (*n_examples*, *n_ae*). For high activity students, it seems that *social* explains, although only marginally significant, a decrease in self-assessment activity (*n_questions*, *n_parsons*) rather than a decrease in the descriptive content (examples and animated examples). *Gauge* shows an opposite effect, although not significant. Figure 71 shows the overall tendency.

A reason for the lower level of activity in the *social* group, which contrast with previous findings (chapter 6), is that adding social comparison features to both coarse- and fine-grained visualization might increase the complexity of the system to the point of discouraging students to use it. I had seen in chapters 7 and 8 that complexity of the visual interface is an important concern. However, more research is needed to explore these issues.

Regarding the effects of the *gauge* feature, results of engagement only shows a marginally significant positive effect of this feature in the number of animations viewed (*n_ae*), which disappears after the 15-activities incentive threshold. *Gauge* only shows effects on navigation indicators, which are discussed in the following section.

9.7.4 Effects on *navigation*

Effects on navigation variables (last 4 rows in Table 47) are noticeable and maintain (although weakened) after the incentive threshold. *Gauge* presents a negative effect on the probability of attempting an activity opened (*prob_attempt*). This means that when exposed to the *gauge*, students tend to open the activities and not attempting them, although they have similar levels (not lower) of attempts.

Notice the positive, although the not significant role of the *gauge* with respect to the mouseover variable. The mouseover activity variable is important to evaluate the *gauge* because the *gauge* activates on this event. Thus *act_mouseover* proxies for the usage of the *gauge*. A clearer picture of the role of the gauge could be analyzed by combining the amount of mouseover and the number of activities open and attempted. The idea is to see to which extent students exposed to the *gauge* used mouseovers to reach activities that they finally attempted. To test this, I computed a rate of mouseover and activity opened, dividing the

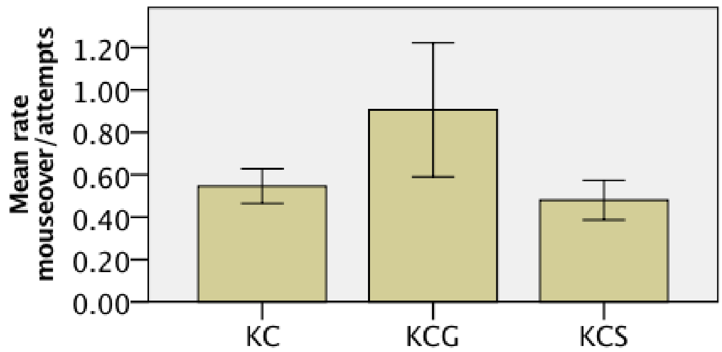


Figure 72: Rate of mouseover and number of activities opened and attempted.

act_mouseover by the number of times the student open an active cell and attempted the content associated. Figure 72 shows a clear advantage of the *KCG* group in this variable. A linear regression found that gauge positively and significantly predicts the rate of mouseover and activity attempts, $\beta_{gauge} = .351$, $p = .014$.

Putting these findings together, the observations point to students using the *gauge* actively to inspect content activities (higher relative level of mouseover) and opening a higher proportion of activities that they finally did not complete (lower *prob_attempt*), compared to the other groups that did not have this feature. Although these findings push to reject H2G (*Fine-Grained OLM complemented with the Learning Gauge helps students to navigate the content of the system more efficiently*), they also suggest that students exposed to *gauge* use this feature to explore the content.

Regarding the social comparison features, this factor shows a mild positive effect on the relative amount of time spent on the interface only when considering all students and disappears after the incentive threshold. Although weak, this observation suggests that students need to spend more time navigating, which added to the observation that *social* does not relate to better performance (effectiveness, success rates), it hints towards the idea of increase of complexity in the interface: students in the social comparison group tend to

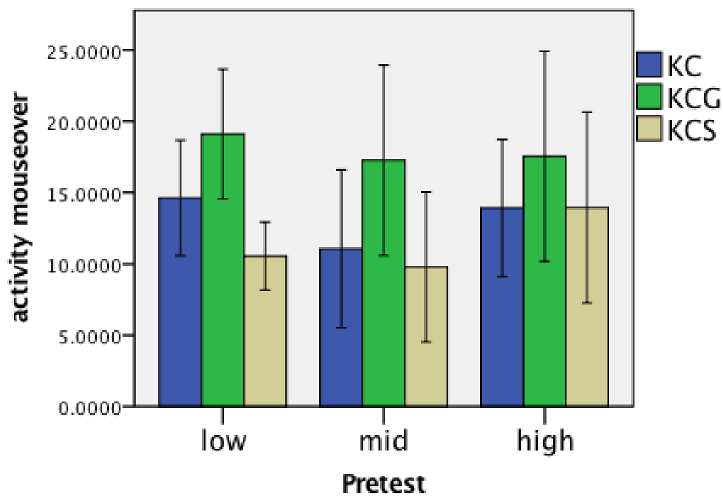


Figure 73: Amount of mouseover on activities by treatment group for different levels of pretest.

spend more time in the interface that does not translate to better success rates.

The amount of mouseover on activities seems to depend on the *social* feature, where I observe a negative effect countered by a positive effect of the interaction *pretest*social* (last row of Table 47). This effect can be seen in Figure 73 where low levels of mouseover are shown for the *KCS* group increasing in the high pretest group.

Figure 73 also show another effect, where *KCG* group shows higher levels of mouseover overall. I verify this relation by repeating the regression model using a forward *stepwise* method. The results showed that only *gauge* enters the regression model explaining positively the amount of mouseover ($\beta_{gauge} = 5.665, p = .002$). In an attempt to isolate the effect of the *gauge* (cancel out the effect of *social*), I then build this regression model only for the groups *KC* and *KCG* (with predictors *pretest*, *gauge*, *pretest*gauge*), and the results were similar: only *gauge* enters the regression ($\beta_{gauge} = 4.435, p = .047$). Using similar analyses to test the effect of the *gauge* in *act_difficulty*, *ratio_gui*, *prob_attempt* and *regularity* did not result in significance. These results suggest that gauge has the positive effect on doing mouseover activities, which is consistent with the functionality of the *gauge* and complements the previous observation about the exploratory behavior that it induces.

9.7.5 Effects on *performance*

Regarding the efficiency and success rates, only *pretest* presents a consistently positive effect on these variables. This positive role of prior knowledge in the performance in self-assessment content is not surprising, since higher pretest students probably solve with little difficulty much of the content in the practice system. No effect is observed regarding *social*, nor *gauge* interface features, and also no effect are observed regarding the interactions of these interface features and the level of prior knowledge.

9.8 ACTIVITY TROUGH THE TERM

The regressions presented in the previous section did not show effects of either *social* not *gauge* in the regularity of activity through the term. This observation contrast with Figure 70 that shows different patterns of activity during the term among the treatment groups. To complement the results related to the regularity of system activity, I now present two analyses. The first analysis looks at the different number of times the students log into the system across groups. The second analysis looks at the term regularity variable conditioned to the level of engagement (mg_completion).

9.8.1 The effects of *Gauge* in *coming back* to the system

In this analysis, I looked at the number of times the students visit the system. The chart in Figure 74 shows the proportion of students who had logged into the system once, twice (come back), three times (come back twice), four and fifth times, among the students who has logged in at least once. Table 48 shows the raw counting values.

There is a significant association between the interface used (treatment) and whether or not students come back at least twice (3 or more loggins) $\chi^2(2) = 8.371$, $p = .015$. The

Table 48: Counts of students who entered the system and came back (has more sessions) once, twice, etc.

	All groups		KC		KCG		KCS	
sessions	N	%	N	%	N	%	N	%
1 or more	396		124		133		139	
2 or more	282	0.712	91	0.734	98	0.737	93	0.669
3 or more	185	0.467	58	0.468	74	0.556	53	0.381
4 or more	131	0.331	40	0.323	54	0.406	37	0.266
5 or more	88	0.222	27	0.218	33	0.248	28	0.201

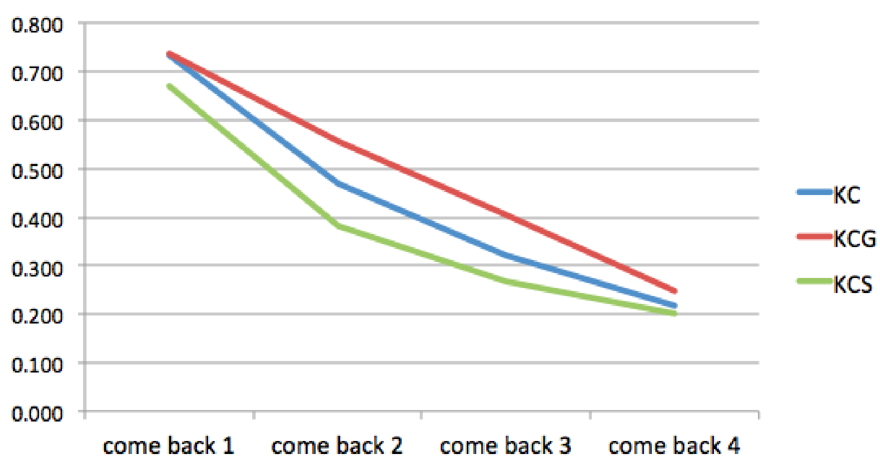


Figure 74: Relative proportions of students that came back to the system (come back 1 = 2 or more sessions, come back 2 = 3 or more sessions, etc.)

effect is moderate, Cramer's V = 1.45. When using *Gauge*, about 56% of students come back to the system, while only 47% and 38% came back at least twice in the *KC* and *KCS*, respectively. Based on odds ratio, the odds of coming back twice to the system were 1.43 times higher if they had the gauge than if using the simple interface (*KC*), and 2 times higher if they had the gauge than if they had the social comparison features (*KCS*). The lower level of retention in the *KCS* group is another evidence towards the idea that social comparison and fine-grained information may increase complexity and discourage students of using the system.

9.8.2 Regularity, level of activity and *gauge*

Regularity in using the system during the term is also conceptually dependent of the level of activity done. It is natural to think that if a student has more activity, it has a higher chance to present higher regularity than another student who has less activity. Following this reasoning, I performed a linear regression on *term_regularity* adding as predictors the completion of activities (*mg_completion*), *pretest*, the system features (*social*, *gauge* as dummies) and their interaction with pretest (*pretest*social*, *pretest*gauge*). Regression models were built using a stepwise method. When running for all students who at least did one activity in the system, only *mg_completion* showed a significant contribution to explain regularity (β = .114, p < .001). However, *gauge* entered the regression significantly when building the regression model for students who went beyond the 15-activities threshold. Results are shown in Table 49.

The analyses of this section show that *gauge* does have a positive effect on motivating students to use the system more regularly in the term. Student exposed to the *gauge* are more likely to come back to the system, thus they become more regular in their use of the system through the term, although this effect is only noticeable for students having higher levels of activity. This observation makes sense because we could expect students to benefit of the complex interface features after gaining some experience using the system.

Table 49: Regressions on regularity including *mg_completion* and features *social* and *gauge*. Only *gauge* entered the regression model after *mg_completion*. Significance (p-values) in parenthesis.

Model	R Square	Sig. F Change	$\beta_{mg\_completion}$	β_{gauge}
1	.032	.007	**.071 (.007)**	-
2	.051	.034	**.062 (.016)**	**.027 (.034)**

9.9 THE ROLE OF SOCIAL COMPARISON ORIENTATION

Exploration of the social comparison orientation as an influential factor determining the system activity is relevant regarding the social features of the system interface. Therefore, the following analyses only consider the groups *KC* and *KCS*. Regressions were built for each engagement variable with predictors *pretest*, dummy variable *social*, Social Comparison Orientation score (*SCO*), and interaction terms *pretest*social* and *SCO*social*. Results are reported in Table 50. Only activity variables that showed some effect of *SCO* are included in the table.

There were no effects of Social Comparison Orientation observed for the completion of the activity, levels of activity in any content except for examples. Also, there was was no observed effect of SCO for self-assessment performance (success rates or effectiveness scores). Only observed is the effect on the number of examples viewed (negative) and the ratio of time in the interface relative to the total time in the system (negative). Regarding the combined effect of *SCO* and *social* features, results only show a positive effect of this interaction in the ratio of time in the interface, which counters the negative effect of *SCO* on this variable. These results do not confirm the previous results obtained in chapter 6, nor provide a strong finding. As I noted before, *SCO* scale lack variability, which could be a reason why it is not a good factor to explain the observations on engagement.

Table 50: Results of regressions (β coefficients and significance) on engagement variables with predictors *pretest*, *social* features, Social Comparison Orientation (*SCO*), and interactions. Only reports on regressions that shows significant or marginally significant effect of *SCO* or *SCO*social*. Significance is marked with the cell background color and with symbols '.' (.1-.05), '*' (.05-.01), '**' (.01-.001), '***' (<.001).

	pretest	social	SCO	pretest*social	SCO*social
n_examples	-1.584	-14.374	**-21.418 ***	12.354	18.607
ratio_gui	0.016	-0.119	**-0.299 ****	0.02	**0.285 .**

9.10 THE ROLE OF LEARNING MOTIVATION

To explore the role of motivation, I build regression models separately for low and high motivated students in which the factors *gauge* and *social* are added as dummy variables. Motivational groups (low/high) for both factors, Mastery Approach and Performance Approach, were created by splitting students by the median value of the motivational factor measured at the initial point (beginning of the term). Table 51 shows the number of students in each of the subgroups after splitting for both Mastery Approach and Performance Approach and Table 52 shows the mean and standard deviation of the system activity variables in both splits.

Regressions run in each of the motivational groups (low/high) help to see local tendencies (local to a motivated group of students), but not the overall influence of the motivational variable. To find global patterns of influence of the motivational variables and its interactions with the interface features (*social*, *gauge*), I ran a second round of regressions on all the system activity variables as dependent variables, and adding all factors and interactions to the regression model using *stepwise* regression method. I call these the *full* models. All

Table 51: Mean and Median of Mastery Approach and Performance Approach factors measured at the beginning of the term.

	Mastery Approach (median=.720)		Performance Approach (median=.610)	
	low	high	low	high
KC	62	39	41	60
KCG	62	49	35	76
KCS	62	56	39	79

the predictors are: *pretest*, *social*, *gauge*, motivation (*MApi* or *PApi*), and the interactions *pretest*social*, *pretest*gauge*, *motivation*pretest*, *motivation*social*, *motivation*gauge*.

9.10.1 The role of Mastery Approach

The results of the regressions on system activity variables and considering Mastery Approach motivation orientation (MAp) are shown in Table 53. Columns 6-11 show results of regressions run separately for *low* and *high* Mastery Approach students. Values correspond to the β coefficients and significance is marked with symbols and background color. Columns 12-15 show the β coefficients and significance of the terms that entered the *stepwise* regression of the second round of regression analyses, or *full* model. In this second round I am interested in seeing the role of the motivation variable (*MAp*) and its interactions, thus other factors that might enter the regression models are not reported in the table.

No influence of *social*, nor *gauge* is noted in the overall completion of activities when looking at each motivational group (*low*, *high*). However, the interaction *MAp*gauge* result in significant positive contribution in the *full* model. Looking at the different type of con-

Table 52: System activity in the low and high groups for Mastery Approach and Performance Approach splits.

	Low MAp		High MAp		Low PAp		High PAp	
	Mean	SD	Mean	SD	Mean	SD	Mean	SD
mg_completion	0.247	0.217	0.275	0.210	0.252	0.208	0.268	0.222
n_questions	24.215	18.950	26.674	21.232	24.115	19.002	26.748	21.118
n_parsons	26.968	29.820	29.965	32.544	28.235	30.726	28.327	31.499
n_examples	7.091	12.090	8.472	12.403	7.475	11.797	7.966	12.780
n_ae	9.774	10.999	10.424	10.510	9.984	10.876	10.150	10.689
term_regularity	0.463	0.082	0.480	0.091	0.468	0.084	0.473	0.088
eff_questions	0.007	0.386	-0.032	0.411	0.002	0.355	-0.024	0.444
eff_parsons	-0.027	0.473	0.012	0.475	-0.037	0.501	0.023	0.436
sr_questions	0.623	0.160	0.623	0.191	0.627	0.170	0.618	0.180
sr_parsons	0.470	0.209	0.516	0.240	0.482	0.215	0.501	0.235
prob_attempt	0.562	0.230	0.546	0.223	0.545	0.229	0.568	0.223
ratio_gui	0.327	0.128	0.314	0.139	0.322	0.130	0.321	0.137
mouseover_act	14.737	17.238	13.896	13.658	14.678	16.360	13.986	15.024
act_difficulty	0.507	0.289	0.434	0.295	0.485	0.299	0.463	0.286

Table 53: Results of regressions on engagement variables considering Mastery Approach measured at the beginning of the term (*MApi*). Columns 2-11 present separate regression analyses for *low* and *high MAp* (split by median). Columns 12-15 show results of regressions analyses, using method *stepwise*, on all students where predictors included *MAp* and its interactions with *social*, *gauge* and *pretest*. Only reporting interactions and main effect of *MAp* when they enter the regression model. Significance is marked with the cell background color and with symbols '.' (.1-.05), '*' (.05-.01), '**' (.01-.001), '***' (<.001).

	Low MApi					High MApi					Interactions with MApi			
	pretest	social	gauge	pre*s	pre*g	pretest	social	gauge	pre*s	pre*g	MAp	MAp*pre	MAp*s	MAp*g
mg_completion	.085	-.083	.014	.145	-.07	-.233	-.036	.054	.263	.223				**.08***
n_questions	7.35	-5.01	-.375	-3.66	-9.52	**-41.52***	**-11.30 .**	-2.42	34.05	32.91				
n_parsons	-3.13	-11.14	-.988	15.59	-6.28	**-44.60 .**	-9.57	-6.75	23.93	**58.00 .**				
n_examples	.115	-3.46	-.21	12.73	2.03	-8.36	-1.29	.262	12.23	15.16				
n_ae	-.149	-4.16	2.43	8.89	-5.07	-3.21	0.86	**7.79***	6.41	-3.95				**5.45****
term_regularity	.019	.002	.033	-.044	-.101	-.061	-.018	-.011	.021	.091				
eff_questions	**.472***	-.015	.083	-.204	-.435	.238	.118	-.046	.200	.124				
eff_parsons	.464	-.006	-.019	-.215	-.05	-.269	-.064	-.209	.262	.584				
sr_questions	.134	.006	-.021	.015	.169	**.334***	.068	.046	-.097	**-.361 .**				
sr_parsons	**.390****	**.103 .**	.074	**-.396 .**	-.219	.158	-.004	-.05	.178	-.022		**.294****		
prob_attempt	.116	.027	-.056	-.134	.094	-.206	-.079	**-.188***	.028	**.400 .**				
act_difficulty	-.067	.007	.003	-.347	-.033	**-.501***	-.083	-.11	.185	.302		**-.394*****		
ratio_gui	-.058	.044	.026	-.011	.091	.092	.06	.048	-.034	-.153			**.042***	
act_mouseover	-1.05	**-8.98***	1.73	**27.08 .**	2.95	**-17.75 .**	-3.03	6.41	12.71	6.36				

tent (questions, parsons, examples, animated examples), it is clear that the main influence of *gauge* expresses in viewing animated examples. In the *high* Mastery Approach group, *gauge* positively and strongly explains the use of this type of content. Overall, it is interesting to notice the general null effect of Mastery Approach orientation on engagement, but conditioned to the presence of *gauge*. This effect is aligned to the design of the *learning gauge*, which has the goal of directing the interpretation of the fine-grained information in the visualization towards the opportunities of learning, thus is theoretically closer to the construct of Mastery orientation. The reason that this effect expresses in the number of animations viewed suggests that this type of content may have an extra importance for students who seek learning.

Regarding *performance* with self-assessment content (rows 7-10 in Table 53), differences observed between *low* and *high MAp* groups in effectiveness scores and success rates are mostly local (within low/high groups). Interestingly the interaction *pretest*gauge* seems to counter the positive effect of *pretest* on questionable success rate (*sr_questions*) in the *high* Mastery Approach group. This suggest that students with higher prior knowledge and motivation tend to decrease their success rate when exposed to *gauge*. This observation supports the idea commented before that for strong students, *gauge* encourages exploration of the content. An observation that also finds support in the negative significant relationship between *gauge* and the probability of attempting activities after opening them (*prob_attempt*), which is analyzed later when I commented on the effects on navigation.

The success rate of parsons problems shows a very different pattern with significant predictors in the *low MAp* group. In the *low* group, pretest shows a strong positive effect, which disappears for highly motivated students. Thus, performance in parsons problems is dependent of pretest, as expected, but only for low motivated students. The interaction *MAp*pretest* in the right side of the table (full model) confirms this.

Variables related to *navigation* in the system say more. Recall that the design of the fine-grained visualization aims navigational support and hypotheses H2 (*Fine-Grained OLM helps students to navigate the content of the system more efficiently*) and H2G (*Fine-Grained OLM complemented with the Learning Gauge helps students to navigate the content of the system more efficiently*) are explicit about this. Regarding the pattern of open and doing activities measured with the probability of activities open and attempted (*prob_attempt*), *gauge* has negative influence compensated by the interaction *gauge*pretest* in the *high* Mastery Approach group. This complement the previous finding rejecting H2G, and attaching the exploration behavior to the highly motivated students.

The difficulty of attempted activities is also influenced by the pretest in the *high MAp* group. The higher the pretest, the lower the difficulty. This observation is confirmed by the negative significant interaction *MAp*pretest* in the full regression model for the difficulty measure. The interpretation of this effect is tricky since different motivational groups capture different segments of the pretest distribution. A non-parametric Mann-Whitney test shows a significant difference on pretest between *low* and *high MAp* groups. Highly motivated students also present higher pretest (M=.222, SE=.016) than lower motivated students (M=.183, SE=.014), Mann-Whitney U = 11,380.5, p=.016.

An interesting observation is the effect of *social* in doing activity mouseover in the *low* Mastery Approach group. Again the effect turned out to be conditioned to pretest: it is negative for *social*, but positive and higher (β is higher) in the interaction *pretest*social*. This means that for low motivated students, the effect of being exposed to social translates to an increment on mouseover activity for higher pretest students, while lower pretest students present a negative trend in this variable.

Another interesting observation is regarding the ratio of time spent on the interface. Even though no significance is observed in regressions within *low/high* MAp groups, the interaction *MAp*social* in the *full* model shows to be a positive significant predictor. This shows that the highly motivated students are who consume more time in the interface with social comparison features, suggesting that these students are able to engage with the complex interface.

The fact that in the full model of the Table 53, Mastery Approach factor does not have any clear effect on the level of mouseover or in the ratio of time in the interface is evidence to reject **H4** *Students with a higher Mastery orientation will use the fine-grained components more.*

9.10.2 The role of Performance Approach

Similarly, than for Mastery Approach (previous analyses), two rounds of regressions were running for Performance Approach orientation. Results are shown in Table 54. Columns 6-11 show results of regressions run separately for *low* and *high* Performance Approach students. Columns 12-15 show the results of the terms that entered the *stepwise* regression in the second round of regressions analyses, or *full* model. In this second round I am interested in seeing the role of the motivation variable (PAp) and its interactions, thus other factors that might enter the regression models are not reported.

Regarding the overall completion of activities in the system, Performance Orientation showed no significant effect when regressing within the *low* and *high* motivational groups, although I can notice a consistent negative effect of pretest in engagement with all types of content in the high PAp group. Similarly *social* shows a consistent negative effect and *gauge* a consistent positive effect in highly performance motivated students. This last relation turns in significance when looking at the *full* model: A significant positive interaction exists between *PAp* and *gauge* in the full regression model (*PAp*g* in Table 54). Same observation is made for Mastery Approach (see previous sub-section), where also this interaction with *gauge* is the strongest. A regression model with only these two interaction terms run in stepwise method show that the stronger prediction is given by *PAp*gauge* and after this term enters the regression, the term *MAp*gauge* does not contribute significantly. It is possible that because of correlation between *MAp* and *PAp* ($Pearson = .411$, $p < .001$), these two interactions terms (*MAp*gauge*, *PAp*gauge*) are introducing the same information to the model, which means that *gauge* produces a positive engagement with the system only for highly motivated students.

Overall, although *social* is not showing observable effects related to Performance Orientation in the amount of activity, it does regarding instructional effectiveness and success rates. Here, the contribution of *pretest* and *social* are observed only in the high-performance Approach group with positive coefficients on the instructional effectiveness of questions and success rates of questions and parsons problems. This effect of *social* was not observed for high Mastery Approach group (previous sub-section). The full model (last columns in Ta-

Table 54: Results of regressions on engagement variables considering Performance Approach measured at the beginning of the term (*PApi*). Columns 2-11 present separate regression analyses for *low* and *high PAp* (split by median). Columns 12-15 show results of regressions analyses, using method *stepwise*, on all students where predictors included *PApi* and its interactions with *social*, *gauge* and *pretest*. Interactions and main effect of *PAp* are reported when they enter the regression model. Significance is marked with the cell background color and with symbols '.' (.1-.05), '*' (.05-.01), '**' (.01-.001), '***' (<.001).

	Low PApi					High PApi					Interactions with PApi			
	pretest	social	gauge	pre*s	pre*g	pretest	social	gauge	pre*s	pre*g	PAp	PAp*p	PAp*s	PAp*g
mg_completion	.076	-.081	.004	.071	-.038	-.101	-.014	.072	.211	.117				**.094 ***
n_questions	-3.73	-7.01	-2.34	3.54	3.55	-19.39	-6.75	.038	15.24	14.20				
n_parsons	-2.28	-5.61	-4.61	.327	25.59	**-35.75 .**	-15.08	-3.18	29.75	23.03			**-11.56 ***	
n_examples	3.49	-2.53	-.38	5.52	2.35	-7.64	-1.60	.655	15.29	12.53				
n_ae	2.70	-2.90	3.42	1.38	-4.92	-3.26	-.217	**6.12 .**	11.35	-4.12				
term_regularity	.053	.002	.024	-.026	-.087	-.075	-.016	.002	-.015	.063				
eff_questions	.034	-.132	.010	.280	.074	**.767 ****	**.295 .**	.113	-.361	-.573		**.426 ****		
eff_parsons	.261	-.120	-.071	.237	-.022	.111	.118	-.101	-.406	.341				
sr_questions	.113	-.026	-.054	.105	.105	**.368 ****	**.124 ***	.105	-.186	-.331				
sr_parsons	**.284 .**	.007	.072	-.051	-.37	**.352 ***	**.141 .**	-.007	-.272	.014				
prob_attempt	-.031	-.029	**-.116 .**	.014	.215	-.006	.007	-.095	-.208	.183				
act_difficulty	-.314	-.118	**-.155 ***	-.125	.378	-.055	.126	.143	-.254	-.37				
ratio_gui	-.043	.029	.035	.154	.052	.019	.062	.029	-.145	-.08				
act_mouseover	-4.75	**-9.15 ***	-.639	22.58	3.96	-5.63	-.700	**10.26 ***	6.68	-1.74				

ble 54) also shows a strong positive effect of the interaction *PAp*pretest* (*PAp*p*) on the effectiveness score of questions (*eff_questions*). This relation is shown in figure 75 where the effectiveness of the questions is plotted for different levels of pretest (Low, Medium and High) and for *low* and *high* Performance Approach.

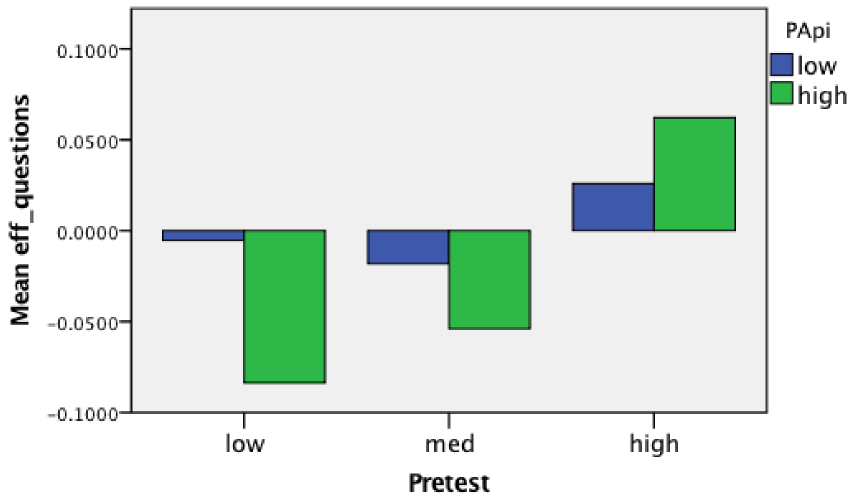


Figure 75: Effectiveness score in questions for low and high Performance Approach at different levels of pretest.

These results are in line with the idea that *social* features are aligned with the Performance orientation of the motivational profile. It also supports the idea that higher performance oriented students, who theoretically tend to compare to others, react to social comparison by completing an activity that they can solve, thus obtaining higher success rates. However, this claim is not well supported by the navigational patterns, where *act_difficulty* does not show that these students performed an easier activity on average.

Regarding patterns of *navigation* in the system, the negative effect of *gauge* on the probability of attempting open activities (*prob_attempt*) is fairly similar (coefficient-wise) in both *low* and *high PAp* group, although the term appears only significant in the low group. Regarding the difficulty score of activities open and attempted, the negative effect of *gauge* observed in the *low* group flips in the *high* Performance Approach oriented group.

However, no interaction resulted in significance levels in the *full* model. Also, *gauge* seems to contribute positively in the number of mouseover activities in the *high PAp* group, while *social* has a negative effect on the same variable in the *low PAp* group. Altogether, these results show, interestingly that the Performance Orientation does not influence the effects of social comparison features (*social*) in navigating the system. Instead, this motivational trait has an impact on the effects associated with the *gauge*, which becomes negative for *low PAp* students and positive for *high PAp* students.

It is surprising the non-existent effect of Performance Approach orientation on system engagement in the *social* group. This seems to contradict the theoretically grounded idea that high Performance oriented students tend to compare to others, thus becoming more engaged with the version of the system that present information of others (recall also that this effect was observed in the study in chapter 6, where a positive significant interaction of *PAp*social* was observed). However, it is possible that Performance orientation still has a positive effect, but hidden behind the overall lower activity observed in the *social* group. To visualize this, Figure 76 shows the difference of activity between individual (*KC+KCG*) and social (*KCS*) groups for different levels of Performance Orientation categorized using the percentile 25, 50, and 75 in *Low* (PApi $\leq .5$), *Medium Low* ($.5 <$ PApi $\leq .61$), *Medium High* ($.61 <$ PApi $\leq .72$), *High* (PApi $> .72$). Error bars in the figure represent 2 SE (standard error of the mean). Higher differences (less activity in *social* group) are observed in the lower 2 bins of Performance Orientation, but the differences become smaller in the higher 2 bins of performance orientation. This supports the idea that even when students tended to do less in the *social* group, higher Performance Approach orientation contributes to counter this effect.

9.11 THE CHANGE OF MOTIVATION

Similarly than in chapter 6, I present here analyses of the change of motivation factors to contribute to answering research question 4 *How does the use of a learning system featuring OLM, OSLM and fine-grained elements affect motivation?*. As argued before, these analyses

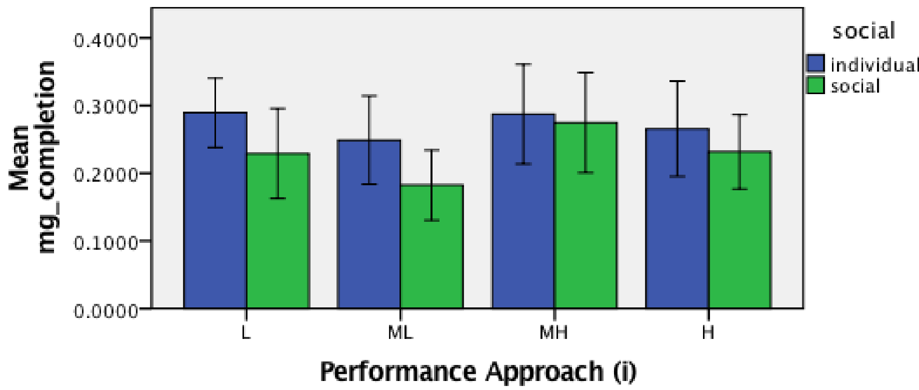


Figure 76: Mean mg_completion across different levels of Performance Approach (initial measure) and between *individual* and *social* group.

search for relationships between the motivational change from the beginning to the end of the term, and the engagement with the system in the different treatment groups (different interfaces). It is not, however, the intention to search for causal relationships, because I understand that the change of motivation of the students is probably due to other more critical experience during the course, rather than the usage of Rich-OLM, that was offered as a complementary and voluntary practice system. Similarly than in chapter 6 Section 6.10, in this section, I included all the motivational factors measured. This is because I am interested in seeing any change in motivational orientations no matter their theoretical relationships or structure (previous sections use only Competency Beliefs and Achievement-Goal Orientation because of their theoretical closeness to the use of the system). These measures are: Fascination (F), Competency Beliefs (CB), Values (V), Mastery Approach (MAp), and Performance Approach (PAp). To simplify the notation, the different measures receive the suffix 'i' or 'f' to refer to the initial or final measure, respectively. E.g *CBf* is Competency Beliefs measured at the end of the term.

A paired sample t-test series of analyses, which results are presented in Table 55, show that all differences are significant, except for Performance Approach Orientation. Interesting is that while Fascination, Competency Beliefs and Values increased, Mastery Approach decreased. Note also that even in the cases of significant difference, the mean difference is not very big. Fascination, Values and Mastery Approach have a mean difference of close to 2%, and Competency Beliefs showed the greater difference with a mean of 21%. These differences are also similar than the ones reported in chapter 6. These analyses considered all students: who used and who did not use the system.

I repeated these analyses selecting students who used the system (at least 1 activity) and students who did not use the system at all. Among who used the system, the difference is that Mastery Approach did not show a significant difference (Mean difference = .002, t=.228, p=.820), and Performance Approach now result in a significant difference (Mean difference=-.036, t=-2.954, p=.003). These students did not decrease their intrinsic motivation orientation, but increased their Performance orientation. Among who did not use the system, significance only occurs in Competency Beliefs (Mean difference=-.192, t=-11.542, p<.001) and Mastery Approach (Mean difference=.052, t=3.530, p=.001), but contrasting with who used the system, there is no significant increase of Fascination, nor on Values. In other words, students who did not engage at all do not increase their motivational traits, and in fact, tend to decrease them (except for Competency Beliefs).

To explore the relationship between the change of motivation and the use of the system in its three different flavors (*KC*, *KCG*, *KCS*) I performed a series of regression analyses considering all students who at least has 1 activity in the system. A regression model is built for each motivational factor measured at the end of the term (e.g., *Ff*). The predictors include first the motivational factor at the beginning of the term (e.g., *Fi*), the dummy variable *social*, the dummy variable *gauge*, the overall amount of activity performed in the system measured with the variable *mg_completion*, and the interaction terms *mg_completion*social* and *mg_completion*gauge*. The results are shown in the Table 56.

In general, only the motivation at the initial point (*Xi* in Table 56 is the main and only predictor of the motivation at the final measure. *Fascination* is the only measure that shows some relation with the factor *social*. Although *social* appeared with a negative effect on *Fascination* (*Ff*), it seems to be compensated with a positive effect of the interaction

Table 55: Paired Samples t-tests for motivation measured at the beginning (i) and at the end (f) of the term.

DV	Mean	SE	t	p
Fi - Ff	-0.022	0.008	-2.805	.005
CBi - CBf	-0.208	0.010	-21.278	<.001
Vi - Vf	-0.026	0.007	-3.921	<.001
MApi - MApf	0.019	0.008	2.331	.020
PApi - PApf	-0.014	0.010	-1.336	.182

Table 56: Coefficients (β) of regressions on motivational factors at the end of the term with predictors motivation at the beginning of the term (Xi column), *social*, *gauge*, *mg_completion* (*mg*) and interactions *mg_completion*social* (*mg*social*) and *mg_completion*gauge* (*mg*gauge*). Significance is marked: '***' means $p < .001$, '**' means $.001 <= p < .01$, '*' means $.01 <= p < .05$, and '.' means $.05 <= p < .1$.

DV	Xi	social	gauge	mg	mg*social	mg*gauge
Ff	**.752 *****	**-.097 ****	-.018	-.084	**.210 .**	-.049
CBf	**.406 *****	-.030	-.024	-112	.126	.058
Vf	**.724 *****	-.012	-.021	.028	-.025	-.024
MApf	**.610 *****	.025	.038	.014	-.142	-.122
PApf	**.502 *****	.029	.012	-.066	-.059	.045

*mg_completion*social* (*mg*social* in the table). To better see this effect, Figure 77 shows the initial and final measures of *Fascination* for different levels of system engagement, using percentiles 25, 50, and 75 as points of cut. Low (L) represent less than 12% of activity. Medium Low (ML) represents between 12% and 20%. Medium High (MH) between 20% and 37%, and High (H) more than 37% of completion. Error bars are 2 standard error of the mean. The chart at the left joins the treatment groups *KC* and *KCG*, which students are not exposed to social features. Chart at the right is for the *social* group (*KCS*).

A few observation scan be made from Figure 77. Consider that since the chart at the left has two groups (KC and KCG) its error bars are smaller. Overall, *Fascination* at final (*Ff*) appeared consistently higher than *Fascination* at the beginning (*Fi*). In the *individual* group (*KC* + *KCG*), the greater difference in initial to final fascination is in the Medium Low level of activity, and High level of activity present lower levels of *Fascination* for both initial and final measures. In contrast, the social group does not present many variations across levels of activity, nor differences between the initial and final measure of *Fascination*. Combining these observations with the first row in Table 56, we can see the negative coefficient of *social* in predicting *Fascination* at final measure means that when *individual* group present a positive change of fascination, this change is null in the social group. Also, while *individual* group presents a decrease of the fascination difference across levels of activity, the *social* group present a slight increase at the higher level of activity, which explain the positive contribution of the interaction *mg_completion*social*.

9.12 CONCLUSIONS OF THIS CHAPTER

In this chapter I present a semester-long classroom study that evaluates an extended version of Mastery Grids, called Rich-OLM which includes a combination of coarse-grained and fine-grained visual representations of the learner model and social comparison features at both levels. The study focuses in answering research question 2 *What are the effects of using a fine-grained OLM on system activity?* and contributes also to research question 1 *What are the effects of OLM with social comparison features (or OSLM) on system activity*

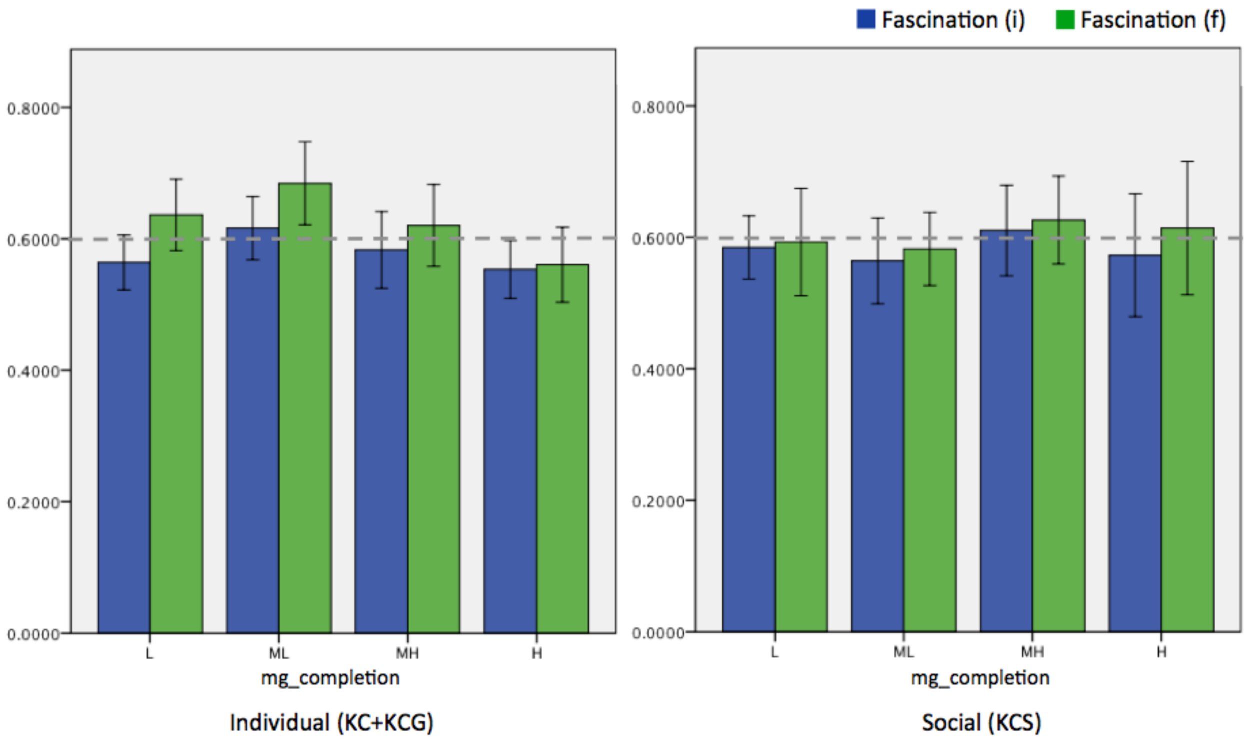


Figure 77: Mean of Fascination level on initial and final measures (i,f) at different levels of completion of activities Low to High (L,ML,MH,H) for *individual* (KC + KCG) and *social* group (KCS).

compared with an individual-view OLM? by contrasting the presence of social comparison features together with the fine-grained visualization, research question 3 *How do individual differences influence system activity within an OLM?*, and research question 4 *How does the use of a learning system featuring OLM, OSLM and fine-grained elements affect motivation?*.

The study was set as a between subjects design in which 3 versions of the Rich-OLM were offered to 3 groups. The version called *KC* included fine-grained and coarse-grained visual representations without social comparison features; the version *KCG* is identical a *KC* and adds a *learning gauge*, which was designed to direct the interpretation of the relation between content activities and fine-grained components towards the idea of learning opportunity. The third version, *KCS*, adds social comparison features to *KC*, and does not include the *gauge*. The reason of these versions was to contrast the addition of the information (social comparison) and the addition of visual aids (*gauge*) in complex visualizations. This configuration allowed me to see the effects of the *gauge* and to test hypothesis **H2G** *Fine-Grained OLM complemented with the Learning Gauge helps students to navigate the content of the system more efficiently.* Hypothesis **H2** *Fine-Grained OLM helps students to navigate the content of the system more efficiently* will be explored in the next chapter where I contrast activity of groups across classroom studies.

Analyses of prior and post knowledge across groups indicate that the three treatments were similar. Also, the treatments were similar in the initial value of motivation. Consistent with previous results (chapter 6), the analyses showed that overall, the completion of activities in the practice system supposes an increase on the posttest obtained by the students, which confirm the general learning benefit of the system. This benefit was however not conditioned to the interface features.

9.12.1 Overall effects on system activity

Regarding engagement with the system, the *social* features seem to be related to less activity when is activated for both coarse- and fine-grained visualizations. Although this effect is mild, it contrasts with previous findings in which the *social* group tends to have higher levels of engagement and goes in counter hypothesis **H1** *Students exposed to an OLM with social* *comparison features increase the level of activity in the system.* A possible reason for this is that adding social comparison features to both coarse- and fine-grained visualization might increase the complexity of the system to the point of discouraging student to use it. I have seen in chapters 7 and 8 that complexity of the visual interface is an important concern. However, deeper research, with a high qualitative component is needed to clear this effect.

The analyses also confirm the observation made in the previous study that in the *social* group, pretest has a role in determining the level of engagement with the system. This effect is however, weaker in this study. Putting these observations together, the activity in the *KCS* group is lower than in the other groups, but increase with pretest. Interestingly, no such relation was observed regarding the group exposed to the *gauge*.

Although the *learning gauge* did not show clear overall effects on engagement and performance variables, it did show an effect of retaining students, making them more likely to come back to the system. *Gauge* also showed a positive effect on motivating students to use the system more regularly in the term, but that this effect is only noticeable for students having higher levels of activity. This observation makes sense because we could expect students to benefit of the complex interface features after gaining some experience using the system.

Analyses on navigation variables show counter-intuitive effects of *learning gauge* in navigation pointing to the rejection of hypothesis **H2G** *Fine-Grained OLM complemented with the Learning Gauge helps students to navigate the content of the system more efficiently.* This was seen in the negative influence of the *gauge* in the probability of attempting activities opened, and the positive effect of this feature in the amount of mouseover and the rate of mouseover and attempts to content. In the *KCG* group, students mouseover more to reach activities they open and attempt. However, this is not necessarily negative finding, because rather than induce efficiency, gauge may support exploration of the content.

9.12.2 Individual differences

Overall, no clear effect of Mastery Approach was observed in the amount of activity in the interface (ratio of time in the interface, the probability of attempt open activities) to support **H4** *Students with a higher Mastery orientation will use the fine-grained components* *more*. However, interactions between system interface features (*social*, *gauge*) and motivation showed a positive role of Mastery Approach in engagement when *gauge* is present. This effect is especially strong in the number of animated examples viewed. This effect is aligned with the design of the *learning gauge*, which has the goal of directing the interpretation of the fine-grained information in the visualization towards the opportunities of learning, thus is theoretically closer to the construct of Mastery orientation. The reason that this effect is expressed in the number of animations viewed suggests that this type of content may have an extra importance for students who seek learning. This observation point positively towards **H4** *Students with a higher Mastery orientation will use the fine-grained components more*, although it has to be noted that the observation regards the engagement with the practice system and not necessarily with the activity within the interface.

Looking at performance in self-assessment content, the effects of *social* and *gauge* changed when adding the motivation factors. First, the strong role of pretest in success rates and effectiveness scores concentrates in the high Performance Approach oriented students. It is in this group that *social* shows some effect. This observation is aligned with the idea that students motivated by performance, which are also more prone to compare to others, engage with the *social* by focusing more in activities that they can solve, thus producing higher success rates.

Motivation influences navigation too. Although *social* group showed a lower level of activity, it presents higher relative levels of time in the interface conditioned to the level of Master orientation. It seems that highly motivated students overcome the complexity of the interface and "play" more with it. However, Mastery orientation is also associated with attempting easier content activities when exposed to *social* features.

Regarding the relations of *gauge* and motivation, Mastery Approach influences the *exploring* effect: the high Mastery Approach students are who tend to open and not attempting activities when exposed to *gauge*. This complements the previous findings and contribute to reject **H2G** (*Fine-Grained OLM complemented with the Learning Gauge helps students to navigate the content of the system more efficiently*): students exposed to *gauge* tend to open more activities, probably out of curiosity, without attempting them (exploration) and this effect amplifies for highly motivated students. *Gauge* also showed an effect when Performance Approach orientation is considered. High-performance oriented students produce more mouseovers when exposed to *gauge*.

Social comparison orientation (SCO) showed to have little influence on the effects of the interface and system activity. High SCO students showed a negative effect on the number of examples viewed and in the ratio of time spent on the interface. However, these effects are canceled in the *social* group, suggesting that social features have in fact a positive effect if students have a high SCO. Although weak, these results help to confirm hypothesis **H5**: *The effects of social comparison features of the system will be stronger for students with higher Social Comparison Orientation.*

9.12.3 Change of motivation

Results showed that motivation changed over the term, although this change is not associated to the different system interfaces, which is consistent with the observations made in chapter 6. When looking separately for students who used the system and students who did not, I observe that while active students tend to increase their motivation, students who did not engage at all did not increase them, and in fact, tend to decrease them (except for Competency Beliefs). As argued before, it is not the aim of my work to establish causal relationships between the use of the system and the change of motivation, thus I claim that the observed effect may correspond to self-selection.

Regarding the relation of the change of motivation, the level of engagement, and the system features, only *Fascination* showed a relation to the presence of *social* features. While students in the groups *KC* and *KCG* increased considerably their Fascination, this increase is not related to the level of engagement, thus the interface does not require an increase of this motivational factor to engage students. A slight difference is observed in the *KCS* group, where overall levels of increase of fascination are null within the students who used the system and present a slight increase in the higher engaged students.

None of the results support hypothesis **H7** *The active use of an OLM with fine-grained features will increase the mastery orientation of the students.* However, a better look at this issue will be performed in the next chapter.

10.0 ANALYSES ACROSS STUDIES

The goal of this chapter is compare the studies presented in chapters 6 and 9. These two studies are very similar: both used Mastery Grids for Python programming with the same set of content activities and topics, but with different variations of the Open Learner Model interface. Recall than in the study presented in chapter 6, Mastery Grids represented a coarse-grained OLM and half of the students were exposed to Mastery Grids without social comparison features (group *individual*), while the other half were exposed to Mastery Grids with social comparison features (group *social*). In the study presented in chapter 9, Mastery Grids was extended with a visualization of the fine-grained information which was developed after a series of studies reported in chapters 7 and 8. This version of Mastery Grids is called Rich-OLM. The study in chapter 9 used the Rich-OLM with three variations: individual view with coarse- and fine-grained visualization (*KC*), adding a gauge which was designed to guide the student's interpretation of fine-grained information towards potential *learning opportunities* (*KCG*), and adding social comparison at the coarse- and fine-grained levels (*KCS*).

Performing comparisons across studies allow me to address research question 2 *What are the effects of using a fine-grained OLM on system activity?* and its hypothesis **H2** *Fine-Grained OLM helps students to navigate the content of the system more efficiently.*

10.1 DIFFERENCES BETWEEN STUDY IMPLEMENTATION

Although both studies were similar, important differences existed between them. First, the organization of course and lectures and the mandatory content that students used was different (although the content in Mastery Grids was the same). In the 2015 study, there were 18 lectures and the main content resource was a PDF textbook written by the lecturer. In the study on 2016, the lectures were reduced to 10, focusing on the main concepts, and the main resource material was an electronic textbook which included some interactive content.

Secondly, the study in 2016 suffered from an initial delay compared to the study in 2015. While in 2015 Mastery Grids was introduced in the same week that students completed the pretest, in 2016 the system was introduced one week after the pretest and without the fine-grained component. The interface variations (*KC*, *KCG*, *KCS*) were activated 2 weeks after that.

Both the lecturer responsible for the course and I have agreed that the differences in the way the system was introduced will necessarily impact engagement with the system activities. We observed from the previous study (2015) that the first weeks concentrate a considerable amount of the activity in the system (see chapter 6, Figure 25). As a result, any analysis of system activity across studies may suffer from distortion by the late start of the study in 2016.

Considering the aforementioned limitations and moderating the potential conclusions, the exploration of differences across studies is still an interesting quest that I present in the next sections. I start looking into prior and post knowledge (pretest/posttest) differences across studies. Then, I compare system activity variables, focusing on the level of activity completion between the studies and between treatment groups first, and later in variables related to navigation and performance in self-assessment. Then, I add the motivation factors into the analysis, finishing with analyses on the change of motivation between studies.

10.2 PRIOR AND POST KNOWLEDGE

In both studies, I used the same pretest and posttest, whose statistics are shown in Table 57 and Figure 78. In 2016, the pretest was slightly lower than in 2016, while the posttest is considerably higher. As a result, the Learning Gain is higher in 2016. The non-parametric Mann-Whitney test shows that these differences are significant for pretest ($p = .007$), posttest

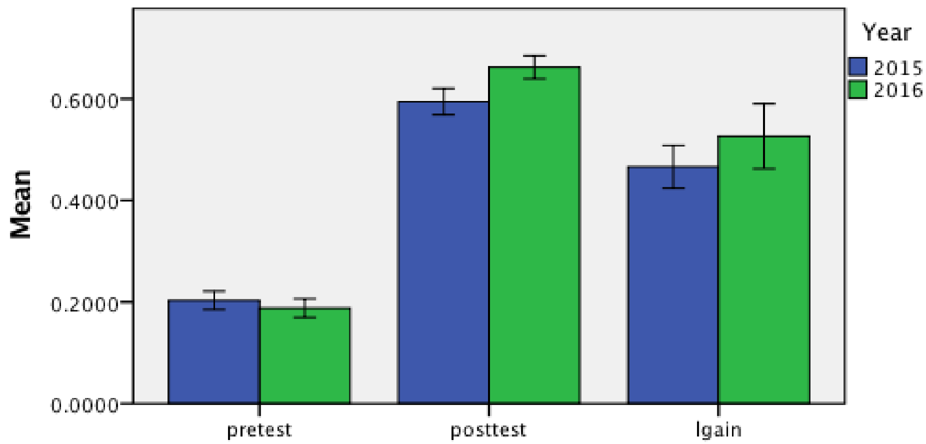


Figure 78: Mean of performance measures in both years.

($p < .001$), and learning gain ($p < .001$).

To better study these effects, I use regression models with *posttest* as dependent variable, and predictors *pretest*, the dummy variable *year* with values 0 indicating the study in 2015, and 1 the study in 2016, and the completion of the system activity (mg_completion). The regression model shows that all predictors have a significant contribution in explaining posttest. While *pretest* explain around 30% of the variation of posttest ($\beta_{pretest} = .301$, $p < .001$), the dummy *year* explains about 7% of posttest variation ($\beta_{year} = .071$, $p < .001$), and *mg_completion* explains almost 16% of posttest variation. In general, students in the study of 2016 reached higher performance, and in both studies, the completion of activity is associated with higher levels of posttest.

10.3 ENGAGEMENT WITH THE PRACTICE SYSTEM

In both studies, system activity was characterized by several variables extracted from the system log. These variables covered three aspects of the activity: *engagement* with the

Table 57: Statistics of performance measures in studies of 2015 and 2016.

		2015	2016
pretest	Mean	0.222	0.202
	SD	0.208	0.208
	SE	0.009	0.009
posttest	Mean	0.602	0.663
	SD	0.262	0.241
	SE	0.013	0.011
lgain	Mean	0.466	0.526
	SD	0.427	0.679
	SE	0.021	0.032

content, *performance* with self-assessment content, and *navigational* patterns through the content and the interface. In the following analyses I focus in the *engagement* dimension, and particularly on the variable *mg_completion*.

The structure of the course in Mastery Grids, the topics defined and the content activities were the same in both years. However, as I mentioned earlier in this chapter, in 2016 the system was enabled later by students. The overall pattern of system activity during the term is shown in figure 79, where the data had been aligned to match the first week after the pretest was delivered (both courses started on early September). Notice how the study in 2016 shows a delay of 2 weeks, and has a similar shape overall.

Table 58 shows the basic statistics of some of the engagement variables in both terms. Figure 80 shows the difference observed in *mg_completion*, favoring the study in 2015. The non-parametric Mann-Whitney shows a significant difference in *mg_completion* between the two years favoring 2015 (Mean Rank = 345.93) over 2016 (Mean Rank=315.62), Mann-Whitney U = 49432.5, p=.041.

Again, as mentioned before, this difference may be due to the early introduction of the system in 2015 (in 2016 system was introduced later), or because of the role of the practice system in the overall course structure (less lectures in 2016, interactive e-textbook in 2016), and not necessarily because of the different system interface features. Supporting this last claim is the fact that, as noted in Table 58, the decrease of the activity in 2016 is in examples and animated examples, and both of these types of content was contained in the electronic textbook on that year (the e-textbook had examples and animated examples, but different than the ones in the practice system).

I have also observed in chapter 9 that the lower activity in 2016 appears in the group with social comparison features (*KCS*). Figure 81, shows this difference between studies. While *individual* groups reached similar levels of completion, the *social* group shows a positive effect in 2015 and a negative effect on 2016. To test this effect I ran a regression analyses on engagement (*mg_completion*) including the predictors *pretest*, *year*, *social*, and the interaction terms *pretest**social*, *pretest**year* and *social**year*. Results of regression, shown in Table 59, confirm the observation and show that being in the *social* group and in the second study (year=1) explains almost 9% of activity reduction. The strongest effect shown in the table

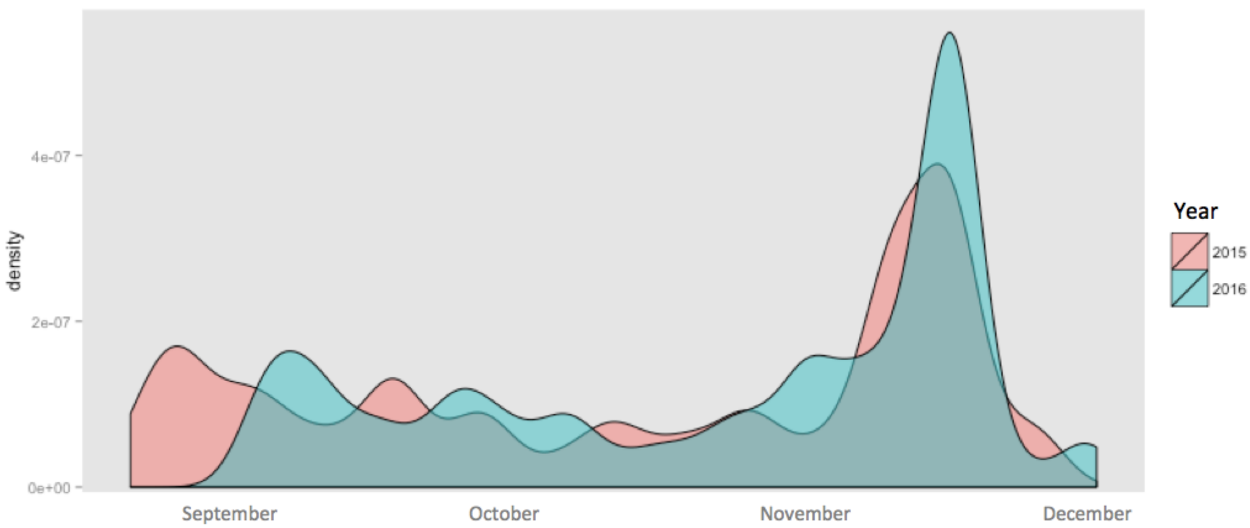


Figure 79: Density plot of activity during the term. Activity in the year 2016 started almost 2 weeks later.

Figure 80: Levels of engagement (mg_completion) on the two studies.

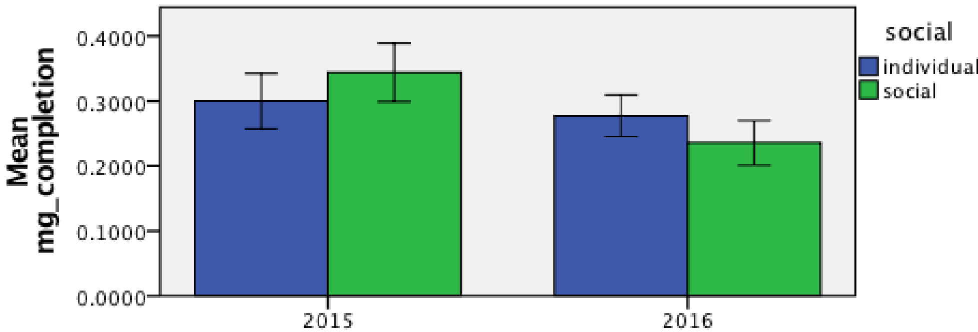


Figure 81: Engagement (mg_completion) of the different treatment groups in the two studies.

is the interaction *pretest*social*, an effect noted in both studies, and that was very clear in the 2015 study. Social features encourage high pretest students to do more in the system.

10.4 NAVIGATION AND PERFORMANCE

The main goal of adding fine-grained information is to give broader support for content navigation, as stated by hypothesis **H2** *Fine-Grained OLM helps students to navigate the content of the system more efficiently*. In the previous section, I have seen differences in terms of engagement in activity within the system among the studies and treatment groups. The main difference is between *social* groups, and no differences appeared between *individual* groups, in terms of engagement. Now I look at the effects of other system activity variables reflecting navigation and performance in self-assessment content. In the next analyses, I only considered the *individual* group of 2015 and the *KC* group of 2016. This is because of the non-existent engagement differences between these groups, which hints that these groups could be more fairly comparable, and because using only these treatment groups allows me to focus on the effects of adding the fine-grained components alone.

Table 58: Statistics of some engagement variables in 2015 and 2016 studies.

		2015	2016
mg_completion	Mean	0.321	0.262
	SD	0.279	0.218
	SE	0.016	0.012
n_questions	Mean	22.661	25.271
	SD	20.685	19.908
	SE	1.149	1.086
n_parsons	Mean	25.991	28.128
	SD	33.945	30.820
	SE	1.886	1.681
n_examples	Mean	17.772	7.923
	SD	16.736	12.573
	SE	0.930	0.686
n_ae	Mean	13.046	10.223
	SD	12.184	10.887
	SE	0.677	0.594
sr_questions	Mean	0.465	0.625
	SD	0.196	0.176
	SE	0.011	0.010
sr_parsons	Mean	0.592	0.494
	SD	0.259	0.228
	SE	0.016	0.013

Table 59: Coefficients (β) results of regression on mg_completion. Significance is marked: '***': p < .001, '**': p between .001 and .01, '*': p between .01 and .05, and '.': p between .05 and .1.

Predictor	Beta
pretest	.045
year	-.001
social	-.007
pretest*social	**.269 ** **
pretest*year	-.056
year*social	**-.088 * **

Table 60 presents the statistics of navigation and performance in self-assessment content of the *individual* groups between studies and the result of linear regression models built on these variables considering the predictors *pretest*, *year* (dummy with 0 indicating 2015 and 1 indicating 2016), and the interaction term *pretest*year*. The table shows a clear effect of the year on success rates and in all navigation variables. This means that adding the fine-grained components in the individual version of Mastery Grids has a noticeable impact on performance and navigation within the system, although it has no impact on the amount of practice activity performed.

Table 60: Mean and Standard Deviation of the navigation and performance in self-asssessment variables between the individual group in 2015 study (N=173) and the KC group in 2016 study (N=104). The second part of the table (right) shows the coefficients of the regressions performed on these variables.

	2015		2016		Regressions		
	Mean	SD	Mean	SD	Beta pre	Beta year	Beta pre*year
eff_questions	-.025	.449	-.003	.444	**.366 ***	.044	-.026
eff_parsons	-.058	.906	.033	.473	.061	.057	.243
sr_questions	.448	.207	.621	.188	**.190 ****	**.178 *****	.017
sr_parsons	.615	.273	.487	.239	**.213 ***	**-.143 ****	.091
prob_attempt	.420	.240	.580	.240	.006	**.168 *****	-.013
ratio_gui	.860	.270	.300	.130	**-.193 ***	**-.615 *****	.189
term_regularity	.436	.083	.467	.083	-.031	**.025 .**	.025

Adding fine-grained information is positively related to the success rate of questions and negatively related to the success rate of parsons, and the effects have a similar magnitude. Interpreting these results is not an easy endeavor. One reason for the difference between success rates of questions and Parson problems may be due to the nature of these two types of self-assessment content. Another reason might be related to the fact that students in the 2015 study were exposed to Parson problems in their mandatory exercise requirements in the course, thus they may have gained experience in solving this type of contents outside their activity with Mastery Grids.

Regarding navigation, the positive influence of *year* on the probability of attempting open activities, and the negative relation observed on the ratio of time spent in the interface suggest that adding the fine-grained component has a positive effect on helping students become more efficient in their navigation through the system, confirming hypothesis **H2**. Students exposed to the *individual* Rich-OLM spent less time in the interface relative to their total time in the system (although they completed similar levels of the practice content), and

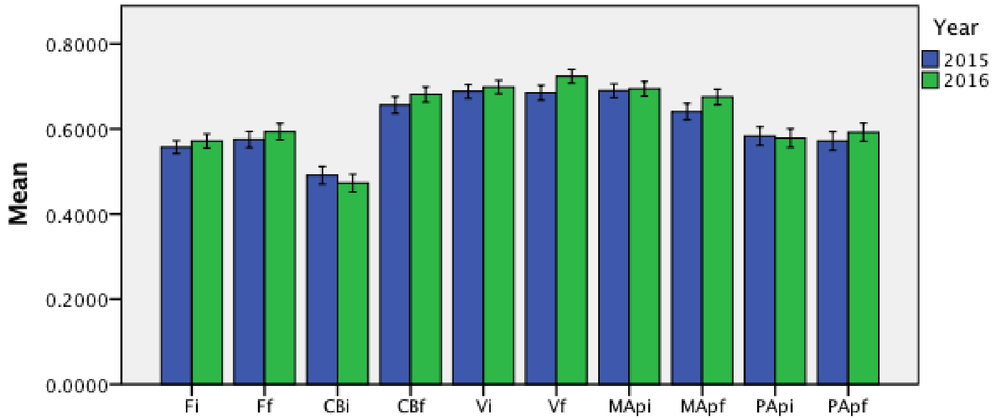


Figure 82: Mean of motivation at initial and final measures (i,f) in both years.

are more prone to attempt the activities they open, than students without the fine-grained visualization features.

10.5 MOTIVATION

Figure 82 and Table 61 show basic statistics of the motivation factors measured at the beginning and at the end of the term in both studies. These statistics consider all students who answer the questionnaires and who finished the course. The non-parametric Mann-Whitney did not show differences between years in the initial measure of any motivational factor, except for Competency Beliefs (p=.021), with this motivational factor being lower in 2016.

In both studies, I looked into the role of Mastery Approach and Performance Approach (measured at the beginning of the term) on engagement with the system when different features are enabled (*social*, *gauge*). There are differences in the findings between the studies. In the 2015 study I observed that the presence of *social* features interacts positively and

Table 61: Statistics of motivation factors measured at the initial and final points in the term in both studies. The last column shows the significant contribution of *year* (being 1 for 2016 study) in predicting the final measure after considering the initial measure. Significance is labeled as '***' for p < .001, '**' for p between .001 and .01, '*' for p between .01 and .05, and '.' for p between .05 and .1.

		2015		2016		Regression
		Initial	Final	Initial	Final	β_{year}
Fascination	Mean	0.563	0.578	0.578	0.595	
	SD	0.159	0.193	0.172	0.202	
	SE	0.007	0.009	0.007	0.010	
Competency Beliefs	Mean	0.505	0.656	0.477	0.682	.032 **
	SD	0.219	0.197	0.215	0.187	
	SE	0.010	0.009	0.009	0.009	
Values	Mean	0.696	0.685	0.703	0.726	.032 ***
	SD	0.168	0.179	0.164	0.170	
	SE	0.007	0.009	0.007	0.008	
Mastery Approach	Mean	0.693	0.641	0.700	0.675	.032 **
	SD	0.167	0.200	0.176	0.192	
	SE	0.007	0.010	0.008	0.009	
Performance Approach	Mean	0.582	0.568	0.584	0.590	.023 .
	SD	0.229	0.226	0.224	0.220	
	SE	0.010	0.011	0.010	0.011	
Social Comparison Orientation	Mean		0.589		0.584	
	SD		0.128		0.118	
	SE		0.006		0.006	

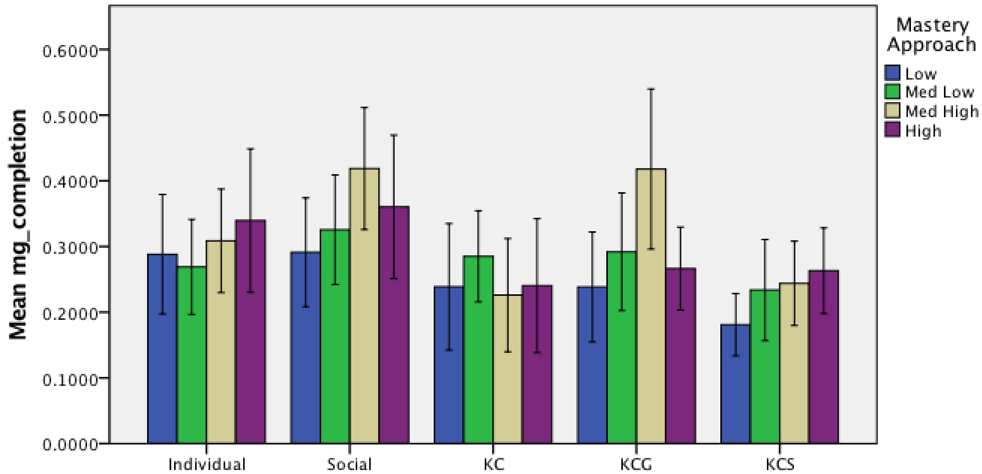


Figure 83: Mean level of completion of activities in each treatment groups across both studies and for different levels of Mastery Approach orientation. Error bars represents 2 standard error of the mean.

significantly with the Performance Orientation (*PAp*social*, $\beta = .351$, $p < .05$), indicating that students in the *social* group engaged correlated in part to their performance orientation (yet it is not correlated to their Mastery Approach orientation). However, this effect was not observed in the 2016 study. Instead, in that study, both motivational factors, but mainly Mastery Approach orientation, had a positive role in the group that was exposed to the *learning gauge*. This effect can be seen in Figure 83, where students with medium-high Mastery Approach in the gauge groups showed a considerably higher level of activity.

10.6 CHANGE OF MOTIVATION

As shown in Table 61 and Figure 82 I observed differences in the change of motivation factors from the initial motivation to the final motivation. To test these differences I built regression models in which the *year* was added as the dummy variable with value 0 for 2015 and 1 for 2016. Models were built separately for each motivational factor measured at the end of the term (final) and included as predictors the motivational factor at the beginning of the term (initial) and the dummy variable *year*. We report the models in which *year* contributes significantly to improve the preliminary model built only with the motivational *initial* measure as predictor. The beta coefficient of *year* and its significant value are reported in the last column of Table 61.

The Competency Beliefs factor was lower at the beginning of the 2016 course, but it increased more than in the 2015 course. Also, in 2016, the increase of Values is significantly higher than in 2015. A positive effect was also observed regarding the Mastery Approach orientation. While in both groups this motivational factor decreased, it decreased less in 2016. Regarding Performance Approach, the regression found a marginal significant positive influence of the year. Performance orientation presents a slight decrease in 2015 (although in chapter 6, I reported no significant difference in the change of Performance orientation), but remained steady in 2016. In summary, the motivational factors developed in more of a positive direction in 2016 than in 2015.

Regarding the differences in motivation change across the treatment groups, these were not observed for students who used the system. This was checked with a series of regressions performed on the final motivation (f) with the predictor being the initial motivation (i) added to the first model, and predictors *mg_completion*, *year* (dummy), *social* (dummy), *gauge* (dummy), and interaction terms *mg_completion*year*, *mg_completion*social*, and *mg_completion*gauge*, added to the second model using the *stepwise* method. Results showed no significant effect associated with engagement (*mg_completion*), nor on the interface variation (*social*, *gauge*), nor on the interactions between them. This means that even though motivation changes, there is no evidence that the use of the practice system in its different flavors has any impact, whether positive or negative, on this change. These observations align towards rejecting, at least in the context of my work and studies, the Hypotheses **H6** *The active use of OLM with social comparison features will increase the performance orientation of the students.* and **H7** *The active use of OLM with fine-grained features will increase the mastery orientation of the students.*

11.0 CONCLUSIONS AND DISCUSSION

From a general perspective, this thesis work has been devoted to studying the effects that Open Learner Models (OLM) with social comparison features and different levels of details (granularity) have in the learning experience of students who are using an online practice and learning system. More specifically, through a series of studies in different settings and involving the domain of computer programming, I have observed how engagement, navigation patterns, and performance in the practice system vary when the system includes visual features enabling group comparison, and visual features extending the coarse-grained OLM with fine-grained information. To deepen the studies, I incorporated the measure of some individual factors which, from a theoretically perspective, are close related to the nature of the studies: prior knowledge, Social Comparison Orientation, Competency-Beliefs (a form of self-efficacy), and motivational orientations framed by the Achievement-Goal Orientation Framework. The thesis contributes to understanding how visual features such as OLM can support the learning process in terms of engaging students to practice and helping them to navigate to useful content.

While previous chapters in this thesis are organized by research study, with each of them contributing to partially answer the research questions, the conclusions in the following sections are organized by research question. Also, hypotheses related to the research questions are also listed in each of the following sections. The text in the following sections may seem redundant at times because conclusions on one research question usually take elements of the other research questions. The main findings are summarized and highlighted in color blocks in each of the following sections.

It is important to notice that in the two main classroom studies, reported in chapters 6 and 9, students were separated into groups with different interface features presented to students with similar levels of prior-knowledge (as measured by the pretest), post-knowledge (posttest) and initial levels of motivation. It is also important to notice that in both studies, the activity in the practice system (Mastery Grids and Rich-OLM), which included the same content material in both studies, showed positive relation to the posttest, after controlling for the pretest, which confirmed the benefit of the learning content. Although the effect of the interventions on learning outcomes is not the center of my work, it is an element that cannot be neglected.

Consistently, in both classroom studies (chapters 6 and 9), the completion of activities in the system is associated with a considerable increase in posttest scores.

11.1 THE EFFECTS OF SOCIAL COMPARISON FEATURES

RQ 1 What are the effects of an OLM with social comparison features (or OSLM) on *System Activity* compared to an individual-view OLM?

H1 *Students exposed to an OLM with social comparison features increase the level of activity in the system.*

In general, the study presented in chapter 6 showed smaller levels of engagement with the practice content of the system than was shown in previous studies in other programming courses, as reported in chapter 4. Thus, the positive effects on engagement associated with the coarse-grained social comparison features was smaller, but still observable, confirming **H1**. Students exposed to coarse-grained social comparison features tended to be engaged earlier in the system, and showed higher levels of activity in the first weeks of the term (chapter 6, section 6.7.2). They not only accomplished more activity, but moved forward through the content faster, as shown in section 6.7.4. The study corroborated previous findings of these effects and the importance of engaging with the practice content: doing more activities in the system is related to higher levels of learning gain throughout the term. The effects on engagement were more noticeable when looking at highly motivated students, specifically with high Performance oriented students. This observation is aligned with the nature of this factor and supports the general idea that by showing the learner models of others, the system feeds a competitive spirit that is stronger among who see such competition as a goal (performance orientation in the achievement-goal framework) of the learning process.

Even though the effects of using social comparison features in OLM have been studied in the past, the study in chapter 6 was performed in a different country than in our previous studies, and in a different domain, so this work contributes to generalizing the findings.

Effects associated with the social comparison features changed when they were combined with the fine-grained components. As observed in the study reported in chapter 9, the group of students exposed to the Rich-OLM with social comparison features, which means a combination of coarse-grained and fine-grained OLM and comparison features at both granularity levels, showed slightly lower levels of engagement than the other groups, which were using the Rich-OLM alone. This contrasts with the overall higher level of engagement of the *social* group in previous studies and in the study of chapter 6. I hypothesize that this negative effect may be due to the increase in complexity of the visual interface. Complexity is an important issue as exposed by students in the studies reported in chapters 7 and 8, thus it deserves more quantitative and qualitative research.

At a coarse-grained level, the social comparison features were associated with higher levels of engagement, confirming hypothesis **H1**. When adding together the fine-grained features and the social comparison features, the effect is reversed and students tended to do less activity. This may be due to the increase in complexity of the visual interface.

11.2 THE EFFECTS OF FINE-GRAINED OLM

RQ 2 What are the effects of fine-grained OLM on *system activity*?

H2 *Fine-Grained OLM helps students to navigate the content of the system more efficiently.*

H2G *Fine-Grained OLM complemented with the Learning Gauge helps students to navigate the content of the system more efficiently.*

To expand the benefits of the Mastery Grids OLM, I enthusiastically decided to develop a visualization that would deliver more information of the learner model than had previously been included in the interface, and which I refer to as fine-grained information. The main reason for adding more details to the OLM was to improve navigational support, i.e., help students navigate the system and find useful content resources. After interviewing students and performing controlled studies contrasting alternative visualizations, I built the *Rich-OLM*. Rich-OLM extends the Mastery Grids by adding a bar-chart representing the knowledge progress on each of the fine-grained components of the domain model, which are also called *concepts* or *knowledge components*, or simply *kcs*. After considering the feedback obtained from participants in past studies, the interface was also built to show the relationship between the fine-grained and coarse-grained components, as well as between the fine-grained components and the content activities. These design features were in response to student requests, which asked for better navigational *affordances* in the interface, indicating what is involved in each topic and what is involved in each activity.

Analyses reported in chapter 10, where I compared groups across the classroom studies with and without the addition of fine-grained components, confirmed the usefulness of the detailed information supporting navigation, at least when the social comparison features are not present. When comparing the groups that were exposed to the *Individual* interface across studies, I noticed that while there were no differences in the level of general engagement (amount of activity), students who used the Rich-OLM showed a higher probability for opening and attempting an activity, and simultaneously invested less time using the interface, in proportion to their total time in the system. These observations suggest that students exposed to fine-grained components seem to more easily find activities that they are willing to complete. These findings support hypothesis **H2**.

Hypothesis **H2** is confirmed: fine-grained features were associated with more efficient navigation, making students reach activities they are willing to attempt easily and in less time.

During the interviews and studies reported in chapters 7 and 8, students also expressed concerns that adding more information to the visual OLM would increase the complexity of the system. This concern is supported in the study reported in chapter 9, in which the addition of both the social comparison features and a fine-grained OLM, i.e., adding a considerable amount of new information to the interface, is associated with a lower level of engagement with system activities.

Anticipating that fine-grained information will amplify complexity because of the information overload that detailed information conveys, a visual feature was added to the Rich-OLM to help interpret the concept information associated with different activities. This visual feature takes the form of a gauge, called the *learning gauge*, which indicates a learning opportunity when an activity is moused over. One of the controlled studies presented in chapter 8 confirms the utility of the *learning gauge* when seeking content to learn. However, when evaluated in a classroom study (chapter 9), the group exposed to the *learning gauge* showed a tendency to open more activities relative to the number of activities completed. I associated this effect with exploration of the content rather than with efficiency in navigation, supporting the rejection of hypothesis **H2G**.

The *learning gauge* was also associated with other effects. It showed mild but positive effects on engagement with the learning content. This engagement is more clear when motivation is included: as motivation increases, engagement increases and this is enhanced when the interface includes the *learning gauge*. This feature also showed the effect of retaining students, making students with high levels of activity even more likely to come back to the system and increase activity in a regular way during the term. This observation makes sense because we could expect students to benefit from the complex interface features after gaining some experience using the system. At that point, the additional complexity of the system is not as vexing to them, while the benefits are still increasing.

These positive effects of the *learning gauge* open an interesting line of research closely related to the area of learning analytics, in which the focus is to deliver the OLM with specific pedagogical intentions. In this sense, my work is preliminary, and overall, the results are encouraging, although not yet conclusive. More research is needed to contrast different visual aids in different domains. One such possible future direction is to provide a simpler visual aid to help interpret the social comparison information at a fine-grained level. This could help to address the increased complexity when these features are combined.

Students exposed to the *learning gauge* showed increased activity, higher regularity during the term, and entered the system more frequently. Results also showed that the *gauge* was associated less with efficient navigation than with an increase in an *exploratory* type of behavior, where students opened many activities without completing them, compared to students who were not exposed to the *gauge*. This last observation supports the rejection of **H2G**.

11.3 THE ROLE OF INDIVIDUAL DIFFERENCES

RQ 3 How do individual differences influence system activity within an OLM?

- **RQ 3.1** How does prior knowledge influence system activity within an OLM?
- **RQ 3.2** How does learning motivation influence system activity within an OLM?
- **RQ 3.3** How does social comparison orientation influence system activity within an OLM?
- **H3** *Social comparison features increase the engagement of students who are highly performance oriented.*
- **H4** *Students with a higher Mastery orientation will use the fine-grained components more.*
- **H5** *The effects of social comparison features of the system will be stronger for students with higher Social Comparison Orientation.*

Regarding **RQ 3.1**, an important finding is that prior knowledge increased engagement with the practice system when social comparison features are present. In the study reported in chapter 6, I noticed that in the *Social* group, higher pretest students engaged significantly more, compared to those in the *Individual* group. Analyses of when the students initially entered the system during the term showed that high pretest students tended to enter the system sooner and account for most of the early activity in the system, probably because the practice content demands less effort for students who have higher prior knowledge. This happened in both the *Individual* and *Social* groups, but in the latter group, the early students tended to maintain their engagement through the term. Although more research is needed to clearly prove this observation, a reasonable explanation is that since students with a higher pretest entered the system earlier, they gained a higher status, which is visually displayed by the social comparison features of the interface. Competition effects then encouraged them to try to maintain their status by continuing completing activity within the system.

Another important observation is that, although I clearly saw that high pretest students became more engaged when having social comparison features, the opposite did not occur for low prior knowledge students. Low prior knowledge students in the *Social* group did not engage less than low prior knowledge students in the *Individual* group, thus there is no evidence that students will become discouraged when finding themselves at a disadvantage when compared to others.

High pretest students engaged early in the system, and when exposed to the social comparison features, they tended to maintain their engagement. While high pretest students became more engaged in the *Social* group, low pretest students in the *Social* group did not engage less compared to the *Individual* group.

Regarding **RQ 3.3** in the studies of chapters 6 and 9, I also included the Social Comparison Orientation scale (SCO), a self-reported measure of how willing and how important it is for students to compare themselves to others. I expected this measure to portray a clear story about engagement and navigation within the practice system when the social comparison features were present. However, the measure did not produce very clear results, thus **H5** was not confirmed. A possible explanation of the low level of discrimination found by this scale is its low variability.

One SCO effect, observed in the chapter 6 study, was an increase in the regular use of the system during the term, where high SCO students tended to become regular in the presence of the social comparison features. Another effect is observed in the chapter 9 study, where SCO produced higher proportions of time spent using the interface when social comparison features were present.

Other individual differences that showed more effects are the Mastery and Performance orientations extracted from the Achievement-Goal Orientation questionnaire. These motivational orientations provide measures of how students oriented their goals when facing learning, such that the *Mastery* goals related to learning, while the *Performance* goals related to scores or compare themselves to others. As expected, higher Performance oriented students showed higher levels of engagement when exposed to social comparison features, confirming **H3**. This effect was clear in the study reported in chapter 6, but became blurry in the chapter 9 study, where I added the fine-grained components. As mentioned earlier, it seems that adding fine-grained detailed information plus social comparison information may be too much, making the system too complex. Also observed in the study of chapter 9, while the *Social* group showed lower levels of activity, it also presented higher relative levels of time spent engaged with the interface, conditioned to the level of Mastery orientation. It appears that highly motivated students overcome the complexity of the interface and "play" more with it.

Performance orientation amplified the effect of the social comparison features, confirming the theoretical basis of this construct and hypothesis **H3**. Students with a higher performance orientation tended to engage more in the Social group.

Regarding the relationship between the *learning gauge* and motivation, it is interesting to note a general null effect associated with Mastery orientation on engagement, but conditioned to the presence of the *gauge*. This effect contributes to support hypothesis **H4**. The effect is also aligned to the design of the *learning gauge*, which has the goal of directing the interpretation of the fine-grained information towards the opportunities of learning. I also observed that the *exploratory* effect of the *gauge*, expressed in the proportion of times they opened activities without attempting them, happened for high Mastery oriented students.

Mastery orientation interacted positively with the *learning gauge*, supporting hypothesis **H4**. High Mastery Approach students tended to explore more activities when exposed to the *gauge* and they also increased their level of activity completion.

Although not explored in this thesis work, the lower general levels of Achievement-Goal Orientations obtained in studies of chapters 6 and 9 may explain the lower general level of engagement with the practice system compared to other previous studies. This opens an interesting line of research into cultural differences that may be associated with different motivational orientations and perhaps, other individual differences.

11.4 THE CHANGE OF MOTIVATION

RQ 4 How does the use of a learning system featuring OLM, OSLM and fine-grained elements affect motivation?

- **H6** *The active use of OLM with social comparison features will increase the performance orientation of the students.*
- **H7** *The active use of an OLM with fine-grained features will increase the mastery orientation of the students.*

Motivation is known to change [Elliot and Murayama, 2008, Moore et al., 2011], and it is expected that motivation will evolve as the semester progresses. Recall that motivational factors were measured at the beginning and at the end of the term. This did not allow me to see the whole pattern of motivation evolution, but did register overall changes in motivation. As happens with learning outcomes, I understand that there could be many factors which influence a change in motivation, including the formal course experience and other content resources and activities performed by students. Thus, I did not expect to see a clear, nor a strong influence from our system, which was accessed voluntarily and complemented the other mandatory exercise system, to change motivation.

In general, Achievement-Goal motivational orientations tended to decrease while Fascination and Competency Beliefs tended to increase from the beginning to the end of the term. Overall, I observed some differences in the changes of motivation between the studies reported in chapters 6 and 9. In summary, the motivational factors developed in a more positive direction in the second study. However, the association between these motivational changes and the use of the system were practically non-existent. Because of the null relationship between use of the system and motivational change, I lean towards rejecting hypotheses **H6** and **H7**. However, as explained before, this null effect is to some extent expected because of the setup of the studies, where the use of the OLM system is complementary and not critical to the learning process during the term. It is important to notice, though, that this null effect also supposes that there is no negative effect associated with the system.

No clear effects (neither positive nor negative) were observed in the change of motivation due to activity within the system and/or interaction with system interface features. Thus, there is no evidence to support either **H6** or **H7**.

11.5 LIMITATIONS AND FUTURE WORK

Several limitations of the work presented in this thesis are due to decisions made while designing the studies. One such limitation decision was to provide the system as a voluntary (and complementary) practice system. This means that self-selection could bias the group that started and continued to use the system. It will be interesting to see what happens if the system is offered in a mandatory way, to see if the interface has positive effects on students who wouldn't be using it otherwise. Future work can explore the role of OLM and smart content delivered in a mandatory and non-mandatory fashion. Moreover, it will be interesting to see a combined system, where mandatory activity might be shown to affect engagement with non-mandatory practice content.

Another limitation in the design of the classroom studies was related to the study reported in chapter 9, where there was no control group to test the differences created by adding the fine-grained components of the Rich-OLM. The initial idea was to compare the results of this study with the previous similar study reported in chapter 6. However, differences in the study setups and in the course where the system were deployed made comparisons between them not completely fair, weakening the results. With this in mind, some comparisons were made in chapter 10, but future work is needed to strengthen the observed effects of the fine-grained components.

The work of this thesis explores several issues related to the inclusion of social comparison features into Open Learner Models. One limitation related to this is that I used only a few interface approaches to represent information about other students (group comparisons at coarse- and fine-grained levels). The effects associated with these comparison features might change if they are designed differently, for example, to stress comparison on specific portions of the activity, or with differently targeted groups of peers, or by adding privacy control elements such as whether to show peer names.

Although classroom studies presented in chapters 6 and 9 were deployed in large courses, they were also limited to a specific population. These studies were deployed in introductory programming courses in a university in Finland, and were certainly biased by the educational culture and motivational orientations of Finnish students. Although I can make some comparisons about motivation and levels of engagement with students from other countries in these past studies, it is necessary to extend the studies containing motivational factors to many more domains and cultural backgrounds. This is an interesting line of research, seeking to understand how the learning experience varies across cultural differences.

The analyses that focus on the change in motivation are limited to the way the motivation was measured and the fact that the use of the system was a complementary role in a bigger learning context (the course). Since motivation can change because of learning experiences, a more advanced study could be designed to see how it changes over the term by measuring different motivational factors at several intervals, and in a more controlled environment. Because of this, I consider my work as a first step into this issue, and thus my results about changes in motivation should only be considered as preliminary and not conclusive.

APPENDIX A

PRETEST AND POSTTEST

Both pretest and posttest to measure knowledge on python programming and applied in studies reported in chapters 6 and 9, used identical 10-questions test that are presented below.

Problem 1. Consider the following code segment:

```
i = 14
j = 2
k = (i + 1) * j
j = 3
```

What is the final value of the variable k (after the line j = 3)?

Problem 2. Consider the following code segment:

```
my_year = 2012
my_text = "Hello, ES17!"
result = 0

if len(my_text) > 20:
    result = 1
    if len(my_text) < 30 and my_year >= 2012:
        result += 5
```

```
else:
    if my_year >= 2000:
        result += 10
    else:
        result += 100
```

What is the final value of the variable result?

Problem 3. For each of the following two code segments, what is the final value of result:

Code segment 1:

```
i = 3
result = 0
while i < 4:
    result = result + i
    i = i + 1
```

Result:

Code segment 2:

```
result = 0
for i in range(5, 0, -1):
    result = result + i


print(result)
```

result:

Problem 4. What would be the output of the following code fragment:

```
data = [0] * 5
for i in range(5):
    data[i] = i * i
data[2] += 1
```

```
print(data[2])
```

Output:

Problem 5. What would be the output of the following code fragment:

```
list1 = []
list1.append(1.1)
list1.append(2.2)
list1.append(3.3)
del list1[0]
for d in list1:
    print(d)
```

Output:

Problem 6. What would be the output of the following code fragment:

```
def calculate(a, b):
    return (1 - b / 100.0) * a


original = 200.0
new1 = calculate(original, 25.0)
new2 = calculate(original, 50.0)
print(new1)
print(new2)
```

Output:

Problem 7. What would be the output of the following code fragment:

```
str1 = "Welcome!"
str2 = ""
i = len(str1) - 1
while i >= 0:
    str2 += str1[i]
```

```
    i = i - 1
print(str2)
```

Output:

Problem 8. Assume that the text file results.txt contains the following two lines (and nothing else):

12;48;30 33;11;50

In that case, what is the output of the following code fragment (we have omitted all error handling to make the program as short as possible):

```
file1 = open("results.txt")
for line in file1:
    points = line.split(";")
    total = int(points[1]) + int(points[2])
    print(total)
file1.close()
```

Output:

Problem 9. What would be the output of the following code fragment:

```
data = ["not known", "45.0"]
for el in data:
    try:
        result = 2 * float(el)
        print(result)
    except ValueError:
        print("Incorrect data")
```

Output:

Problem 10. Consider the class Rectangle defined as follows:

```
class Rectangle:
    def __init__(self, x, y, height, width):
```

```
        self.x = x
        self.y = y
        self.height = height
        self.width = width

    def get_height(self):
        return self.height

    def get_width(self):
        return self.width

    def magnify(self, ratio):
        self.height = self.height * ratio
        self.width = self.width * ratio

    # some other methods
```

What would be the output of the following code fragment using the new method?

```
my_box = Rectangle(50, 40, 10, 10)
my_box.magnify(3)
print(my_box.get_height())
print(my_box.get_width())
```

Output:

APPENDIX B

MOTIVATION QUESTIONNAIRE

B.1 LEARNING ACTIVATION

The Learning Activation questionnaire used in this thesis is a reduced version of the questionnaire developed by [Moore et al., 2013] and keeps only 3 factors: Fascination, Competency Beliefs, and Values.

B.1.1 Fascination

There are 4 items measuring fascination. The items use a 4 point scale with different phrasing as presented in Table 62

Table 62: Items of the motivation questionnaire corresponding to the factor Fascination.

	Item	Answers
1	In general, I find programming:	Very boring, Boring, Interesting, Very Interesting
	In general, thinking about working on programming tasks, I will:	
2	Enjoy it	NO!, no, yes, YES
3	Love it	NO!, no, yes, YES
	Please fill in the circle that represents how YOU feel about programming.	
4	I want to know all I can about programming	NO!, no, yes, YES

B.1.2 Competency Beliefs

Competency Beliefs were measured with 5 items. The answers for all of them were in a 5 point scale where the extremes and middle point were labeled as “I’m sure I CANNOT do it”, “I’m not sure I can do it”, and “I’m sure I CAN do it”. The items are shown in Table 63

Table 63: Items of the motivation questionnaire corresponding to the factor *Competency Beliefs*. Answers were requested in a 5 point likert scale.

1	I can answer all the questions on a programming class test or exam.
2	I can figure out how to finish a programming class project at home.
3	I can find and understand what I am looking for on website that has programming or code information on it.
4	If a group of students is having a discussion about the code of an assignment I could participate actively.
5	If I were working on a programming class project I could find useful books in a library and read them on my own.

B.1.3 Values

Values were measured with 5 items shown in Table 64.

Table 64: Items of the motivation questionnaire corresponding to the factor Values.

	Item	Answers
	Please fill in the circle that represents how YOU feel about programming.	not important, a little important, important, very important
1	How important is it for you to learn about programming for your future career?	
	Please fill in the circle that represents how YOU feel about programming.	NO!, no, yes, YES
2	Do you think programming is useful for making the world a better place to live?	
3	Do you think programming is useful in your life?	
4	I think programming will be useful for me in the future.	
5	I think programming ideas are valuable.	

B.2 ACHIEVEMENT-GOAL ORIENTATION

Achievement-Goal Orientations include 4 factors: Mastery Approach, Mastery Avoidance, Performance Approach, and Performance Avoidance. These are measured with a 12-items (3 items for each factor) developed by Elliot [Elliot and McGregor, 2001]. Answers were collected in a 7-point likert scale showing only the extremes and middle point: "Not at all true of me", "Unsure", "Very true of me". Table 65 shows the 12 items of the questionnaire in the order in which they are presented. The corresponding factor (which is not shown in the questionnaire) is shown in the table at the left side.

Table 65: Achievement-Goal questionnaire items. The corresponding factor is shown in the left column.

Factor	Item
Mastery Avoidance	I strive to avoid an incomplete understanding of the course material.
Mastery Approach	My goal is to learn as much as possible
Performance Avoidance	My aim is to avoid doing worse than other students.
Performance Approach	My goal is to perform better than the other students.
Mastery Approach	My aim is to completely master the material presented in class.
Mastery Approach	I strive to understand the content of the course as thoroughly as possible.
Mastery Avoidance	My aim is to avoid learning less than I possibly could.
Performance Avoidance	My goal is to avoid performing poorly compared to others.
Performance Approach	I strive to do well compared to other students.
Mastery Avoidance	My goal is to avoid learning less than it is possible to learn.
Performance Avoidance	I strive to avoid performing worse than others.
Performance Approach	My aim is to perform well relative to other students.

APPENDIX C

SOCIAL COMPARISON ORIENTATION QUESTIONNAIRE

Social Comparison Orientation is measured with the INCOM questionnaire developed by [Gibbons and Buunk, 1999]. The answers are collected in a 5-point likert scale ("Strongly disagree", "Disagree", "Neither agree nor disagree", "Agree", "Strongly Agree"). Items of the questionnaire are listed in Table 66.

Table 66: Items of the INCOM questionnaire that measures Social Comparison Orientation.

Item
I often compare myself with others with respect to what I have accomplished in life
If I want to learn more about something I try to find out what others think about it
I always pay a lot of attention to how I do things compared with how others do things
I often compare how my loved ones (boy or girlfriend family members etc.) are doing with how others are doing
I always like to know what others in a similar situation would do
I am not the type of person who compares often with others
If I want to find out how well I have done something I compare what I have done with how others have done.
I often try to find out what others think who face similar problems as I face
I often like to talk with others about mutual opinions and experiences
I never consider my situation in life relative to that of other people
I often compare how I am doing socially (e.g., social skills popularity) with other people.

APPENDIX D

CONCEPT SPACE

Table 67: List of concepts used in the Java programming version of Mastery Grids. On the left side are the associated topics (PART 1).

Topic	KCs
Variables	Addition, Multiplication, Simple Assignment, String Literal, String Variable, println, print
Primitive Data Types	double Type, Explicit Type Casting, Char Type, int type, float type, Integer.parseInt, Implicit conversion, Double.parseDouble
Arithmetic Operations	Post Increment, Post Decrement, Pre Decrement, Modulus, Pre Increment, Math.pow, Math.ceil, Math.abs, Math.sqrt, Subtraction, Multiply assignment, Math.round, Division, Add Assignment
Strings	String Type, String Literal Method, String Creation, Concatenation, substring, replace, length, equalsIgnoreCase, equals, charAt, String Initialization
Constants	Constant Initialization, Constant, Constant Invocation
Decisions	if else if, if else, if, switch, Break Statement, Default Clause, Case Clause
Boolean Expressions	Or, Equal, Greater Equal than, Greater than, Less Equal than, Less than, Not Equal, Not, And, Object Equality, boolean Type, null
Loops	do-while, for, while, Nested Statement
Objects	Object Creation

Table 68: List of concepts used in the Java programming version of Mastery Grids. On the left side are the associated topics (PART 2).

Topic	KCs
Classes	Private Field Specifier, Constructor Specifier, Public Field Specifier, Reference this, return, Static Field Specifier, Class Field, Object Method Call, Instance Field, final specifier, Instance Field Initialization, Instance Field Invocation, Constructor Definition, Class Constant
Arrays	for each, Array Initializer, Array Variable, Array Length, Array Initialization, Array Element, Array Type, Array Creation
Two-dimensional Arrays	MultiDimensional Array
ArrayList	ArrayList, add, get, set, size, remove
Inheritance	Method Overriding, Polymorphic Object, Overriding toString, Overriding equals, Method Inheritance, Extends Specification, Inheritance Polymorphism, Object, Super Constructor Call, Reference super, Super to Subclass, Super Method Call
Interfaces	Interface Polymorphism, Interface Definition, Abstract Method, Method Implementation, Interface to Class, implements Specification

Table 69: List of concepts used in the Python programming version of Mastery Grids. On the left side are the associated topics.

Topic	KCs
Variables	Unary Subtraction, Floor Division, Pow, Subtraction, Multiplication, Modulo, Int, Addition, Division, Augmented Assign, Assign
Strings	Slice, String
Comparison	Not Equal, Greater or equal, Equal, Greater than, Less than, Less or equal
Logical Operators	And, Not, Or, TRUE, FALSE
If Statements	If
Loops	For Loop, While loop, Continue
Lists	In, Index, List
Output Formatting	Float
Dictionary	Dictionary, None
Functions	Function Definition, Return, Argument
Classes and Objects	Attribute, Import From, Class Definition, Alias
exceptions	Try Except, Exception Handler, Try, Raise

BIBLIOGRAPHY

[Ames, 1992] Ames, C. (1992). Classrooms: Goals, structures, and student motivation. *Journal of educational psychology*, 84(3):261.

[Bandura, 1986] Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Prentice-Hall, Inc.

[Brusilovsky, 2003] Brusilovsky, P. (2003). Adaptive navigation support in educational hypermedia: the role of student knowledge level and the case for meta-adaptation. *British Journal of Educational Technology*, 34(4):487–497.

[Brusilovsky, 2004] Brusilovsky, P. (2004). Knowledgetree: A distributed architecture for adaptive e-learning. In *13th International World Wide Web Conference, WWW 2004*, pages 104–113. ACM Press.

[Brusilovsky et al., 2013] Brusilovsky, P., Baishya, D., Hosseini, R., Guerra, J., and Liang, M. (2013). Knowledgezoom for java: A concept-based exam study tool with a zoomable open student model. In *Advanced Learning Technologies (ICALT), 2013 IEEE 13th International Conference on*, pages 275–279. IEEE.

[Brusilovsky et al., 2004a] Brusilovsky, P., Chavan, G., and Farzan, R. (2004a). Social adaptive navigation support for open corpus electronic textbooks. In *Adaptive Hypermedia and Adaptive Web-Based Systems*, pages 24–33. Springer.

[Brusilovsky et al., 2014a] Brusilovsky, P., Edwards, S., Kumar, A., Malmi, L., Benotti, L., Buck, D., Ihantola, P., Prince, R., Sirkiä, T., Sosnovsky, S., et al. (2014a). Increasing adoption of smart learning content for computer science education. In *Proceedings of the Working Group Reports of the 2014 on Innovation & Technology in Computer Science Education Conference*, pages 31–57. ACM.

[Brusilovsky et al., 2014b] Brusilovsky, P., Edwards, S., Kumar, A., Malmi, L., Benotti, L., Buck, D., Ihantola, P., Prince, R., Sirki, T., Sosnovsky, S., Urquiza, J., Vihavainen, A., and Wollowski, M. (2014b). Increasing adoption of smart learning content for computer science education. In *Proceedings of the Working Group Reports of the 2014 on Innovation & Technology in Computer Science Education Conference*, pages 31–57. ACM.

[Brusilovsky et al., 2011] Brusilovsky, P., Hsiao, I.-H., and Folajimi, Y. (2011). Quizmap: open social student modeling and adaptive navigation support with treemaps. In *Towards Ubiquitous Learning*, pages 71–82. Springer.

[Brusilovsky and Millán, 2007] Brusilovsky, P. and Millán, E. (2007). User models for adaptive hypermedia and adaptive educational systems. In *The adaptive web*, pages 3–53. Springer-Verlag.

[Brusilovsky et al., 2015] Brusilovsky, P., Somyürek, S., Guerra, J., Hosseini, R., and Zadorozhny, V. (2015). The value of social: Comparing open student modeling and open social student modeling. In *User Modeling, Adaptation and Personalization*, pages 44–55. Springer.

[Brusilovsky et al., 2016] Brusilovsky, P., Somyurek, S., Guerra, J., Hosseini, R., Zadorozhny, V., and Durlach, P. (2016). Open social student modeling for personalized learning. *IEEE Transactions on Emerging Topics in Computing*, 4(3):450–461.

[Brusilovsky and Sosnovsky, 2005] Brusilovsky, P. and Sosnovsky, S. (2005). Engaging students to work with self-assessment questions: A study of two approaches. In *ACM SIGCSE Bulletin*, volume 37, pages 251–255. ACM.

[Brusilovsky et al., 2004b] Brusilovsky, P., Sosnovsky, S., and Shcherbinina, O. (2004b). Quizguide: Increasing the educational value of individualized self-assessment quizzes with adaptive navigation support. In *World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, volume 2004, pages 1806–1813.

[Brusilovsky et al., 2005] Brusilovsky, P., Sosnovsky, S., and Yudelson, M. (2005). Ontology-based framework for user model interoperability in distributed learning environments. In *E-Learn: World Conference on E-Learning in Corporate, Government, Healthcare, and Higher Education*, pages 2851–2855. Association for the Advancement of Computing in Education (AACE).

[Brusilovsky et al., 2010] Brusilovsky, P., Sosnovsky, S., Yudelson, M. V., Lee, D. H., Zadorozhny, V., and Zhou, X. (2010). Learning sql programming with interactive tools: From integration to personalization. *ACM Transactions on Computing Education (TOCE)*, 9(4):19.

[Brusilovsky and Yudelson, 2008] Brusilovsky, P. and Yudelson, M. (2008). From WebEx to NavEx: Interactive Access to Annotated Program Examples. *Proc. of the IEEE*, 96(6):990–999.

[Buchs and Butera, 2009] Buchs, C. and Butera, F. (2009). Is a partners competence threatening during dyadic cooperative work? it depends on resource interdependence. *European Journal of Psychology of Education*, 24(2):145–154.

[Bull, 2004] Bull, S. (2004). Supporting learning with open learner models. In *Proceedings of 4 th Hellenic Conference in Information and Communication Technologies in Education*. Citeseer.

[Bull, 2012] Bull, S. (2012). Preferred features of open learner models for university students. In *International Conference on Intelligent Tutoring Systems*, pages 411–421. Springer.

[Bull et al., 2016] Bull, S., Brusilovsky, P., Araujo, R., and Guerra, J. (2016). Individual and peer comparison open learner model visualisations to identify what to work on next. In *24th Conference on User Modeling, Adaptation and Personalization*. CEUR.

[Bull et al., 2010] Bull, S., Gakhal, I., Grundy, D., Johnson, M., Mabbott, A., and Xu, J. (2010). Preferences in multiple-view open learner models. In *European Conference on Technology Enhanced Learning*, pages 476–481. Springer.

[Bull et al., 2015] Bull, S., Johnson, M. D., Masci, D., and Biel, C. (2015). Integrating and visualising diagnostic information for the benefit of learning. *Measuring and Visualizing Learning in the Information-Rich Classroom*, page 167.

[Bull and Kay, 2007] Bull, S. and Kay, J. (2007). Student models that invite the learner in: The smili:() open learner modelling framework. *International Journal of Artificial Intelligence in Education*, 17(2):89–120.

[Bull and Kay, 2010] Bull, S. and Kay, J. (2010). Open learner models. In *Advances in intelligent tutoring systems*, pages 301–322. Springer.

[Bull and Kay, 2013] Bull, S. and Kay, J. (2013). Open learner models as drivers for metacognitive processes. In *International Handbook of Metacognition and Learning Technologies*, pages 349–365. Springer.

[Bull and Kay, 2016] Bull, S. and Kay, J. (2016). Smili: a framework for interfaces to learning data in open learner models, learning analytics and related fields. *International Journal of Artificial Intelligence in Education*, 26(1):293–331.

[Bull and Mabbott, 2006] Bull, S. and Mabbott, A. (2006). 20000 inspections of a domain-independent open learner model with individual and comparison views. In *International Conference on Intelligent Tutoring Systems*, pages 422–432. Springer.

[Bull et al., 2003] Bull, S., McEVOY, A. T., and Reid, E. (2003). Learner models to promote reflection in combined desktop pc/mobile intelligent learning environments. In *Proceedings of Workshop on Learner Modelling for Reflection, Supplementary Proceedings of the 11th International Conference*, volume 5, pages 199–208.

[Bull and McKay, 2004] Bull, S. and McKay, M. (2004). An open learner model for children and teachers: inspecting knowledge level of individuals and peers. In *Intelligent tutoring systems*, pages 646–655. Springer.

[Bull and Vatrapu, 2011] Bull, S. and Vatrapu, R. (2011). Supporting collaborative interaction with open learner models: Existing approaches and open questions. *Proceedings of the Computer Supported Collaborative Learning (CSCL) 2011*.

[Burguillo, 2010] Burguillo, J. C. (2010). Using game theory and competition-based learning to stimulate student motivation and performance. *Computers & Education*, 55(2):566–575.

[Buunk and Gibbons, 2007] Buunk, A. P. and Gibbons, F. X. (2007). Social comparison: The end of a theory and the emergence of a field. *Organizational Behavior and Human Decision Processes*, 102(1):3–21.

[Cialdini et al., 1999] Cialdini, R. B., Wosinska, W., Barrett, D. W., Butner, J., and Gornik-Durose, M. (1999). Compliance with a request in two cultures: The differential influence of social proof and commitment/consistency on collectivists and individualists. *Personality and Social Psychology Bulletin*, 25(10):1242–1253.

[Ciani et al., 2010] Ciani, K. D., Middleton, M. J., Summers, J. J., and Sheldon, K. M. (2010). Buffering against performance classroom goal structures: The importance of autonomy support and classroom community. *Contemporary Educational Psychology*, 35(1):88–99.

[Cimolino et al., 2004] Cimolino, L., Kay, J., and Miller, A. (2004). Concept mapping for eliciting verified personal ontologies. *International Journal of Continuing Engineering Education and Life Long Learning*, 14(3):212–228.

[Conejo et al., 2011] Conejo, R., Trella, M., Cruces, I., and Garcia, R. (2011). Ingrid: A web service tool for hierarchical open learner model visualization. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 406–409. Springer.

[Corbett and Bhatnagar, 1997] Corbett, A. T. and Bhatnagar, A. (1997). Student modeling in the act programming tutor: Adjusting a procedural learning model with declarative knowledge. In *User modeling*, pages 243–254. Springer.

[Davis et al., 2017] Davis, D., Jivet, I., Kizilcec, R. F., Chen, G., Hauff, C., and Houben, G.-J. (2017). Follow the successful crowd: raising mooc completion rates through social comparison at scale. In *LAK*, pages 454–463.

[de la Fuente Valentín and Solans, 2014] de la Fuente Valentín, L. and Solans, D. B. (2014). Am i doing well? a4learning as a self-awareness tool to integrate in learning management systems. *Campus Virtuales*, 3(1):32–40.

[Dimitrova et al., 2001] Dimitrova, V., Self, J., and Brna, P. (2001). *Applying interactive open learner models to learning technical terminology*. Springer.

[Duan et al., 2010] Duan, D., Mitrovic, A., and Churcher, N. (2010). Evaluating the effectiveness of multiple open student models in eer-tutor.

[Dweck, 2000] Dweck, C. S. (2000). *Self-theories: Their role in motivation, personality, and development*. Psychology Press.

[Elliot and McGregor, 2001] Elliot, A. J. and McGregor, H. A. (2001). A 2× 2 achievement goal framework. *Journal of personality and social psychology*, 80(3):501.

[Elliot and Murayama, 2008] Elliot, A. J. and Murayama, K. (2008). On the measurement of achievement goals: Critique, illustration, and application. *Journal of Educational Psychology*, 100(3):613.

[Elliott and Dweck, 1988] Elliott, E. S. and Dweck, C. S. (1988). Goals: an approach to motivation and achievement. *Journal of personality and social psychology*, 54(1):5.

[Epp and Bull, 2015] Epp, C. D. and Bull, S. (2015). Uncertainty representation in visualizations of learning analytics for learners: current approaches and opportunities. *IEEE Transactions on Learning Technologies*, 8(3):242–260.

[Falakmasir et al., 2012] Falakmasir, M. H., Hsiao, I.-H., Mazzola, L., Grant, N., and Brusilovsky, P. (2012). The impact of social performance visualization on students. In *Advanced Learning Technologies (ICALT), 2012 IEEE 12th International Conference on*, pages 565–569. IEEE.

[Festinger, 1954] Festinger, L. (1954). A theory of social comparison processes. *Human relations*, 7(2):117–140.

[Field, 2012] Field, A. (2012). *Discovering Statistics Using R*. Sage.

[Fulantelli et al., 2013] Fulantelli, G., Taibi, D., and Arrigo, M. (2013). A semantic approach to mobile learning analytics. In *Proceedings of the First International Conference on Technological Ecosystem for Enhancing Multiculturality*, pages 287–292. ACM.

[Gibbons and Buunk, 1999] Gibbons, F. X. and Buunk, B. P. (1999). Individual differences in social comparison: development of a scale of social comparison orientation. *Journal of personality and social psychology*, 76(1):129.

[Grant and Dweck, 2003] Grant, H. and Dweck, C. S. (2003). Clarifying achievement goals and their impact. *Journal of personality and social psychology*, 85(3):541.

[Guerra et al., 2016] Guerra, J., Hosseini, R., Somyürek, S., and Brusilovsky, P. (2016). An intelligent interface for learning content: Combining an open learner model and social comparison to support self-regulated learning and engagement. IUI.

[Harper et al., 2010] Harper, Yan Chen, F. M., Konstan, J., and Li, S. X. (2010). Social comparisons and contributions to online communities: A field experiment on movielens. *The American economic review*, pages 1358–1398.

[Higashi et al., 2017] Higashi, R. M., Schunn, C. D., and Flot, J. B. (in press 2017). Different underlying motivations and abilities predict student versus teacher persistence in an online course. *Educational Technology Research and Development*, pages 1–23.

[Hosseini and Brusilovsky, 2013] Hosseini, R. and Brusilovsky, P. (2013). Javaparser: A fine-grain concept indexing tool for java problems. In *The First Workshop on AI-supported Education for Computer Science (AIEDCS 2013)*, pages 60–63. University of Pittsburgh.

[Hosseini et al., 2015a] Hosseini, R., Hsiao, I.-H., Guerra, J., and Brusilovsky, P. (2015a). Off the beaten path: The impact of adaptive content sequencing on student navigation in an open social student modeling interface. In *Artificial Intelligence in Education*, pages 624–628. Springer International Publishing.

[Hosseini et al., 2015b] Hosseini, R., Hsiao, I.-H., Guerra, J., and Brusilovsky, P. (2015b). What should i do next? adaptive sequencing in the context of open social student modeling. In *Design for Teaching and Learning in a Networked World*, pages 155–168. Springer.

[Hsiao et al., 2013] Hsiao, I. H., Bakalov, F., Brusilovsky, P., and Knig-Ries, B. (2013). Progressor: social navigation support through open social student modeling. *New Review of Hypermedia and Multimedia*, 19(2):112–131.

[Hsiao and Brusilovsky, 2012] Hsiao, I.-H. and Brusilovsky, P. (2012). Motivational social visualizations for personalized e-learning. In *21st Century Learning for 21st Century Skills*, pages 153–165. Springer.

[Hsiao et al., 2010] Hsiao, I.-H., Sosnovsky, S., and Brusilovsky, P. (2010). Guiding students to the right questions: adaptive navigation support in an e-learning system for java programming. *Journal of Computer Assisted Learning*, 26(4):270–283.

[Huang et al.,] Huang, Y., Guerra, J., and Brusilovsky, P. A data-driven framework of modeling skill combinations for deeper knowledge tracing. In *Proc. of the 9th Intl. Conf. on Educational Data Mining*.

[Huang et al., 2016] Huang, Y., Guerra, J., and Brusilovsky, P. (2016). Modeling skill combination patterns for deeper knowledge tracing. In *Proceedings of the 6th Workshop on Personalization Approaches in Learning Environments (PALE 2016)*. 24th conference on User Modeling, Adaptation, and Personalization (UMAP 2016), CEUR workshop proceedings, this volume.

[Huguet et al., 2001] Huguet, P., Dumas, F., Monteil, J. M., and Genestoux, N. (2001). Social comparison choices in the classroom: Further evidence for students' upward comparison tendency and its beneficial impact on performance. *European Journal of Social Psychology*, 31(5):557–578.

[Ihantola and Karavirta, 2011] Ihantola, P. and Karavirta, V. (2011). Two-Dimensional Parson's Puzzles: The Concept, Tools, and First Observations . *Journal of Information Technology Education: Innovations in Practice*, 10:1–14.

[Kay and Lum, 2005] Kay, J. and Lum, A. (2005). Exploiting readily available web data for scrutable student models. In *AIED*, pages 338–345.

[Kay et al., 2006] Kay, J., Maisonneuve, N., Yacef, K., and Reimann, P. (2006). The big five and visualisations of team work activity. In *Intelligent tutoring systems*, pages 197–206. Springer.

[Kerly and Bull, 2008] Kerly, A. and Bull, S. (2008). Childrens interactions with inspectable and negotiated learner models. In *Intelligent Tutoring Systems*, pages 132–141. Springer.

[Khan and Pardo, 2016] Khan, I. and Pardo, A. (2016). Data2u: scalable real time student feedback in active learning environments. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*, pages 249–253. ACM.

[Kumar and Maries, 2007] Kumar, A. and Maries, A. (2007). The effect of open student model on learning: A study. *FRONTIERS IN ARTIFICIAL INTELLIGENCE AND APPLICATIONS*, 158:596.

[Law et al., 2015] Law, C. Y., Grundy, J., Cain, A., and Vasa, R. (2015). A preliminary study of open learner model representation formats to support formative assessment. In *Computer Software and Applications Conference (COMPSAC), 2015 IEEE 39th Annual*, volume 2, pages 887–892. IEEE.

[Law et al., 2017] Law, C. Y., Grundy, J. C., Cain, A., Vasa, R., and Cummaudo, A. (2017). User perceptions of using an open learner model visualisation tool for facilitating self-regulated learning. In *ACE*, pages 55–64.

[Lazarinis and Retalis, 2007] Lazarinis, F. and Retalis, S. (2007). Analyze me: Open learner model in an adaptive web testing system. *IJ Artificial Intelligence in Education*, 17(3):255–271.

[Linton and Schaefer, 2000] Linton, F. and Schaefer, H.-P. (2000). Recommender systems for learning: building user and expert models through long-term observation of application use. *User Modeling and User-Adapted Interaction*, 10(2-3):181–208.

[Liu and Stasko, 2010] Liu, Z. and Stasko, J. (2010). Mental models, visual reasoning and interaction in information visualization: A top-down perspective. *IEEE transactions on visualization and computer graphics*, 16(6):999–1008.

[Loboda and Brusilovsky, 2008] Loboda, T. D. and Brusilovsky, P. (2008). Adaptation in the context of explanatory visualization. In *European Conference on Technology Enhanced Learning*, pages 250–261. Springer.

[Loboda et al., 2014] Loboda, T. D., Guerra, J., Hosseini, R., and Brusilovsky, P. (2014). Mastery grids: An open source social educational progress visualization. In *Open Learning and Teaching in Educational Communities*, pages 235–248. Springer.

[Long and Aleven, 2013a] Long, Y. and Aleven, V. (2013a). Active learners: Redesigning an intelligent tutoring system to support self-regulated learning. In Hernandez-Leo, D., Ley, T., Klamma, R., and Harrer, A., editors, *8th European Conference on Technology Enhanced Learning (EC-TEL 2013)*, volume 8095 of *Lecture Notes in Computer Science*, page 490495.

[Long and Aleven, 2013b] Long, Y. and Aleven, V. (2013b). Supporting students self-regulated learning with an open learner model in a linear equation tutor. In *International Conference on Artificial Intelligence in Education*, pages 219–228. Springer.

[Long and Aleven, 2017] Long, Y. and Aleven, V. (2017). Enhancing learning outcomes through self-regulated learning support with an open learner model. *User Modeling and User-Adapted Interaction*, pages 1–34.

[Mabbott and Bull, 2004] Mabbott, A. and Bull, S. (2004). Alternative views on knowledge: Presentation of open learner models. In *International Conference on Intelligent Tutoring Systems*, pages 689–698. Springer.

[Mabbott and Bull, 2006] Mabbott, A. and Bull, S. (2006). Student preferences for editing, persuading, and negotiating the open learner model. In *Intelligent tutoring systems*, pages 481–490. Springer.

[Mabbott et al., 2007] Mabbott, A., Bull, S., et al. (2007). Comparing student-constructed open learner model presentations to the domain. *FRONTIERS IN ARTIFICIAL INTELLIGENCE AND APPLICATIONS*, 158:281.

[Maries and Kumar, 2008] Maries, A. and Kumar, A. (2008). The effect of student model on learning. In *Advanced Learning Technologies, 2008. ICALT'08. Eighth IEEE International Conference on*, pages 877–881. IEEE.

[Mazzola and Mazza, 2010] Mazzola, L. and Mazza, R. (2010). Gvis: a facility for adaptively mashing up and representing open learner models. In *Sustaining TEL: From Innovation to Learning and Practice*, pages 554–559. Springer.

[Mitrovic and Martin, 2002] Mitrovic, A. and Martin, B. (2002). Evaluating the effects of open student models on learning. In *Adaptive Hypermedia and Adaptive Web-Based Systems*, pages 296–305. Springer.

[Mitrovic and Martin, 2007] Mitrovic, A. and Martin, B. (2007). Evaluating the effect of open student models on self-assessment. *International Journal of Artificial Intelligence in Education*, 17(2):121–144.

[Moore et al., 2011] Moore, D. W., Bathgate, M. E., Chung, J., and Cannady, M. A. (2011). Technical report: Measuring activation and engagement. Technical report, Activation Lab, Learning Research and Development Center, University of Pittsburgh, Pittsburgh, PA (USA).

[Moore et al., 2013] Moore, D. W., Bathgate, M. E., Chung, J., and Cannady, M. A. (2013). Measuring and evaluating science learning activation. *Dimensions*, November/December.

[Niederle and Vesterlund, 2011] Niederle, M. and Vesterlund, L. (2011). Gender and competition. *Annu. Rev. Econ.*, 3(1):601–630.

[OḰeefe et al., 2013] OḰeefe, P. A., Ben-Eliyahu, A., and Linnenbrink-Garcia, L. (2013). Shaping achievement goal orientations in a mastery-structured environment and concomitant changes in related contingencies of self-worth. *Motivation and Emotion*, 37(1):50–64.

[Paas and Van Merriënboer, 1993] Paas, F. G. and Van Merriënboer, J. J. (1993). The efficiency of instructional conditions: An approach to combine mental effort and performance measures. *Human Factors: The Journal of the Human Factors and Ergonomics Society*, 35(4):737–743.

[Papanikolaou, 2015] Papanikolaou, K. A. (2015). Constructing interpretative views of learners interaction behavior in an open learner model. *IEEE Transactions on Learning Technologies*, 8(2):201–214.

[Papanikolaou et al., 2003] Papanikolaou, K. A., Grigoriadou, M., Kornilakis, H., and Magoulas, G. D. (2003). Personalising the interaction in a web-based educational hypermedia system: the case of inspire. *User Modeling and User Adapted Interaction*, 13(3):213–267. Questions: interaction between goals. Sequencing? Self-choice? Al-la chapter?

[Parsons and Haden, 2006] Parsons, D. and Haden, P. (2006). Parson's programming puzzles: A fun and effective learning tool for first programming courses. In *Proceedings of the 8th Australasian Conference on Computing Education - Volume 52*, ACE '06, pages 157–163, Darlinghurst, Australia, Australia. Australian Computer Society, Inc.

[Pérez-Marín et al., 2007] Pérez-Marín, D., Alfonseca, E., Rodríguez, P., and Pascual-Nieto, I. (2007). A study on the possibility of automatically estimating the confidence value of students knowledge in generated conceptual models. *Journal of Computers*, 2(5):17–26.

[Pirolli and Card, 1999] Pirolli, P. and Card, S. (1999). Information foraging. *Psychological review*, 106(4):643.

[Rueda et al., 2003] Rueda, U., Larrañaga, M., Ferrero, B., Arruarte, A., and Elorriaga, J. (2003). Study of graphical issues in a tool for dynamically visualising student models. *Learner Modelling for Reflection*, page 1.

[Schunk, 1990] Schunk, D. H. (1990). Goal setting and self-efficacy during self-regulated learning. *Educational psychologist*, 25(1):71–86.

[Shaffer, 2016] Shaffer, C. (2016). Opendsa: An interactive etextbook for computer science courses. In *Proceedings of the 47th ACM Technical Symposium on Computing Science Education*, pages 5–5. ACM.

[Shi and Cristea, 2016] Shi, L. and Cristea, A. I. (2016). Learners thrive using multifaceted open social learner modeling. *IEEE MultiMedia*, 23(1):36–47.

[Shi et al., 2014] Shi, L., Cristea, A. I., and Hadzidedic, S. (2014). Multifaceted open social learner modelling. In *Advances in Web-Based Learning–ICWL 2014*, pages 32–42. Springer.

[Shneiderman, 1996] Shneiderman, B. (1996). The eyes have it: A task by data type taxonomy for information visualizations. In *Visual Languages, 1996. Proceedings., IEEE Symposium on*, pages 336–343. IEEE.

[Siemens and Long, 2011] Siemens, G. and Long, P. (2011). Penetrating the fog: Analytics in learning and education. *EDUCAUSE review*, 46(5):30.

[Sirkiä, 2016] Sirkiä, T. (2016). Jsvee kelmu: Creating and tailoring program animations for computing education. In *2016 IEEE Working Conference on Software Visualization (VISSOFT)*, pages 36–45.

[Sirkiä and Sorva, 2015] Sirkiä, T. and Sorva, J. (2015). How do students use program visualizations within an interactive ebook? In *Proceedings of the Eleventh Annual International Conference on International Computing Education Research*, ICER '15, pages 179–188, New York, NY, USA. ACM.

[Sosnovsky and Brusilovsky, 2015] Sosnovsky, S. and Brusilovsky, P. (2015). Evaluation of topic-based adaptation and student modeling in quizguide. *User Modeling and User-Adapted Interaction*, 25(4):371–424.

[Upton and Kay, 2009] Upton, K. and Kay, J. (2009). Narcissus: group and individual models to support small group work. In *User Modeling, Adaptation, and Personalization*, pages 54–65. Springer.

[Vassileva, 2008] Vassileva, J. (2008). Toward social learning environments. *Learning Technologies, IEEE Transactions on*, 1(4):199–214.

[Vassileva and Sun, 2007] Vassileva, J. and Sun, L. (2007). An improved design and a case study of a social visualization encouraging participation in online communities. *Groupware: Design, Implementation, and Use*, pages 72–86.

[Verbert et al., 2014] Verbert, K., Govaerts, S., Duval, E., Santos, J. L., Van Assche, F., Parra, G., and Klerkx, J. (2014). Learning dashboards: an overview and future research opportunities. *Personal and Ubiquitous Computing*, 18(6):1499–1514.

[Weber and Brusilovsky, 2001] Weber, G. and Brusilovsky, P. (2001). Elm-art: An adaptive versatile system for web-based instruction. *International Journal of Artificial Intelligence in Education (IJAIED)*, 12:351–384.

[Wigfield and Eccles, 2000] Wigfield, A. and Eccles, J. S. (2000). Expectancy–value theory of achievement motivation. *Contemporary educational psychology*, 25(1):68–81.

[Wolters et al., 1996] Wolters, C. A., Shirley, L. Y., and Pintrich, P. R. (1996). The relation between goal orientation and students' motivational beliefs and self-regulated learning. *Learning and individual differences*, 8(3):211–238.

[Yudelson et al., 2007] Yudelson, M., Brusilovsky, P., and Zadorozhny, V. (2007). A user modeling server for contemporary adaptive hypermedia: An evaluation of push approach to evidence propagation. In Conati, C., McCoy, K., and Paliouras, G., editors, *11th International Conference on User Modeling, UM 2007*, volume 4511 of *Lecture Notes in Computer Science*, pages 27–36. Springer Verlag.

[Zimmerman, 1990a] Zimmerman, B. J. (1990a). Self-regulated learning and academic achievement: An overview. *Educational psychologist*, 25(1):3–17.

[Zimmerman, 1990b] Zimmerman, B. J. (1990b). Self-regulating academic learning and achievement: The emergence of a social cognitive perspective. *Educational psychology review*, 2(2):173–201.

[Zimmerman, 2000] Zimmerman, B. J. (2000). Self-efficacy: An essential motive to learn. *Contemporary educational psychology*, 25(1):82–91.