Assessment of Motivation in Online Learning Environments

Mihaela Cocea

Masters of Science

School of Informatics

National College of Ireland

Supervisor: Stephan Weibelzahl

Submitted to the Higher Education and Training awards Council

October 2007

Declaration

I, Mihaela Cocea, declare that this thesis is submitted by me in partial fulfillment of the requirement for the degree of Masters of Science by research, is entirely my own work except where otherwise accredited. It has not at any time either whole or in part, been submitted for any other educational award.

Mihaela Cocea

We have passed the knowledge, skill and competence of the candidate Mihaela Cocea by way of viva voce examination using the thesis, submitted in partial fulfillment of the requirements for the degree of Masters by Research, as evidence and find that it meets with the standard determined by the Higher Education and Training Awards Council. We recommend that the degree be awarded.

Examiners' signatures

Examiners' names, date

Abstract

Educational software strives to meet the learners' needs and preferences in order to make learning more efficient; the complexity is considerable and many aspects are taken into consideration. However, most systems do not consider the learner's motivation for tailoring teaching strategies and content, while its great impact on learning is generally acknowledged. A number of attempts have been undertaken to accommodate the learner's motivational states, mostly by means of design. Others started from motivational assessment, using log file analysis or self-assessment.

Building upon previous work, in this dissertation we propose a two-step approach for a complete motivational diagnosis, using both sources of information: log files and self-reports. Thus, the first step aims to identify disengaged learners unobtrusively while the second envisages a dialog with the disengaged learners that would include assessment of several motivational characteristics related to learning.

Three studies were conducted in order to investigate and validate disengagement detection. They demonstrated that disengagement can be predicted at a very good level (up to 97%) from attributes related to basic events like reading pages and taking tests. For self-assessment, a questionnaire was built partly from validated instruments and partly from created items. Two studies were conducted in order to investigate the validity and the reliability of this questionnaire. The results show that it is reliable and valid.

Keywords: motivation, e-Learning, data mining, adaptive systems, user modeling

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

Acknowledgements

I would like to thank my supervisor, Stephan Weibelzahl for the support and guidance throughout my degree. I have learned a lot from him from research point of view: I learned to justify decisions and choices and to be attentive to details; I also considerably improved my scientific writing and presentation skills. He gave me the opportunity to co-organize a summer school, which was a great experience, and also to participate to conferences and share my ideas and work with other people in the field. I would like to thank Prof. Steven Hite for his feedback and his encouragement that made me feel that I was on the right path. Both Stephan and Steve are great examples of how to be a mentor and I have learned from them not only scientific skills, but also human relations skills. I would also like to thank Scott Bateman and Christopher Brooks from ARIES Lab, University of Saskatchewan for facilitating the access to the iHelp system log files.

Chapter 1

Introduction

Motivation has always been important for learning. It is present in the learning process and has a great influence on its course. In the classroom context the students who are not motivated are detected and handled by human teachers. They have a variety of sources for inferring the motivational status of a learner, like visual cues or facial expressions, sources that are not readily available to an online learning system.

For a long time motivational considerations were pretty much ignored when designing e-Learning environments. Two factors that contributed to this exclusion are: 1) motivation, like all affective issues is hard to understand, structure and formalise, a prerequisite for a system implementation; 2) the presence of a theoretical separation between cognitive and affective processes and the image of a disparate functionality, where the influence of affectivity on cognition was ignored.

However, there have been several attempts to include motivation in e-Learning systems and currently, the influence of motivation on cognition is acknowledged and taken into consideration. The first attempts to include motivation in educational systems were through design; it was considered that a certain way of presenting the material to be learned would increase motivation and no personal characteristics of the learners were taken into account. Other approaches followed the line of personalization in order to adapt to personal features. Together with the development of the adaptive system, another factor that offered more possibilities for the inclusion of motivation in e-Learning was the usage of logged information. This source might be under certain circumstances even more accurate and rich than the one possessed by human teachers: all the actions of learners when interacting with the system. This information has been used in recent research for motivational assessment and intervention. These two factors created a shift in the way motivation was considered in e-Learning: the design was not sufficient anymore and assessment of motivation became necessary.

1.1. e-Learning and Personalisation

Personalisation aims to identify the learner's characteristics and adapt the content or the form of presentation based on these characteristics. Systems enhanced with personalisation typically have three components:

1) the *learner model* that stores the learner's characteristics;
2) the *domain model* that comprises the structure of the domain (the concepts and relations between them);
3) the *adaptation model* that contains a set of rules that combine the information from the learner model with the information from the domain module and decides whether and how the information in the user model is to be changed and what content and/ or how it should be presented to the learner.

Depending on the degree of control of the user over the adaptation process, two types of personalization can be distinguished:

i) In *adaptable systems* the user has the control over initialisation, selection and result of adaptation; these systems use the learner profile, but only the learner can change/ update it;
ii) *Adaptive systems* are able to automatically identify the learners and their characteristics and provide different content based on them.

The most frequent characteristic used for adaptation in e-Learning is knowledge. Other learner characteristics often used are: goals, interest, preferences, etc.

Knowledge is most frequently represented as an overlay of the domain model: for each concept in the domain model there is an estimation value of the learner's knowledge stored with that concept. The adaptation model takes into account the current state of the learner's knowledge when performing adaptation and also updates the estimations of the learner's level of knowledge.

The goal of the learner indicates what he/ she wants to achieve by using the systems. They might be interested in starting from the basics (no previous knowledge assumed), they might want to skip to more complex concepts or to focus on a particular sub-theme; they could also be interested in evaluating their current knowledge and finding the areas where they still need to learn. One specific characteristic of learning goals is that they may vary from one session to another.

Interest varies considerably between learners and even for the same learner while using the educational software. Preferences vary as well, though not as much as interest; they may refer to presentation style (font types, pictures, colours, navigation style) or to content (type of information or links). These characteristics are difficult to estimate by the system. In most systems the learner has to input this type of data manually.

Personalisation based on motivational characteristics would bring great benefit for learners, especially for the ones who experience difficulties in their learning. This raises the question of how to find out about these motivational characteristics and integrate them in the learning process.

1.2. Motivation in e-Learning

Several attempts to integrate motivation in the online learning process have been experimented. These attempts could be grouped in three categories depending on the focus: 1) design, 2) learner's actions or 3) learner's self-assessment. The influence of design on motivation is acknowledged (Keller, 2006) and most of the times taken into consideration when creating an online learning system. The learner's actions preserved in log files have been relatively recently discovered as a valuable source of information and several approaches to motivation detection and intervention have used log file analysis. Learner's self-assessment has been used for a long time in classroom context, but also in e-Learning, where it has been proved to be reliable and a valuable and accurate source of motivational information (Beal et. al, 2006).

An important advantage of log file analysis over self-assessment approaches is the unobtrusiveness of the assessment process – that would be similar to the classroom situation where a teacher observes that a learner is not motivated without interrupting his/ her activities. However motivational aspects are not always visible and thus undetectable by observation. This would suggest using self-assessment that would also ensure the accuracy of information.

We propose an approach that combines the two sources of motivational information in order to balance the advantage of unobtrusiveness with the importance of accurate information. Thus, the first step of our proposed approach – i.e. disengagement detection – would ensure that a learner would be interrupted only when he/ she appears to be disengaged. Once disengagement is detected, the system initiates a dialog and invites the learner to give information about his/ her motivation, after ensuring that the learner agrees with the diagnosis of the system.

1.3. Research goals

As mentioned previously, the aim of this dissertation is motivational diagnosis of learners in on-line learning environments. Accordingly, the main research goals are:

1. To explore the possibility of detecting disengagement unobtrusively, using the actions of learners registered in log files. The following list illustrates the range of questions associated with this goal: Is it possible at all to detect disengagement at a reasonable level of accuracy? Are there particular actions of the learners that would indicate engagement/ disengagement? Are these actions common to e-

Learning systems? This last question can be refined to: Is this approach applicable to other e-Learning systems? How to apply it: is it a matter of training new data or are there more complicated adjustments required?

To address this research goal, three studies were conducted, all presented in Chapter 4:

(a) a pilot study, where the possibility of predicting disengagement is investigated;
(b) the "core" study, where we refine the approach based on findings from the pilot study and where we identify the actions of learners that indicate engagement/ disengagement.
(c) a validation study, where we apply the refined approach resulted from the "core" study to data from another e-Learning system.

Three additional studies were conducted in order to refine the disengagement prediction approach:

(a) validation of reading speed attributes study, where two new attributes related to reading speed were proved to be valuable for prediction;
(b) patterns of disengagement study, where two patters are distinguished and their prediction is investigated;
(c) exclusion of exploratory sequences study, where we investigate the influence on prediction of the elimination of exploratory sequences.

2. To build an instrument for assessment of motivation that would measure motivational concepts related to the background theory: Social Cognitive Learning Theory (Bandura, 1986): Are there any existing instruments that would fit our purpose? If not, how to build this instrument? How to satisfy constraints related to length of instrument (we are dealing with de-motivated learners) and validity and reliability of measurement at the same time?

This research goal is addressed in Chapter 5, where we present:

(a) how the instrument was created;

(b) two studies conducted in order to investigate the validity and reliability of the instrument.

1.4. Structure of dissertation

The dissertation is structured as follows:

Chapter 1. The problem and the context are introduced in this chapter; the solution proposed and the research goals are also presented.

Chapter 2. The research background is presented, including both theory and related work. Several trends about motivation in e-Learning are identified and studies related to each of these are presented. Two motivational theories related to motivation for learning are described, with an emphasis on the one used in our research.

Chapter 3. In this chapter the research framework is explained. We propose a two-step approach to motivational diagnosis, based on previous approaches and the theoretical background presented in Chapter 2. The research questions are identified and the methodology used for addressing them is introduced.

Chapter 4. This chapter presents our work related to the first step of our research framework: detection of disengagement. Three studies were conducted for exploring, defining and validating the disengagement detection approach; three additional studies were conducted in order to refine the approach.

Chapter 5. In this chapter we present the work related to the second step of our research framework: the dialog with the learner that includes the assessment of motivation by means of self-reports. The steps to the final instrument used for this assessment are presented, followed by two studies conducted in order to verify its reliability and validity.

Chapter 6. This chapter binds together the results of the previous two chapters. The structure and functionality of an adaptive system enhanced with two components corresponding to the two steps of our research is presented. Several practical issues are discussed.

Chapter 7. This chapter summarizes the results of our research, presents future work directions and concludes the dissertation.

Chapter 2

Theoretical Background and Related Work

One of the most important themes in psychology of learning is motivation. As the trends of learning theories changed, new information about motivation was achieved. The evolution from behaviourism to cognitive psychology brought new motivational factors: if behaviourism reduced motivation to external stimulation, cognitive psychology placed the "motor" internally and also introduced a shift from seeing motivation as a state/ trait to a functional/ process view. The evolution from cognitive psychology to social constructivism brought equilibrium between the previous views, explaining how people learn in a social context (external stimuli), internalizing it differently, depending on the internal cognitive structure and thus constructing the learning material.

Most of the new approaches to motivation developed from these theories – especially from constructivist theories (i.e. self-efficacy, self-regulation, meta-cognition etc.) – were explored in the classroom context both to verify theories and to build interventions to increase motivation. There are some learning theories that were developed (more or less) especially for technology use, but without direct implications on motivation; for example cognitive flexibility theory and cognitive load theory are used in instructional design for complex domains in order to prevent learners' cognitive overload.

Being different from the classical learning in the classroom, e-learning may require different approaches to increase motivation, given the fact that interactions between human teachers and students or between students are either absent or very different from the ones in classroom.

As e-Learning is increasingly popular and as two of its main problems are drop-out and the quality of learning, motivation in online learning environments is a key to solving these problems. This chapter presents the background of motivational research in e-Learning and is structured in two main sections: 1) the theoretical background and 2) related research. The theoretical background presents briefly one of the theories used in related work and extensively the theory used in our research.

2.1. Motivation in Education

Two theoretical perspectives are presented here: 1) ARCS model (Keller, 1987) and 2) Social Cognitive Learning Theory (Bandura, 1986), with a larger focus on the second one, as it constitutes the theoretical foundation of our research.

2.1.1. ARCS Model

ARCS stands for Attention, Relevance, Confidence and Satisfaction. Gaining and retaining the learner's attention is necessary for efficient learning, relevance (of the learning content) is a condition for attention and motivation, confidence determines the level of effort invested in learning and satisfaction refers to the reward gained from the learning experience.

Keller's theory (1979) includes several concepts related to instruction and learning; it identifies the major variables of individual behaviour and instructional design related to effort and performance. The individual behaviour is influenced by personal factors and environment and the theory describes the effect of this influence on three categories of outputs: effort, performance and consequences.

Effort refers to the engagement of the person in actions aimed at accomplishing the learning task; it is a direct indicator of motivation. Effort is affected by three factors:

1) Values or motives that refer to the relation between the needs and beliefs of the individual, on one hand, and the choices of action, on the other hand;

2) Expectancy, which describes how the learner's expectancies on success or failure influence behaviour;
3) The design and management of instruction.

Performance refers to the accomplishment and is an indirect indicator of motivation. It is influenced by:

1) individual knowledge, abilities and skills;
2) learning design and management;
3) the effort invested by the individual into the task.

Consequences refer to outcomes, which can be intrinsic (e.g. emotional responses) or extrinsic (e.g. material objects). The consequences play an important role in motivation, feeding back into motives and values of the individual.

2.1.2. Social Cognitive Learning Theory

Social Cognitive Learning Theory (SCLT) sees the human behaviour as being determined by personal (cognitive/ affective/ motivational) factors and environmental conditions. This theory is based on a "triadic reciprocity" (Bandura 1986, p.18) between behaviour, cognitive factors and environmental situations. One of its assumptions is that personal determinants are not necessarily unconscious and that people can consciously change their conditions – they can influence their motivation and performance: the results of human behaviour can be changed by influencing one of the three elements of the triad. Thus, SCLT offers a framework for enhancing human learning.

2.1.2.1. Key concepts

Social Cognitive Learning Theory is an alternative for ARCS. The key concepts used are self-efficacy and self-regulation which are very much related to other concepts like goal orientation and attribution.

Self-efficacy

Self-efficacy is a self-judgment of one's capabilities to perform a task at a certain level of performance. Bandura defines it as "people's judgments of their capabilities to organize and execute courses of action required to attain designated types of performances. It is not concerned with the strategies one has but with judgments of what one can do with whatever strategies one possesses." (Bandura, 1986, p. 391).

Self-efficacy determines the amount of effort that the person is willing to invest and also the degree of perseverance in face of adversity (Bandura, 1977; 1986).

The determinants of self-efficacy, which are at the same time ways of changing a person's percept of efficacy, are: 1) mastery experience, 2) vicarious experience, 3) verbal persuasion and 4) physiological state (Bandura, 1977; 1986).

Self-regulation

Within the frame of Social Cognitive Learning Theory, self-regulation is seen as a socialization process taking place in the "triadic determinism" presented earlier. The theory was extended to include goals and expectations as motivational stimuli to self-control of behaviour, directed to change the person or the situation.

The role of self-efficacy on the self-regulation process has been explored in depth, i.e. Bandura (1982), Bandura and Schunk (1981), Bandura and Cervone (1983). Self-judgment on capability and perception of efficacy affects motivation and behaviour: even when skills are developed and the person is motivated to use them, it may not mean that the person will use them in any situation. For the self-regulation process the belief in one's capability is as powerful as the actual possession of skills.

Zimmerman (1989) analyzed self-regulation in the context of academic learning, describing self-regulated learners as metacognitively, motivationally and behaviourally active participants in their own learning processes. Self-regulated learning involves three aspects, according to Zimmerman (1990): self-regulation of learning strategies, self-efficacy perception of performance and commitment to academic goal.

There are three sub-processes of self-regulation: self-monitoring, self-judgment and self-reaction. Self-monitoring assesses quality, rate, quantity and originality of work or effort. The information acquired from self-monitoring indicates to the person whether to change some aspects of behaviour. Self-judgment refers to the assessment of the distance between the course of action and the settled goal. It is influenced by the type of standard (fixed or normative), the value of the goal, the properties of the goal (i.e. proximity, level of difficulty, specificity etc.) and attribution for performance which is described below.

Knowledge about one's performance is not sufficient for subsequent behavioural adjustments (Kazdin, 1974). It is the affective response – the self-reaction – to the self-judgment that motivates future actions. If the response is positive, self-efficacy will increase and will make similar (or more difficult) engagements more probable. If the response is negative, the change in the level of self-efficacy depends on its previous level: if self-efficacy was high, the person will probably consider that he/she didn't put enough effort which will motivate him/ her to persist and put more effort to improve performance; if self-efficacy was low, the self-judgment will probably confirm lack of ability and, thus, further effort will not be invested.

Goal orientation

Shrunk (1990) has developed a model relating goal-setting and self-efficacy in the academic context: self-efficacy for goal attainment is influenced by abilities, previous experience, attitudes towards learning, instruction and social context. Working on a task, students observe their performance in terms of goal approach and, depending on their observation, they continue or they change something in their approach of the task. If the progress is considered satisfactory, self-efficacy is likely to increase and attaining one goal usually leads to setting new challenging goals.

In our research we will start from Achievement Theory taxonomy of goal-orientations that includes two types: learning/mastery and performance goal orientation, the second one comprising two other subtypes – approach and avoidance.

Mastery goals are related to the intrinsic value of learning (Meece and Holt, 1993). Ames showed in his research work several aspects connected to mastery goals – students with such goals: 1) believe that perseverance and effort are likely to lead to success (Ames, 1992a); 2) they are concerned about improving their competencies, about developing new skills and understanding their work (Ames, 1992b); 3) they possess effective learning strategies that are related to self-regulation behaviour (Ames, 1992c; Tuckman and Sexton, 1991; Zimmerman and Martinez-Pons, 1990).

Performance goals are related to extrinsic value of learning: keeping a good image in front of others. Thus, *approach performance goals* are about doing better than others or better than the normative standard, while *avoidance performance goals* are about not appearing less able than others or not able to reach normative standards.

Social variables have also an influence on learning: there are *social goals* directed towards acceptance and status in the peer group, these ones being related to performance goal type; there are also *social goals* directed to learning, personal improvement and effort – related to mastery goal type (Anderman and Anderman, 1999). We will not look into these types of goals, although it would be interesting to consider them, especially for collaborative learning environments.

Beal and Lee (2005) added one more type of goal-orientation: ***"disengaged"***. Students from this category "do not really care about doing well in school or learning the material; their goal is simply to get through the activity." (p. 4). We include this category in our research, because the learners with this goal are the most likely to get disengaged when learning. Having knowledge about the learners with such goals may contribute to an earlier detection of disengagement during the activities and also provide information valuable for selecting appropriate interventional strategies; for example, as these learners are not interested in learning or performance, appropriate strategies for the other learning goal would not be applicable.

Attribution Theory

Attribution Theory offers a framework for understanding how learners explain their performance – success or failure. There are three dimensions included: locus of control (Rotter, 1966), control vs. non-control and stability over time. When combining the first two dimensions, the options as displayed in Table 2.1 are obtained.

Locus of control can be internal or external. People with internal locus of control attribute their results to themselves while people with external locus of control attribute their results to external factors. The control vs. non-control dimension refers to how controllable the situation is perceived to be. Stability over time is related to the more or less permanent character of the result.

Table 2.1 Dimensions of Attribution Theory

	Internal	**External**
Control	Effort	Task difficulty
Non-control	Ability	Luck

Stability over time has been associated with performance goals, while the view that situation can be changed is associated with learning goals (Driscoll, 2000).

Perceived characteristics of the task

These characteristics are important in the frame of Social Cognitive Theory as they are concerned with how the learner perceives the task rather than with objective characteristics. Usually, a learning task is built to be suited to many learners, trying to balance the characteristics in order to maximize suitability for as many learners as possible.

As we are interested in personalization of learning, these characteristics of the tasks as perceived by the learners are important because they could play an important role in the selection of interventional strategies to motivate the learner.

This important role is due to the connection between these perceived characteristics of the task and other variables considered. For example, they are very much related to attribution of performance: if the task is perceived as being difficult, depending on locus of control and control vs. non-control dimension, the attribution will be to task difficulty or (bad) luck. Also, there is a relation with self-efficacy – i.e. if a task is not challenging enough for a certain level of self-efficacy, the learner will not be motivated to start or complete the task. Again, depending on the degree of perceived controllability of the task, the process of self-regulation will be more or less "precise".

2.1.2.2. Social Cognitive Learning Theory in e-Learning research

The following paragraphs review the research related to the Social Cognitive Learning Theory in online environments and especially work related to the concept of self-efficacy as a key term in the theory and also in our research.

In 2001, King reported that self-efficacy research is not present in the context of asynchronous distance learning. In 2003, Miltiadou and Savenye reached the same conclusion about online environments.

Most of the studies regarding self-efficacy and computers are about use of technology. Several studies (e.g. Ertmer et al., 1994; Busch, 1995 etc.) showed that a higher perceived self-efficacy for using computers leads to a higher probability of using them. Ertmer et al. (1994) distinguished between positive and negative experience as determinant of self-efficacy for computer usage. Anxiety related to computer use was found as being a major obstacle in computer usage in education (e.g. Hakkinen, 1995; McInerney et al., 1994; Reed and Overbaugh, 1993).

Some studies related to self-efficacy were undertaken in relation to use of technology at undergraduate level (e.g. Karsten and Roth, 1998a, 1998b; Langford and Reeves, 1998). These and other studies concluded that high computer self-efficacy is correlated to increased performance in computer courses.

The number of studies that look into self-efficacy in online learning is limited. For example, Lim (2001) proved that self-efficacy is an accurate predictor of learner's satisfaction in the context of web-based courses. In 2002, Bandura mentioned that the information technology tools available to learners in online courses are efficient only if the learners have self-efficacy for regulating their learning behaviour. Wang and Newlin (2002) studied self-efficacy for online technologies in a web-based course and found that it was a good predictor of performance. Research conducted by Holcomb et al. (2004) concluded that there are no gender differences for self-efficacy beliefs regarding the use of technology.

However, there are some results that contradict the ones presented above. Joo et al. (2000) showed that self-efficacy for self-regulation did not directly predict performance, but influenced it indirectly. King (2001) did not find any differences in achievement between students with high or low self-regulation. Lee and Witta (2001) demonstrated that self-efficacy for online technologies was not a significant predictor of performance in class; however, measured at the end of the semester, it was a significant predictor of performance. Also, they found that self-efficacy for course content was not a significant predictor of performance in the online course. DeTure (2004) found as well that self-efficacy for online technologies was a poor predictor of success.

The above mentioned research studies contain issues that can "disqualify" their results. Thus, in their study Joo, Bang and Choi (2000) found that self-efficacy was not significant for performance prediction in a web-based test, but it was significant in the context of a written test. Lee and Witta (2001) used a small sample – 16 students and thus, the generalisability of results is questionable; also their result about self-efficacy for course content and performance is contradicted by results obtained by Wang and Newlin (2002): self-efficacy for course content correlated with the final exam performance. In his research DeTure (2004) had a sample of self-selected students; also, the statistical mode for online technologies self-efficacy for his sample was the highest score on the scale he used for measuring it, which leads to a shadow of doubt about the variety of data and thus, to the results of the regression prediction.

The presented studies and results lead to the conclusion that research on self-efficacy and self-regulation in online learning environments show "positive" results related to the two concepts from Social Cognitive Learning Theory. However, this research direction is at a very early stage and further research is needed in order to conclude about how the theory applies to online learning environments.

In a review article about motivational techniques in e-learning, Hodges (2004) comments: "It is clear that self-efficacy is at the heart of motivation. When designing learning experiences, one should take this into consideration and make every effort to increase the students' self-efficacy" (p. 6).

2.2. Assessment of motivation in e-Learning

Research on motivation in online learning could be classified in the three groups as introduced in the previous section: 1) design, 2) learner's actions, and 3) self-assessment. Here we will present some approaches for each group.

2.2.1. Design

Initially instructional design assumed that the user is interested in the information presented and was concerned with producing efficient and effective instruction. The efficiency dimension referred to the use of time and resources, while effectiveness dimension refers to the quality and the results at individual level. Motivation seemed to be assumed in the effectiveness dimension as it is hard to assume effectiveness when instruction is not appealing or when there is no interest in the material to be learned.

Motivational design (Keller, 2006) refers to instructional design enhanced with motivation. More specifically, it refers to "the process of arranging resources and procedures to bring about changes in motivation" (Keller 2006, p.3). Four models of motivational design have been suggested: person-centred, environmentally-centred, interaction-centred and omnibus.

Person-centred models are based on psychological theories about motivational dimensions of human personality. *Environmentally-centred* models assume that human behaviour can be explained by the influence of the environment on human volition. *Interaction-centred* models are grounded in psychological theories that take into account both personality characteristics and environments, between them being a reciprocal influence. The *omnibus* models combine instructional design with motivational design and they seem to be the complete solution to instructional goals.

Both Keller's ARCS, used in various studies on motivation for e-Learning, and the Social Cognitive Theory, the ground of our research, are theories that fall into the category of interaction-centred models.

2.2.2. Learners' actions

Three approaches are particularly relevant for the first step of our research: a) a rule-based approach, b) a focus of attention approach and c) a factorial analysis approach.

The rule-based approach (de Vicente and Pain, 2002) envisages inferring motivational states from two sources: the *interactions* of the students with the tutoring system and their *motivational traits*. Human tutors were asked to infer motivational states of learners' using a 10 question quiz. They had access to replays of the learners' interactions with the system and to their motivational traits. 85 rules were initially produced and they were reduced to 64 after eliminating the overlaps.

The focus of attention approach (Qu et al., 2005b) aimed to infer three aspects of motivation: *confidence*, *confusion* and *effort*, from the learner's focus of attention and several attributes related to the learners' actions: time to perform the task, time to read the paragraph related to the task, the time for the learner to decide how to perform the task, the time when the learner starts/ finishes the task, the number of tasks the learner has finished with respect to the current plan (progress), the number of unexpected tasks performed by the learner which are not included in the current learning plan and number of questions asking for help.

The factorial analysis approach (Zhang et al., 2003) focused on two motivational aspects: *attention* and *confidence*. Results showed that it is possible to group the user's actions by means of factorial analysis and thus, to distinguish between relevant actions that predict attention and relevant actions that predict confidence.

These three approaches are particularly relevant for our research because they are based on ARCS model which is close to Social Cognitive Learning Theory. However, there are several other approaches using learner's actions to investigate learner's behaviour in relation to some personal characteristics related to motivation. Five such approaches are presented in the following paragraphs.

Engagement tracing (Beck, 2004) is an approach based on Item Response Theory that proposes the estimation of the probability of a correct response given a specific response time for modelling disengagement; two methods of generating responses are assumed: blind guessing when the student is disengaged and an answer with a certain probability of being correct when the student is engaged. The model also takes into account individual differences in reading speed and level of knowledge.

A dynamic mixture model combining a hidden Markov model with Item Response Theory was proposed in (Johns and Woolf, 2006). The dynamic mixture model takes into account: student proficiency, motivation, evidence of motivation, and a student's response to a problem. The motivation variable can have three values: a) *motivated*, b) unmotivated and exhausting all the hints in order to reach the final one that gives the correct answer: *unmotivated-hint* and c) unmotivated and quickly guessing answers to find the correct answer: *unmotivated-guess*.

A Bayesian Network has been developed (Arroyo and Woolf, 2005) from log-data in order to infer variables related to learning and attitudes towards the tutor and the system. The log-data registered variables like problem-solving time, mistakes and help requests.

A latent response model (Baker et al., 2004) was proposed for identifying the students that game the system. Using a pretest–posttest approach, the gaming behaviour was classified in two categories: a) with no impact on learning and b) with decrease in learning gain. The variables used in the model were: student's actions and probabilistic information about the student's prior skills.

The same problem of gaming behaviour was addressed in (Walonoski and Heffernan, 2006a), an approach that combines classroom observations with logged actions in order to detect gaming behaviour manifested by guessing and checking or hint/ help abuse. Prevention strategies have been proposed (Walonoski and Heffernan, 2006b): two active interventions for the two types of gaming behaviour and a passive intervention. When a student was detected to manifest one of the two gaming behaviours, a message was displayed to the student encouraging him/her to try harder, ask the teacher for help or pursue other suitable actions. The passive intervention had no triggering mechanism and consisted in providing visual feedback on student's actions and progress that was continuously displayed on screen and available for viewing by the student and teacher.

All these approaches have the advantage of unobtrusively monitoring the learners' behaviour and identifying patterns associated with motivational issues. However they differ from our proposed approach in two aspects. First, the environments used include only test-type activities, while we are interested in learning-type activities as well. Second, the domain is math, which is rather technical and also a special domain which does not allow easy generalization for other areas; the domain considered in our research is HTML which is at the border between technical and non-technical subjects and thus, may allow an easier generalization across domains.

2.2.3. Self-assessment

Although self-assessment has been very much used in classroom learning, in this section we will present only three studies related to e-Learning, as they are of particular relevance for our research.

In a study investigating why students are disengaged rather than when they get disengaged, self reports on motivation have been used (Beal et al., 2006). The tutoring system used was an ITS for high school mathematics. The 10 item instrument for motivation included two questions for each of the following constructs: math self-efficacy, beliefs that math is important to learn, liking of math, expected success in math and difficulty of math. There was also an item about the students' belief about math ability – whether it is a native and fixed ability or a skill that can be enhanced by effort. Among the findings of this study, the most important is about the relation between motivation and the usage of the ITS divided in 4 types: independency, help abuse, guessing and learning. There were very few students with high math self-efficacy, who liked math and thought that math is important to learn; these students were most likely to solve the ITS math problems independently. Most of the students had low math motivation; those with low self-efficacy, who didn't like math and had low expectation of success were more likely to use the ITS in a way that suggested that they were putting effort to learn (reading the problems and using the hints).

In summary, low motivated students that do not perform well in classroom situation are more likely to benefit from ITS instruction because they have access and they are actually using the help provided; as opposed to classroom situation, this help is private.

In a study already mentioned in the previous section (Arroyo and Woolf, 2005) a survey was used to detect students' attitudes and motivation. These included the following aspects: i) the student perception of the tutor: if they learned how to use it, whether they liked using it, whether they found it helpful and whether they would use it again; ii) the interactions with the tutor - if they used audio for explanations; iii) attitudes towards help and learning: how seriously they tried to learn using the tutor, if they just wanted to get over with the task, if they felt challenged to see how many answers they could get right, if they did not care about the help available, if they tried to solve problems independently and use help only when really needed, if they used help to see other approaches even if they solved the problem, if they used help because they did not want to enter a wrong answer. Relations between these variables and the observed behaviour were analyzed and these relations were used to build a Bayesian network to infer attitudes and motivation.

Self-reports on motivation and mood have been used as valuable information for adapting the instruction (Beal and Lee, 2005). A pedagogical model has been proposed to consider cognitive skill as well as motivation for adapting instruction in the manner of human tutors. The self-reports included the following aspects: a) learning goal orientation, b) incremental-entity beliefs – when the ability for a subject (math in this case) is consider to be fixed or to be improvable by effort, c) mathematics motivation: self-efficacy, value of math and enjoyment of math, and d) daily mood reports. The pedagogical model adapted the instruction considering the information from these reports along with the cognitive skills information.

While the first mentioned study looked at why the learners are disengaged and the second used self-assessment for developing an automatic way of inferring motivation, the third study is closer to the purpose of self-assessment from our proposed approach: to use the information in order to adapt the instruction according to the motivational status of the learner.

2.3. Chapter summary

In this chapter we outlined two theories of motivation for learning: ARCS Model and Social Cognitive Learning Theory, the later being the theoretical background of the research presented in this dissertation. Thus the motivational characteristics that will be part of the motivational learner model are: self-efficacy, self-regulation, goal-orientation (mastery/ performance approach/ performance avoidance/ disengaged), perceived characteristics of the task and attribution (locus of control and stability/ instability).

The background of motivational research was presented, with examples of studies from three categories depending on the focus: design, learner's actions and self assessment. If the early attempts to include motivation in e-Learning took motivation into consideration only through design, currently assessment of motivation from the learner's actions and/ or self-report are used.

Chapter 3

Research questions and proposed methodology

In our research we aim to combine the advantage of unobtrusive assessment with the accuracy of self-assessment and thus designed our approach as a two-step process. Previous research (e.g. de Vicente and Pain, 2003; Qu et al., 2005a) targeted very specific motivational states, such as relevance, confidence, effort, satisfaction, etc. This target was possible due to the nature of the task: tests/ quizzes. However, the learning process includes also exploring and reading activities that we wanted to take into consideration along with the test-type activities. Given the more open and thus, less controllable nature of the reading activities, targeting very specific motivational characteristics would not be feasible. Thus, we were interested in a general indicator, i.e. engagement, that would tell us if the learner is focused on the learning or not.

3.1. Research outline

The goal of our research is motivation diagnosis seen as a two-step process to be followed by personalized intervention. A schematic representation of the research framework is presented in Figure 3.1.

For the first step – detection of disengagement, our research question is: *What actions in the learner behaviour can predict disengagement?* The method used to answer this question is analysis of log files. The learner's actions are used as input while the level of disengagement is predicted as output. In order to validate the findings comparison studies are conducted.

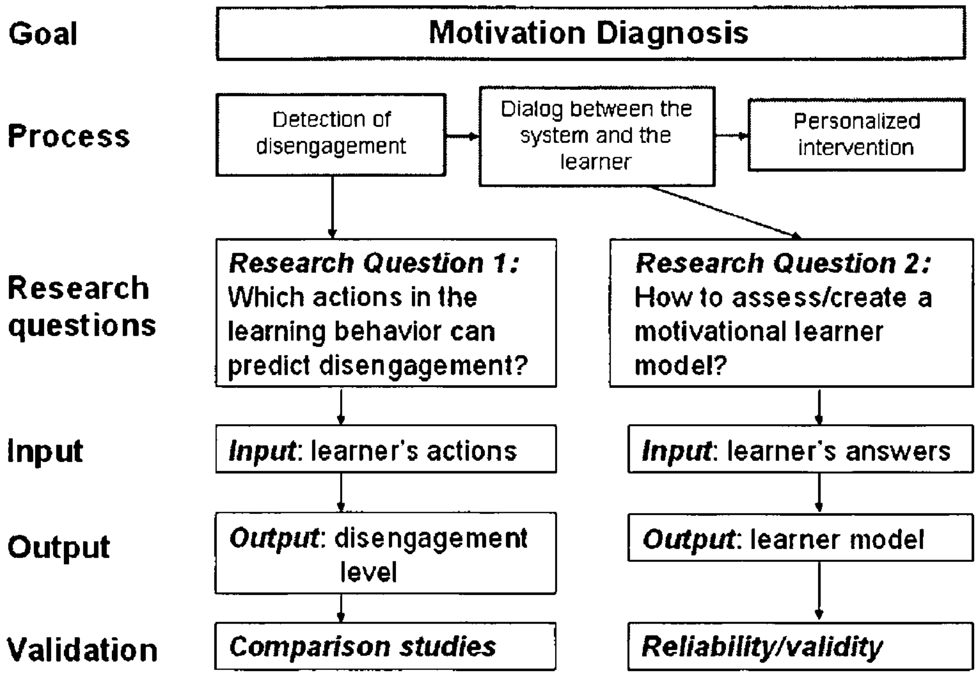


Figure 3.1 Research project outline.

For the second step – the dialog with the learner, the research question is: *How to assess/ create a motivational learner model?* The input for the learner model would be the learner's answers to the questions included in the dialog and the output would be the motivational profile of the learner. In order to validate the assessment through the dialog, reliability and validity issues are addressed.

3.2. Disengagement detection

Disengagement is a key concept for the first step of the proposed research. In this section we review the concept of engagement as used in research in general and in our research in particular, and we also describe in detail the methodology intended for answering the research question of the first step of our approach.

3.2.1. Engagement

Although there is no specific definition for engagement as a psychological concept, there are two theories that refer to it. One is the flow theory (Csikszentmihalyi, 1997) and the other one is theory of engagement (Shneiderman et al., 1995).

Flow theory (Csikszentmihalyi, 1997) describes the state of flow which appears when several characteristics are met. Among these characteristics are: 1) clear goals; 2) concentrating and focusing; 3) balance between ability level and challenge; 4) a sense of personal control, etc. Concentrating and focusing refers to engagement in the same meaning as used in our research.

The theory of engagement (Shneiderman et al., 1995) emerged around the mid nineties in the context of teaching in electronic and distance education environments. This theory stresses the importance of being engaged in learning activities and the authors mention two ways of increasing engagement: collaboration and interaction with other learners and meaningful tasks. Our approach does not consider engagement as it is seen by this theory, but in its "simple" meaning of being focused on learning.

In order to see engagement in the context of motivation and other related concepts associated with motivation, we describe the relation between engagement and some of these concepts: 1) engagement can be influenced by *interest*, as people tend to be more engaged in activities they are interested in; thus, interest is a determinant of engagement; 2) *effort* is closely related to interest in the same way: more effort is invested if the person has interest in the activity; the relation between engagement and effort can be resumed by: engagement can be present with or without effort; if the activity is pleasant (and/or easy), engagement is possible without effort; in the case of more unpleasant (and/or difficult) activities, effort might be required to stay engaged; 3) the difference between engagement and *focus of attention*, as it is used in research is that focus of attention refers to attention through a specific sensorial channel (e.g. visual focus), while engagement refers to the entire mental activity (involving in the same time perception, attention, reasoning, volition and emotions); 4) in relation to *motivation*, engagement is just one aspect indicating that, for a reason or another, the person is motivated to do the activity he/she is engaged in, or the other way, if the person is disengaged, he/she may not be motivated to do the activity; in other words, engagement is an indicator of motivation.

3.2.2. Methodology

To detect engagement, and even more important disengagement, log file analysis is used. The actions registered in log files are inspected and attributes related to them are established for the analysis. Several data mining methods applicable to the database structure and types of variables are employed.

The approach for this first research question builds upon Qu et al. (2005a) approach. Rather than directly inferring particular motivational states from the observed behaviour, we propose to use behavioural cues as indicators for disengagement detection. These indicators may relate to the concept of self-regulation and may include: browsing fast rather than reading, skipping sections, non-systematic progression, and answering questions quickly (in less time than the minimum required time for at least reading the questions), etc. Other indicators could be how often and how insistently the learner seeks help from peers/ instructor; also if the learner is searching external content for a related topic it may be a sign of getting lost in the course content; it may also be a sign of an elaboration cognitive strategy (Pintrich and Schunk, 2002).

Perhaps the most intuitive and easy to use indicator is time. It is interesting that de Vicente and Pain (2003) used it to infer confidence or lack of interest, while Qu et al. (2005a) used it to infer effort. Time is probably a component of each of the three mentioned aspects, but is not sufficient to infer any of them. We also consider time as a general indicator of disengagement: a too short or a too long focus on an issue may indicate "problems". Of course, both could be due to other factors: a too short time spent on a task might be explained by a good knowledge and exceeding time could be justified by factors like breaks or deep thought. These situations can be clarified by asking the learner.

A pilot study is conducted on a limited number of log files in order to determine if disengagement prediction is possible at a satisfactory level based on the established attributes. If successful, further investigations will be conducted in order to answer the research question: what are the relevant attributes?

To validate the results, a similar analysis based on the same types of attributes found relevant will be performed on data from another e-Learning system.

3.3. Assessment of Motivation

The second research question will be approached based on a dialog with the learner. The dialog deals with the following aspects: a) Inform and explain to the learner about the dialog: the learners identified to be disengaged will be informed by the system that it has detected disengagement, ask the learner if he/ she agrees with this "diagnosis" and, in case of agreement, will inform the learner about the following questions that he/she will be asked in order to provide personalised intervention; b) Ask the learners about their self-efficacy, self-regulation, goal-orientation, attribution of their performance and perceived characteristics of the task performed.

To elicit the level of self-efficacy, self-regulation and goal-orientation adapted versions of existing questionnaires are used and items are created for the assessment of attribution and perceived characteristics of the task. To elicit the attribution of their performance, learners will have to rate the contribution towards their learning outcome of each of the following: (lack of) ability; (lack of) effort; (bad) luck; task (reasonable/ hard) difficulty. From the attribution choices we will infer the locus of control and the stability/ instability dimension. Initially we considered several perceived characteristics of the task to be included in the assessment: difficulty, cognitive interest, sensory interest (structure/ presentation), controllability and challenge. Later on, considering the length of the instrument and the fact that the learner completing it would already be disengaged, we decided to measure only perceived task difficulty. This information together with the other measured aspects would be included in a learner model.

An experiment is conducted to investigate the reliability and construct validity of the adapted self-efficacy, self-regulation and goal-orientation scales. The construct validity of the attribution measurement is assured by the fact that the options given for answering are from the theory of attribution (Heider, 1958; Weiner, 1974). The participants will be required to complete two instruments: one including the validated scales and one including the adapted ones; the data analysis covers: 1) differences between the two instruments; 2) reliability coefficients values, especially for the created items (attribution and perceived task difficulty).

3.4. Chapter summary

In this chapter we have presented our research framework that includes two steps: the first one aims at an unobtrusive diagnosis of disengagement and the second one aims at a learner model of motivation through self-assessment.

The research question and the proposed methodology were presented for each of the two steps. For the first step in the proposed approach, i.e. disengagement detection, the goal is to identify the actions of learners that indicated disengagement; in order to reach this goal log file analysis is used. For the second step, i.e. assessment of motivational characteristics by means of dialog with the learner, the purpose is to find a way to obtain the motivational model of a learner; the methodology used in order to achieve this purpose is self-assessment.

Chapter 4

Disengagement detection

In our research framework, the first step is unobtrusive disengagement detection by means of log file analysis. Three studies were conducted for defining and validating disengagement detection: a pilot study, a "core" study and a validation study, each presented in the following subsections. Three additional studies were conducted in order to refine the disengagement prediction: validation of reading speed attributes, patterns of disengagement, and exclusion of exploratory sequences.

4.1. Pilot study

This small-scale study was conducted in order to investigate the possibility of predicting disengagement at a satisfactory level, based on attributes related to events registered in log files.

The log files used in our pilot study were from an online learning environment called HTML-Tutor, which is a web interactive learning environment based on NetCoach (Weber et al., 2001). HTML-Tutor offers an introduction to HTML and publishing on the Web; it is online and can be accessed freely. Beyond the log-files, there is no information about the users available; accordingly, they could be of any age and using the system for different purposes.

A list of possible events that are recorded by HTML-Tutor is presented in Table 4.1 together with the derived attributes used in the analysis.

Table 4.1 Logged events and the derived attributes used in the analysis

Events	Parameters/ Attributes
Goal	The selected goal (from a list of 12 goals)
Preferences	Number; Time spent selecting them
Reading pages	Number of pages; average time reading pages
Pre-tests	Number of pre-tests; average time; number of correct answers; number of incorrect answers
Tests	Number of tests; average time; number of correct answers, number of incorrect answers
Hyperlink, Manual, Help, Glossary, Communication, Search, Remarks, Statistics, Feedback	For each of these: Number of times accessed; average time

From the basic log data presented in Table 4.1, four indicators or attributes with higher level of information were calculated: performance on tests, the time spent reading, the number of accessed pages and the time spent solving tests. A description of these attributes and the way they were calculated is presented in Table 4.2. These derived indicators are used in the analysis.

Table 4.2 Derived attributes to be used in the analysis

Attribute	Description
Performance	Percentage of correctly answered tests (calculated as number of correct tests divided by total number of performed tests)
TimeReading	Time spent on pages (calculated as the sum of the time spent on each page accessed) in a session
NoPages	The number of accessed pages
TimeTests	The time spent performing tests (calculated as the sum of time spent on each test)

The information was aggregated in order to create a database with the same indicators for every user and to give meaning to the raw data. Basically only two events with their average times are considered (reading and taking tests) because none of the other events seem to have occurred, and thus, were not registered in the log files considered for analysis.

Waikato Environment for Knowledge Analysis (WEKA) (Witten at al., 1999) was used for the analysis. The chosen method was decision trees based on a variation of C4.5 algorithm (Quinlan, 1993), J48. Other methods could have been used, such as naïve Bayes classifier or regression. We chose decision trees and J48 algorithm because it provides classification and prediction, and also intelligible output through a graphical representation. Thus, the users' activities can be characterized in terms of the attributes generated from the log files data (classification) and the predictability can be examined in order to see if such log file data can be used for motivation prediction.

In order to use the data for decision tree learning, each user has been assigned a motivational "status": engaged or disengaged. The criteria used for this assignment is described in detail in the next section, where the "core" study is presented. The distribution of the 20 learners comprised 10 engaged and 10 disengaged people.

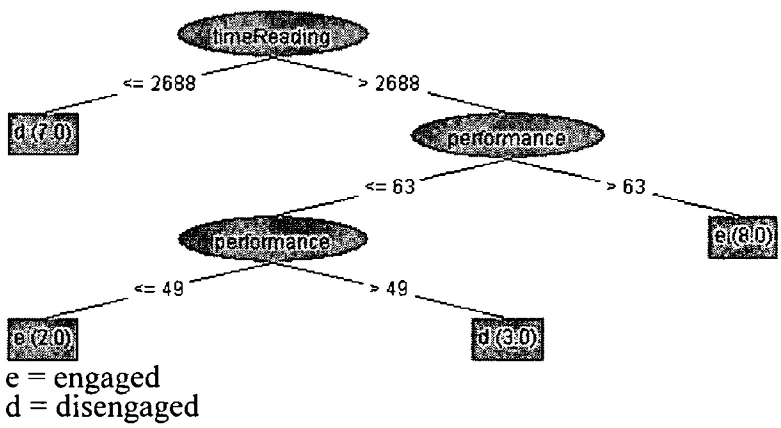


Figure 4.1 Pilot study: decision tree for engagement level

The generated decision tree is shown in Figure 4.1. The most important attribute for predicting motivation is, according to this decision tree, the time spent reading (timeReading): users that spend less then 2688 seconds (approximately 45 minutes) are classified as disengaged; if the time spent reading exceeds 2688 seconds, performance is the second attribute to be used in classifying learners. Thus, if performance ratio is above 63%, users are classified as engaged. Otherwise, the same attribute, performance is used to classify learners as engaged if the ratio does not exceed 49% or as disengaged, otherwise.

Summarizing the information from the decision tree, four categories of learners have been identified:

1) learners who spend less then approximately 45 minutes reading; they are classified as disengaged;
2) learners who spend more then 45 minutes reading and with a performance that exceeds 63%; these learners are classified as engaged;
3) learners who spend more then 45 minutes reading and with a performance between 49% and 63%; they are classified as disengaged;
4) learners who spend more then 45 minutes reading and with a performance below 49%; they are classified as engaged.

The confusion matrix is presented in Table 4.3. It shows the quality of the decision tree using fourfold cross-validation.

Table 4.3 The confusion matrix with fourfold cross-validation

		Predicted		
		Engaged	**Disengaged**	**Total**
Actual	**Engaged**	8	2	10
	Disengaged	3	7	10
	Total	11	9	20

The elements in the matrix show the number of test examples for which the actual class is the row and the predicted class is the column. The diagonals of the confusion matrix indicate 75% of correctly classified examples and 25% examples classified incorrectly. Thus, we can state that the quality of the decision tree is quite good. Looking at the disengaged learners as they are our main interest, we see a lower rate of correct classification: 70% of the disengaged students are correctly classified.

Although this study has been conducted with a limited number of subjects, some interesting remarks can be made. The decision tree finds a particularly refined category of disengaged students: learners who spend a considerable time reading (above 45 minutes) and with a performance between 49% and 63%. Trying to give some meaning to these figures, a possible interpretation is the following: the fact that these learners have an average performance gives them a medium level of confidence; they go on reading, as they know they could improve their knowledge and performance, but knowing that they already have a medium or good knowledge level makes them invest less effort in learning. Also, the results outline two categories of engaged learners that spend considerable time reading (over 45 minutes): 1) the learners with a performance lower than 49%; 2) the learners with a performance greater than 63%. The engagement in both cases could be explained by the learners' desire to acquire more knowledge or just a better performance. From this perspective, it would be interesting to investigate the type of goal orientation of the learner (mastery or performance).

Perhaps the most valuable information about this study, besides the fact that disengagement can be predicted at a reasonably good level, is that the results cannot tell anything about the users' level of motivation within the first 45 minutes. According to the decision tree, a user could be qualified as engaged or disengaged only after 45 minutes and, by that time, a demotivated user would probably have already logged out. Thus, it is of no benefit to know this information if there is no possibility to intervene. So, in order to be able to intervene on time, it is required to have information about the level of motivation in less time. This is also supported by the known fact that motivation can fluctuate at short periods of time.

Thus, the lesson learned from this pilot study is to analyze the user's activity for shorter time periods, e.g. 10-15 minutes, and to extract the level of motivation for those specific periods. By this approach information about the level of motivation could be updated at every 10-15 minutes and thus, have the possibility to intervene before the user would log out.

4.2. "Core" study

The core study was conducted with log files from the same e-Learning system and taking into consideration the findings from the pilot study. Accordingly, we split the sessions into sequences of 10 minutes. In this study we used the log files of 48 subjects who spent between 1 and 7 sessions on HTML-Tutor, each session varying between 1 and 92 sequences. The database included 1015 entries (i.e. sequences), of which 943 were of exactly 10 minutes and 72 varied between 7 and 592 seconds.

Table 4.4 Frequency of events registered in log files

Events/attributes	Frequency of appearances (in 1015 sequences)
Goal	59
Preferences	7
Reading pages	850
Pre-tests	14
Tests	458
Hyperlinks	245
Manual	7
Help	11
Glossary	76
Communication	6
Search	27
Remarks	6
Statistics	8
Feedback	4

The HTML-Tutor events/attributes were already presented when describing the pilot study. In relation to these events we present also their frequency of occurrence in Table 4.4, as it represents one of the criteria for selecting the attributes that are relevant for disengagement prediction. Thus, if an event or attribute is found relevant for prediction, but its frequency is low, this may indicate that this particular event or attribute is not very valuable because of its low occurrence; considering this attribute in the prediction algorithm may introduce an extra cost in terms of sparsity of data and of computational complexity that might not be worth it. Thus, we can notice from Table 4.4 that three events comprise almost 90% of the activity of learners: reading pages, taking tests and hyperlinks. When discussing the attributes found relevant for prediction we will refer to the frequency of the event they are related to.

4.2.1. Expert ratings on Level of Engagement

Each sequence of 10 minutes was assigned a value or code: engaged (e), neutral (n) or disengaged (d). The assignment was done by experts who had access only to the unprocessed log files (split into sequences of 10 minutes) containing all events. In the pilot study (where we analyzed sessions instead of sequences) we had only 2 categories: engaged and disengaged. Because we introduced the 10 minutes sequences, in some cases it was hard to decide whether overall the learner was engaged or disengaged. Thus, we introduced a third category: neutral. A detailed presentation of the criteria used for this rating is presented in Table 4.5, which contains the instructions given to a second coder in order to verify the reliability of the ratings.

Table 4.5 Instructions for level of engagement rating.

Timeframes for HTML Tutor
- Necessary time for reading a page: varies from 30 sec. to a maximum of 4-5 minutes.
- Necessary time for a test: varies from just a few seconds to a maximum of 3-4 minutes.

Engaged (e)	**Disengaged (d)**	**Neutral (n)**
Spending reasonable time on pages and tests given the characteristics of HTML Tutor Examples of patterns: - people focused reading – spend most of the time reading and less on other tasks - people focused on taking tests - spend most of the time taking tests and less on other tasks - people that read and take tests - spend most of the time reading and taking tests	Spending too much time on pages/tests Moving fast through pages/tests Automatic logouts Examples of patterns: - spend more than reasonable time on just one or a few tasks - move fast though the same / different tasks	Hard to decide if overall (for the 10 minutes) the person is engaged or disengaged E.g.: for approximately half of the time the person seems engaged and for the other half seems disengaged E.g.: can't decide if overall the person is moving too fast through pages or the amount of time spent on pages is reasonable

The investigation conducted in order to verify the coding reliability included two steps: 1) *Informal assessment*, conducted using only 10 sequences; the ratings based on the given instructions were discussed to prevent different results due to instruction vagueness or suggestibility; the percent agreement was 80% (only 2 different ratings from 10); the kappa measurement of agreement was .60 (p=.038) and the Krippendorff's alpha (Hayes and Krippendorff, 2006) was .60 as well; 2) *Second expert rating.* A second rater coded 100 sequences randomly sampled from the 1015 entries in the data set; the instructions used for the informal assessment were expanded with typical situations or patterns for each case. Table 4.5 includes the instructions given to the second rater.

The second expert rating resulted in a rater agreement of 92% (only eight different ratings from 100; in further discussion between the raters the eight disagreements were resolved) with a kappa measurement of agreement of .826 (p<.01) and Krippendorff's alpha of .8449. Although the percent agreement is high, we can see that kappa and Krippendorff's alpha have lower values. The percent agreement is not always the best indicator for agreement as it tends to be too liberal, while Cohen's Kappa and Krippendorff's alpha are known to be more conservative (Lombard et al., 2003). For the last two coefficients, values above .80 indicate high inter-coder reliability. Thus, the instructions presented in Table 4.5 enabled us to establish a learner's lever of engagement within a ten minutes sequence in an objective and reliable manner.

4.2.2. Analysis

In order to perform the analysis, Waikato Environment for Knowledge Analysis (WEKA) (Witten et al., 1999) was used again. Several methods were compared to find which one is best for our purpose and to see if results are consistent over different methods. We present here trials used only on a reduced data set of 943 entries obtained from the 1015 entries data set by eliminating the entries with time per sequence shorter than 10 minutes. In order to explore the effect of the number of attributes included, we created three different data sets: 1) DS-30 that includes all attributes; 2) DS-10 which includes ten attributes related to the following events: reading pages, tests, hyperlinks and glossary and 3) DS-6 which includes six attributes related only to reading pages and tests.

From the 943 entries 679 (72%) were used for training and 264 (28%) for testing. In previous research (Cocea and Weibelzahl, 2007a) we considered the entries independent and did not control if entries from the same students were both in training and testing, which may have introduced a "positive" bias to the results. The analysis presented here includes this control. We tried to be as close as possible to the classical balance between training and testing: 66% and 33%. Because the number of entries per student varied considerably, we applied the following procedure: we have calculated the number of entries per each student and where several students had the same number of entries two thirds were selected for training and one third for testing; in the case of close, but not exactly the same values for the number of entries the same principle was applied.

The analysis included eight methods (Witten and Frank, 2005): (a) Bayesian Nets with K2 algorithm and maximum 3 parent nodes (BN); (b) Logistic regression (LR); (c) Simple logistic classification (SL); (d) Instance based classification with IBk algorithm (IBk); (e) Attribute Selected Classification using J48 classifier and Best First search (ASC); (f) Bagging using REP (reduced-error pruning) tree classifier (B); (g) Classification via Regression (CvR) and (h) Decision Trees with J48 classifier based on Quilan's C4.5 algorithm (Quinlan, 2003) (DT). These methods represent the most commonly used techniques for the given kind of data: nominal data for the predicted variable and numeric data for the predictors.

4.2.3. Results

The results are displayed in Table 4.6, including: percentage of correctly classified instances, the true positive (TP) rate, false positive (FP) rate, precision and recall for disengaged class, and the mean absolute error; d prime values were also included.

The results indicate very good level of prediction on all datasets, ranging from 85% to 93%. The best performance is obtained on DS-30 with classification via regression.

For the disengaged class, the TP rate varies from 0.91 to 0.96 and the FP rate varies from 0.12 to 0.29. The very similar results obtained from different methods and trials shows consistency of prediction and of the attributes used for prediction.

The high TP rate and relatively low FP rate indicate a very good level of prediction and a good discrimination; the d-prime values are between 2.36 and 2.88 for DS-30, between 2.20 and 2.68 for DS-10 and between 2.11 and 2.55 for DS-6. D-prime values above 2 show that engagement levels can be accurately distinguished and identified.

Table 4.6 "Core" Study: experiment results summary

		BN	LR	SL	IBk	ASC	B	CvR	DT
DS-30	**%correct**	90.15	89.77	91.29	89.02	88.64	90.91	92.80	88.26
	TP rate	0.92	0.91	0.93	0.91	0.95	0.94	0.96	0.94
	FP rate	0.12	0.13	0.12	0.12	0.22	0.15	0.13	0.21
	Precision	0.94	0.93	0.94	0.94	0.89	0.93	0.93	0.90
	Recall	0.92	0.91	0.93	0.91	0.95	0.94	0.96	0.94
	Error	0.10	0.11	0.11	0.08	0.10	0.12	0.11	0.10
	d'	2.58	2.46	2.65	2.51	2.41	2.59	2.88	2.36
DS-10	**%correct**	90.91	89.39	89.02	89.02	86.36	89.02	91.67	88.26
	TP rate	0.94	0.93	0.93	0.93	0.94	0.94	0.95	0.94
	FP rate	0.13	0.17	0.19	0.15	0.26	0.20	0.16	0.22
	Precision	0.93	0.92	0.91	0.93	0.88	0.90	0.92	0.89
	Recall	0.94	0.93	0.93	0.93	0.94	0.94	0.95	0.94
	Error	0.10	0.12	0.12	0.08	0.13	0.12	0.12	0.12
	d'	2.68	2.43	2.35	2.51	2.20	2.40	2.64	2.33
DS-6	**%correct**	89.77	86.36	87.12	84.85	90.91	89.77	90.91	90.15
	TP rate	0.94	0.94	0.95	0.93	0.94	0.93	0.94	0.93
	FP rate	0.18	0.29	0.28	0.24	0.16	0.17	0.16	0.16
	Precision	0.91	0.86	0.87	0.89	0.92	0.92	0.92	0.92
	Recall	0.94	0.94	0.95	0.93	0.94	0.93	0.94	0.93
	Error	0.12	0.14	0.14	0.14	0.11	0.13	0.12	0.11
	d'	2.47	2.11	2.23	2.18	2.55	2.43	2.55	2.47

The highest percentage of correctly predicted instances was obtained using Classification via Regression (CvR) on all data sets, with a maximum for DS-30: 92.80%. The percentage for DS-6 is only slightly lower, 90.91%, a decrease of less than 2% resulting from eliminating 24 attributes. Considering the Minimum Description Length (MDL) principle, the frequency of events, the sparsity of data and the computational complexity, we argue for the use of the six attributes in the prediction model. The confusion matrix for DS-6 with classification via regression is displayed in Table 4.7.

Table 4.7 The confusion matrix for dataset DS-6 using CvR

		Predicted			
		Disengaged	**Neutral**	**Engaged**	**Total**
Actual	**Disengaged**	165	0	10	175
	Engaged	12	0	75	87
	Neutral	1	0	1	2
	Total	178	0	86	264

We notice that none of the neutral instances were correctly classified: one has been classified as disengaged and one as engaged. The same situation occurred in the previous study (Cocea and Weibelzahl, 2007a) where we used 10 fold cross validation and did not control the students' distribution on training and testing. Two possible explanations would be: 1) the small number of neutral instances and 2) the fact that a distinction that seems very hard for human raters, which was the reason for the introduction of the neutral category, could be easier for the computer. Given these results, we argue that only two categories should be used for the level of engagement: engaged and disengaged.

The Bayesian Network from DS-6, displayed in Figure 4.2, has an interesting structure: Number of False (TNoF) and Correct (TNoCor) Answers to Tests feed into the Number of Tests (Tests). Which itself, together with Average Time on Tests (AvgTimeT) feeds into Average Time spend on Pages (AvrTimeP). All of them also feed directly into the Level of Engagement (Eng/Diseng), i.e., the Bayesian Network structured the attributes in a semantically meaningful way.

In order to see which attributes are more important for prediction, we used three different single attribute evaluation methods with ranking (Witten and Frank, 2005, pp. 424-425) as search method for attribute selection: (a) Chi Squared Attribute evaluation (ibid., p.302, p.324): computes the chi-square statistic of each attribute with respect to the class; (b) Information Gain Attribute Evaluation (ibid., p.99, p.423): evaluates the attributes based on information gain; (c) OneR Attribute Evaluation (ibid., pp. 84-85, p.423): used OneR methodology to evaluate attributes; OneR stands for one-rule and it generates a one-level decision tree expressed in the form of a set of rules that all test one particular attribute.

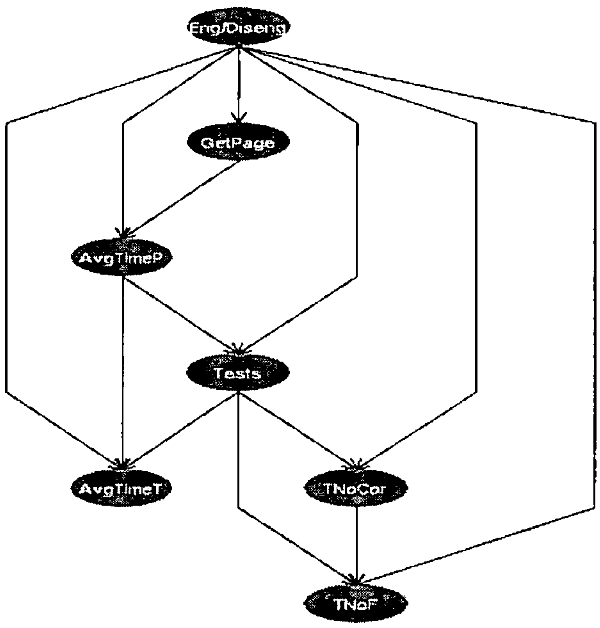


Figure 4.2 Bayesian Network from dataset DS-6

We present the ranking only for the first eight attributes out of 30, as they correspond to the most frequent events. The first two methods delivered the same ranking: Average time/ Pages, Number of pages, Tests, Average time/ Tests, Number of correctly answered testes, Number of incorrectly answered testes, Average time/ Hyperlinks and Number of hyperlinks. OneR resulted in the same ranking for the first four attributes, followed by: Number of incorrectly answered testes, Average time/Hyperlinks, Number of correctly answered testes and Number of hyperlinks.

In order to see how the prediction is influenced by the attribute selection, we used three trials and two experimental conditions. Trial 1 included all actions, Trial 2 comprised only the following actions: reading pages, taking test and following hyperlinks (top three actions found using frequency counting that covered 90% of the total number of actions) and Trial

3 included just two actions: reading pages and taking tests. The two experimental conditions are: 1) with attribute selection prior to prediction and 2) without attribute selection.

The results from the three trials in the two experimental conditions are compared in terms of: (a) percentage correct for overall prediction, meaning for all levels of engagement and (b) true positives and false positives rate for disengagement.

The same tool, WEKA was used for the analysis. Several methods were experimented and similar results were found. We present here only two of them: one that had the best results for overall prediction for all three levels of engagement, classification via regression (CVR), and one that had the best results for the disengagement prediction, Bayesian Networks (BN). The results for these two methods are presented in Table 4.8. The high TP rate and relatively low FP rate indicate a very good level of prediction and a good discrimination (the d-prime values are between 2.11 and 2.32).

Table 4.8 Predictions of engagement level with and without attribute selection, using Classification via Regression and Bayesian Networks

			Trial 1		Trial 2		Trial 3	
			CVR	**BN**	**CVR**	**BN**	**CVR**	**BN**
No attribute selection		**%correct**	87.64	87.07	88.10	87.00	87.21	86.68
		TP rate	0.92	0.93	0.92	0.93	0.91	0.93
		FP rate	0.20	0.23	0.18	0.22	0.18	0.24
Attribute selection	**Chi-square**	**%correct**	87.75	87.79	88.10	87.47	87.25	86.70
		TP rate	0.93	0.94	0.92	0.93	0.91	0.93
		FP rate	0.21	0.24	0.18	0.22	0.18	0.24
	Info gain	**%correct**	87.70	87.80	88.10	87.44	87.25	86.67
		TP rate	0.93	0.94	0.92	0.93	0.91	0.92
		FP rate	0.21	0.25	0.18	0.22	0.18	0.24
	OneR	**%correct**	87.69	87.55	88.03	87.36	87.20	86.70
		TP rate	0.93	0.93	0.92	0.93	0.91	0.93
		FP rate	0.21	0.24	0.18	0.22	0.18	0.24

Comparing the results we notice that there is no significant difference between the results obtained using the three different trials. The tables shows a better prediction for both percentage correct and true positives rate with attribute selection for the first trial (for both classification methods: CVR and BN); for the second trial there is better prediction with attribute selection for BN, while for CVR is constant for the first two cases and decreases for OneR attribute selection; for the third trial there are both increases and decreases with attribute selection for CVR and BN. However these variations are not statistically significant. From all trials, attribute selection increases most the prediction in Trial 1, which includes 30 attributes. As not all of them are relevant, an increase is to be expected when attribute selection is performed prior to prediction.

Using again the three attribute evaluation methods with ranking as search method for attribute selection, we can see the ranking among the 6 attributes from Trial 3, attributes related to reading and taking tests. The ranking is the same as the one obtained when using all attributes, for all three methods. Thus, according to chi-square and information gain ranking the most valuable attribute is average time spent on pages, followed by the number of pages, number of tests, average time spent on tests, number of correctly answered tests and number of incorrectly answered tests. OneR ranking differs only in the position of the last two attributes: number of incorrectly answered tests comes before number of correctly answered tests.

Summarizing the results from the "core" study, we have shown that the level of engagement can be predicted at a very good level, e.g. 91% using classification via regression; disengagement can be even better: 94% using the same method. The analysis included 943 sequences of 10 minutes from 48 users, showing that a general indicator of the motivational level could be predicted from very basic data commonly recorded in log files, such as events related to reading pages and taking tests. To validate these results, we conducted a validation study, presented in the following section.

4.3. Validation study

In order to validate our approach for engagement prediction presented above we analyzed data from a second system: iHelp, the University of Saskatchewan web-based learning environment. This system includes two web based applications designed to support both learners and instructors throughout the learning process: the iHelp

Discussion system and iHelp Learning Content Management System. The latter is designed to deliver online courses to students working at a distance, providing course content (text and multimedia) and quizzes/surveys. The students' interactions with the system are preserved in a machine readable format.

The same type of data about the interactions was selected from the registered information in order to perform the same type of analysis as the one performed with HTML Tutor data. An HTML course was also chosen in order to prevent differences in results caused by differences in subject matter.

We used logged data from all 21 students studying the selected course, meaning a total of 218 sessions and 735 sequences, 513 of exactly 10 minutes and 222 less than 10 minutes. Again, only the 513 sequences of exactly 10 minutes were used in the analysis.

4.3.1. Attributes description

In the analysis several attributes mainly related to reading pages and quizzes events were used. These attributes are presented in Table 4.9. The terms tests and quizzes will be used interchangeably; they refer to the same type of assessment, except that in HTML they are called tests and in iHelp they are named quizzes.

Also, iHelp provides the score to quizzes that have been finalized on a scale from 1 to 100 as opposed to HTML-Tutor that provides correctness or incorrectness of answer per each question.

Table 4.9 iHelp: the attributes used for analysis

Codes (as used in analysis)	Attributes
NoPages	Number of pages read / accessed
AvgTimeP	Average time spent reading
NoQuestions	Number of questions from quizzes/ surveys
Score	Scores sum of the taken tests
NoPpP	Number of pages above the threshold established for maximum time required to read a page
NoPpM	Number of pages below the threshold established for minimum time to read a page

Two new attributes were introduced for this analysis, attributes that were not considered for HTML Tutor: the number of pages above and below a certain time threshold; they are described in the subsequent section. As described in the instructions given to the second rater from the "core" study, two patterns of disengagement were observed: spending too much time on a page/test and moving fast through pages/tests. The two new attributes are related to these patterns and were introduced in order to help with their identification which would potentially improve disengagement detection.

4.3.2. Level of engagement

The level of engagement was established using the same approach as in the "core" study, adding two extra rules related to the two additional attributes regarding number of pages that are above or below a threshold, depending on the time required for reading.

At first we intended to use the average time spent on each page across all users, as suggested by (Martinez, 2003), but analyzing the data, we have seen that some pages are accessed by a very small number of users, sometimes only one, a problem encountered in other research as well (Farzan and Brusilovsky, 2005). Thus, we decided to use the average reading speed known to be in between 200 and 250 words per minute (ReadingSoft.com, TurboRead.com). The distribution for the 664 pages accessed by the students is displayed in Table 4.10.

Table 4.10 iHelp: time intervals for reading and the number of pages in each interval.

Time interval	No of pages
500-550	3
400-500	2
300-400	5
200-300	41
100-200	145
<100	468

Some pages of the course include images and videos. However, only 4 of the 21 students attempted to watch videos; the number of attempts and the corresponding times spent watching videos are displayed in Table 4.11.

Table 4.11 iHelp: number of attempts and time spent watching videos grouped by subject.

Subject	No of attempts	Time (sec.)
S1	1	3.47
S2	1	162
S3	9	1.16
	2	2.31
	1	94.91
S4	8	1.16
	2	2.31

Given the distribution of pages presented in Table 4.10 and taking into consideration that very few students used videos, on one hand and the fact that there are individual differences in reading speed and also that some learners go through the material more than once, on the other hand, we determined 420 seconds as the maximum time required to read a page. This does not cover the five pages that need more that 400 seconds to be read, but considering that they represent less than 1% of the total 664 pages and that most of the pages are require less than 100 seconds (70%), we considered 420 seconds to be more appropriate. For the minimum threshold for reading a page we agreed on 5 seconds. Note that these thresholds are somewhat arbitrary. However, as outlined under future perspectives, these thresholds may be derived more precisely.

In the "core" study, the level of engagement was established by human experts that looked at the log files and established the level of engagement for sequences of 10 minutes or less, in a similar way to de Vicente and Pain (2002). The same procedure was applied for iHelp, considering also the two rules aforementioned.

Accordingly, the level of engagement was determined for each sequence of 10 minutes. If in a sequence the learner spent more than 420 seconds on a page, we considered that he/she was disengaged during that sequence. Related to pages accessed less than 5 seconds, we agreed to consider a user disengaged if 2/3 of the total number of pages accessed in a sequence were below 5 seconds.

With HTML Tutor, three level of engagement were used: engaged, disengaged and neutral. Neutral was used for situations when raters found it hard to decide whether the user was engaged or disengaged. With iHelp, this difficulty was not encountered.

4.3.3. Analysis and results

Using the attributes described above, an analysis was conducted in order to investigate engagement prediction with iHelp and compare the results with the ones from HTML-Tutor.

The same environment and methods as the ones used in our previous research were employed and two datasets were used: (i) Dataset 1 including all attributes and (ii) Dataset 2 obtained from Dataset 1 by eliminating the two additional attributes (NoPpP, NoPpM). Dataset 2 was considered in order to compare the results with the ones from HTML Tutor. Table 4.12 presents the datasets with the corresponding attributes.

Table 4.12 The validation study: datasets used in the experiment

Dataset	Attributes
Dataset 1	NoPages, AvgTimeP, NoQuestions, AvgTimeQ, Score, NoPpP, NoPpM
Dataset 2	NoPages, AvgTimeP, NoQuestions, AvgTimeQ, Score

As in the core study, we controlled the distribution of instances in order to not have instances from the same user in the training and the testing set. From the 513 instances 348 (68%) were used for training and 165 for testing (32%). The results are presented in Table 4.13.

Table 4.13 The validation study: experiment results summary

		BN	LR	SL	IBk	ASC	B	CvR	DT
Dataset 1	**%correct**	97.50	97.93	97.99	97.87	97.38	97.50	97.44	97.75
	TP rate	0.97	0.97	0.97	0.96	0.95	0.96	0.96	0.97
	FP rate	0.02	0.01	0.01	0.01	0.00	0.01	0.01	0.01
	Precision	0.98	0.99	0.99	0.99	1.00	0.99	0.99	0.99
	Recall	0.97	0.97	0.97	0.96	0.95	0.96	0.96	0.97
	Error	0.04	0.03	0.04	0.03	0.05	0.04	0.04	0.03
	d'	3.93	4.21	4.21	4.08	3.97	4.08	4.08	4.21
Dataset 2	**%correct**	85.62	85.92	85.56	85.44	84.77	85.80	85.37	85.07
	TP rate	0.77	0.77	0.76	0.78	0.76	0.77	0.75	0.76
	FP rate	0.06	0.05	0.05	0.07	0.06	0.05	0.04	0.07
	Precision	0.93	0.93	0.94	0.91	0.92	0.93	0.94	0.92
	Recall	0.77	0.77	0.76	0.78	0.76	0.77	0.75	0.76
	Error	0.23	0.21	0.22	0.20	0.24	0.23	0.24	0.23
	d'	2.29	2.38	2.35	2.25	2.26	2.38	2.42	2.18

The results displayed in Table 4.13 show good levels of prediction varying between 85% and 98%; the true positives rates for disengaged class varies between 75% and 97%. The results for Dataset 1 which includes the two new attributes are better than the ones for Dataset 2 with approximately 12% for the correct percentage for all classes and with approximately 20% for the true positives rate for disengaged, suggesting that a considerable improvement has been brought by the two new attributes. As in the results for HTML Tutor, the similarity of results obtained from different methods shows the consistency of prediction and of the attributes used for prediction.

The highest percentage of correctly predicted instances was obtained using Simple Logistic classification on Dataset 1: 97.99%. The confusion matrix for this result is presented in Table 4.14. Focusing on the disengaged learners we see that the same method performs best (together with three other methods) on the same dataset: 97%. The confusion matrix shows that none of the engaged students is predicted as disengaged and that three disengaged students are predicted as engaged. This shows that engaged students are correctly identified and that they won't be interrupted for an intervention that is not required, but also it shows that some disengaged students are not identified as such and thus they would not receive intervention.

Table 4.14 The confusion matrix for best method (SL) on Dataset1

		Predicted		
		Disengaged	**Engaged**	**Total**
Actual	**Disengaged**	78	3	81
	Engaged	0	84	84
	Total	78	87	165

Investigating further the information gain brought by the two additional attributes, attribute ranking using information gain ranking filter as attribute evaluator was performed and the following ranking was found: NoPpP, NoPages, AvgTimeP, NoPpM, AvgTimeQ, Score and NoQuestions. Thus, the two new attributes are valuable for prediction and improve not only the prediction values, but also the processing time for the required calculations.

4.4. HTML-Tutor and iHelp: Results Comparison

HTML-Tutor and iHelp are similar in the type of events registered in log files, but they are different in one thing that is reflected in the data: the degree of access restriction. HTML-Tutor is a free tutorial available for everyone, no matter if they are enrolled in a course or not, while the iHelp materials are available only to registered students. This difference is reflected on the number of engaged vs. disengaged sequences: with HTML-tutor the majority of the sequences are labelled as disengaged (610 out of 943, meaning approximately 65%) while with iHelp the two classes are almost equally distributed (there are 253 sequences of disengaged out of 513, meaning almost 49%).

Looking at the results obtained for the two systems, on one hand we see a lower performance for iHelp (see results on Dataset 2 in Table 4.13, Section 4.3.3) compared to HTML-Tutor (see table 4.6 in Section 4.2.3) when the same attributes are used and a better performance for iHelp (see results on Dataset 1 in Table 4.13, Section 4.3.3) when two new attributes are used with iHelp. A summary of the comparison between the two systems, including prediction values and attribute ranking, is presented in Table 4.15. The difference in prediction values between the two iHelp datasets shows that the two new attributes contribute to a clearer (and faster) distinction between the two levels of engagement and thus, it would be expected that the usage of similar attributes with HTML-Tutor would increase the prediction values.

Table 4.15 Similarities and dissimilarities between iHelp and HTML Tutor

Characteristic	iHelp	HTML Tutor
Prediction based on reading and tests attributes	85% with similar attributes to HTML-Tutor 97% with two additional attributes	84-91%
Attribute ranking	1. Number of pages above a threshold 2. Number of pages 3. Average time spent reading 4. Number of pages below a threshold 5. Average time spent on quizzes 6. Score 7. Number of questions from quizzes	1. Average time spent on pages 2. Number of pages 3. Number of tests 4. Average time spent on tests 5. Number of correct answers 6. Number of incorrect answers

With both iHelp and HTML Tutor some patterns in the disengaged users' behaviour were distinguished: a) the disengaged students that click fast through pages without reading them and b) the disengaged students that spend long time on a page, (far) exceeding the needed time for reading that page. Two of the previous approaches mentioned in Section 2.3.2 also present some patterns, with the difference that those patterns are related only to learners' behaviour when answering quizzes. Thus, we find a similarity between blind guessing in Beck (2005) or unmotivated-guess in Johns and Woolf (2006), on one hand, and the fast click through pages, on the other hand, as both reflect students' rush and lack of attention. Knowledge about these patterns would be useful for a more targeted intervention.

4.5. Disengagement prediction refinement

In this section three studies conducted in order to refine the prediction model are presented. The first study investigates with HTML the prediction value of the two additional attributes introduced with iHelp, the second study investigates the possibility to predict the two patterns of disengagement introduced previously, and the third investigates the effect of eliminating exploratory sequences.

4.5.1. Validation of reading speed attributes

This study was conducted in order to investigate the effect on HTML-Tutor data of the reading speed attributes introduced with iHelp in the validation study. As the new attributes increased considerably the prediction with iHelp, the next natural step is to see if the same effect occurs with HTML-Tutor.

4.5.1.1. Design

For each sequence of 10 minutes in the HTML Tutor log data, the two attributes used with iHelp were added: the number of pages exceeding the 420 seconds threshold and the number of pages below five seconds. We compared the predictions obtained after adding these attributes with the predictions obtained without them. All three databases were considered in this study: DS-30, DS-10 and DS-6 attributes. The study design is presented in Table 4.16, with the corresponding datasets. Our hypothesis is that the two additional attributes will improve the overall and especially the disengagement prediction level.

Table 4.16 Validation of reading speed attributes study design

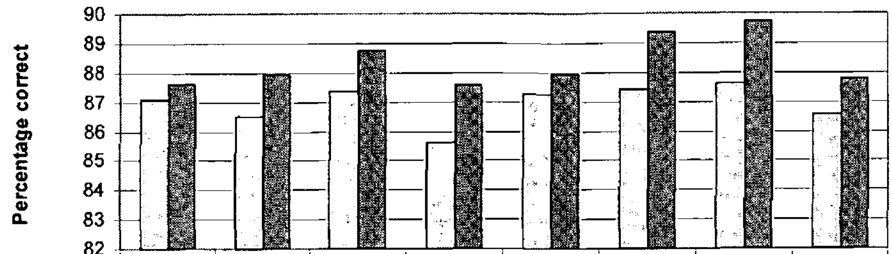
	30 attributes	10 attributes	6 attributes
With original attributes	DS-30	DS-10	DS-6
With the 2 additional attributes	DS-30+2	DS-10+2	DS-6+2

We chose a different notation of the databases used in the experiment in order to distinguish them from the ones used in the previous study.

4.5.1.2. Analysis and results

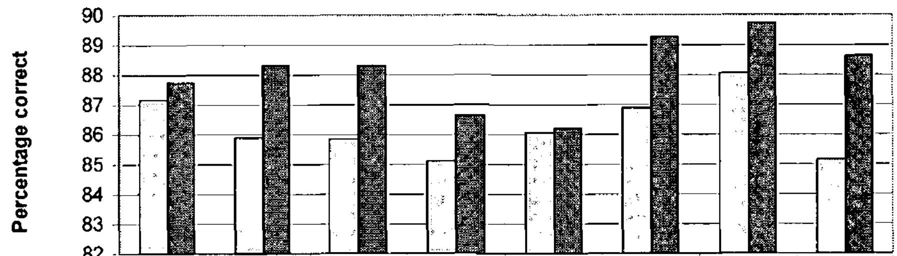
Like in the previous studies, we used Waikato Environment for Knowledge Analysis (WEKA) for the analysis and the eight prediction methods introduced in Section 4.2.2. The experiment was done using 10-fold stratified cross validation iterated 10 times.

The results are grouped in six figures. Figure 4.3 presents the percentage correct comparison between the original 30 attributes database (DS-30) with the same database that includes the two additional attributes (DS-30+2). Figure 4.4 and Figure 4.5 present the same comparison starting form the original 10 attributes and 6 attribute database. The subsequent three figures present the comparison for the true positive rates.



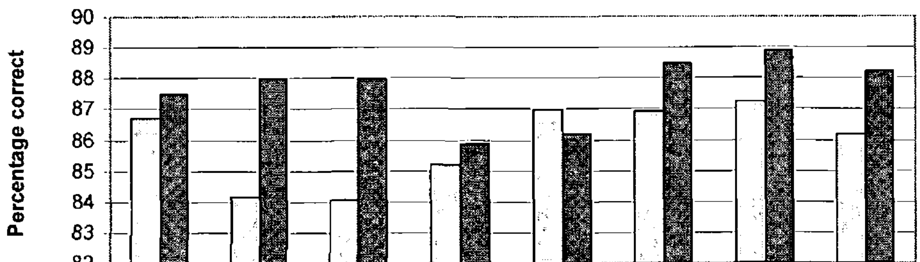
Methods	BN	LR	SL	IBk	B	ASC	CvR	DT
□ DS-30	87.07	86.52	87.33	85.62	87.24	87.41	87.64	86.58
■ DS-30+2	87.60	87.92	88.74	87.54	87.91	89.38	89.71	87.79

Figure 4.3 Percentage correct for original database with 30 attributes (DS-30) and the same database with the two additional attributes (DS-30+2)



Methods	BN	LR	SL	IBk	B	ASC	CvR	DT
□ DS-10	87.18	85.88	85.82	85.13	86.03	86.87	88.07	85.16
■ DS-10+2	87.72	88.31	88.33	86.62	86.22	89.25	89.76	88.63

Figure 4.4 Percentage correct for original database with 10 attributes (DS-10) and the same database with the two additional attributes (DS-10+2)



Methods	BN	LR	SL	IBk	B	ASC	CvR	DT
□ DS-6	86.68	84.15	84.05	85.18	86.95	86.90	87.21	86.20
■ DS-6+2	87.48	87.97	87.95	85.89	86.18	88.46	88.92	88.21

Figure 4.5 Percentage correct for original database with 6 attributes (DS-6) and the same database with the two additional attributes (DS-6+2)

In order to investigate the existence of significant differences, a normality test (Kolmogorov-Smirnov) was computed first in order to decide what test to be used for comparison: a parametric or a non-parametric one. The Kolmogorov-Smirnov Test compares an observed cumulative distribution function to a theoretical cumulative distribution. Large significance values (>0.05) indicate that the observed distribution corresponds to the theoretical distribution, in our case, the normal distribution. The significance values are displayed in Table 4.17. The figures in bold italic outline the datasets that are not normally distributed.

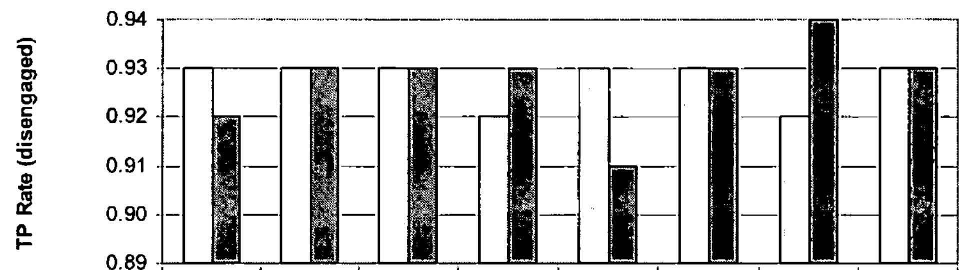
When the distributions for both databases are normal, paired t-test is used for comparison and when one or both distributions are not normal, Wilcoxon test is applied. More specifically, for BN, LR, B, ASC, and DT t-test is applied, while for SL, IBk, and CvR Wilcoxon test is used for some of the datasets. Paired t-test compares the means of two distributions that represent the same group. Wilcoxon test is a non-parametric test that detects differences in the distributions of two related variables. For both tests, a significance value less than 0.05 indicates a significant difference between the two distributions. The significance levels are presented in Table 4.17. The numbers in bold outline significant differences.

Table 4.17 Percentage correct: cells denote the significance values for Kolmogorov-Smirnov normality test respectively for the appropriate comparison test (either t-test for normal distribution or Wilcoxon test otherwise)

	BN	LR	SL	IBk	B	ASC	CvR	DT
DS-30	0.272	0.265	***0.007***	***0.045***	0.141	0.328	0.345	0.117
DS-30+2	0.426	0.414	0.178	0.704	0.164	0.398	0.139	0.470
Comparison	0.213	**0.000**	**0.001**	**0.000**	0.114	**0.000**	**0.000**	**0.003**
DS-10	0.333	0.320	0.310	0.086	0.362	0.167	0.113	0.693
DS-10+2	0.320	0.555	0.289	0.629	0.667	0.376	0.269	0.507
Comparison	0.202	**0.000**	**0.000**	**0.005**	0.663	**0.000**	**0.014**	**0.000**
DS-6	0.077	0.435	0.280	0.371	***0.032***	0.050	***0.022***	0.085
DS-6+2	0.473	0.381	0.374	0.702	0.482	0.257	0.095	0.161
Comparison	0.055	**0.000**	**0.000**	**0.000**	0.118	**0.000**	**0.000**	**0.000**

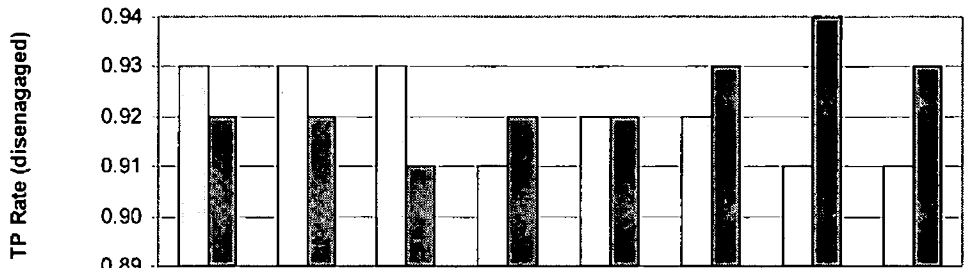
For all three database pairs, there are significant differences for 6 out of 8 methods: LR, SL, IBk, ASC, CvR and DT. In all cases, the percentage correct is higher for the databases with the two additional attributes. Thus, we consider that in the case of overall prediction, our hypothesis was confirmed.

The same comparison has been performed for the true positives (TP) rate for disengaged, as we are especially interested in identifying the disengaged learners. The results are displayed in Figure 4.6, 4.7 and 4.8. The significance values for the normality and comparison tests are presented in Table 4.18.



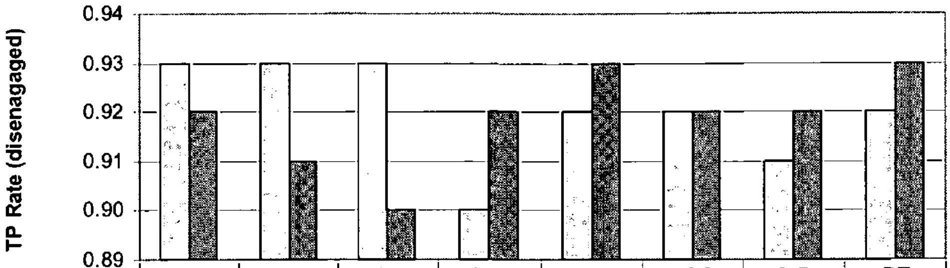
Methods	BN	LR	SL	IBk	B	ASC	CVR	DT
DS-30	0.93	0.93	0.93	0.92	0.93	0.93	0.92	0.93
DS-30+2	0.92	0.93	0.93	0.93	0.91	0.93	0.94	0.93

Figure 4.6 True positive rate (disengaged) for original database with 30 attributes (DS-30) and the same database with the two additional attributes (DS-30+2)



Methods	BN	LR	SL	IBk	B	ASC	CVR	DT
DS-10	0.93	0.93	0.93	0.91	0.92	0.92	0.91	0.91
DS-10+2	0.92	0.92	0.91	0.92	0.92	0.93	0.94	0.93

Figure 4.7 True positive rate (disengaged) for original database with 10 attributes (DS-10) and the same database with the two additional attributes (DS-10+2)



Methods	BN	LR	SL	IBk	B	ASC	CvR	DT
□ DS-6	0.93	0.93	0.93	0.90	0.92	0.92	0.91	0.92
▩ DS-6+2	0.92	0.91	0.90	0.92	0.93	0.92	0.92	0.93

Figure 4.8 True positive rate (disengaged) for original database with 6 attributes (DS-6) and the same database with the two additional attributes (DS-6+2)

In two situations, Figure 4.6: SL and Fig. 4.7: B, it appears in the graph that the true positive rate for the two databases (DS-30 and DS-30+2 in Figure 4.6; DS-10 and DS-10+2 in Figure 4.7) has the same value: 0.93 in Figure 4.6 and 0.92 in Figure 4.7. At the same time for these cases in Table 4.18 it appears that the differences for each of the two pairs of databases are significantly different. This is explained by the fact that the figures displayed are rounded to two digits. In fact, the values are: for Figure 4.6, DS-30: 0.9343; Figure 4.6, DS-30+2: 0.9267, Figure 4.7, DS-10: 0.9153 and Figure 4.7, DS-10+2: 0.9248.

Table 4.18 True positive rate for disengaged: significance values for Kolmogorov-Smirnov normality test and for comparison tests (t-test for normal distribution and Wilcoxon test otherwise)

	BN	LR	SL	IBk	B	ASC	CvR	DT
DS-30	0.127	***0.014***	***0.004***	0.082	***0.048***	0.099	0.158	0.089
DS-30+2	0.081	0.181	0.195	0.200	0.068	0.289	***0.008***	0.170
Comparison	0.436	0.538	**0.040**	**0.009**	**0.002**	0.750	**0.002**	0.193
DS-10	0.168	***0.045***	***0.023***	0.161	0.165	0.086	0.057	0.295
DS-10+2	0.256	0.196	0.374	0.093	***0.001***	0.260	0.129	0.263
Comparison	0.507	**0.049**	**0.002**	0.053	**0.016**	0.098	**0.000**	**0.000**
DS-6	0.079	0.186	0.113	0.232	***0.047***	0.160	0.071	0.181
DS-6+2	0.077	0.278	0.285	***0.040***	***0.001***	0.300	0.403	0.027
Comparison	0.555	**0.000**	**0.000**	**0.000**	**0.007**	0.475	**0.001**	**0.000**

For the TP rate there are less significant differences and not all of them are in favour of the two additional attributes. Thus in the case of the first pair (DS-30 vs. DS-30+2), the new attributes brought significant increase only for CvR and IBk methods, in the case of the second pair (DS-10 vs. DS-10+2), for CvR, DT and B (and almost significant for IBk: 0.053) and in the case of the third pair (DS-6 vs. DS-6+2), for IBk, B, CvR and DT. Thus, we can conclude that for disengagement prediction, our hypothesis was not confirmed for several cases.

From this study we conclude that the two new attributes are valuable for prediction, even if there are some cases where they seem to cause a decrease in the true positive rate values for disengaged class. Considering that the two new attributes are related to disengagement and that based on them, rules for labelling disengagement were used, the results are somehow surprising. We expected to have an increase in the true positive rate for the disengaged class when using the new attributes. Trying to clarify these results, in the study conducted for prediction of the two patterns, we also looked at the impact of the attributes on the predictions.

4.5.2. Patterns of disengagement

As in the previous study, in most of the cases an increase in the overall prediction was noticed, we considered that it is best to keep the two additional attributes for the following experiments. In the case of TP rate for disengagement, as there were situations when the prediction decreased, for HTML-Tutor we decided to have two trials: with and without the two additional attributes and, thus, investigate possible explanations for the results from the previous study.

The main purpose of this experiment was to investigate the possibility to predict two different patterns of disengagement: 1) fast browsing through pages/ tests, denoted as DF: "disengaged-fast" and 2) long time spent on the same page/ test, denoted as DL: "disengaged-long". Even if the names are not expressing opposite situations, as one may expect, these names were chosen because they express the corresponding behaviour accurately. The investigation was conducted with both HTML-Tutor and iHelp.

4.5.2.1. HTML-Tutor

We started from the same six databases from the validation of speed attributes study and used four levels of engagement: engaged, neutral, "disengaged long", and "disengaged fast". In order to distinguish the datasets used in this study compared to the previous one, the "L/F" label was added on the names of the datasets to indicate that "disengaged long" and "disengaged fast" patterns are used.

The sequences were coded as "disengaged-long" or "disengaged-fast" using the same rules that were introduced with iHelp validation study presented in Section 4.3.2:

- if in a sequence the learner spent more that 420 seconds (7 minutes) on a page or test, the sequence was coded DL ("disengaged-Long");
- if in a sequence 2/3 of the total number of pages were below 5 seconds, the sequence was coded DF ("disengaged-fast").

The same maximum threshold was used as with iHelp because all pages from HTML-Tutor require less that 400 seconds to be read and all the other arguments mentioned in Section 4.3.1. The minimum threshold, 5 seconds, was also the same; this minimum threshold has been used also in other studies (e.g. Farzan and Brusilovsky, 2005), and there seems to be an agreement about this minimal time to process the information on a page regardless if the time is spent to read the page or to look for other links.

From the total of 945 sequences of 10 minutes, 646 were DL and only 21 DF. Thus, as there were two few instances of DF, we focused on the DL pattern. The same software and methods were used for the analysis; 10-fold cross validation iterated 10 times was applied. Table 4.19 shows the percent correct and TP for DL for all databases.

Good prediction levels have been obtained for the overall prediction (percent correct), with values between 85.16 and 89.22, values slightly lower than the ones obtained when disengagement was only one category (see Table 4.6 from Section 4.2.3 – results from the "core" study) and than the ones presented in the validation of reading speed attributes study (Figures 4.3, 4.4 and 4.5 in Section 4.5.1.2), which was expected due to the introduction of the two patterns. On the other hand, the TP rates for DL, with the two additional attributes, with values between 0.89 and 0.95, are higher than the previous results from both the "core" study and the reading speed attributes validation study.

Table 4.19 HTML Tutor predictions of engagement levels when the two disengaged patterns, DL and DF, are considered; true positive rate is displayed only for DL.

		BN	LR	SL	IBk	ASC	B	CvR	DT
DS-30-L/F	**%correct**	84.33	86.31	87.14	84.66	87.12	86.81	87.16	86.10
	TP rate	0.89	0.94	0.95	0.93	0.95	0.94	0.94	0.94
DS-30+2-L/F	**%correct**	86.68	87.50	88.32	85.82	87.68	88.27	89.12	87.53
	TP rate	0.93	0.95	0.94	0.94	0.93	0.94	0.95	0.95
DS-10-L/F	**%correct**	83.40	85.96	85.69	84.37	86.66	86.37	87.47	85.20
	TP rate	0.88	0.93	0.94	0.92	0.93	0.94	0.94	0.92
DS-10+2-L/F	**%correct**	86.94	87.63	87.96	85.80	85.83	88.65	89.22	88.27
	TP rate	0.94	0.93	0.93	0.93	0.95	0.95	0.95	0.95
DS-6-L/F	**%correct**	83.06	83.90	84.00	82.41	86.95	86.52	86.73	85.86
	TP rate	0.89	0.92	0.93	0.91	0.93	0.93	0.93	0.92
DS-6+2-L/F	**%correct**	86.33	87.01	87.16	85.16	85.97	87.81	88.44	87.83
	TP rate	0.94	0.92	0.91	0.94	0.95	0.94	0.94	0.95

Figure 4.9 displays the distribution of percent correct for the best method (CvR) on DS-6-L/F. Most values are between 86% and 93%, and all of them are above 81%. The vertical lines in the figure are due to fractional percent correct values of the 95 test cases, for example 85/95 is approximately 89%. More common results for a certain value of percentage correct are visible in the higher frequency of dots along the vertical lines.

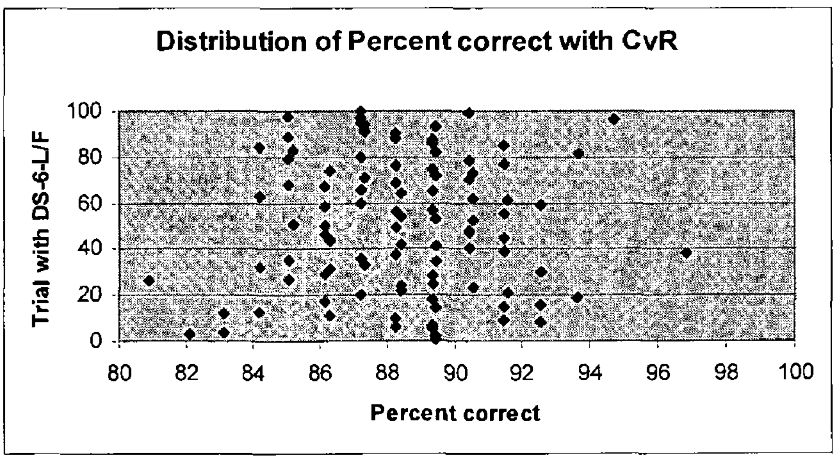


Figure 4.9 Distribution of Percent correct with CvR on DS-6-L/F.

For the TP rate for DL on DS-6-L/F and DS-6+2-L/F using CvR we notice close values: 0.94 with the two additional attributes and 0.93 without them. The graphs displayed in Figure 4.10 and 4.11 show that the distributions have more or less the same range, but the values are distributed differently.

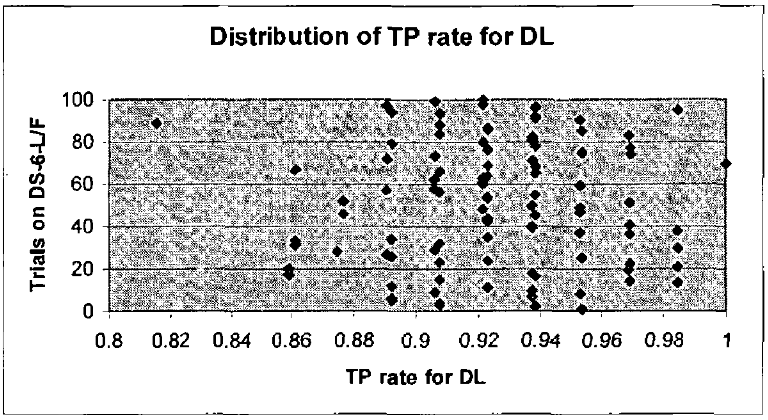


Figure 4.10 Distribution of TP rate for DL using CvR on DS-6-L/F (without the two additional attributes).

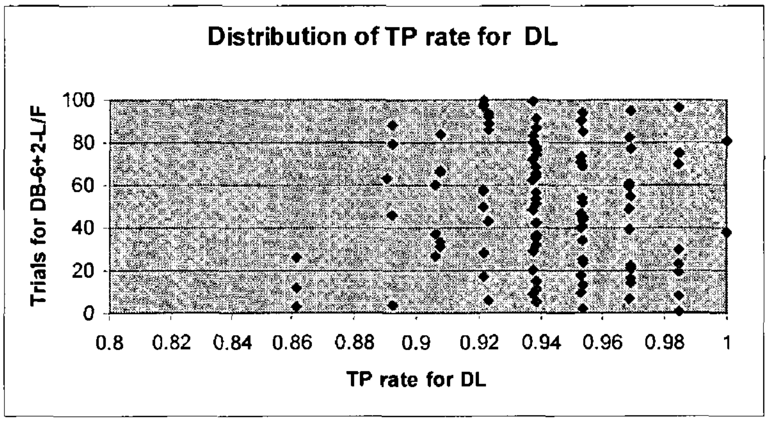


Figure 4.11 Distribution of TP rate for DL using CvR on DS-6+2-L/F (with the two additional attributes).

In order to see if there are significant differences between the two distributions, we applied the same procedure as in the validation of the reading speed attributes study. The normality and comparison test results for percentage correct are presented in Table 4.20 and the ones for TP rate in Table 4.21.

Table 4.20 Percentage correct: significance values for Kolmogorov-Smirnov normality test and for comparison test (t-test or Wilcoxon)

	BN	LR	SL	IBk	ASC	B	CvR	DT
DS-30-L/F	.152	.192	***.018***	.654	.133	.326	***.036***	.153
DS-30+2-L/F	.204	.353	.126	.681	.181	.349	.204	.348
Comparison	**.000**	**.000**	**.000**	**.000**	**.018**	**.000**	**.000**	**.000**
DS-10-L/F	.627	.663	.449	.222	.128	.245	.269	.193
DS-10+2-L/F	.103	.329	.317	.206	.594	.074	.251	.148
Comparison	**.000**	**.000**	**.000**	**.000**	**.010**	**.000**	**.000**	**.000**
DS-6-L/F	.396	.214	.322	.338	.268	.391	.577	.193
DS-6+2-L/F	.253	.558	.383	.168	.203	.518	.377	.327
Comparison	**.000**	**.000**	**.000**	**.000**	**.004**	**.000**	**.000**	**.000**

The three comparisons from Table 4.20 are displayed graphically in the Figure 4.12, 4.13 and 4.14. Table 4.20 shows significant differences for all three pairs of datasets and for all method and the figures show that almost all differences are in favour of the datasets with the new attributes (except ASC for DS-10+2-L/F and DS-6+2-L/F), meaning that the prediction tends to be better on DS-30+2-L/F, DS-10+2-L/F and DS-6+2-L/F compared to their correspondent datasets without the new attributes.

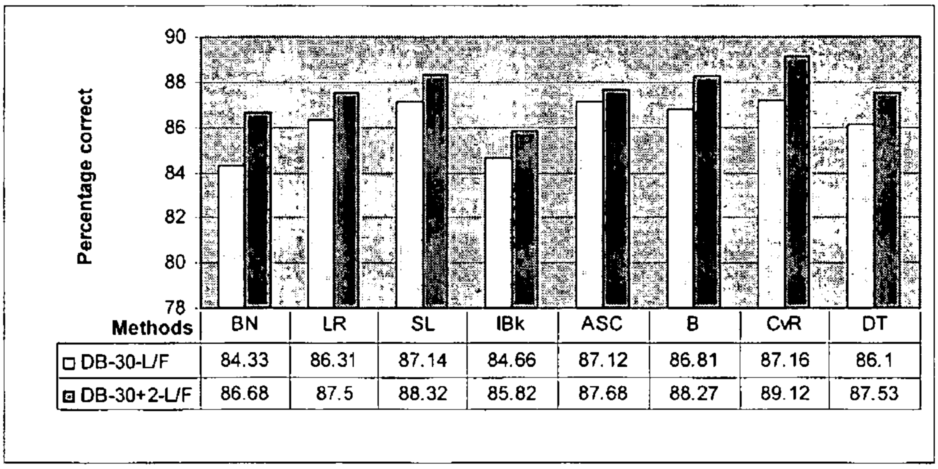
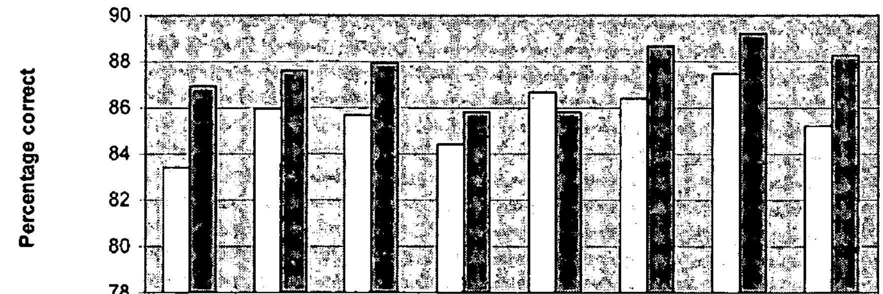
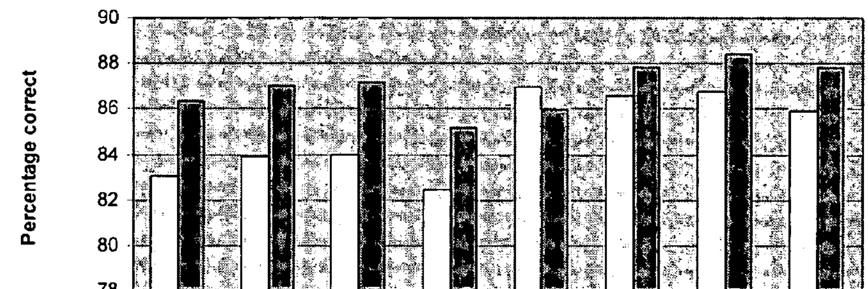


Figure 4.12 Percentage correct comparison between DS-30-L/F and DS-30+2-L/F.



Methods	BN	LR	SL	IBk	ASC	B	CvR	DT
DS-10-L/F	83.4	85.96	85.69	84.37	86.66	86.37	87.47	85.2
DS-10+2-L/F	86.94	87.63	87.96	85.8	85.83	88.65	89.22	88.27

Figure 4.13 Percentage correct comparison between DS-10-L/F and DS-10+2-L/F.



Methods	BN	LR	SL	IBk	ASC	B	CvR	DT
DS-6-L/F	83.06	83.9	84	82.41	86.95	86.52	86.73	85.86
DS-6+2-L/F	86.33	87.01	87.16	85.16	85.97	87.81	88.44	87.83

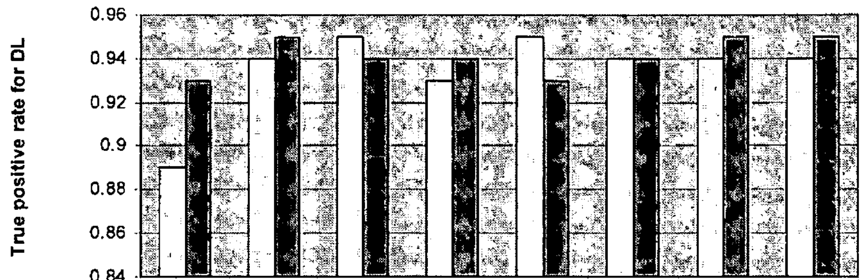
Figure 4.14 Percentage correct comparison between DS-6-L/F and DS-6+2-L/F.

Table 4.21 TP rate for DL: significance values for Kolmogorov-Smirnov normality test and for comparison test (t-test or Wilcoxon)

	BN	LR	SL	IBk	ASC	B	CvR	DT
DS-30-L/F	.324	***.012***	***.031***	***.030***	***.002***	.193	.136	.105
DS-30+2-L/F	.229	***.017***	***.030***	.060	.296	.083	***.032***	***.003***
Comparison	**.000**	.081	**.041**	**.000**	**.000**	.674	**.000**	**.005**
DS-10-L/F	.111	.215	.224	.255	.094	***.017***	.102	***.005***
DS-10+2-L/F	.053	.053	.075	***.036***	***.000***	***.020***	***.000***	***.004***
Comparison	**.000**	.307	.157	**.000**	**.000**	**.001**	**.000**	**.000**
DS-6-L/F	.319	.323	.188	.290	***.003***	.070	.274	.179
DS-6+2-L/F	***.029***	.198	.333	.222	***.000***	.061	***.019***	.106
Comparison	**.000**	**.015**	**.000**	**.000**	**.000**	**.018**	**.000**	**.000**

For the first pair, DS-30-L/F and DS-30+2-L/F, Table 4.21 displays significant differences for six methods and non-significant differences for two methods (LR and B).

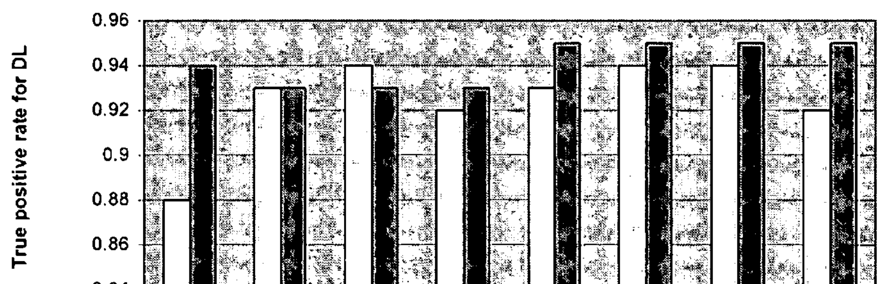
From the significant differences, four (BN, IBk, CvR and DT) are in favour of the datasets with the new attributes and two (SL and ASC) are in favour of the datasets without the new attributes, as it can be see in Figure 4.15.



Methods	BN	LR	SL	IBk	ASC	B	CvR	DT
DB-30-L/F	0.89	0.94	0.95	0.93	0.95	0.94	0.94	0.94
DB-30+2-L/F	0.93	0.95	0.94	0.94	0.93	0.94	0.95	0.95

Figure 4.15 True positive rate for DL comparison between DS-30-L/F and DS-30+2-L/F.

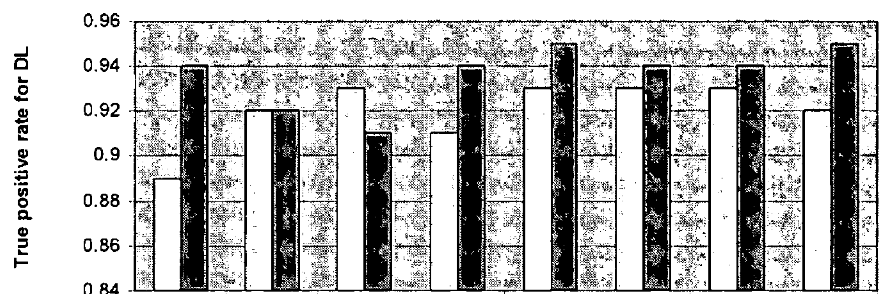
For the second pair, DS-10-L/F and DS-10+2-L/F, Table 4.21 shows six significant differences and two non-significant differences (LR and SL). In Figure 4.16 we can see that predictions are better for the dataset with the new attributes (DS-10+2-L/F) compared to the one without the new attributes (DS-10-L/F) for all six methods for which the differences are significant.



Methods	BN	LR	SL	IBk	ASC	B	CvR	DT
DS-10-L/F	0.88	0.93	0.94	0.92	0.93	0.94	0.94	0.92
DS-10+2-L/F	0.94	0.93	0.93	0.93	0.95	0.95	0.95	0.95

Figure 4.16 True positive rate for DL comparison between DS-10-L/F and DS-10+2-L/F.

For the third pair, DS-6-L/F and DS-6+2-L/F, Table 4.21 shows significant differences for all methods. Figure 4.17 shows that for six of the total of the eight methods the true positive rate is higher for the dataset with the new attributes (DS-6+2-L/F) compared to the one without the new attributes (DS-6-L/F). For LR in Figure 4.17 it appears that the values are the same, but Table 4.21 displays significant differences; as in a previous situation this is explained by the fact that the results were rounded to two digits. Looking at the values with four digits, for DS-6-L/F the value is 0.9238 and for DS-6+2-L/F is 0.9150. Thus, for LR the true positive rate is higher for the dataset without the new attributes.



Methods	BN	LR	SL	IBk	ASC	B	CVR	DT
DS-6-L/F	0.89	0.92	0.93	0.91	0.93	0.93	0.93	0.92
DS-6+2-L/F	0.94	0.92	0.91	0.94	0.95	0.94	0.94	0.95

Figure 4.17 True positive rate for DL comparison between DS-6-L/F and DS-6+2-L/F.

Summarizing the results of this study, we notice that, on one hand, the introduction of the two patterns brought a small decrease of the correct percentage of predictions up to 5% and, on the other hand, it introduced a higher positive rate for DL. Because of the small number of DF sequences, the prediction of this pattern was not possible. Comparing the results from the datasets with the reading speed attributes with the results from the datasets without them, like in the previous study, for some cases the true positive rate was lower for the datasets with the new attributes. However, this happened only in 4 cases out of 24 and in 16 cases true positive rates were higher for the datasets with the new attributes. Considering also the results from the previous study, we can conclude that the new attributes are more valuable for the correct prediction of "disengaged-long" than for prediction of disengagement.

4.5.2.2. iHelp

For this study, different datasets from the ones in the validation study were used. The validation study was done initially without the information about the scores on quizzes (due to technical problems we did not have access to the database with the scores); a lower number of instances were analyzed at that point. The results were reported in (Cocea and Weibelzahl, 2007b). For the study presented here, we used these initial datasets, that we denoted DS-all and DS-600. As at a later point we had access to the information about the scores on quizzes, we updated the results for the validation study.

In order to distinguish the datasets used in this study from the following one, we added "L/F" to indicate that "disengaged long" and "disengaged fast" patterns are included. From the total of 450 sequences, 169 were DL and 82 were DF. DS-all-L/F includes all instances, while DS-600-L/F includes only sequences of exactly 10 minutes (340 with 161 DL and 8 DF). Both datasets include all attributes. Because DS-600-L/F contained only 8 DF instances, we investigated only the overall and DL prediction of this dataset. The larger number of DF instances in DS-all-L/F compared to DS-600-L/F indicates that the learners that are "disengaged fast" tend to spend less than 10 minutes on the system. In other words, this pattern may indicate that the learner is about to leave the system.

The same tool and methods were used, as well as the 10-fold stratified cross validation iterated 10 times. The results are presented in Table 4.22.

The percent correct for DS-all-L/F has good values: from 88.87 to 91.13; for DS-600-L/F, the values are higher: from 93.14 to 94.58. The TP rate for DL has values from 0.91 to 0.93 with DS-all-L/F and from 0.93 to 0.94 with DS-600-L/F. The TP rate for DF (only DS-all-L/F) has satisfactory values given the small number of entries: from 0.73 to 0.85. For all trials and methods, the mean absolute error is between 0.05 and 0.11. The d prime values are extremely good for both DL and DF, indicating a good discrimination of both patterns.

Table 4.22 iHelp predictions of engagement levels with the two disengaged patterns, DL and DF.

		BN	LR	SL	IBk	ASC	B	CvR	DT
DS-all-L/F	**%correct**	89.27	91.13	91.13	88.87	88.98	90.22	90.62	89.73
	TP rate DL	0.91	0.92	0.91	0.92	0.91	0.91	0.91	0.91
	FP rate DL	0.01	0.02	0.02	0.04	0.02	0.01	0.02	0.02
	d'	3.67	3.46	3.39	3.16	3.46	3.67	3.46	3.46
	TP rate DF	0.73	0.84	0.85	0.76	0.74	0.79	0.81	0.80
	FP rate DF	0.04	0.04	0.04	0.04	0.04	0.04	0.04	0.04
	d'	2.36	2.75	2.79	2.46	2.39	2.56	2.63	2.59
	Error	0.11	0.09	0.09	0.08	0.11	0.10	0.10	0.10
DS-600-L/F	**%correct**	93.14	94.58	94.40	94.13	93.76	93.90	94.28	93.81
	TP rate DL	0.93	0.94	0.93	0.94	0.93	0.93	0.93	0.93
	FP rate DL	0.02	0.03	0.02	0.04	0.02	0.02	0.02	0.03
	d'	3.53	3.44	3.53	3.31	3.53	3.53	3.53	3.53
	Error	0.08	0.06	0.07	0.05	0.08	0.07	0.07	0.07

The distribution of percent correct on DS-all-L/F for one of the best performing methods, SL, is presented in Figure 4.18, where we can see that most values fall between 86 and 96. These values are lower than the original results (see Table 4.13 from Section 4.3.3 – the validation study) where no distinction between the two disengagement patterns was done.

The distribution of TP rate for DL includes values from 0.70 to 1 (Figure 4.19), with most values above 0.86. Again, compared to the original results, the prediction performance decreased.

Figure 4.20 displays the distribution of TP rates for DF. The results obtained with most values above 0.75 and 19 cases (out of 100) with value 1, meaning exact prediction. We were surprised to find these values, considering the low number of instances for DF.

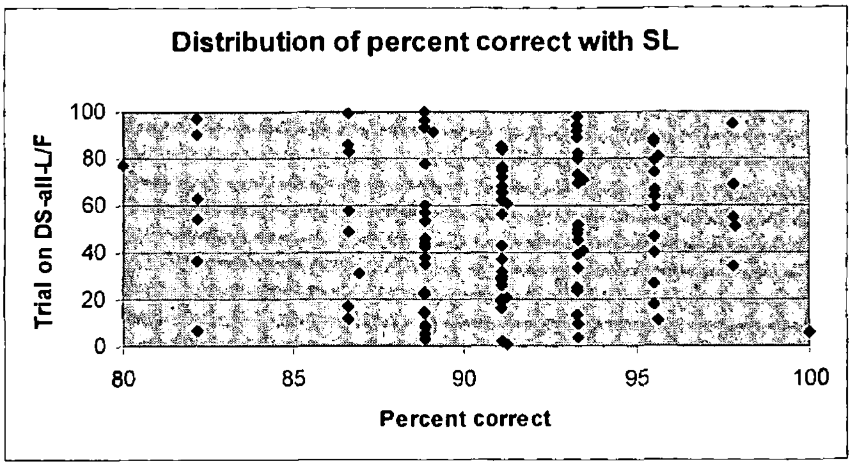


Figure 4.18 Distribution of percent correct with SL on DS-all-L/F.

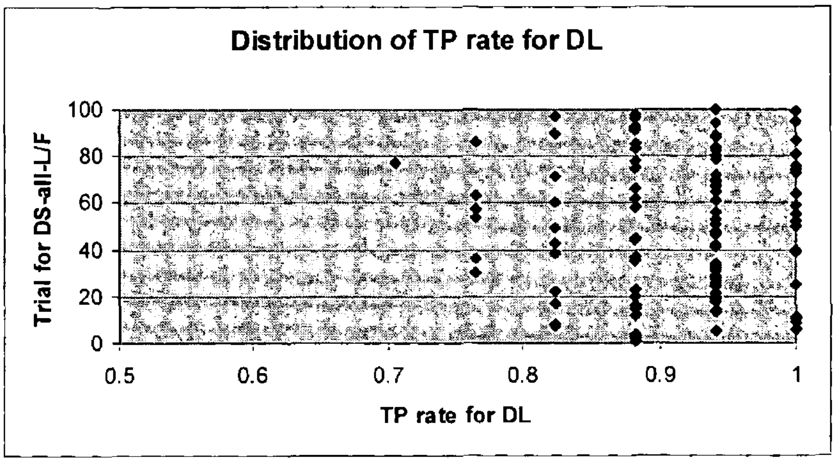


Figure 4.19 Distribution of true positives rate for DL on DS-all-L/F, using SL method.

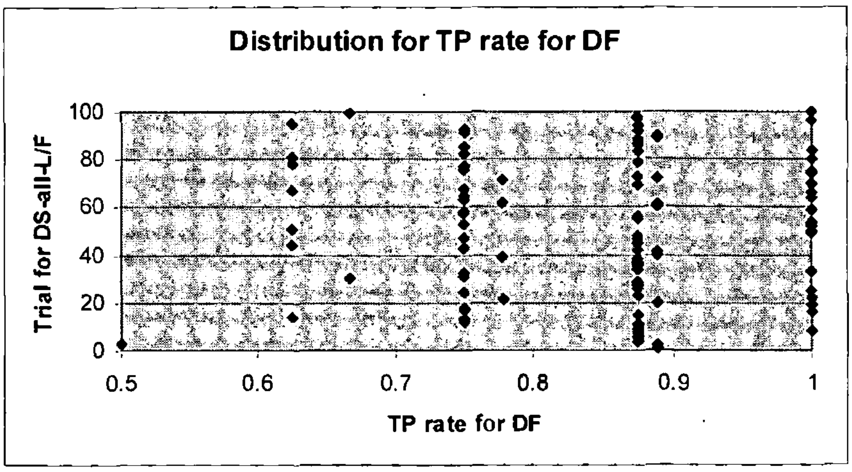


Figure 4.20 Distribution of true positives rate for DF on DS-all-L/F, using SL method.

This study showed good prediction and discrimination of the two patterns, "disengaged-long" and "disengaged short" with the data from iHelp, with higher values for "disengaged long" and lower, but satisfactory values for "disengaged short".

4.5.3. Exclusion of the exploratory sequences

Besides the two patterns investigated in the previous study, we observed with both HTML and iHelp that on the first entry to the system, the learners tend to have an exploratory behaviour, meaning that they click on the menu options and also on the links to the main chapters of the course. This behaviour of getting familiarized with the system is different from what was observed with the following sequences, when the learners seem to focus on the content. Given this difference, the presence of the initial sequences where the exploratory behaviour occurs in the analysis may negatively influence the results. This study was conducted in order to explore the influence of the exclusion of these exploratory sequences on prediction values; both systems, HTML-Tutor and iHelp, were considered.

4.5.3.1. HTML-Tutor

From the 943 sequences, the 65 representing the first sequence of the first session were eliminated. Thus, the database used for analyses included 878 instances. We looked at all the datasets, i.e. DS-30+2, DS-10+2 and DS-6+2, with (labelled "dl/df/e/n") or without (labelled "d/e/n") the two patterns; all datasets contained the reading speed attributes.

The results are displayed in Table 4.23. Looking at the percentage correct, we observe the following:

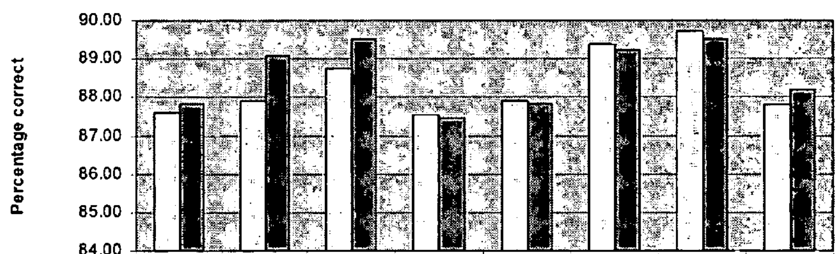
- compared to results from the validation of the reading speed attributes (no patterns included):

- Datasets with all attributes: the values are pretty similar across the board; in four cases the values are higher if the exploratory sequences are not considered. In four other cases the values are higher with the exploratory sequences included (see Figure 4.21);
- Datasets with 10 attributes: the values are lower when the exploratory sequences are excluded for most of the cases, i.e. six out of eight (see Figure 4.22);
- Datasets with 6 attributes: the same situation as for the datasets with 10 attributes (see Figure 4.23).

Table 4.23 HTML-Tutor: Prediction results without the exploratory sequences

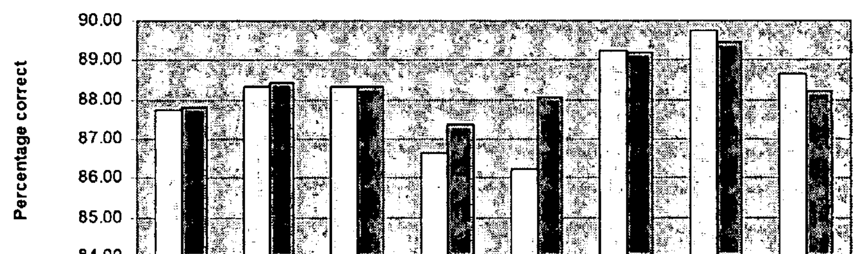
		BN	LR	SL	IBk	ASC	B	CvR	DT
DS-30+2 (d/e/n)	**%correct**	87.82	89.07	89.53	87.47	87.82	89.21	89.53	88.21
	TP rate d	0.93	0.94	0.94	0.93	0.94	0.93	0.94	0.93
	FP rate d	0.21	0.23	0.20	0.20	0.23	0.18	0.18	0.21
	Error	0.10	0.10	0.10	0.09	0.11	0.10	0.10	0.10
	d'	2.28	2.29	2.40	2.32	2.29	2.39	2.47	2.28
DS-30+2 (dl/df/e/n)	**%correct**	86.83	88.39	88.83	85.73	87.59	88.84	89.35	88.92
	TP rate DL	0.93	0.95	0.95	0.93	0.95	0.95	0.95	0.96
	FP rate DL	0.19	0.20	0.18	0.18	0.21	0.16	0.16	0.16
	Error	0.08	0.08	0.08	0.07	0.09	0.09	0.08	0.08
	d'	2.35	2.49	2.56	2.39	2.45	2.64	2.64	2.75
DS-10+2 (d/e/n)	**%correct**	87.81	88.43	88.31	87.35	88.06	89.18	89.45	88.19
	TP rate d	0.93	0.93	0.93	0.92	0.94	0.93	0.94	0.93
	FP rate d	0.21	0.21	0.19	0.19	0.22	0.18	0.18	0.21
	Error	0.10	0.10	0.11	0.09	0.11	0.11	0.10	0.10
	d'	2.28	2.28	2.35	2.28	2.33	2.39	2.47	2.28
DS-10+2 (dl/df/e/n)	**%correct**	86.83	87.98	88.01	85.98	87.59	88.92	89.34	88.88
	TP rate DL	0.93	0.94	0.94	0.93	0.95	0.95	0.96	0.96
	FP rate DL	0.19	0.21	0.17	0.16	0.21	0.16	0.16	0.16
	Error	0.08	0.08	0.08	0.07	0.09	0.09	0.08	0.08
	d'	2.35	2.36	2.51	2.47	2.45	2.64	2.75	2.75
DS-6+2 (d/e/n)	**%correct**	87.76	87.89	87.53	86.10	88.01	88.23	88.52	87.86
	TP rate d	0.93	0.92	0.91	0.93	0.94	0.93	0.92	0.94
	FP rate d	0.23	0.18	0.17	0.23	0.22	0.19	0.17	0.24
	Error	0.11	0.11	0.11	0.10	0.11	0.11	0.11	0.11
	d'	2.21	2.32	2.30	2.21	2.33	2.35	2.36	2.26
DS-6+2 (dl/df/e/n)	**%correct**	87.06	87.35	87.44	84.59	87.63	88.19	88.64	88.35
	TP rate d	0.94	0.93	0.93	0.93	0.95	0.94	0.94	0.96
	FP rate DL	0.21	0.16	0.14	0.20	0.21	0.17	0.16	0.18
	Error	0.09	0.09	0.09	0.08	0.09	0.09	0.09	0.09
	d'	2.36	2.47	2.56	2.32	2.45	2.51	2.55	2.67

- compared to the results from the pattern detection study:
    - o Datasets with all attributes: the values are higher when the exploratory sequences are excluded, in six cases out of eight (see Figure 4.24);
    - o Datasets with 10 attributes: the values are higher when the exploratory sequences are excluded for most of the cases, i.e. seven out of eight (see Figure 4.25);
    - o Datasets with 6 attributes: the same situation as for the datasets with 10 attributes (see Figure 4.26).



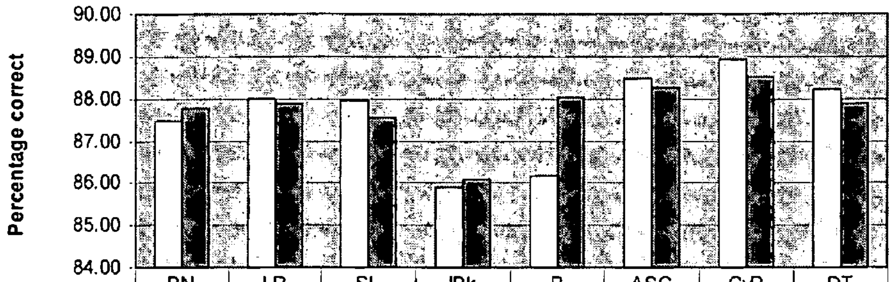
Methods	BN	LR	SL	IBk	B	ASC	CvR	DT
DS-30+2	87.60	87.92	88.74	87.54	87.91	89.38	89.71	87.79
DS-30+2 (d/e/n)	87.82	89.07	89.53	87.47	87.82	89.21	89.53	88.21

Figure 4.21 Percentage correct comparison between DS-30+2 and DS-30+2 (d/e/n).



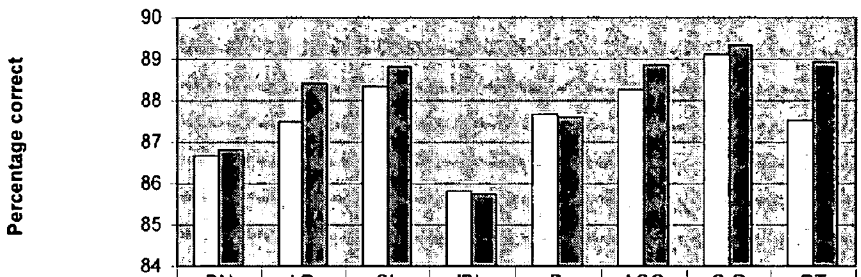
Methods	BN	LR	SL	IBk	B	ASC	CvR	DT
DS-10+2	87.72	88.31	88.33	86.62	86.22	89.25	89.76	88.63
DS-10+2 (d/e/n)	87.81	88.43	88.31	87.35	88.06	89.18	89.45	88.19

Figure 4.22 Percentage correct comparison between DS-10+2 and DS-10+2 (d/e/n).



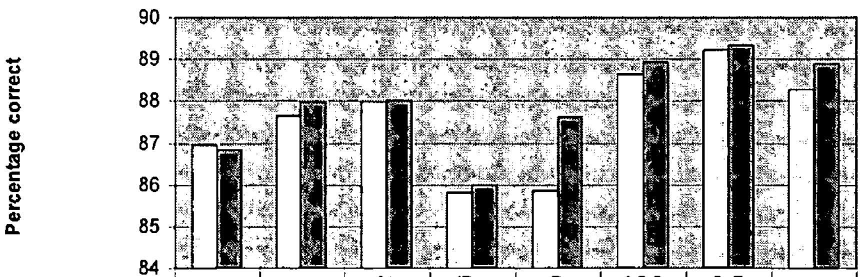
Methods	BN	LR	SL	IBk	B	ASC	CVR	DT
□ DS-6+2	87.48	87.97	87.95	85.89	86.18	88.46	88.92	88.21
▣ DS-6+2 (d/e/n)	87.76	87.89	87.53	86.1	88.01	88.23	88.52	87.86

Figure 4.23 Percentage correct comparison between DS-6+2 and DS-6+2 (d/e/n).



Methods	BN	LR	SL	IBk	B	ASC	CVR	DT
□ DS-30+2-L/F	86.68	87.5	88.32	85.82	87.68	88.27	89.12	87.53
▣ DS-30+2 (dl/df/e/n)	86.83	88.39	88.83	85.73	87.59	88.84	89.35	88.92

Figure 4.24 Percentage correct comparison between DS-30+2-L/F and DS-30+2 (dl/df/e/n).



Methods	BN	LR	SL	IBk	B	ASC	CVR	DT
□ DS-10+2-L/F	86.94	87.63	87.96	85.8	85.83	88.65	89.22	88.27
▣ DS-10+2 (dl/df/e/n)	86.83	87.98	88.01	85.98	87.59	88.92	89.34	88.88

Figure 4.25 Percentage correct comparison between DS-10+2-L/F and DS-10+2 (dl/df/e/n).

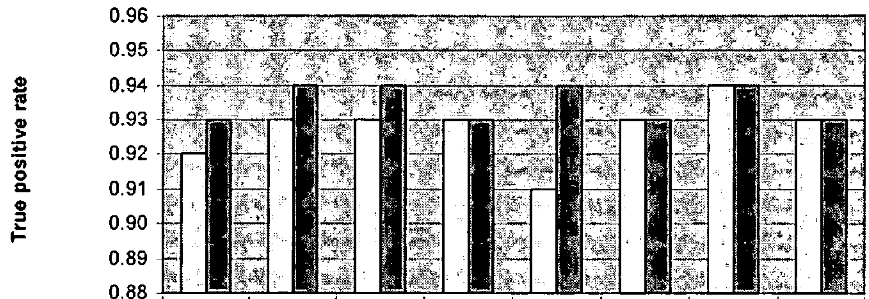
Methods	BN	LR	SL	IBk	B	ASC	CVR	DT
□ DS-6+2-L/F	86.33	87.01	87.16	85.16	85.97	87.81	88.44	87.83
▣ DS-6+2 (dl/df/e/n)	87.06	87.35	87.44	84.59	87.63	88.19	88.64	88.35

Figure 4.26 Percentage correct comparison between DS-6+2-L/F and DS-30+2 (dl/df/e/n).

Looking at the true positive rate (for disengaged or "disengaged-long"), we observe the following:

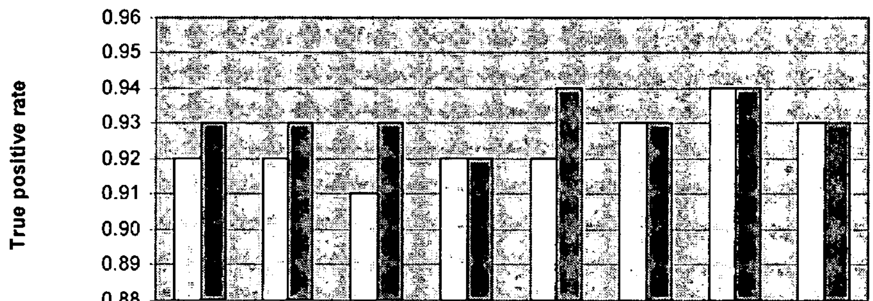
- compared to results from the validation of the reading speed attributes:
  - o Datasets with all attributes: the values are the same in four cases and in the other four the values are higher when the exploratory sequences are excluded (see Figure 4.27);
  - o Datasets with 10 attributes: the same situation as for the datasets with 30 attributes (see Figure 4.28);
  - o Datasets with 6 attributes: in one case the values are the same; for the other seven the values are higher when the exploratory sequences are excluded (see Figure 4.29);
- compared to the results from the pattern detection study:
  - o Datasets with all attributes: the values are higher when the exploratory sequences are excluded in four cases; in one case the opposite situation is encountered; for the other three cases, the values are the same (see Figure 4.30);
  - o Datasets with 10 attributes: the same situation as for the datasets with 30 attributes (see Figure 4.31);

- Datasets with 6 attributes: in four cases the values are the same; in one cases the value is higher when the exploratory sequences are included; for the remaining three cases, the values are higher when the exploratory sequences are excluded (see Figure 4.32).



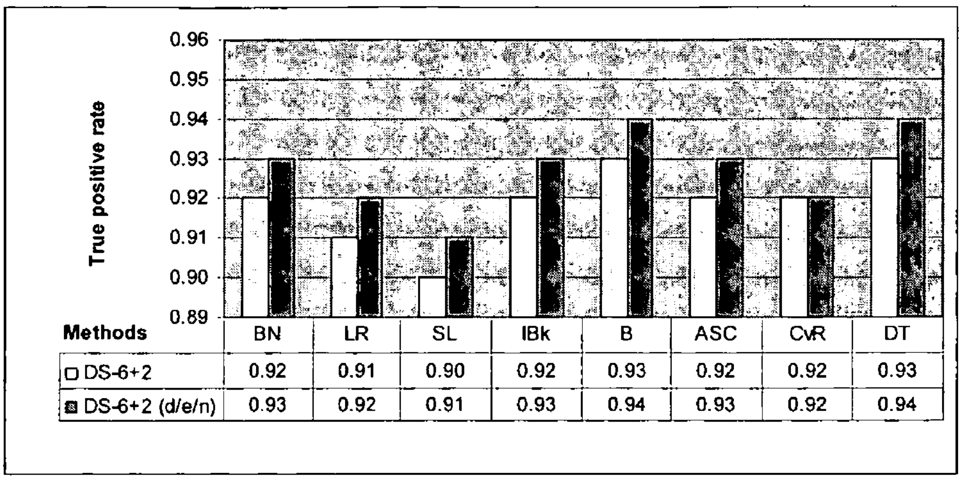
Methods	BN	LR	SL	IBk	B	ASC	CVR	DT
DS-30+2	0.92	0.93	0.93	0.93	0.91	0.93	0.94	0.93
DS-30+2 (d/e/n)	0.93	0.94	0.94	0.93	0.94	0.93	0.94	0.93

Figure 4.27 True positive rate for disengagement (d) comparison between DS-30+2 and DS-30+2 (d/e/n).



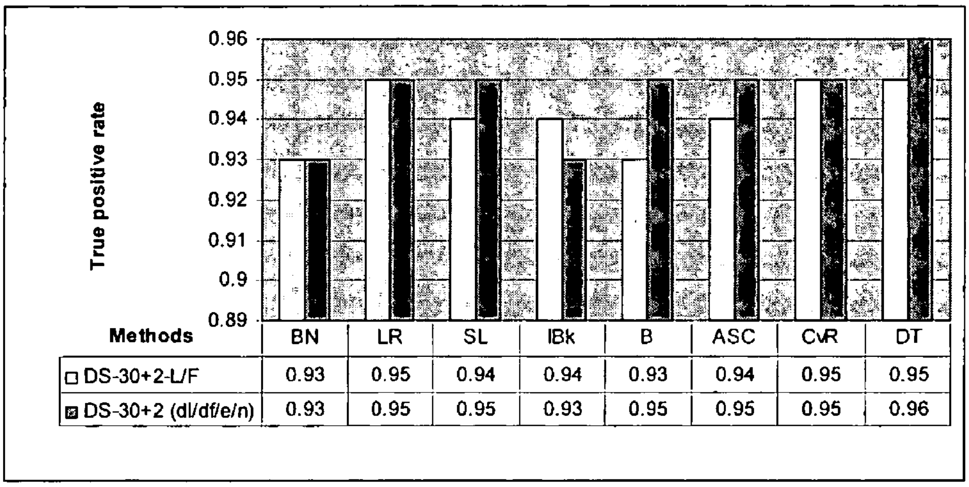
Methods	BN	LR	SL	IBk	B	ASC	CVR	DT
DS-10+2	0.92	0.92	0.91	0.92	0.92	0.93	0.94	0.93
DS-10+2 (d/e/n)	0.93	0.93	0.93	0.92	0.94	0.93	0.94	0.93

Figure 4.28 True positive rate for disengagement (d) comparison between DS-10+2 and DS-10+2 (d/e/n).



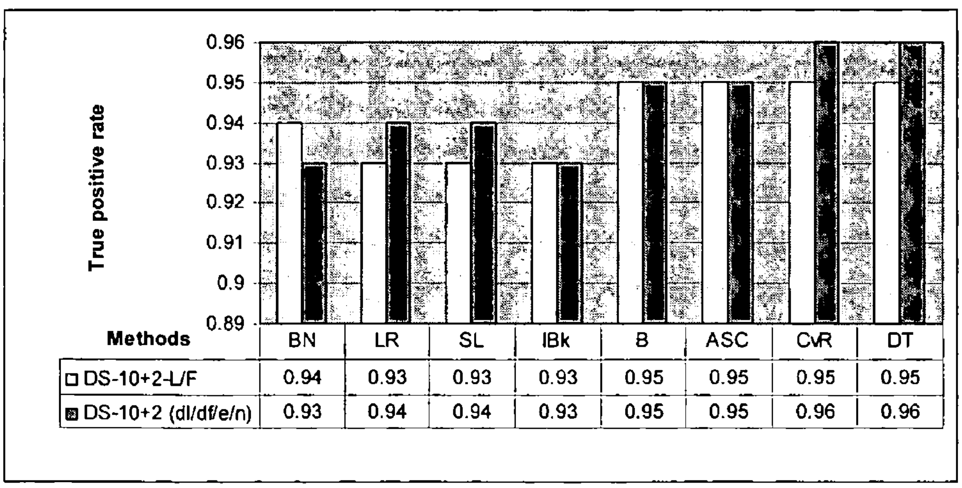
Methods	BN	LR	SL	IBk	B	ASC	CVR	DT
DS-6+2	0.92	0.91	0.90	0.92	0.93	0.92	0.92	0.93
DS-6+2 (d/e/n)	0.93	0.92	0.91	0.93	0.94	0.93	0.92	0.94

Figure 4.29 True positive rate comparison for disengagement (d) between DS-6+2 and DS-6+2 (d/e/n).



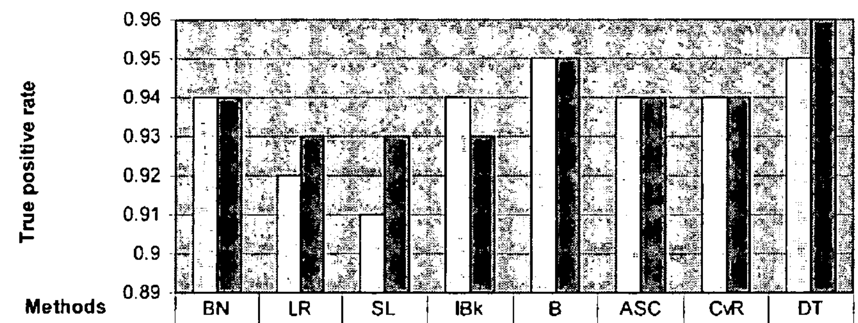
Methods	BN	LR	SL	IBk	B	ASC	CVR	DT
DS-30+2-L/F	0.93	0.95	0.94	0.94	0.93	0.94	0.95	0.95
DS-30+2 (dl/df/e/n)	0.93	0.95	0.95	0.93	0.95	0.95	0.95	0.96

Figure 4.30 True positive rate for DL comparison between DS-30+2-L/F and DS-30+2 (dl/df/e/n).



Methods	BN	LR	SL	IBk	B	ASC	CVR	DT
DS-10+2-L/F	0.94	0.93	0.93	0.93	0.95	0.95	0.95	0.95
DS-10+2 (dl/df/e/n)	0.93	0.94	0.94	0.93	0.95	0.95	0.96	0.96

Figure 4.31 True positive rate for DL comparison between DS-10+2-L/F and DS-10+2 (dl/df/e/n).



	BN	LR	SL	IBk	B	ASC	CVR	DT
DS-6+2-L/F	0.94	0.92	0.91	0.94	0.95	0.94	0.94	0.95
DS-6+2 (dl/df/e/n)	0.94	0.93	0.93	0.93	0.95	0.94	0.94	0.96

Figure 4.32 True positive rate for DL comparison between Ds-6+2-L/F and DS-6+2 (dl/df/e/n).

Summarizing the results, overall an increase of the prediction values is observed, which tends to support the fact that excluding the exploratory sequences influences the prediction and thus, suggesting that the training should not include the exploratory sequences. There are several ways of taking in consideration this information and details about possible approaches are given in Section 7.3 – Further work.

4.5.3.2. iHelp

Like in the previous study, two datasets were used: DS-all including all sequences and DS-600 including only sequences of exactly 10 minutes. Both datasets included the new attributes and the two patterns of disengagement: "disengaged-long" and "disengaged-fast". To distinguish these datasets from the ones used in the patterns of disengagement study, "dl/df/e" was added to the names of the datasets.

From dataset DS-all, 11 exploratory sequences were excluded, while from DS-600, only 3 such sequences were eliminated. This indicates that in 8 cases out of 11 the learners spent less than 10 minutes on their first login to the system.

The results are displayed in Table 4.24. Comparing them with the ones from the patterns of disengagement study (see Table 4.22, Section 4.5.2.1) and focusing on percentage correct, we observe the following:

- for DL-all: in five cases the values are higher when the exploratory sequences are included and in three cases the values are higher when the exploratory sequences are excluded – see Figure 4.33;
- for DS-600: for all eight methods the values are higher when the exploratory sequences are excluded – see Figure 4.34.

Table 4.24 iHelp: Prediction results without the exploratory sequences

		BN	LR	SL	IBk	ASC	B	CvR	DT
DS-all (dl/df/e)	**%correct**	88.48	91.48	91.46	88.79	89.18	90.16	90.53	89.57
	TP rate DL	0.92	0.93	0.92	0.92	0.92	0.92	0.92	0.92
	FP rate DL	0.01	0.02	0.02	0.04	0.02	0.02	0.02	0.02
	d'	3.73	3.53	3.46	3.16	3.46	3.46	3.46	3.46
	TP for DF	0.71	0.81	0.81	0.70	0.71	0.75	0.76	0.76
	FP for DF	0.05	0.04	0.04	0.04	0.04	0.04	0.04	0.04
	d'	2.20	2.63	2.63	2.28	2.30	2.43	2.46	2.46
	Error	0.08	0.07	0.07	0.06	0.09	0.08	0.07	0.08
DS-600 (dl/df/e)	**%correct**	93.53	94.98	94.63	94.47	94.06	94.27	94.66	94.33
	TP rate DL	0.93	0.94	0.94	0.94	0.94	0.94	0.94	0.94
	FP rate DL	0.02	0.02	0.02	0.03	0.02	0.02	002	0.02
	Error	0.07	0.06	0.07	0.05	0.07	0.07	0.06	0.06
	d'	3.53	3.61	3.61	3.44	3.61	3.61	3.61	3.61

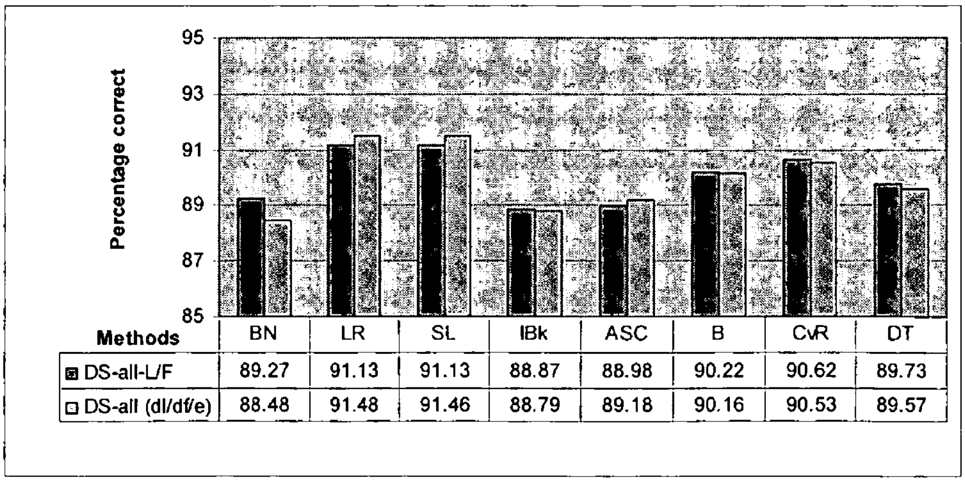
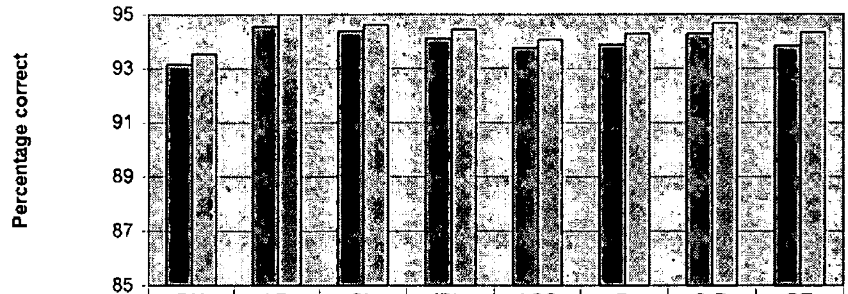


Figure 4.33 Percentage correct comparison between DS-all-L/F and DS-all (dl/df/e).

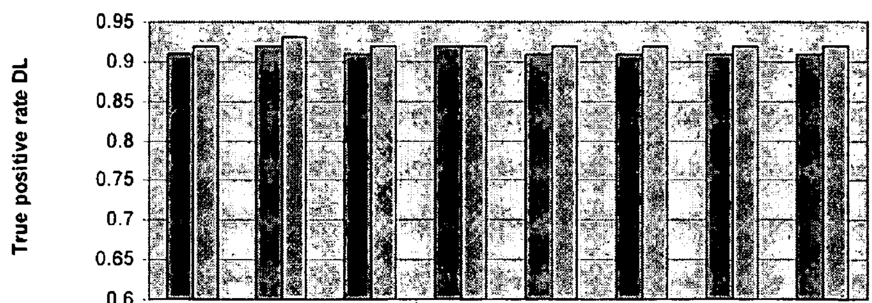


Methods	BN	LR	SL	IBk	ASC	B	CvR	DT
DS-600-L/F	93.14	94.58	94.4	94.13	93.76	93.9	94.28	93.81
DS-600 (dl/df/e)	93.53	94.98	94.63	94.47	94.06	94.27	94.66	94.33

Figure 4.34 Percentage correct comparison between DS-600-L/F and DS-600 (dl/df/e).

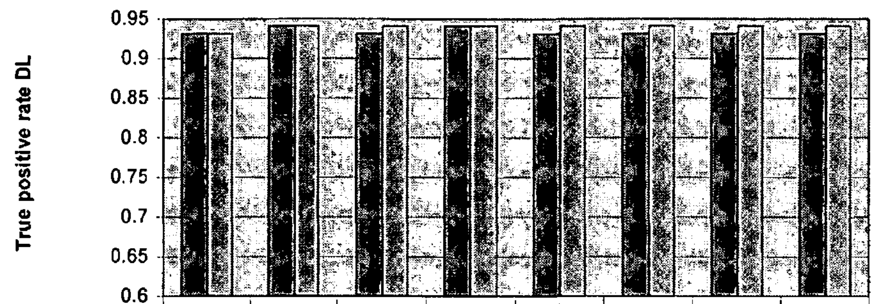
Comparing the results from Table 4.24 with the ones from the patterns of disengagement study (see Table 4.22, Section 4.5.2.1) and focusing on true positive rate for DL, the following can be observed:

- for DL-all: in seven cases the values are higher when the exploratory sequences are excluded and in one case the values are the same – see Figure 4.35;
- for DS-600: for five methods the values are higher when the exploratory sequences are excluded and for the other three the values are the same – see Figure 4.36.



Methods	BN	LR	SL	IBk	ASC	B	CvR	DT
DS-all-L/F	0.91	0.92	0.91	0.92	0.91	0.91	0.91	0.91
DS-all (dl/df/e)	0.92	0.93	0.92	0.92	0.92	0.92	0.92	0.92

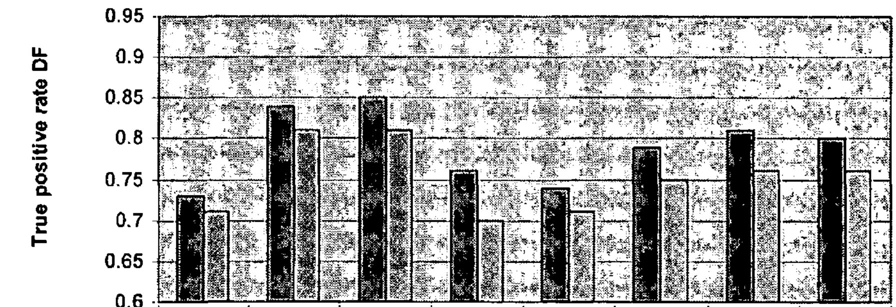
Figure 4.35 True positive rate for DL comparison between DS-all-L/F and DS-all (dl/df/e).



Methods	BN	LR	SL	IBk	ASC	B	CVR	DT
DS-600-L/F	0.93	0.94	0.93	0.94	0.93	0.93	0.93	0.93
DS-600 (dl/df/e)	0.93	0.94	0.94	0.94	0.94	0.94	0.94	0.94

Figure 4.36 True positive rate for DL comparison between DS-600-L/F and DS-600 (dl/df/e).

Comparing the results from Table 4.24 with the ones from the patterns of disengagement study (see Table 4.22, Section 4.5.2.1) and focusing on true positive rate for DF (only dataset DS-all), a decrease is observed for all methods when the exploratory sequences are excluded – see Figure 4.37.



Methods	BN	LR	SL	IBk	ASC	B	CVR	DT
DS-all-L/F	0.73	0.84	0.85	0.76	0.74	0.79	0.81	0.8
DS-all (dl/df/e)	0.71	0.81	0.81	0.7	0.71	0.75	0.76	0.76

Figure 4.37 True positive rate for DF comparison between DS-all-L/F and DS-all (dl/df/e).

Summarizing the results from this study, we observe an increase for percentage correct and true positive rate for DL and a decrease for true positive rate for DF when the exploratory sequences are excluded. Considering that from the 11 exploratory sequences that were eliminated 10 were DF, the fact that the already small number of DF sequences was reduced even more may explain the decrease in the true positive rate for DF. On the other hand, the elimination of these sequences brought an increase of the overall predictive values and of the one for DL and thus suggests that the exclusion of the exploratory sequences is overall more beneficial.

4.6. Discussion

The studies presented in this chapter defined, validated and refined a general approach for disengagement prediction for e-Learning systems. This approach is simple and the information needed is related to actions that take place in most learning environments: reading pages and taking tests and thus, can be generalized to other systems, the validation study being an example for that.

The refinement studies showed that the approach can be improved by using two attributes that are system-specific; they would need to be changed for each particular system, but the benefit of using these attributes would considerably increase prediction and also decrease processing time.

One of the refinement studies showed that two patterns of disengagement can be distinguished: "disengaged-long" and "disengaged-fast", information that is valuable for personalized and appropriate intervention. Comparing these patterns to the ones reported by related research we have found a correspondence for "disengaged-fast pattern", as mentioned in Section 4.4, but we did not find any correspondence for the "disengaged-long" pattern. Looking at the related approaches, we have seen that they are using a rather technical domain, math, and that the systems includes only test-type activities; in a system where only test-type activities are available, it is more likely to have a "disengaged-fast" pattern and in a system were learning activities are present, it is more likely to have a "disengaged-long" pattern, as the interaction with the system for the two types of activities are considerably different. Thus, our approach extends the previous ones by:

1) including learning-type activities, as well as testing-type activities and thus, identifying a new disengagement pattern associated with them: "disengagement-long";
2) using a domain that is at the junction between technical and non-technical domains: HTML; it is known that technical domains such as math or programming are more "controllable" compared to non-technical domains; the fact that our approach is using HTML increases the chance of generalizing it to other domains, including the non-technical ones.

The last study reported for the refinement of the prediction considered the particular case on what we called exploratory sequences. These are characterized by an exploratory behaviour and usually occur at the first login of a learner to the system. Given the fact that this behaviour is quite different and seems somehow "chaotic" compared to the one observed in the subsequent actions, we have explored their influence on the prediction values. We expected that their exclusion would improve prediction. While we were able to demonstrate this effect for most cases, we observed a decrease in the prediction of the "disengaged-fast" pattern. This might be explained by the small number of DF instances and the fact that it dropped even more by the exclusion of the exploratory sequences. Looking at the overall results for both HTML-Tutor and iHelp, the exclusion of these sequences has more benefits than drawbacks.

Thus, we proposed a simple approach for disengagement prediction that extends previous approaches by including learning-type activities besides the test-type ones; this approach gives very good results using attributes from only two actions: reading pages and taking tests, and can distinguish between two patterns of disengagement: "disengaged-long" and "disengaged-fast". Its simplicity and the characteristic of the chosen domain, HMTL, make it easier to generalize across systems and domains.

Chapter 5

Assessment of motivation

After detecting disengagement a more detailed assessment would be required for a personalized and efficient intervention. The dialog would start with an "introduction" explaining the interruption (disengagement detection), presenting the purpose of this dialog (a more detailed assessment for personalized intervention) and asking the user if they agree to answer some questions in order to obtain their motivational profile.

For the actual assessment we started from validated questionnaires for the concepts we wanted to measure and in the case of no available validated instruments, we created items. The "Patterns of Adaptive Learning Scales (PALS)" questionnaire (Midgley et al., 2000) was used for self-efficacy and goal orientation, IQ Learn (IQ-Research group, 2001) was used for self-regulation and items were created for perceived task difficulty, attribution (including locus of control and stability/ instability dimensions) and disengaged category of goal-orientation.

PALS (Midgley et al., 2000) was developed at University of Michigan for assessing the relation between the learning environment and the students' motivation, affect and behaviour. For the student scales it includes: 1) personal achievement goal orientations; 2) perceptions of teacher's goals; 3) perceptions of the goal structures in the classroom; 4) achievement-related beliefs, attitudes, and strategies; and 5) perceptions of parents and home life. Goal-orientation falls into the first category while academic self-efficacy falls into the fourth category. Besides student scales, PALS included scales for teachers.

The first version of PALS was developed in 1997 and refined in 2000. We used the items from the refined version.

IQ Learn is part of the IQ FORM project coordinated by the University of Helsinki for the Finnish Virtual University. The IQ Learn tool was designed to increase students' self-knowledge. The tool includes some tests grouped in three categories: 1) forethought of learning, 2) strategies for learning, and 3) learning skills. From these three categories that correspond to the process of self-regulation (Zimmerman, 2000) we focused on the strategies for learning. They included four aspects: i) time management, ii) self management, iii) persistency, and iv) help-seeking strategies. The items from the first three sub-categories were used; help-seeking strategies were excluded because they are not related directly to motivation. They could be valuable in terms of personalized intervention, more specifically, by considering the students preference for one of the two strategies: coping on their own or cooperating/ discussing with other students. Even if there would be a benefit from knowing this preference, including items related to help-seeking strategies would mean to diverge from the purpose of the assessment and would make it longer; thus, we decided not to include it.

Three items were created for disengaged goal orientation category; one was meant to express the idea of disengagement on its own and the other two were meant to express the same idea by opposition to the other categories of goal orientation: mastery and performance.

Created items for perceived task difficulty included two questions about the difficulty of the course content and the difficulty of the assignments.

Attribution questions were created on the structure of Weiner's theory (1974) because it was the most timesaving measurement from which we could infer four values: internal locus of control (effort and ability), external locus of control (difficulty and luck) and stability (difficulty and ability) and instability (effort and luck). We used a 6-point scale for these items in order to enforce a distinctive selection. In the case of equal values for external and internal locus of control or stability and instability, a possible solution is to ask an additional question that would shift the balance towards one of the two.

The created items were constructed based on theory, which potentially contributes to their validity. However, in order to establish their validity, further investigations are necessary.

Appendix A.1 presents the aggregated instrument with some modified items from the original scales to fit online learning (PALS was developed for classroom learning).

As already mentioned one aspect to be considered for the assessment is its length. Most validated questionnaires have considerable number of items, but they also need a lot of time to be completed. In order to have a short, but valid and reliable assessment, a subset of items was selected from the validated questionnaires and in the case of created items minimum numbers were considered. In the next section the creation of the final instrument is described.

5.1. Instrument construction

For the selection of items we used expert-ranking. Four experts were asked to select a number of items for each concept and then to rank them. Table 5.1 presents for each concept the number of items from the original questionnaires and the number of items requested for selection corresponding to the number of items included in the final questionnaire.

Table 5.1 Number of items in the original and final questionnaires per concept

Concept	No items original questionnaire	No items final questionnaire
Self-efficacy	5	3
Self-regulation		
Time-management	4	2
Self-management	4	2
Persistency	4	2
Goal orientation		
Mastery	5	3
Performance approach	5	3
Performance avoidance	4	2

The four experts selected the requested number of items from the original questionnaires and ranked them. By combining frequency of selection with the mean rank values for each item an integrated ranking was derived. The items were selected in the descending order of integrated ranking. Three experts were asked to comment on the created items and some changes were made accordingly. The final questionnaire is presented in Appendix A.2.

With this final instrument two experiments were conducted to verify its reliability and to see if similar results are obtained when using it compared to using the full instrument (concurrent validity).

5.2. Study 1

This study was conducted in order to investigate the validity and reliability of the constructed instrument. The design, including the objectives, the participants, the task and the procedure, and the results are presented.

5.2.1. Study design

Objectives. The objective of this study was twofold: 1) to see if the results obtained with the two instruments (the long and the shot versions) are similar (not significantly different) and 2) to investigate the reliability of the short version questionnaire.

Participants. 20 first-year students participated in this study. They were offered 5% for their continuous assessment mark for their Applications Software module as an incentive.

Task. The participants were asked to complete the two questionnaires. Both of them were completed online.

Design. A correlational study design was chosen. The initial plan was to have a break of a week between the two questionnaires. However, due to unexpected variance in class attendance and associated delays in completing the two parts, changes in the design were necessary.

Procedure. In the first week of the experiment (3 weeks before the end of semester) the students were informed that they would receive 5% for their continuous assessment mark if they complete two questionnaires – one during the current week and the other one during the following week. The initial plan was to have the first questionnaire (the short version) online and available for one week only and then to post online the second questionnaire (the original version). By the end of the first week only 8 students completed the first questionnaire and at this point we decided to leave it online along with the second questionnaire in the hope that more students would complete it. By the end of the semester, 20 students completed the two questionnaires, of which 11 completed both versions on the same day (Group 1), 8 completed them with some time in between (Group 2) and 1 completed the long questionnaire first and the short version after that. The data from this participant were excluded from the analysis because results have shown an order effect, as presented below.

5.2.2. Results

The data were analyzed in three ways: First, the two groups were compared to see if the comparison between the questionnaires should be done separately or as a homogeneously group. Second, the measurements with the two questionnaires were compared and third, reliability of the selected and the created items was investigated.

5.2.2.1. Comparison between Group 1 and Group 2

Because the time between the two questionnaires varied for the two groups, different results could be due to the different time lag. In the case of significant differences between the two groups, the comparison between the two versions of the questionnaire should be done per group; otherwise, the analysis should be conducted on data from all participants.

In order to decide what type of test (parametric or non-parametric) to use, we verified if the distributions for the dependent variables per group are significantly different from normal distributions. Kolmogorov-Smirnov test was used for this. The results can be seen in Appendix B.1. (B.1.1 for Group 1 and B.1.2 for Group 2) and they show that all variables have a Normal distribution. Thus, parametric tests can be used, more specifically, independent t-test. A summary of the results for the short questionnaire only is presented in Table 5.2. Complete results as delivered by SPSS can be found in Appendix B.2.1.

Table 5.2 Independent t-test results for the comparison between Group 1 (same day) and Group 2 (time in between)

	Time/ Group	N	Mean	t	df	p
Self Efficacy 1	Same day	11	4.15	-0.162	17	.873
	Time in between	8	4.21			
Time Management 1	Same day	11	2.50	1.312	17	.207
	Time in between	8	1.88			
Self Management 1	Same day	11	3.23	-0.157	17	.877
	Time in between	8	3.31			
Persistency 1	same day	11	2.86	0.274	17	.787
	time in between	8	2.75			
Self Regulation 1	same day	11	2.86	0.593	17	.561
	time in between	8	2.65			
Mastery GO 1	same day	11	4.24	-0.307	17	.763
	time in between	8	4.33			
Performance Approach GO 1	same day	11	1.55	-2.211	8.64	.056
	time in between	8	2.71			
Performance Avoidance GO 1	same day	11	1.86	-2.015	17	.060
	time in between	8	3.00			
Disengaged GO 1	same day	11	3.18	0.027	17	.979
	time in between	8	3.17			
Perceived Task Difficulty 1	same day	11	2.91	-0.209	17	.837
	time in between	8	3.00			
Internal Locus of Control 1	same day	11	4.18	-0.404	17	.691
	time in between	8	4.38			
External Locus of Control 1	same day	11	3.09	-0.883	17	.389
	time in between	8	3.44			
Stability 1	same day	11	3.77	-0.545	17	.593
	time in between	8	4.00			
Instability 1	same day	11	3.50	-0.705	17	.490
	time in between	8	3.81			

The result for Performance Approach GO1 from Table 5.2 differs from the other ones in the degrees of freedom (df) value because the two groups, i.e. *same day* and *time in between*, have different variances (indicated by the low significant value for Levene test – see Appendix B.2.1) and t is calculated based on a different formula.

The results show no significant differences between the two groups, which indicates that we should continue the analysis using the data from all participants regardless of the group they belong to. From all variables we notice that for two of them the differences are closer to the significance threshold compared to the other variables: Performance Approach GO1 ($p = 0.056$) and Performance Avoidance GO1 ($p = 0.060$). For these two an analysis per group as well as per total might disclose different results.

5.2.2.2. Comparison between the two questionnaires

Given that there are no significant differences between the two groups, as shown above, the following analyses treat the complete sample as a coherent group. The combined dataset was checked for normality of the distribution. The results can be found in Annex B.1.3. They show that all variables are normally distributed. Thus, parametric tests are to be used, in this case, paired t-test. A summary of results are presented in Table 5.3 and full results can be found in Appendix B.2.2.

The results show no significant difference for all variables except for self-efficacy; the results for time management are also close to being significant. The means for self-efficacy show that self-efficacy is significantly higher when measured with the short questionnaire compared to the long questionnaire. Looking at the data of the student that answered the long questionnaire first we noticed that self-efficacy is higher for this questionnaire, which indicates that order may have an influence; on this basis the data from this participant was excluded.

Performing the same analysis per group for the two variables that had almost significant differences for the group comparison (Performance Approach and Performance Avoidance), same type of results as for all participants were obtained, meaning no significant differences.

Table 5.3 Paired t-test results for the comparison between the two questionnaires

		Mean	t	df	p
Pair 1	Self Efficacy 1	4.17	4.616	19	.000
	Self Efficacy 2	3.60			
Pair 2	Time Management 1	2.28	-2.030	19	.057
	Time Management 2	2.63			
Pair 3	Self Management 1	3.23	-0.896	19	.382
	Self Management 2	3.43			
Pair 4	Persistency 1	2.80	-0.163	19	.873
	Persistency 2	2.83			
Pair 5	Self Regulation 1	2.77	-1.613	19	.123
	Self Regulation 2	2.96			
Pair 6	Mastery GO 1	4.22	1.093	19	.288
	Mastery GO 2	4.03			
Pair 7	Performance Approach GO 1	2.08	-0.127	19	.900
	Performance Approach GO 2	2.11			
Pair 8	Performance Avoidance GO 1	2.38	0.159	19	.875
	Performance Avoidance GO 2	2.35			
Pair 9	Disengaged GO 1	3.18	-1.410	19	.175
	Disengaged GO 2	3.38			
Pair 10	Perceived Task Difficulty 1	2.95	t cannot be computed because the standard error of the difference is 0.		
	Perceived Task Difficulty 2	2.95			
Pair 11	Internal Locus of Control 1	4.25	1.112	19	.280
	Internal Locus of Control 2	3.98			
Pair 12	External Locus of Control 1	3.28	-0.237	19	.815
	External Locus of Control 2	3.33			
Pair 13	Stability 1	3.90	0.326	19	.748
	Stability 2	3.83			
Pair 14	Instability 1	3.63	0.578	19	.570
	Instability 2	3.48			

The analysis presented above shows that if measured with the short questionnaire, the variables of interest would have the same results as if measured with the long questionnaire for all variables except self-efficacy. There are several possible explanations for the difference between self-efficacy values measured with the two instruments. One possible explanation is the influence of the first questionnaire on the answers given to the second one. A second possible explanation could be related to the position of self-efficacy items at the beginning of the instrument. Another possible explanation would be group specificity.

Before presenting how we addressed this problem, a few comments on item order are presented. There are two general approaches for questionnaire items order (Dillman, 2000; Tourangeau et al., 2000): 1) presentation of items in random order and 2) presentation of items by topic. We chose the second approach because respondents tend to give more meaningful answers when questions flow in a logical order with items grouped together by topic, while question in random order are likely to frustrate respondents.

Regardless of the order case, the first items of the questionnaire should draw the respondents' interest, be representative for the main topic and engage them with the questionnaire. That is why self-efficacy items were placed first: they are likely to raise interest because they ask for a personal general evaluation of "competency"; self-efficacy is a central concept in the motivational framework; and on the same grounds of personal evaluation they are likely to engage the respondent.

In order to investigate the possible causes of the observed difference between self-efficacy values obtained with the two questionnaires, we planned a second study starting from the possible influence of the order of items in the questionnaire because this would need the most complex design compared with the others and because the other two could be somehow included in this complex design. Thus, if the results for self-efficacy vary between the two questionnaires without any influence of the place of self-efficacy items in the first questionnaire, then the first possible explanation mentioned would apply: the results of self-efficacy for the second questionnaire are influenced by the first one. If there are no differences between self-efficacy measurements using the two questionnaires and different positions for the self-efficacy items in the short questionnaire, then the third possible explanation mentioned previously would apply: the differences obtained for self-efficacy would be due to group characteristic. Thus, the second study was designed to investigate if the position of items in the structure of the questionnaire would cause different results.

5.2.2.3. Reliability measures

In order to investigate the reliability of the measurement using the short questionnaire we calculated Alpha Cronbach for the following variables: self-efficacy, self-regulation (and subscales), goal orientation types and perceived task difficulty, and correlation for items related to attribution: difficulty, effort, luck and ability.

The values of Alpha Cronbach for the mentioned variables are displayed in Table 5.4. Although not very high, the values for alpha Cronbach are good, with the exception of Persistency. For the created scales, i.e. disengaged goal orientation and perceived task difficulty, the values are satisfactory given the small number of participants.

Table 5.4 Study 1: Alpha Cronbach values

Scale	Alpha Cronbach
Self-efficacy	.743
Time management	.759
Self-management	.865
Persistency	.321
Self-regulation	.748
Mastery GO	.429
Performance Approach GO	.757
Performance Avoidance GO	.760
Disengaged	.807
Perceived Task Difficulty	.746

For the items related to attribution we used Pearson correlation as a measure of reliability (consistency and stability). The results are displayed in Table 5.5.

Table 5.5 Study 1: Reliability for attribution items.

	r	p
Difficulty	.665	.001
Effort	.290	.229
Luck	.680	.002
Ability	.536	.022

Only three correlations out of four are significant, and for those, the strength of correlations is medium. Thus, further investigation using more subjects is necessary to conclude something certain about the reliability of these items.

5.3. Study 2

This study was conducted to investigate the existence of an order effect for the self-efficacy items and also to expand the inquiry about reliability. The design of the study, including objectives, participants, task and procedure, as well as results are presented.

5.3.1. Study design

Objectives. The main objective of this study was to investigate if the order effect is responsible for the significant difference obtained for self-efficacy in the previous study. Another objective was to see if a larger number of subjects would improve the reliability for the variables related to attribution especially, but also for the other variables.

Participants. 30 participants took part in the experiment and only 22 completed both questionnaires entirely. The experiment was advertised on the psychology online experiments web page of the University of Zurich (Reips, 2002). To motivate participation a full report of the results and their interpretation was offered at the end of the experiment. At the beginning participants could see a sample report. Two screenshots of this report are displayed in Figure 5.1. The number of participants per each group in the experiment is presented in the design section below.

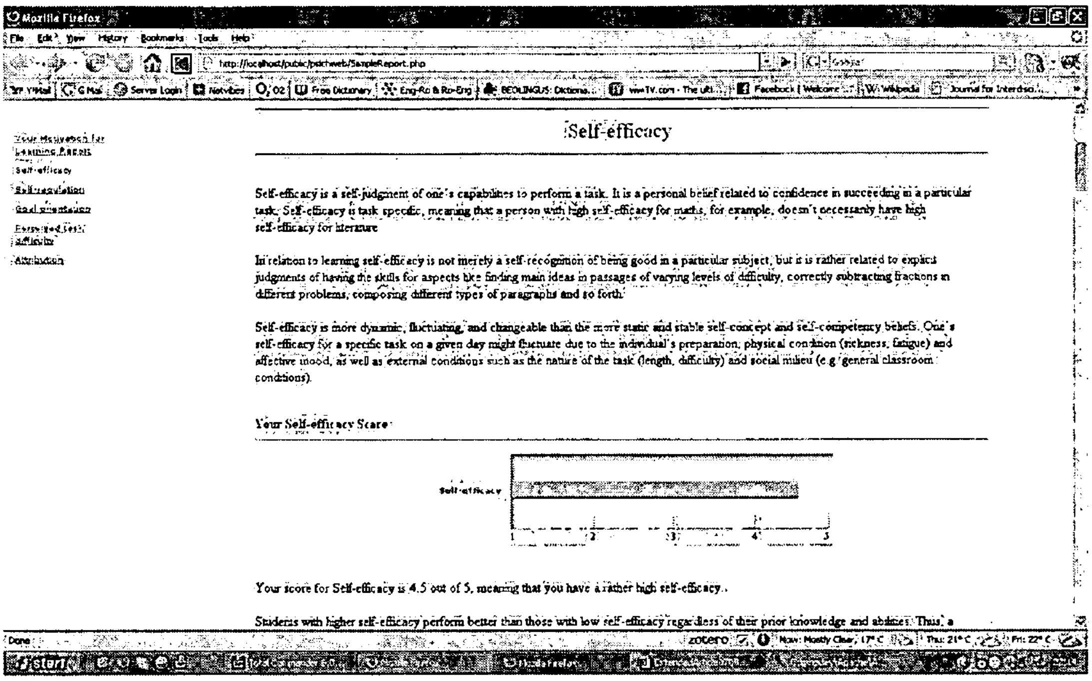
Task. The participants were asked to complete the two questionnaires.

Design. The design includes three groups depending on the position of self-efficacy items in the short questionnaire: beginning and end, middle, and end. This is illustrated in Table 5.6, which includes also the umber of participants in each group.

Table 5.6 Study 2: experimental design.

	Position of self-efficacy items in the short questionnaire		
	Beginning and end	**Middle**	**End**
No of participants	7	8	7

a)



b)

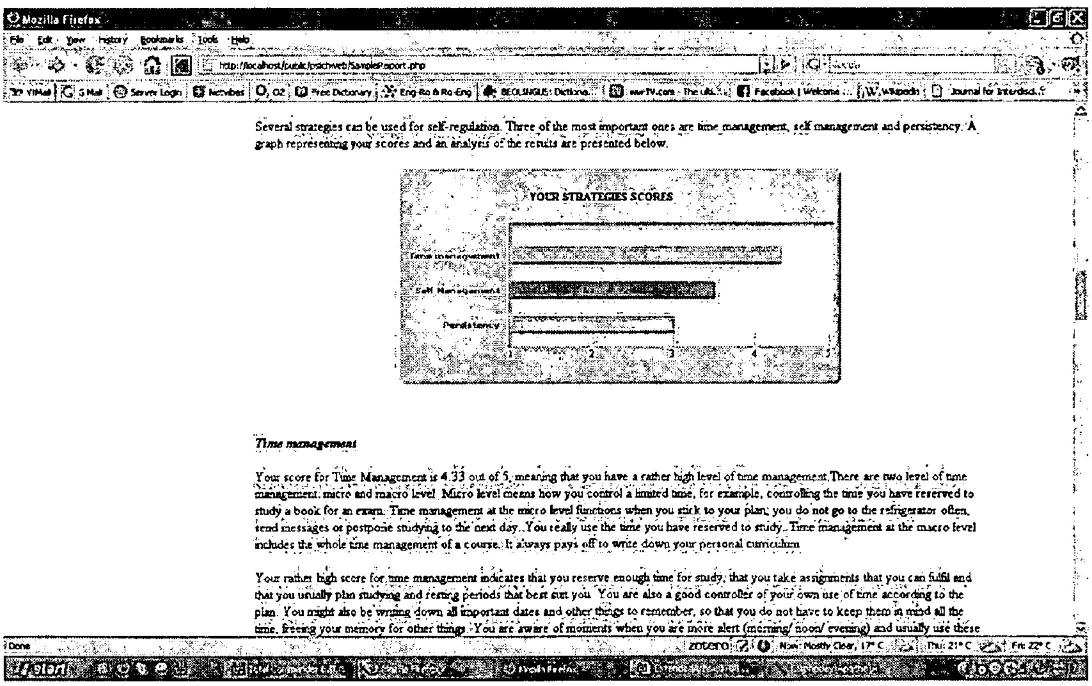


Figure 5.1 Screenshots of the sample report: a) self-efficacy; b) self-regulation strategies.

Procedure. The experiment took place from the beginning of June 2007 to mid August. The participants found out about the experiment either from the Web Experimental

Psychology Lab page or from emails sent to invite people to participate. The experiment started with a presentation of the study, mentioning the two questionnaires to be completed and the report to be received on completion of the study. For the first questionnaire one question per page (except for the attribution items) was presented, while for the second questionnaire questions were presented as a topic group per page. Figure 5.2 presents two screenshots from the first and the second questionnaire.

a)

b)

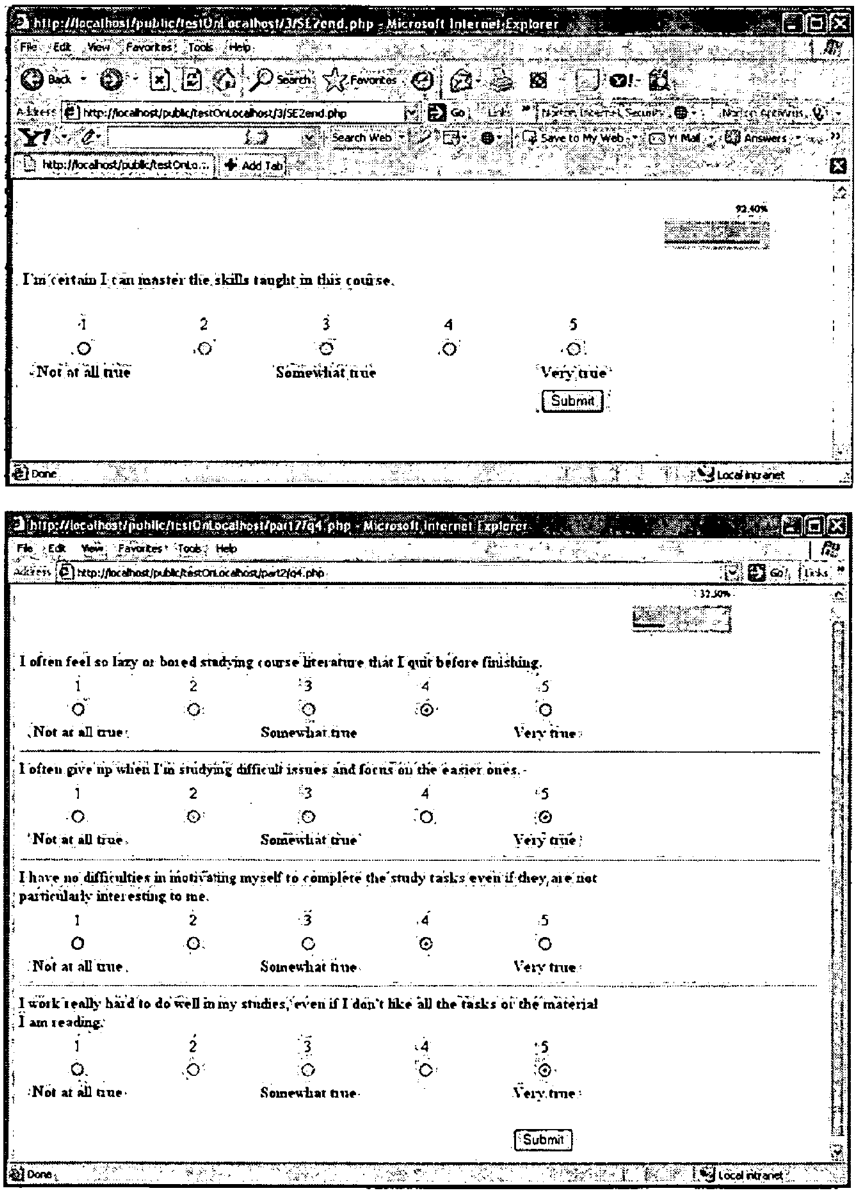


Figure 5. Screen shots: a) first questionnaire; b) second questionnaire.

5.3.2. Results

Results are grouped in two categories: 1) effect of order on self-efficacy and 2) reliability measures.

5.3.2.1. Effect of order on self-efficacy

As in the previous study, we verified the normality of the distributions in order to establish the type of tests to be used. Results of Kolmogorov-Smirnov test are displayed in Appendix C.1. (C.1.1. for group "beginning and end", C.1.2. for group "middle" and C.1.3. for group "end"). They show that all variables have a normal distribution. Thus, to investigate the effect of order on self-efficacy, paired t-test was used on all two-pair groups. A summary of results is presented in Table 5.7 and full results can be found in Appendix C.2.

We also looked at differences between the four groups (the three groups from second study plus the group from the first study) for self-efficacy for each of the two questionnaires, using one-way ANOVA and Turkey as post-hoc test. A summary of results is presented in Table 5.8 and full results are listed in Appendix C.3.

Table 5.7 Paired t-test for the three groups corresponding to the position of self-efficacy items in the short questionnaire.

Group		Mean	t	df	p
Beginning and end	Self efficacy 1	4.50	-1.911	7	.098
	Self efficacy 2	4.70			
Middle	Self efficacy 1	4.21	-1.262	7	.248
	Self efficacy 2	4.35			
End	Self efficacy 1	4.58	1.387	7	.208
	Self efficacy 2	4.50			

Table 5.8 One-way ANOVA results

	F	p
Self-efficacy 1	.837	.482
Self-efficacy 2	6.964	.001

The table shows no significant difference between the four groups for self-efficacy measured with the short questionnaire and a significant difference for the self-efficacy measured with the long questionnaire. Turkey test – see Table 5.9 and Appendix C.2 – shows that the only significant differences are between the "beginning" group (the one from the first study) and each of the other three.

Table 5.9 Turkey test for multiple comparison – extract

Dependent Variable	(I) group	(J) group	Mean Difference (I-J)	Std. Error	Sig.
Self efficacy 2	beginning	beginning and end	-1.13	0.292	.002
		Middle	-0.78	0.292	.051
		End	-0.93	0.292	.014

These results are also reflected in the homogeneous groups defined by Turkey procedure. Thus, for self-efficacy measured by the long questionnaire two homogeneous groups can be distinguished: one corresponding to the participants from the first study and one corresponding to the three groups from the second study.

The lack of significant differences between the three groups corresponding to different positions of self-efficacy items within the short questionnaire and the significant difference between each of the three groups and the group from the first study indicates that group characteristic is the most probable cause of the differences in self-efficacy values observed in the first study.

5.3.2.2. Reliability measures

In order to analyze the reliability of the scales, the same measures as in the previous study were used. The values of Alpha Cronbach are displayed in Table 5.10.

Same as is the previous study, the values are good overall, with one exception: Persistency, which has a very low value indicating that there is definitely a problem. For the created items, i.e. disengaged goal orientation and perceived task difficulty, the values are above .75, which is satisfactory given the small scale of the study. Looking at the values for alpha Cronbach obtained from the participants of both studies, Persistency is the one that stands out as the lowest value.

The reliability indicators for the items related to attribution are displayed as correlation values in Table 5.11.

Table 5.10 Study 2: Alpha Cronbach values

	Alpha Cronbach (participants to second study)	Alpha Cronbach (all participants)
Self-efficacy	.791	.769
Time management	.709	.733
Self-management	.524	.695
Persistency	.085	.346
Self-regulation	.553	.621
Mastery GO	.448	.429
Performance Approach GO	.887	.856
Performance Avoidance GO	.634	.705
Disengaged	.770	.839
Perceived Task Difficulty	.803	.778

Table 5.11 Study 2: Reliability for attribution items.

	Participants from study 2 only		Participants in both studies	
	r	p	r	p
Difficulty	.814	.000	.668	.000
Effort	.785	.000	.455	.003
Luck	.904	.000	.791	.000
Ability	.760	.000	.612	.000

Compared to the previous study, the values indicate higher reliability; for all items the correlations are significant at .01 level and the strengths are medium to high. If all participants are considered, all correlations remain significant at .01 level, but the strengths are all medium.

5.4. Discussion

We argue that the results from the two studies indicate that the short version of the questionnaire can be used for assessing motivation. The first study has shown no differences between the two questionnaires, except for self-efficacy; the reliability values were rather low for one of the concepts (persistency) and one of the items related to attribution (effort). The second study has shown that the differences in self-efficacy observed in the first study are due to group characteristic and that reliability is generally satisfactory with one exception (persistency).

Content and construct validity of the items from standard questionnaire are assured by the fact that they were extracted from validated instruments. The fact that there are no significant differences between the short questionnaire and the long one (an established instrument), as observed in the second study indicates that the short instrument is valid (concurrent validity). With regard to the reliability of the short instrument, alpha Cronbach as an indicator of internal consistency showed satisfactory results with only one exception: persistency.

For the created questions content and construct validity are ensured by the fact that items were constructed from theory with simple and straightforward operationalisation. The stability and consistency (both related to reliability) of these items are given by the fact that there is no difference between the repeated measurements performed with the two questionnaires (same items in both instruments). The internal consistency was measured only for some of the created items, i.e. disengaged category of goal-orientation and perceived task difficulty, and the values are fairly good.

To conclude, the validity and reliability of the instrument is satisfactory and thus, the instrument can be used for motivational assessment; only the results for Persistency should be considered with care due to the low reliability of this sub-scale.

Chapter 6

Integration of Assessment Components

The two steps of our research project are meant to provide a complete motivation diagnosis and thus, deliver a detailed motivational profile of the learner. The engagement monitoring component would keep track of the learner's engagement and the system would initiate the dialog only when disengagement is detected. The information from the two components (engagement monitoring component and dialog component) would feed into the learner model which is the basis for the adaptation component. This process is displayed in Figure 6.1.

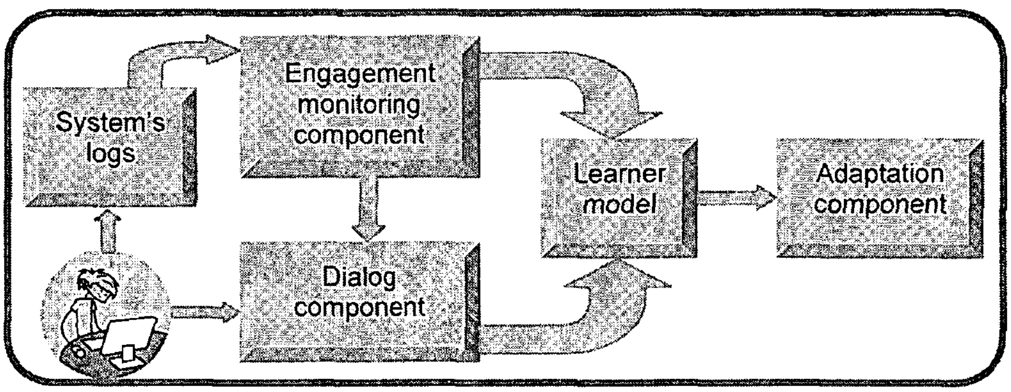


Figure 6.1 Integration of engagement monitoring component and dialog component.

6.1. Engagement monitoring component

Besides identifying the disengaged learners, the engagement component monitoring could be used for keeping track of the evolution of engagement levels of the learner, with starting times and duration of periods. This information could help with detection of possible personal patterns (e.g. long periods of engagement followed by short periods of disengagement, alternating periods of engagement and disengagement of similar length, etc.) that could be valuable to tutors.

Also, as one study presented in Chapter 4 Section 4.5.2 showed, two patterns of disengagement can be distinguished:

1) "disengaged long" characterized by long periods of time spent on one or very few pages;
2) "disengaged fast" characterized by fast browsing through pages.

These two patterns can be distinguished at good levels and thus, this information could be taken into consideration when choosing an interventional strategy. For example, strategies targeting making the learner move to a different topic would be more appropriate for "disengaged long" learners than for the "disengaged short" users that would benefit more from strategies targeting keeping the learner's focus longer on the learning material. This distinction could be useful for excluding some strategies from the list of all possibilities. Thus, following the previous example, strategies targeting keeping the learner's focus longer on the learning material would be an inappropriate interventional strategy for "disengaged long".

One aspect to consider in a real system about the engagement monitoring component is the prediction method to be used. Several methods were used with the two systems we investigated, i.e. HTML-Tutor and iHelp, and the results showed that different methods provide best predictions: classification via regression for HTML-Tutor and simple logistic classification for iHelp. Among the several alternatives for deciding the method(s) to be used, two seem more advantageous in terms of time and space complexity:

a) an initial test followed by the choice of the best three performing ones and usage of the average;
b) an initial test of all methods, followed by the choice of the best performing one.

Another aspect to be considered is the time frame for engagement/ disengagement detection. In our studies we used sequences of 10 minutes resulted from splitting the sessions. In a real system, a sliding-window approach would be more useful and thus, an update at every minute, for example, would contribute to an early detection of disengagement.

6.2. Dialog component

The aim of the dialog component is a very difficult one: to engage a disengaged learner in a dialog with the system in order to measure several motivational characteristics to be used for intervention. The dialog would start by informing the learner that disengagement was detected and about the purpose and benefits of the dialog, followed by questions meant to verify whether the learner agrees with the system's diagnosis of disengagement or not, and to check his/ her willingness to answer questions about his/ her motivation for learning.

In our research we did not address issues like the design of the dialog, formulation of the phrases in the dialog, and how it would be perceived by learners; we focused only on the measurement and its quality. In Chapter 5 we presented the instrument to be used for this measurement and two studies that investigated reliability and validity issues. The results showed that the instrument is valid and reliable and that only one subscale has low reliability: persistency.

6.3. Learner model

The learner model would include information from the engagement monitoring component and the dialog component. The engagement component monitoring would provide information on aspects like:

a) disengagement periods with starting times and duration;
b) corresponding pattern, i.e. "disengaged long" or "disengaged fast", for each period of disengagement.

The dialog component would provide values for the following motivational characteristics:

a) self-efficacy;
b) self-regulation and the three strategies: time management, self management and persistency;
c) goal orientation – one of the following: mastery/ performance approach / performance avoidance / disengaged;
d) perceived task difficulty;
e) attribution, including locus of control and stability/ instability dimensions.

Based on the information described above, decisions about interventional strategies can be made. As mentioned in Chapter 2, one of the reasons for the choice of Social Cognitive Learning Theory as theoretical basis for our research was its numerous possibilities of influencing the learning process.

Updating the Learner Model needs to be addressed in a real system. The information from the engagement monitoring component would be updated automatically, while for the motivational characteristics, several possibilities can be considered:

1) only the learner can update the motivational characteristics either by simply modifying the values or by completing the questionnaire again;

2) only the tutor can update the motivational characteristics of the learner by changing their values;
3) both learner and tutor can update the motivational characteristics, in the manner of open learner models (Bull, 2004), with various possible scenarios:
   a) the tutor can update the values and the learner has to agree with the change before the modification takes place;
   b) the learner can update the values and the tutor has to agree with the change before the modification takes place;
   c) two separate models can be used: one for the learner that can be updated only by him/ her; one for the tutor, that can be updated only by the tutor; the tutor should be able to visualize the model updated by the learner, but the learner should not be able to visualize the model updatable only by the tutor.
4) the update could be done automatically.

6.4. Chapter summary

In this chapter we integrated the results from the two steps of our proposed research approach for motivational diagnosis; we presented how these two steps are related and how they would work in a real system. Several practical issues were discussed: prediction method choice and sliding window approach for disengagement detection and learner model update for assessment of motivation.

Chapter 7

Conclusions and Further Work

This chapter summarizes the research presented in this thesis underlining the main findings of our approach on motivational assessment. Some methodological issues from our research are discussed and future research directions are outlined.

7.1. Summary of Research and Findings

The interest of the research community in motivational issues in e-Learning started to emerge only fairly recently and, as presented in Chapter 2, has resulted in several approaches that seem quite disparate, focusing on different motivational aspects, using different theories and different methodology. Thus, the research on motivation in e-Learning generally addresses narrow issues with applicability restricted to test-type activities and to a specific learning system.

The research presented in this dissertation is similar to previous research in some aspects: we address a narrow issue, i.e. assessment of motivation, and we use a similar methodology, but also aims to extend the applicability to learning-type as well as test-type activities and to e-Learning systems in general. We proposed a two steps approach for a complete motivational diagnosis:

1) disengagement detection from the learner's interactions with the system;
2) assessment of motivation by means of a dialog with the learner.

Motivation assessment has been addressed in previous research using two sources of information: the learner's actions and self-assessment, both used in our approach to benefit from the advantages of each: unobtrusiveness for the former and accuracy for the later.

The first step in our research, disengagement detection, was addressed using similar methodology to related work, as presented in Chapter 3. However there are some differences in our approach given by the targeted activities. As already mentioned, most research on motivation focused on test-type activities, while our research envisaged the learning-type activities as well. Test-type activities are more specific and thus, easier to monitor compared to learning-types activities. We started from previous work and built on it in order to extend the applicability to learning-type activities. Thus, we conducted several studies, described in detail in Chapter 4, to achieve this purpose:

1) A pilot study, where we investigated the possibility of detecting disengagement at a satisfactory level. Data from HTML-Tutor, a web-based interactive environment, was used and disengagement detection proved to be possible and satisfactory. One other finding has proved to be particularly relevant, contributing to a refinement of the approach for the following studies; this finding is related to the "unit of analysis". In this pilot study we used session as unit of analysis and investigated engagement status for the entire session; although prediction proved to be possible and satisfactory, we found out that it is not very useful, because by the time we diagnosed disengagement, the disengaged learner would have already left the system; thus we decide to change the unit of analysis from complete sessions to 10-minute sequences.
2) The "core" study, where our approach was refined and the relevant attributes for prediction were identified. These attributes are related to two events, i.e. reading pages and taking tests, and they are:
    a. Average time spent reading;
    b. Number of pages (accessed/ read);
    c. Number of tests

   d. Average time spent on tests
   e. Number of correctly answered tests
   f. Number of incorrectly answered tests

   The correct percentage values of prediction varied between 86% and 91% across different methods.

3) The validation study, where we used a different system, iHelp, in order to validate the approached refined in the "core" study. Using similar attributes, detection of disengagement was demonstrated to be possible at slightly lower levels compared to HTML-Tutor: 84% to 85% percentage correct.

   In this study two new attributes were introduced:
   a. Number of pages above the threshold established for maximum time required to read a page;
   b. Number of pages below the threshold established for minimum time to read a page;

   Including these two new attributes in the analysis, the prediction performance improves significantly, going up to 98% correct prediction.

4) Three studies were dedicated to the prediction approach refinement:
   a. The validation of reading attributes study, where we investigated the effect of introducing the reading speed attributes on HTML-Tutor. Results showed an increase of prediction values and thus, suggest the inclusion of reading speed attributes in the prediction approach.
   b. The patterns of disengagement study, where we investigated the possibility to predict two pattern of disengagement: "disengaged-long", characterized by long time spent on a page or test and "disengaged-fast", characterized by fast browsing through pages. Results showed that both patterns can be detected at reasonable levels of prediction: around 90% for "disengaged-long" and around 80% for "disengaged-fast".
   c. Exclusion of exploratory sequences study, where we investigated whether the exclusion of exploratory sequences improve prediction. An improvement of

prediction value was observed suggesting that disengagement monitoring should start after the learner has explored the system.

The second step of our approach, assessment of motivation, was addressed using self-reports, as described in detail in Chapter 5. In relation to this step in our research framework, we conducted several investigations:

1) An instrument was constructed partly from validated questionnaires (for self-efficacy, self-regulation and goal-orientation) and partly from created items (for disengaged goal-orientation, perceived task difficulty and attribution). Because our target group was disengaged learners, one of the important aspects considered when building the instrument was its length. Although several other motivational characteristics would have been valuable, such as social goals and perceived characteristics of the task (challenge, controllability, cognitive interest, etc.), we left them out in order to achieve an instrument with a reasonable length. Thus, two or three items per scale were selected from the validated questionnaires and minimum numbers of questions were created: 3 for disengaged goal-orientation, 2 for perceived task difficulty and 4 for attribution (from which four scales are inferred: internal locus of control, external locus of control, stability and instability).
2) Study 1 was conducted in order to investigate the reliability and validity of the constructed instrument. Results from the original scales were compared with the results from the ones selected for the constructed instrument and significant differences were observed only for self efficacy. Reliability values were satisfactory for most of the items from the validated questionnaires and for disengaged goal-orientation and perceived task difficulty, but they were quite low for attribution.
   In order to investigate the possible causes of the significant difference observed for self-efficacy, a new study was conducted. Reliability was also investigated in the new study.
3) The main purpose of Study 2 was to investigate the possible factors that cause the significant difference for self-efficacy. It was proved that the characteristics of the

group that participated in Study 1 were responsible for the observed differences. Reliability was also investigated for this study and for all participants (from Study 1 and Study 2) and satisfactory values were obtained, with one exception: persistency subscale of self-regulation, for which further investigation is required.

We consider that the results presented in this dissertation are a good start for a general approach on motivation diagnosis that can be applied to learning-type as well as test-type activities, and to different e-Learning systems. In the following sections we will consider some criticism of the work presented here and we will also give some research directions for further work that could improve our approach on motivation diagnosis.

7.2. Criticism

In this section we comment on some methodological issues that, in hindsight, could have been done better for both steps of our approach: disengagement detection and assessment through dialog with the learner.

In relation to the first step of our research, disengagement detection, we would like to comment on a particular methodological issue: the distribution of students across training and testing sets. In some of the studies presented in this dissertation (the pilot study and the prediction refinement studies) cross validation was used with no control over the distribution of students, which means that a positive bias could have been introduced by the fact that sequences from the same learner are present in training as well as testing sets. The main studies presented, i.e. the "core" study and the validation study, took this issue into consideration and learners were separated accordingly into training and testing sets.

Another issue related to the first step of our proposed approach is the somehow arbitrary way of establishing the minimum and maximum time required to read a page. The arbitrary way of determining the thresholds was due to low access rate per page across users. However, a more systematic way of calculating these thresholds is possible and it would include: a) a confidence interval calculated from the mean and standard deviation of times spent per page when the page has been accessed by five users at least; b) the use of the arbitrary thresholds when the page has been accessed by less than five users. This approach would not be entirely systematic, but it would contribute to a more precise control over the real time spent by users on pages or tests.

In relation to the second step in our approach, assessment of motivation through dialog, we would like to comment on the following aspects: a) item order; b) small number of participants; c) appropriateness of motivational learner model.

In Study 1, when we observed the difference in self-efficacy and considered one of the possible causes to be item order, we thought that this problem could have been prevented by using random order of items instead on topic-grouped items, which was our choice for the presentation of the questionnaire. Study 2 proved that order was not the cause of the observed differences, but group characteristics. Thus, we argue for the presentation of questions along topic-groups.

In both studies the number of participants was quite small. Thus, we are aware of the limits in the generalisability of results, but we argue that the lack of significant differences between the original questionnaires and our instrument, and the satisfactory values for reliability, are a good enough start for using this instrument and investigate further on its adequacy for motivational assessment.

In relation to the content of the learner model, its appropriateness is derived from the theoretical background. Thus, this model is generic and can be used on any subject matter and any learner. This might apparently decrease its applicability in a real system. In the same time, the concepts used in the motivational model vary between subject matters: different motivational profiles for the same user can be obtained for different domains. Thus, we argue that the chosen learner model provides a good balance between generality and specificity.

7.3. Further work

The research presented in this dissertation focuses on building a general approach for motivational diagnosis in e-Learning environments, but is yet to be tested in a real system and several aspects need further investigation. A list of suggestions is presented below.

1. **Design of the dialog with the learner**

   In the research presented in this thesis we focus on the validity and reliability of the motivational self-assessment, but a very important issue is the design of the dialog in which the assessment is embedded. Interesting aspects to investigate include: the responses of disengaged learners to the initiative of the system; whether they complete the assessment or not; whether they update their motivational profile (if an open learner model is used), etc.

2. **Investigating the complete assessment approach**

   We looked at the two assessment components separately and a study that would use them both would disclose more about how the combination works.

3. **Extend the approach to other domains, categories of users and format of learning materials**

   Our investigation looked at only one domain, i.e. HTML, and it would be interesting how the approach should be adapted to a different domain. As already mentioned in Chapter 4, Section 4.6, an interesting aspect to investigate is the way the approach would apply for both technical and non-technical domains.

   It would also be necessary to study how the approach would be adapted to other learners; we have already seen in the first study for the dialog component that group characteristics can influence results.

   Another aspect worth investigating is how the approach could be adapted to different types of formats used for the learning materials (in our research, only text and images were considered; videos were also available, but rarely used).

4. **Exploratory behaviour**

   We have shown that the exclusion of the first 10 minutes sequence at the very first login, when exploratory behaviour tends to occur, improves prediction of disengagement. However, for some users this behaviour could occur for less than 10 minutes, for more than 10 minutes or at subsequent logins to the system, i.e. at the beginning of each sessions, and thus, further investigation would be required in order to establish from which point the learner is using the system and thus, to begin monitoring his/ her activity.

5. **The content of the motivational learner model**

   Further investigations may be conducted in relation to the content of the motivational learner model. For collaborative learning environments it would be interesting to include in the assessment the social goals (goal orientation) and the help-seeking strategies (self-regulation) and investigate how they affect the interactions with the peers, but also with the learning materials.

6. **Open motivational learner model**

   One research direction already suggested is the use of open learner models. Interesting aspects to investigate are: the impact of having an open leaner model of motivation; whether students like this idea and whether they find it useful; the update of the model (by changing values, by re-assessment, by negotiation with the tutor), etc.

7.4. Conclusions

In this dissertation we addressed one of the most important issues of learning in general and e-Learning in particular: motivation; more specifically we focus on assessment of motivation, a crucial step for systems that would adapt to the motivation of their users.

The contributions of the research presented here are twofold, corresponding to the two steps of our proposed approach for motivational diagnosis: disengagement detection and assessment of motivational characteristics:

1) Detection of disengagement. We have shown that disengagement can be detected using attributes related to basic events like reading ages and taking tests that are common to most e-Learning systems.

   The proposed approach has the advantage of disengagement detection during learning time as opposed to only testing time which is the focus of most of the work on motivation in e-Learning. Thus, our approach gives the possibility to intervene earlier in the learning process, when difficulties arise and thus, to prevent future more complex and serious problems like dropping out.

2) Assessment of motivation. We have constructed an instrument for this purpose, partly from existing validated questionnaires and partly from created items. The two studies conducted for the validation of this instrument have shown satisfactory levels of validity and reliability.

Integrating the two components, a complete motivational diagnosis is provided, using two sources of information: an objective one – the actions of the learner and a subjective one – the self-assessment. Only the first has the advantage of unobtrusiveness, but the disadvantage of restricted information; the subjective source has the advantage of accurate information on motivational status, but used alone would mean interruption of the learners that are engaged in their learning and that do not need intervention. Thus, combining the two, we benefit from the advantages of both and eliminate the disadvantages of using only one of them.

Thus, we consider that our approach is a good starting point for motivational assessment and that further research would lead to a refined approach generally applicable to e-Learning systems.

Bibliography

Ames, C. (1992a). Classrooms: Goals, structures, and student motivation. *Journal of educational psychology*, 84(3), 261-271.

Ames, C. (1992b). Achievement goals and adaptive motivation patterns: The role of the environment. In G. Roberts (Ed.), *Motivation in sport and exercise*, Champaign, IL: Human Kinetics, 161-176.

Ames, C. (1992c). Achievement goals and classroom motivation climate. In J. Meece and D. Schunk (Eds.) *Students' perceptions in the classroom*, Hillsdale, NJ: Erlbaum, 327-348.

Anderman, L. H., Anderman E. M. (1999). Social predictors of changes in students' achievement goal orientations. *Contemporary Educational Psychology*, 25, 21-37.

Arroyo, I., Woolf, B.P. (2005). Inferring learning and attitudes from a Bayesian Network of log file data. In C. K. Looi, G. McCalla, B. Bredeweg, and J. Breuker (Eds.), *Artificial Intelligence in Education, Supporting Learning through Intelligent and Socially Informed Technology*, Amsterdam: IOS Press, 33-40.

Baker, R., Corbett, A., Koedinger, K. (2004). Detecting Student Misuse of Intelligent Tutoring Systems. In J. C. Lester, R. M. Vicari and F. Paraguau (eds), *Proceedings of the Seventh International Conference on Intelligent Tutoring Systems, Lecture Notes in Computer Science (LNCS)*, vol. 3220, 2004, 531–540. Springer, Berlin/ Heidelberg.

Bandura, A. (1977). *Social Learning Theory*. New York: General Learning Press.

Bandura, A. (1982). Self-efficacy mechanism in human agency. *The American Psychologist*, 37(2), 122–147.

Bandura, A. (1986). *Social foundations of thought and action: A social cognitive theory.* Englewood Cliffs, NJ: Prentice Hall.

Bandura, A. (2002). Growing primacy of human agency in adaptation and change in the electronic era. *European Psychologist*, 7, 1-16.

Bandura, A., Cervone, D. (1983). Self-evaluative and self-efficacy mechanisms governing the motivational effects of goal systems. *Journal of personality and social psychology*, 45(5), 1017 - 1028.

Bandura, A., Schunk, D. H. (1981). Cultivating competence, self-efficacy, and intrinsic interest through proximal self-evaluating. *Journal of Personality and Social Psychology*, 41, 586-598.

Beal, C.R., Lee, H. (2005). Creating a pedagogical model that uses student self reports of motivation and mood to adapt ITS instruction. In *Proceedings of the Workshop on Emotion and Motivation in Educational Software* (EMES). Amsterdam: IOS Press. Retrieved on 10/06/2006 from http://k12.usc.edu/Pubs/.

Beal, C.R., Qu, L., Lee, H. (2006). Classifying Learner Engagement through Integration of Multiple Data Sources. In *Proceedings of the 21st National Conference on Artificial Intelligence*. Menlo Park: AAAI Press. Retrieved on 10/06/2006 from http://k12.usc.edu/Pubs/.

Beck, J. E. (2005). Engagement tracing: Using response times to model student disengagement. In C. K. Looi, G. McCalla, B. Bredeweg, and J. Breuker (eds.) *Artificial Intelligence in Education: Supporting Learning through Intelligent and Socially Informed Technology*, 88–95. IOS Press, Amsterdam.

Beck, J. E. (2004). Using response times to model student disengagement. In *Proceedings of the ITS2004 Workshop on Social and Emotional Intelligence in Learning Environments*; Retrieved on 10/06/2006 from www.cs.cmu.edu/~listen/pdfs/affect%20crc.pdf.

Bull, S. (2004). Supporting Learning with Open Learner Models. In *Proceedings of 4th Hellenic Conference with International Participation: Information and Communication Technologies in Education*, Athens, Greece. *Keynote*. Retrieved on 11/06/07 from http://www.eee.bham.ac.uk/bull/.

Busch, T. (1995). Gender differences in self-efficacy and attitudes toward computers. *Journal of Educational Computing Research*, 12(2), 147.

Cocea, M., Weibelzahl, S. (2007a). Eliciting motivation knowledge from log files towards motivation diagnosis for Adaptive Systems. In Conati, C., McCoy, K., Paliouras, G. (eds.) *User Modelling 2007. Proceedings of 11th International Conference, UM2007, Lecture Notes in Artificial Intelligence (LNAI)*, vol. 4511, 197-206. Springer, Berlin.

Cocea, M., Weibelzahl, S. (2007b). Cross-System Validation of Engagement Prediction from Log Files. In Duval, E., Klamma, R. and Wolpers, M. (Eds.) *Creating New Learning Experiences on a Global Scale, Second European Conference on Technology Enhanced Learning, EC-TEL 2007, Crete, Greece, September 2007 Proceedings, Lecture Notes in Computer Science (LNCS)*, vol. 4753, 14-25, Springer Berlin / Heidelberg.

Csikszentmihalyi, Mihaly (1997). *Finding Flow: The Psychology of Engagement with Everyday Life*, BasicBooks, New York, 1997.

De Vicente, A., Pain, H. (2002). Informing the Detection of the Students' Motivational State: an empirical Study. In Cerri, S. A., Gouarderes, G., Paraguau, F: *Intelligent Tutoring Systems, 6th International Conference, Lecture Notes in Computer Science (LNCS)*, vol. 2363, 933-943. Berlin: Springer-Verlag.

De Vicente, A., Pain, H. (2003). Validating the Detection of a Student's Motivational State. In Mendez Vilas, A.; Mesa Gonzalez, J. A.; Mesa Gonzalez, J. (eds): *Proceedings of the Second International Conference on Multimedia Information & Communication Technologies in Education* (m-ICTE2003), number 15 of "Sociedad de la Informacion" Series, Volume III, 2004-2008. Merida, Junta de Extremadura (Consejeria de Educacion, Ciencia y Tecnologia)

DeTure, M. (2004). Cognitive style and self-efficacy: Predicting student success in online distance education. *The American Journal of Distance Education*, 18(1), 21-38.

Dillman, D. A. (2000). *Mail and Internet Surveys: The Tailored Design Method.* Second Edition. New York: John Wiley and Sons.

Driscoll, M. P. (2000). *Psychology of learning for instruction* (2nd ed.). Needham Heights, MA: Allyn and Bacon.

Ertmer, P. A., Evenbeck, E., Cennamo, K. S., Lehman, J. D. (1994). Enhancing self-efficacy for computer technologies through the use of positive classroom experiences. *Educational Technology, Research and Development*, 42(3), 45-62.

Farzan, R., Brusilovsky, P. (2005). Social navigation support in E-Learning: What are real footprints. In *IJCAI'05 Workshop on Intelligent Techniques for Web Personalization*, Edinburgh, U.K., 49–56.

Hakkinen, P. (1995). Changes in computer anxiety in a required computer course. *Journal of Research on Computing in Education*, 27(2), 141-153.

Hayes, A. F., Krippendorff, K. (2006): Answering the call for a standard reliability measure for coding data. *Communication Methods and Measures*. Retrieved November 8, 2006 from http://www.comm.ohio-state.edu/ahayes/ .

Heider, F. (1958). *The Psychology of Interpersonal Relations*, N.Y.: John Wiley and Sons.

Hodges, C. B. (2004). Designing to Motivate: Motivational Techniques to Incorporate in E-Learning Experiences, *The Journal of Interactive Online Learning*, Vol. 2, No 3.

Holcomb, L. B., King, F. B., Brown, S. W. (2004). Student traits and attributes contributing to success in online courses: Evaluation of University online courses. *The Journal of Interactive Online Learning*, 2(3), 1-17.

IQ-Research group (2001). *Two test sets adapted from motivation tests developed by P. Pintrich and P. Ruohotie, "Learning experiences and motivation" and "Learning strategies"*. Department of Education, University of Helsinki.

Johns, J., Woolf, B. (2006). A Dynamic Mixture Model to Detect Student Motivation and Proficiency. *Proceedings of the Twenty-first National Conference on Artificial Intelligence and the Eighteenth Innovative Applications of Artificial Intelligence Conference* (AAAI-06), July 16-20, 2006, Boston, Massachusetts, USA.

Joo, Y., Bong, M., Choi, H. (2000). Self-efficacy in web-based instruction. *Educational Technology Research and Development*, 48(2), 5-17.

Karsten, R., Roth, R. (1998a). The relationship of computer experience and computer self-efficacy to performance in introductory computer literacy courses. *Journal of Research on Computing in Education*, 31 (1), 14-22.

Karsten, R., Roth, R. (1998b). Computer self-efficacy: A practical indicator of student computer competency in introductory IS courses. *Informing Science*, 1 (3), 61-68.

Kazdin, A. E. (1974). Self-monitoring and behavior change. In M.J. Mahoney and C.E. Thoresen (Eds.), *Self-control: Power to the person.* Monterey, California: Brooks/Cole, 218-246.

Keller, J. M. (1979). Motivation and instructional design: A theoretical perspectives. *Journal of Instructional Development*, 2(4), 26-34.

Keller, J. M. (1987). Development and use of the ARCS model of instructional design. *Journal of Instructional Development*, 10(3), 2-10.

Keller, J. M. (2006). *What is motivational design?* Retrieved on 11/06/2007 from http://www.arcsmodel.com/.

King, F.B. (2001). Asynchronous distance education courses employing Web-based instruction: Implications of individual study skills self-efficacy and self-regulated learning. Doctoral dissertation, University of Connecticut, 2001. *Dissertations Abstracts International*, 62, 912.

Langford, M., Reeves, T.E. (1998). The relationship between computer self-efficacy and personal characteristics of the beginning information systems student. *Journal of Computer Information Systems*, 38 (4), 41-45.

Lee, C., Witta, E. L. (2001). Online students' perceived self-efficacy: Does it change? *Proceedings of National Convention the Association for Educational Communications and Technology (AECT)*, Atlanta, GA, November 8-12, 2001, 228-233.

Lim, C. K. (2001). Computer self-efficacy, academic self-concept, and other predictors of satisfaction and future participation of adult distance learners. *The American Journal of Distance Education*, 15(2), 41-51.

Lombard, M., Snyder-Duch, J., Campanella Bracken, C. (2003): *Practical Resources for Assessing and Reporting Intercoder Reliability in Content Analysis Research.* Retrieved on 11/06/2007 from http://www.temple.edu/mmc/reliability.

Martinez, M. (2003). High Attrition Rates in e-Learning: Challenges, Predictors, and Solutions. *The e-Learning Developers Journal*, July, 2003. Retrieved on 10/06/2006 from http://www.elearningguild.com.

McInerney, V., McInerney, D. M., Sinclair, K. E. (1994). Student teachers, computer anxiety and computer experience. *Journal of Educational Computing Research*, 7 (1), 27-50.

Meece, J. L., Holt, K. (1993). A pattern analysis of students' achievement goals. *Journal of Educational Psychology*, 85, 582-590.

Midgley, C., Maehr, M. L., Hruda, L. Z., Anderman, E., Anderman, L., Freeman, K. E., Gheen, M., Kaplan, A., Kumar, R., Middleton, M. J., Nelson, J., Roeser, R., Urdan, T. (2000). *Manual for the Patterns of Adaptive Learning Scales (PALS)*. Ann Arbor, MI: University of Michigan.

Miltiadou, M., Savenye, W. (2003). Applying social cognitive constructs of motivation to enhance student success in online distance education. *AACE Journal,* 11(1), 78-95. Norfolk, VA: AACE.

Pintrich, P., Schunk, D. (2002). *Motivation in Education: Theory, Research and Applications*, 2nd edition, Englewood Cliffs, NJ: Prentice-Hall.

Qu, L., Wang N., Johnson, W. L. (2005a). Detecting the Learner's Motivational States in an Interactive Learning Environment. In C.-K. Looi et al. (Eds.), *Proceedings of Artificial Intelligence in Education*, IOS Press, 547-554.

Qu, L., Wang N., Johnson, W. L. (2005b). Using Learner Focus of attention to Detect Learner Motivation Factors. In Ardissono, L., Brna, P.; Mitrovic, A. (Eds.): *Proceedings of User Modelling 2005, Lecture Notes in Computer Science (LNCS)*, vol. 3538, 70-73. Springer-Verlag Berlin Heidelberg.

Quinlan, R. (1993). *C4.5: Programs for Machine Learning*, Morgan Kaufmann Publishers, San Mateo, CA.

ReadingSoft.com found at HYPERLINK, http://www.readingsoft.com .

Reed, W.M., Overbaugh, R.C. (1993). The effects of prior experience and instructional format on teacher education students' computer anxiety and performance. *Computers in the Schools*, 9(2/3), 75-89.

Reips, U.-D. (2002). Standards for Internet-based experimenting. *Experimental Psychology*, 49 (4), 243-256.

Rotter, J.B. (1966). Generalized expectancies for internal versus external control of reinforcement, *Psychological Monographs*, 80, 1-28.

Schunk, D, H. (1990). Goal setting and self-efficacy during self-regulated learning. *Educational Psychologist*, 25, 71-86.

Shneiderman, B., Alavi, M., Norman, K., Borkowski, E. (1995). Windows of opportunity in electronic classrooms, *Communications of the ACM*, 38(11), 19-24.

Tourangeau, R., et al. (2000). *The Psychology of Survey Response*. Cambridge, UK: Cambridge University Press.

Tuckman, B.W., Sexton, T.L. (1991). The effect of teacher encouragement on student self-efficacy and motivation for self-regulated performance. *Journal of Social Behavior and Personality*, 6, 137-146.

TurboRead Speed Reading found at HYPERLINK, http://www.turboread.com .

Walonoski, J., Heffernan, N. T. (2006a). Detection and Analysis of Off-Task Gaming Behavior in Intelligent Tutoring Systems. In Ikeda, Ashley & Chan (Eds.). *Proceedings of the Eighth International Conference in Intelligent Tutoring Systems, Lecture Notes in Computer Science (LNCS)*, vol. 4053, 382-391. Springer-Verlag: Berlin.

Walonoski, J., Heffernan, N. T. (2006b). Prevention of Off-Task Gaming Behaviour within Intelligent Tutoring Systems. In Ikeda, Ashley & Chan (Eds.). *Proceedings of the Eighth International Conference in Intelligent Tutoring Systems, Lecture Notes in Computer Science (LNCS)*, vol. 4053, 722-724. Springer-Verlag: Berlin.

Wang, A. Y., Newlin, M. H. (2002). Predictors of web-student performance: The role of self-efficacy and reasons for taking an on-line class. *Computers in Human Behavior*, 18, 151-163.

Weber, G., Kuhl, H.-C., Weibelzahl, S. (2001). Developing adaptive internet based courses with the authoring system NetCoach. In *Hypermedia: Openness, Structural Awareness, and Adaptivity* (LNAI 2266), Springer, Berlin, 226-238.

Weiner, B. (1974). *Achievement motivation and attribution theory*. Morristown, N.Y.: General learning Press.

Witten, I. H., Frank, E., Trigg, L.E, Hall, M., Holmes, G., Cunningham, S.J. (1999). *Weka: Practical machine learning tools and techniques with Java implementations.* Proc ICONIP/ ANZIIS/ANNES99 Future Directions for Intelligent Systems and Information Sciences, 192-196, Dunedin, New Zealand.

Witten, I.H., Frank, E. (2005). *Data mining. Practical Machine Learning Tools and Techniques.* Second Edition, Morgan Kauffman Publishers, Elsevier.

Zhang, G., Cheng, Z., He, A., Huang, T. (2003). A WWW-based Learner's Learning Motivation Detecting System. In *Proceedings of International Workshop on "Research Directions and Challenge Problems in Advanced Information Systems Engineering"*, Honjo City, Japan, September 16–19, http://www.akita-pu.ac.jp/system/ KEST2003/ . Retrieved 05/02/06

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of Educational Psychology*, 81(3), 329-339.

Zimmerman, B. J., Martinez-Pons, M. (1990). Student differences in self-regulated learning: Relating grade, sex, and giftedness to self-efficacy and strategy use. *Journal of Educational Psychology*, 82(1), 51-59.

Zimmerman, B.J. (2000). Attaining self-regulation. A social cognitive perspective. In Boekaerts, M., Pintrich, P. & Zeidner, M. (Eds.). *Handbook of self-regulation*, Academic Press.

Zimmerman, B.J. (1990). Self-regulating academic learning and achievement: The emergence of a social cognitive perspective. *Educational Psychology Review*, 2 (2), 173–201.

A.1. Initial instrument: aggregated instrument from validated questionnaires and created items

1. Self-Efficacy (from PALS)

I'm certain I can master the skills taught in this course.
I'm certain I can figure out how to do the most difficult work.
I can do almost all the work in the course if I don't give up.
Even if the work is hard, I can learn it.
I can do even the hardest work in this course if I try.

2. Self Regulation (from IQ Learn – section Strategies in Learning)

Time Management
Even in a tough situation I can stick to the schedule I have made for myself.
I stick to a certain timetable when I'm studying.
I use the time that I have reserved for studying efficiently.
I always stick to the study schedule that I have made.

Self Management
I try to accommodate my own style of studying so that it would match with the requirements of each course.
Before a study assignment, I often go through its different steps in my mind.
I set learning goals to be able to direct my studies.
After a study assignment I often think about how I did and how I could improve my performance.

Persistency
I often feel so lazy or bored studying course literature that I quit before finishing.
I often give up when I'm studying difficult issues and focus on the easier ones.
I have no difficulties in motivating myself to complete the study tasks even if they are not particularly interesting to me.
I work really hard to do well in my studies, even if I don't like all the tasks or the material I am reading.

3. Goal Orientation (from PALS, except disengaged category)

Mastery
It's important to me that I learn a lot of new concepts.
One of my goals is to learn as much as I can.
One of my goals is to master a lot of new skills.
It's important to me that I thoroughly understand my work.
It's important to me that I improve my skills.

Performance approach
It's important to me that other students think I am good at my work.
One of my goals is to show others that I'm good at my work.
One of my goals is to show others that work is easy for me.
One of my goals is to look smart in comparison to the other students.
It's important to me that I look smart compared to others.

Performance avoidance
It's important to me that I don't look stupid.
One of my goals is to keep others from thinking I'm not smart.
It's important to me that my lecturer/tutor doesn't think that I know less than others.
One of my goals in class is to avoid looking like I have trouble doing the work.

Disengaged (created)
I'm just interested in passing the course.
I don't care about how stupid I look compared to others as long as I pass the course.
I don't care about what I learned, as long as I pass the course.

For all previous questions a five point Likert-type scale is used (from 1 = not at all true to 5 = very true).

4. Perceived Task Difficulty (created)

Please estimate how difficult you find the content of this subject/course/module:
Too easy (1) -- Very difficult (5)

Please estimate how difficult you find the tests/ assignments of this subject:
Too easy (1) -- Very difficult (5)

5. Attribution (Created)

On the attached scale estimate the degree in which each of the following aspects contributed to your learning outcomes:

	Not at all					Very much
- difficulty of the subject	1	2	3	4	5	6
- effort you put in	1	2	3	4	5	6
- the luck you had at exams/ tests	1	2	3	4	5	6
- your ability for this type of subject	1	2	3	4	5	6

A.2. Final instrument

1. Self-Efficacy

I'm certain I can master the skills taught in this course.
I can do almost all the work in the course if I don't give up.
I can do even the hardest work in this course if I try.

2. Self Regulation

Time Management
I always stick to the study schedule that I have made.
I use the time that I have reserved for studying efficiently.

Self Management
I set learning goals to be able to direct my studies.
Before a study assignment, I often go through its different steps in my mind.

Persistency
I often give up when I'm studying difficult issues and focus on the easier ones.
I work really hard to do well in my studies, even if I don't like all the tasks or the material I am reading.

3. Goal Orientation

Mastery
One of my goals is to master a lot of new skills.
It's important to me that I thoroughly understand my work.
It's important to me that I improve my skills.

Performance approach
It's important to me that other students in my class think I am good at my work.
One of my goals is to show others that I'm good at my work.
One of my goals is to look smart in comparison to the other students.

Performance avoidance
It's important to me that my lecturer/tutor doesn't think that I know less than others.
One of my goals is to keep others from thinking I'm not smart.

Disengaged
I am only interested in passing the course.
As long as I pass the course, I don't care about how stupid I look compared to others.
I don't care about what I learned, as long as I pass the course.

For all previous questions a five point Likert-type scale is used (from 1 = not at all true to 5 = very true).

4. Perceived Task Difficulty

Please estimate how difficult you find the content of this subject/course/module:
Too easy (1) -- Very difficult (5)

Please estimate how difficult you find the tests/ assignments of this subject:
Too easy (1) -- Very difficult (5)

5. Attribution

Using the scale below, estimate how much each of the following aspects contributed to your learning outcomes:

	Not at all					Very much
- difficulty of the subject	1	2	3	4	5	6
- effort you put in	1	2	3	4	5	6
- the luck you had at exams/ tests	1	2	3	4	5	6
- your ability for this type of subject	1	2	3	4	5	6

Appendix B.1.1. Group 1 (both questionnaires completed on the same day)

		SE 1	TM1	SM1	P 1	SR 1	M 1	PAp 1	PAv 1	D1	PTD1	ILC	ELC	S1	InS1
N		11	11	11	11	11	11	11	11	11	11	11	11	11	11
Normal Parameters a, b	Mean	4.152	2.500	3.227	2.864	2.864	4.242	1.546	1.864	3.182	2.909	4.182	3.091	3.773	3.409
	Std. Dev.	0.736	1.183	1.292	1.051	0.906	0.685	0.563	1.267	1.393	0.861	1.189	0.735	0.905	0.970
Most Extreme Differences	Absolute	0.148	0.209	0.180	0.272	0.222	0.187	0.233	0.298	0.175	0.276	0.139	0.256	0.145	0.281
	Positive	0.140	0.209	0.115	0.272	0.222	0.163	0.233	0.298	0.123	0.276	0.122	0.153	0.128	0.281
	Negative	-0.148	-0.108	-0.180	-0.133	-0.129	-0.187	-0.166	-0.248	-0.175	-0.226	-0.139	-0.256	-0.145	-0.174
Kolmogorov-Smirnov Z		0.492	0.694	0.596	0.901	0.738	0.620	0.773	0.988	0.582	0.916	0.461	0.850	0.480	0.931
Asymp. Sig. (2-tailed)		0.969	0.722	0.869	0.391	0.648	0.837	0.589	0.284	0.888	0.371	0.984	0.465	0.975	0.351

		SE 2	TM2	SM2	P2	SR2	M2	PAp2	PAv2	D2	PTD2	ILC2	ELC2	S2	InS2
N		11	11	11	11	11	11	11	11	11	11	11	11	11	11
Normal Parameters a, b	Mean	3.527	2.614	3.523	3.091	3.076	4.018	1.855	2.000	3.455	2.909	3.864	3.227	3.682	3.409
	Std. Dev.	0.840	0.817	0.984	0.910	0.757	0.827	1.230	1.289	1.447	0.861	1.206	1.232	1.146	0.970
Most Extreme Differences	Absolute	0.288	0.137	0.146	0.172	0.106	0.155	0.244	0.287	0.195	0.276	0.308	0.231	0.269	0.281
	Positive	0.288	0.137	0.146	0.158	0.106	0.148	0.218	0.287	0.143	0.276	0.308	0.231	0.269	0.281
	Negative	-0.142	-0.112	-0.127	-0.172	-0.102	-0.155	-0.244	-0.219	-0.195	-0.226	-0.146	-0.096	-0.185	-0.174
Kolmogorov-Smirnov Z		0.955	0.456	0.483	0.571	0.351	0.515	0.808	0.953	0.646	0.916	1.023	0.765	0.894	0.931
Asymp. Sig. (2-tailed)		0.322	0.986	0.974	0.900	1.000	0.954	0.531	0.324	0.797	0.371	0.246	0.602	0.401	0.351

a Test distribution is Normal.
b Calculated from data.

SE = Self-efficacy	SR = Self Regulation	D = Disengaged	S = Stability
TM = Time Management	M = Mastery Goal	PTD = Perceived Task Difficulty	InS = Instability
SM = Self Management	PAp = Performance Approach	ILC = Internal Locus of Control	1 = short version;
P = Persistency	PAv = Performance Avoidance	ELC = External Locus of Control	2 = long (initial) version

B.1.2. – Group 2 (questionnaires completed with time in between)

		SE1	TM1	SM1	P1	SR1	M1	PAp1	PAv1	D1	PTD1	ILC1	ELC1	S1	InS1
N		8	8	8	8	8	8	8	8	8	8	8	8	8	8
Normal Parameters [a, b]	Mean	4.208	1.875	3.313	2.750	2.646	4.333	2.708	3.000	3.167	3.000	4.375	3.438	4.000	3.813
	Std. Dev.	0.776	0.744	0.961	0.598	0.587	0.563	1.408	1.134	0.943	1.035	0.744	0.980	0.886	0.884
Most Extreme Differences	Absolute	0.223	0.193	0.252	0.162	0.203	0.223	0.145	0.170	0.320	0.250	0.192	0.217	0.250	0.291
	Positive	0.154	0.193	0.252	0.162	0.203	0.223	0.145	0.093	0.320	0.167	0.183	0.206	0.250	0.291
	Negative	-0.223	-0.175	-0.199	-0.162	-0.136	-0.223	-0.126	-0.170	-0.180	-0.250	-0.192	-0.217	-0.130	-0.209
Kolmogorov-Smirnov Z		0.630	0.546	0.714	0.459	0.573	0.631	0.411	0.482	0.906	0.707	0.542	0.614	0.707	0.823
Asymp. Sig. (2-tailed)		0.823	0.927	0.688	0.985	0.897	0.821	0.996	0.974	0.385	0.699	0.930	0.845	0.699	0.507

		SE2	TM2	SM2	P2	SR2	M2	PAp2	PAv2	D2	PTD2	ILC2	ELC2	S2	InS2
N		8	8	8	8	8	8	8	8	8	8	8	8	8	8
Normal Parameters [a, b]	Mean	3.625	2.531	3.219	2.406	2.719	4.075	2.250	2.688	3.333	3.000	4.063	3.438	3.938	3.563
	Std. Dev.	0.539	1.161	0.589	0.681	0.710	0.763	0.826	0.943	0.926	1.035	0.496	0.678	0.904	0.776
Most Extreme Differences	Absolute	0.213	0.218	0.191	0.193	0.107	0.204	0.197	0.151	0.234	0.250	0.300	0.366	0.278	0.218
	Positive	0.143	0.218	0.145	0.108	0.107	0.141	0.128	0.099	0.234	0.167	0.300	0.366	0.222	0.162
	Negative	-0.213	-0.154	-0.191	-0.193	-0.096	-0.204	-0.197	-0.151	-0.234	-0.250	-0.200	-0.259	-0.278	-0.218
Kolmogorov-Smirnov Z		0.603	0.617	0.541	0.547	0.304	0.577	0.557	0.428	0.663	0.707	0.849	1.034	0.785	0.616
Asymp. Sig. (2-tailed)		0.860	0.841	0.931	0.926	1.000	0.893	0.915	0.993	0.772	0.699	0.467	0.235	0.569	0.842

a Test distribution is Normal.
b Calculated from data.

SE = Self-efficacy	SR = Self Regulation	D = Disengaged	S = Stability
TM = Time Management	M = Mastery Goal	PTD = Perceived Task Difficulty	InS = Instability
SM = Self Management	PAp = Performance Approach	ILC = Internal Locus of Control	1 = short version;
P = Persistency	PAv = Performance Avoidance	ELC = External Locus of Control	2 = long (initial) version

B.1.3. – All participants

		SE1	TM1	SM1	P1	SR1	M1	PAp1	PAv1	D1	PTD1	ILC1	ELC1	S1	InS1
N		19	19	19	19	19	19	19	19	19	19	19	19	19	19
Normal Parameters a, b	Mean	4.175	2.237	3.263	2.816	2.772	4.281	2.035	2.342	3.175	2.947	4.263	3.237	3.868	3.632
	Std. Dev.	0.732	1.046	1.135	0.869	0.776	0.621	1.138	1.313	1.193	0.911	1.005	0.839	0.879	0.940
Most Extreme Differences	Absolute	0.170	0.169	0.171	0.221	0.188	0.206	0.259	0.215	0.190	0.161	0.144	0.202	0.177	0.242
	Positive	0.138	0.169	0.171	0.221	0.188	0.154	0.259	0.215	0.190	0.161	0.144	0.178	0.177	0.242
	Negative	-0.170	-0.118	-0.110	-0.121	-0.113	-0.206	-0.182	-0.153	-0.178	-0.155	-0.134	-0.202	-0.138	-0.146
Kolmogorov-Smirnov Z		0.741	0.735	0.744	0.962	0.818	0.900	1.127	0.938	0.828	0.703	0.626	0.880	0.773	1.056
Asymp. Sig. (2-tailed)		0.643	0.653	0.638	0.313	0.515	0.393	0.158	0.343	0.499	0.707	0.828	0.420	0.588	0.214

		SE2	TM2	SM2	P2	SR2	M2	PAp2	PAv2	D2	PTD2	ILC2	ELC2	S2	InS2
N		19	19	19	19	19	19	19	19	19	19	19	19	19	19
Normal Parameters a, b	Mean	3.568	2.579	3.395	2.803	2.925	4.042	2.021	2.289	3.404	2.947	3.947	3.316	3.789	3.474
	Std. Dev.	0.713	0.947	0.835	0.872	0.740	0.779	1.071	1.179	1.225	0.911	0.956	1.017	1.032	0.874
Most Extreme Differences	Absolute	0.167	0.170	0.125	0.160	0.091	0.150	0.170	0.169	0.213	0.161	0.162	0.218	0.199	0.172
	Positive	0.167	0.170	0.125	0.146	0.091	0.136	0.146	0.169	0.099	0.161	0.162	0.218	0.199	0.172
	Negative	-0.087	-0.098	-0.108	-0.160	-0.066	-0.150	-0.170	-0.137	-0.213	-0.155	-0.108	-0.168	-0.160	-0.136
Kolmogorov-Smirnov Z		0.726	0.743	0.543	0.698	0.399	0.656	0.742	0.739	0.929	0.703	0.707	0.948	0.867	0.751
Asymp. Sig. (2-tailed)		0.668	0.640	0.929	0.714	0.997	0.783	0.641	0.646	0.354	0.707	0.699	0.329	0.439	0.626

a Test distribution is Normal.
b Calculated from data.

SE = Self-efficacy	SR = Self Regulation	D = Disengaged	S = Stability
TM = Time Management	M = Mastery Goal	PTD = Perceived Task Difficulty	InS = Instability
SM = Self Management	PAp = Performance Approach	ILC = Internal Locus of Control	1 = short version;
P = Persistency	PAv = Performance Avoidance	ELC = External Locus of Control	2 = long (initial) version

Appendix B.2.1. Independent t-test: comparison between Group 1 (same day) and Group 2 (time in between)

Group Statistics

	time	N	Mean	Std. Deviation	Std. Error Mean
Self Efficacy 1	same day	11	4.1515	.73581	.22185
	time in between	8	4.2083	.77619	.27442
Time Management 1	same day	11	2.5000	1.18322	.35675
	time in between	8	1.8750	.74402	.26305
Self Management 1	same day	11	3.2273	1.29158	.38943
	time in between	8	3.3125	.96130	.33987
Persistency 1	same day	11	2.8636	1.05097	.31688
	time in between	8	2.7500	.59761	.21129
Self Regulation 1	same day	11	2.8636	.90621	.27323
	time in between	8	2.6458	.58715	.20759
Mastery GO 1	same day	11	4.2424	.68461	.20642
	time in between	8	4.3333	.56344	.19920
Performance Approach GO 1	same day	11	1.5455	.56318	.16981
	time in between	8	2.7083	1.40789	.49776
Performance Avoidance GO 1	same day	11	1.8636	1.26671	.38193
	time in between	8	3.0000	1.13389	.40089
Disengaged GO 1	same day	11	3.1818	1.39335	.42011
	time in between	8	3.1667	.94281	.33333
Perceived Task Difficulty 1	same day	11	2.9091	.86076	.25953
	time in between	8	3.0000	1.03510	.36596
Internal Locus of Control 1	same day	11	4.1818	1.18896	.35849
	time in between	8	4.3750	.74402	.26305
External Locus of Control 1	same day	11	3.0909	.73547	.22175
	time in between	8	3.4375	.97970	.34638
Stability 1	same day	11	3.7727	.90453	.27273
	time in between	8	4.0000	.88641	.31339
Instability 1	same day	11	3.5000	1.00000	.30151
	time in between	8	3.8125	.88388	.31250
Self Efficacy 2	same day	11	3.5273	.84035	.25337
	time in between	8	3.6250	.53918	.19063
Time Management 2	same day	11	2.6136	.81673	.24625
	time in between	8	2.5313	1.16065	.41035
Self Management 2	same day	11	3.5227	.98396	.29668
	time in between	8	3.2188	.58915	.20830
Persistency 2	same day	11	3.0909	.91017	.27443
	time in between	8	2.4063	.68057	.24062
Self regulation 2	same day	11	3.0758	.75687	.22821
	time in between	8	2.7188	.70982	.25096
Mastery GO 2	same day	11	4.0182	.82682	.24930
	time in between	8	4.0750	.76298	.26976
Performance Approach GO 2	same day	11	1.8545	1.22993	.37084
	time in between	8	2.2500	.82635	.29216
Performance Avoidance GO 2	same day	11	2.0000	1.28938	.38876
	time in between	8	2.6875	.94255	.33324
Disengaged GO2	same day	11	3.4545	1.44739	.43641
	time in between	8	3.3333	.92582	.32733
Perceived Task Difficulty 2	same day	11	2.9091	.86076	.25953
	time in between	8	3.0000	1.03510	.36596
Internal Locus of Control 2	same day	11	3.8636	1.20605	.36364
	time in between	8	4.0625	.49552	.17519
External Locus of Control 2	same day	11	3.2273	1.23215	.37151
	time in between	8	3.4375	.67810	.23975
Stability 2	same day	11	3.6818	1.14614	.34557
	time in between	8	3.9375	.90386	.31956
Instability 2	same day	11	3.4091	.97000	.29247
	time in between	8	3.5625	.77632	.27447

Independent Samples Test

	Levene's Test for Equality of Variances		t-test for Equality of Means						
								95% Confidence Interval of the Difference	
	F	Sig.	t	df	Sig. (2-tailed)	Mean Difference	Std. Error Difference	Lower	Upper
Self Efficacy 1	.049	.828	-.162	17	.873	-.057	.350	-.795	.681
			-.161	14.734	.874	-.057	.353	-.810	.697
Time Management 1	1.123	.304	1.312	17	.207	.625	.476	-.380	1.630
			1.410	16.755	.177	.625	.443	-.311	1.561
Self Management 1	1.193	.290	-.157	17	.877	-.085	.542	-1.229	1.059
			-.165	16.970	.871	-.085	.517	-1.176	1.005
Persistency 1	2.081	.167	.274	17	.787	.114	.415	-.761	.989
			.298	16.273	.769	.114	.381	-.693	.920
Self Regulation 1	2.769	.114	.593	17	.561	.218	.367	-.557	.993
			.635	16.854	.534	.218	.343	-.507	.942
Mastery GO 1	.480	.498	-.307	17	.763	-.091	.296	-.716	.534
			-.317	16.659	.755	-.091	.287	-.697	.515
Performance Approach GO 1	7.743	.013	-2.499	17	.023	-1.163	.465	-2.145	-.181
			-2.211	8.642	.056	-1.163	.526	-2.360	.034
Performance Avoidance GO 1	.039	.846	-2.015	17	.060	-1.136	.564	-2.326	.054
			-2.052	16.157	.057	-1.136	.554	-2.309	.037
Disengaged GO 1	1.135	.302	.027	17	.979	.015	.571	-1.189	1.219
			.028	16.955	.978	.015	.536	-1.117	1.147
Perceived Task Difficulty 1	.431	.520	-.209	17	.837	-.091	.435	-1.009	.827
			-.203	13.433	.842	-.091	.449	-1.057	.875
Internal Locus of Control 1	1.935	.182	-.404	17	.691	-.193	.478	-1.202	.816
			-.434	16.736	.670	-.193	.445	-1.132	.746
External Locus of Control 1	1.369	.258	-.883	17	.389	-.347	.392	-1.175	.481
			-.843	12.450	.415	-.347	.411	-1.239	.546
Stability 1	.092	.765	-.545	17	.593	-.227	.417	-1.107	.652
			-.547	15.425	.592	-.227	.415	-1.111	.656
Instability 1	.007	.934	-.705	17	.490	-.313	.443	-1.248	.623
			-.720	16.245	.482	-.313	.434	-1.232	.607
Self Efficacy 2	2.219	.155	-.288	17	.777	-.098	.340	-.815	.619
			-.308	16.824	.762	-.098	.317	-.767	.572
Time Management 2	.239	.631	.182	17	.858	.082	.452	-.872	1.036
			.172	11.872	.866	.082	.479	-.962	1.126
Self Management 2	.961	.341	.775	17	.449	.304	.392	-.523	1.131
			.839	16.545	.414	.304	.362	-.462	1.070
Persistency 2	.142	.711	1.789	17	.091	.685	.383	-.123	1.492
			1.876	16.963	.078	.685	.365	-.085	1.455
Self regulation 2	.010	.923	1.041	17	.312	.357	.343	-.366	1.080
			1.052	15.800	.308	.357	.339	-.363	1.077
Mastery GO 2	.235	.634	-.153	17	.880	-.057	.372	-.842	.729
			-.155	15.929	.879	-.057	.367	-.836	.722
Performance Approach GO 2	.508	.486	-.786	17	.442	-.395	.503	-1.456	.665
			-.838	16.942	.414	-.395	.472	-1.392	.601
Performance Avoidance GO 2	.561	.464	-1.276	17	.219	-.688	.539	-1.824	.449
			-1.343	16.990	.197	-.688	.512	-1.768	.393
Disengaged GO2	2.464	.135	.207	17	.838	.121	.585	-1.113	1.356
			.222	16.814	.827	.121	.546	-1.031	1.273
Perceived Task Difficulty 2	.431	.520	-.209	17	.837	-.091	.435	-1.009	.827
			-.203	13.433	.842	-.091	.449	-1.057	.875
Internal Locus of Control 2	10.451	.005	-.438	17	.667	-.199	.454	-1.158	.760
			-.493	14.096	.630	-.199	.404	-1.064	.666
External Locus of Control 2	1.210	.287	-.435	17	.669	-.210	.483	-1.230	.810
			-.475	16.080	.641	-.210	.442	-1.147	.727
Stability 2	1.312	.268	-.522	17	.608	-.256	.489	-1.288	.777
			-.543	16.832	.594	-.256	.471	-1.249	.738
Instability 2	.050	.825	-.369	17	.717	-.153	.416	-1.031	.724
			-.382	16.779	.707	-.153	.401	-1.000	.694

B.2.2. Paired t-test: comparison between the two questionnaires

Paired Samples Statistics

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Self Efficacy 1	4.1667	20	.71361	.15957
	Self Efficacy 2	3.6000	20	.70785	.15828
Pair 2	Time Management 1	2.2750	20	1.03205	.23077
	Time Management 2	2.6250	20	.94416	.21112
Pair 3	Self Management 1	3.2250	20	1.11774	.24993
	Self Management 2	3.4250	20	.82358	.18416
Pair 4	Persistancy 1	2.8000	20	.84915	.18988
	Persistancy 2	2.8250	20	.85494	.19117
Pair 5	Self Regulation 1	2.7667	20	.75587	.16902
	Self regulation 2	2.9583	20	.73474	.16429
Pair 6	Mastery GO 1	4.2167	20	.66907	.14961
	Mastery GO 2	4.0300	20	.76026	.17000
Pair 7	Performance Approach GO 1	2.0833	20	1.12845	.25233
	Performance Approach GO 2	2.1100	20	1.11539	.24941
Pair 8	Performance Avoidance GO 1	2.3750	20	1.28631	.28763
	Performance Avoidance GO 2	2.3500	20	1.17932	.26370
Pair 9	Disengaged GO 1	3.1833	20	1.16215	.25986
	Disengaged GO2	3.3833	20	1.19587	.26741
Pair 10	Perceived Task Difficulty 1	2.9500[a]	20	.88704	.19835
	Perceived Task Difficulty 2	2.9500[a]	20	.88704	.19835
Pair 11	Internal Locus of Control 1	4.2500	20	.98006	.21915
	Internal Locus of Control 2	3.9750	20	.93857	.20987
Pair 12	External Locus of Control 1	3.2750	20	.83469	.18664
	External Locus of Control 2	3.3250	20	.99041	.22146
Pair 13	Stability 1	3.9000	20	.86754	.19399
	Stability 2	3.8250	20	1.01664	.22733
Pair 14	Instability 1	3.6250	20	.91587	.20479
	Instability 2	3.4750	20	.85031	.19013

a. The correlation and t cannot be computed because the standard error of the difference is 0.

Paired Samples Correlations

		N	Correlation	Sig.
Pair 1	Self Efficacy 1 & Self Efficacy 2	20	.702	.001
Pair 2	Time Management 1 & Time Management 2	20	.699	.001
Pair 3	Self Management 1 & Self Management 2	20	.505	.023
Pair 4	Persistancy 1 & Persistancy 2	20	.674	.001
Pair 5	Self Regulation 1 & Self regulation 2	20	.746	.000
Pair 6	Mastery GO 1 & Mastery GO 2	20	.435	.055
Pair 7	Performance Approach GO 1 & Performance Approach GO 2	20	.650	.002
Pair 8	Performance Avoidance GO 1 & Performance Avoidance GO 2	20	.841	.000
Pair 9	Disengaged GO 1 & Disengaged GO2	20	.856	.000
Pair 11	Internal Locus of Control 1 & Internal Locus of Control 2	20	.336	.147
Pair 12	External Locus of Control 1 & External Locus of Control 2	20	.475	.034
Pair 13	Stability 1 & Stability 2	20	.412	.071
Pair 14	Instability 1 & Instability 2	20	.139	.558

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Self Efficacy 1 - Self Efficacy 2	.567	.549	.123	.310	.824	4.616	19	.000
Pair 2	Time Management 1 - Time Management 2	-.350	.771	.172	-.711	.011	-2.03	19	.057
Pair 3	Self Management 1 - Self Management 2	-.200	.999	.223	-.667	.267	-.896	19	.382
Pair 4	Persistancy 1 - Persistancy 2	-.025	.688	.154	-.347	.297	-.163	19	.873
Pair 5	Self Regulation 1 - Self regulation 2	-.192	.531	.119	-.440	.057	-1.61	19	.123
Pair 6	Mastery GO 1 - Mastery GO 2	.187	.764	.171	-.171	.544	1.093	19	.288
Pair 7	Performance Approach GO 1 - Performance Approach GO 2	-.027	.938	.210	-.466	.413	-.127	19	.900
Pair 8	Performance Avoidance GO 1 - Performance Avoidance GO 2	.025	.702	.157	-.304	.354	.159	19	.875
Pair 9	Disengaged GO 1 - Disengaged GO2	-.200	.634	.142	-.497	.097	-1.41	19	.175
Pair 11	Internal Locus of Control 1 - Internal Locus of Control 2	.275	1.106	.247	-.243	.793	1.112	19	.280
Pair 12	External Locus of Control 1 - External Locus of Control 2	-.050	.945	.211	-.492	.392	-.237	19	.815
Pair 13	Stability 1 - Stability 2	.075	1.029	.230	-.407	.557	.326	19	.748
Pair 14	Instability 1 - Instability 2	.150	1.160	.259	-.393	.693	.578	19	.570

Appendix C.1.1. "Beginning and end" Group

		SE1	TM1	SM1	P1	SR1	M1	PAp1	PAv1	D1	PTD1	ILC1	ELC1	S1	InS1
N		9	9	9	9	9	9	9	9	9	9	9	9	9	9
Normal Parameters a, b	Mean	4.444	2.889	3.611	3.722	3.407	4.407	2.852	2.556	1.667	3.722	5.056	3.556	4.667	3.944
	Std. Dev.	0.577	1.140	1.083	0.441	0.514	0.494	1.180	1.074	1.333	0.905	0.682	1.102	0.707	0.846
Most Extreme Differences	Absolute	0.205	0.239	0.158	0.248	0.224	0.226	0.228	0.228	0.358	0.287	0.199	0.232	0.208	0.363
	Positive	0.168	0.239	0.158	0.248	0.224	0.226	0.228	0.228	0.358	0.157	0.199	0.232	0.208	0.363
	Negative	-0.205	-0.144	-0.127	-0.196	-0.221	-0.218	-0.217	-0.163	-0.309	-0.287	-0.139	-0.196	-0.185	-0.189
Kolmogorov-Smirnov Z		0.616	0.717	0.474	0.745	0.672	0.679	0.683	0.685	1.074	0.862	0.597	0.697	0.623	1.088
Asymp. Sig. (2-tailed)		0.842	0.683	0.978	0.635	0.758	0.746	0.739	0.736	0.199	0.448	0.868	0.716	0.833	0.187

		SE2	TM2	SM2	P2	SR2	M2	PAp2	PAv2	D2	PTD2	ILC2	ELC2	S2	InS2
N		8	8	7	7	8	7	7	7	7	7	7	7	7	7
Normal Parameters a, b	Mean	4.700	2.938	3.500	3.714	3.500	4.343	2.086	2.571	1.095	3.714	4.786	3.214	4.500	3.500
	Std. Dev.	0.428	1.050	0.764	0.699	0.839	0.650	0.799	0.886	0.252	0.488	0.636	0.567	0.408	0.500
Most Extreme Differences	Absolute	0.384	0.226	0.214	0.318	0.165	0.249	0.233	0.213	0.504	0.435	0.245	0.362	0.214	0.270
	Positive	0.241	0.226	0.123	0.318	0.165	0.156	0.233	0.213	0.504	0.279	0.245	0.362	0.214	0.270
	Negative	-0.384	-0.149	-0.214	-0.155	-0.117	-0.249	-0.224	-0.117	-0.353	-0.435	-0.184	-0.210	-0.214	-0.270
Kolmogorov-Smirnov Z		1.085	0.640	0.567	0.842	0.465	0.660	0.617	0.564	1.335	1.151	0.648	0.957	0.567	0.714
Asymp. Sig. (2-tailed)		0.190	0.807	0.905	0.477	0.982	0.777	0.841	0.909	0.057	0.141	0.796	0.319	0.905	0.688

a Test distribution is Normal.
b Calculated from data.

SE = Self-efficacy	SR = Self Regulation	D = Disengaged	S = Stability
TM = Time Management	M = Mastery Goal	PTD = Perceived Task Difficulty	InS = Instability
SM = Self Management	PAp = Performance Approach	ILC = Internal Locus of Control	1 = short version;
P = Persistency	PAv = Performance Avoidance	ELC = External Locus of Control	2 = long (initial) version

Appendix C.1.2. "Middle" Group

		SE1	TM1	SM1	P1	SR1	M1	PAp1	PAv1	D1	PTD1	ILC1	ELC1	S1	InS1
N		8	9	9	9	9	9	9	8	8	8	8	8	8	8
Normal Parameters [a, b]	Mean	4.208	2.944	3.333	3.778	3.352	4.296	3.574	2.625	1.708	3.000	5.188	3.188	4.375	4.000
	Std. Dev.	0.925	0.846	0.866	1.093	0.562	0.512	1.182	1.217	0.700	0.655	0.884	0.961	0.991	0.845
Most Extreme Differences	Absolute	0.304	0.201	0.239	0.206	0.268	0.163	0.198	0.196	0.219	0.152	0.196	0.202	0.186	0.152
	Positive	0.196	0.201	0.110	0.206	0.179	0.163	0.114	0.196	0.219	0.152	0.179	0.202	0.186	0.125
	Negative	-0.304	-0.189	-0.239	-0.202	-0.268	-0.137	-0.198	-0.179	-0.189	-0.152	-0.196	-0.112	-0.175	-0.152
Kolmogorov-Smirnov Z		0.859	0.604	0.717	0.618	0.804	0.489	0.594	0.555	0.620	0.431	0.554	0.572	0.527	0.430
Asymp. Sig. (2-tailed)		0.451	0.859	0.683	0.839	0.538	0.971	0.873	0.918	0.837	0.992	0.918	0.899	0.944	0.993

		SE2	TM2	SM2	P2	SR2	M2	PAp2	PAv2	D2	PTD2	ILC2	ELC2	S2	InS2
N		8	8	8	8	8	8	8	8	8	8	8	8	8	8
Normal Parameters [a, b]	Mean	4.350	2.938	3.531	3.625	3.365	4.200	2.925	3.031	1.500	3.125	4.875	3.250	4.438	3.688
	Std. Dev.	0.943	0.863	0.807	0.886	0.667	0.535	1.291	0.647	0.436	0.582	0.876	0.926	0.943	0.998
Most Extreme Differences	Absolute	0.271	0.194	0.143	0.181	0.269	0.250	0.224	0.269	0.274	0.290	0.291	0.231	0.151	0.255
	Positive	0.245	0.194	0.143	0.181	0.269	0.250	0.224	0.269	0.249	0.210	0.291	0.231	0.150	0.255
	Negative	-0.271	-0.139	-0.135	-0.148	-0.177	-0.183	-0.153	-0.183	-0.274	-0.290	-0.209	-0.144	-0.151	-0.167
Kolmogorov-Smirnov Z		0.767	0.548	0.405	0.512	0.760	0.707	0.635	0.762	0.774	0.820	0.822	0.655	0.428	0.720
Asymp. Sig. (2-tailed)		0.599	0.925	0.997	0.956	0.610	0.699	0.816	0.608	0.587	0.511	0.509	0.785	0.993	0.678

a Test distribution is Normal.
b Calculated from data.

SE = Self-efficacy	SR = Self Regulation	D = Disengaged	S = Stability
TM = Time Management	M = Mastery Goal	PTD = Perceived Task Difficulty	InS = Instability
SM = Self Management	PAp = Performance Approach	ILC = Internal Locus of Control	1 = short version;
P = Persistency	PAv = Performance Avoidance	ELC = External Locus of Control	2 = long (initial) version

Appendix C.1.3. "End" Group

		SE1	TM1	SM1	P1	SR1	M1	PAp1	PAv1	D1	PTD1	ILC1	ELC1	S1	InS1
N		8	8	8	8	8	8	8	8	8	8	8	8	8	8
Normal Parameters a, b	Mean	4.583	2.563	3.188	3.938	3.229	4.292	3.083	2.750	1.625	3.063	4.813	2.875	4.063	3.625
	Std. Dev.	0.496	0.776	0.594	0.678	0.398	0.603	1.282	1.069	0.677	0.863	0.594	0.791	0.563	0.744
Most Extreme Differences	Absolute	0.300	0.218	0.326	0.241	0.188	0.233	0.221	0.283	0.225	0.236	0.249	0.241	0.216	0.308
	Positive	0.255	0.162	0.251	0.241	0.187	0.225	0.221	0.283	0.225	0.154	0.164	0.241	0.216	0.182
	Negative	-0.300	-0.218	-0.326	-0.172	-0.188	-0.233	-0.138	-0.121	-0.178	-0.236	-0.249	-0.188	-0.159	-0.308
Kolmogorov-Smirnov Z		0.847	0.616	0.921	0.681	0.531	0.659	0.624	0.799	0.638	0.668	0.704	0.681	0.611	0.872
Asymp. Sig. (2-tailed)		0.469	0.842	0.364	0.743	0.941	0.777	0.830	0.545	0.811	0.763	0.705	0.742	0.849	0.433

		SE2	TM2	SM2	P2	SR2	M2	PAp2	PAv2	D2	PTD2	ILC2	ELC2	S2	InS2
N		8	8	8	8	8	8	8	7	7	7	7	7	7	7
Normal Parameters a, b	Mean	4.500	2.406	3.531	3.688	3.208	4.300	2.675	2.786	1.524	3.214	4.500	3.429	4.286	3.643
	Std. Dev.	0.545	0.533	0.725	0.530	0.456	0.623	1.219	1.510	0.813	0.809	0.577	1.097	0.756	0.900
Most Extreme Differences	Absolute	0.320	0.277	0.244	0.172	0.239	0.244	0.339	0.224	0.312	0.263	0.235	0.240	0.219	0.294
	Positive	0.209	0.277	0.161	0.153	0.239	0.189	0.339	0.224	0.312	0.176	0.235	0.189	0.219	0.203
	Negative	-0.320	-0.242	-0.244	-0.172	-0.158	-0.244	-0.165	-0.158	-0.260	-0.263	-0.193	-0.240	-0.149	-0.294
Kolmogorov-Smirnov Z		0.906	0.783	0.689	0.486	0.675	0.691	0.960	0.592	0.825	0.695	0.623	0.636	0.579	0.778
Asymp. Sig. (2-tailed)		0.384	0.572	0.730	0.972	0.753	0.726	0.316	0.875	0.505	0.719	0.833	0.814	0.891	0.580

a Test distribution is Normal.
b Calculated from data.

SE = Self-efficacy	SR = Self Regulation	D = Disengaged	S = Stability
TM = Time Management	M = Mastery Goal	PTD = Perceived Task Difficulty	InS = Instability
SM = Self Management	PAp = Performance Approach	ILC = Internal Locus of Control	1 = short version;
P = Persistency	PAv = Performance Avoidance	ELC = External Locus of Control	2 = long (initial) version

Appendix C.1.4. All participants

		SE1	TM1	SM1	P1	SR1	M1	PAp1	PAv1	D1	PTD1	ILC1	ELC1	S1	InS1
N		25	26	26	26	26	26	26	25	25	25	25	25	25	25
Normal Parameters [a, b]	Mean	4.413	2.808	3.385	3.808	3.333	4.333	3.173	2.640	1.667	3.280	5.020	3.220	4.380	3.860
	Std. Dev.	0.640	0.917	0.864	0.763	0.485	0.516	1.204	1.075	0.933	0.855	0.714	0.969	0.781	0.797
Most Extreme Differences	Absolute	0.180	0.148	0.139	0.195	0.100	0.171	0.097	0.209	0.243	0.200	0.169	0.150	0.161	0.230
	Positive	0.180	0.148	0.139	0.195	0.077	0.164	0.096	0.209	0.243	0.148	0.151	0.150	0.150	0.230
	Negative	-0.180	-0.121	-0.130	-0.126	-0.100	-0.171	-0.097	-0.116	-0.237	-0.200	-0.169	-0.130	-0.161	-0.206
Kolmogorov-Smirnov Z		0.901	0.753	0.710	0.995	0.510	0.871	0.493	1.045	1.213	1.001	0.844	0.749	0.805	1.152
Asymp. Sig. (2-tailed)		0.391	0.622	0.695	0.275	0.957	0.434	0.968	0.225	0.106	0.269	0.474	0.629	0.536	0.141

		SE2	TM2	SM2	P2	SR2	M2	PAp2	PAv2	D2	PTD2	ILC2	ELC2	S2	InS2
N		24	24	23	23	24	23	23	22	22	22	22	22	22	22
Normal Parameters [a, b]	Mean	4.517	2.760	3.522	3.674	3.358	4.278	2.583	2.807	1.379	3.341	4.727	3.295	4.409	3.614
	Std. Dev.	0.662	0.845	0.730	0.689	0.654	0.578	1.141	1.026	0.557	0.662	0.703	0.854	0.718	0.801
Most Extreme Differences	Absolute	0.267	0.180	0.144	0.166	0.131	0.199	0.190	0.198	0.343	0.250	0.263	0.181	0.141	0.147
	Positive	0.233	0.180	0.090	0.166	0.131	0.120	0.190	0.198	0.343	0.197	0.263	0.181	0.131	0.147
	Negative	-0.267	-0.112	-0.144	-0.102	-0.095	-0.199	-0.092	-0.112	-0.248	-0.250	-0.146	-0.137	-0.141	-0.131
Kolmogorov-Smirnov Z		1.310	0.882	0.693	0.795	0.640	0.953	0.909	0.929	1.608	1.170	1.235	0.848	0.663	0.691
Asymp. Sig. (2-tailed)		0.065	0.418	0.724	0.553	0.807	0.324	0.380	0.354	0.011	0.129	0.095	0.469	0.772	0.726

a Test distribution is Normal.
b Calculated from data.

SE = Self-efficacy	SR = Self Regulation	D = Disengaged	S = Stability
TM = Time Management	M = Mastery Goal	PTD = Perceived Task Difficulty	InS = Instability
SM = Self Management	PAp = Performance Approach	ILC = Internal Locus of Control	1 = short version;
P = Persistency	PAv = Performance Avoidance	ELC = External Locus of Control	2 = long (initial) version

Appendix C.2. Paired t-test

group = beginning and end

Paired Samples Statistics[a]

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Self efficacy 1	4.5000	8	.43644	.15430
	Self efficacy 2	4.7000	8	.42762	.15119

a. group = beginning and end

Paired Samples Correlations[a]

		N	Correlation	Sig.
Pair 1	Self efficacy 1 & Self efficacy 2	8	.765	.027

a. group = beginning and end

Paired Samples Test

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Self efficacy 1 Self efficacy 2	-.20000	.29601	.10465	-.44747	.04747	-1.911	7	.098

a. group = begining and end

group = middle

Paired Samples Statistics[a]

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Self efficacy 1	4.2083	8	.92475	.32695
	Self efficacy 2	4.3500	8	.94264	.33327

a. group = middle

Paired Samples Correlations[a]

		N	Correlation	Sig.
Pair 1	Self efficacy 1 & Self efficacy 2	8	.942	.000

a. group = middle

Paired Samples Test[a]

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Self efficacy 1 Self efficacy 2	-.14167	.31761	.11229	-.40719	.12386	-1.262	7	.248

a. group = middle

group = end

Paired Samples Statistics[a]

		Mean	N	Std. Deviation	Std. Error Mean
Pair 1	Self efficacy 1	4.5833	8	.49602	.17537
	Self efficacy 2	4.5000	8	.54511	.19272

a. group = end

Paired Samples Correlations[a]

		N	Correlation	Sig.
Pair 1	Self efficacy 1 & Self efficacy 2	8	.951	.000

a. group = end

Paired Samples Test[a]

		Paired Differences					t	df	Sig. (2-tailed)
		Mean	Std. Deviation	Std. Error Mean	95% Confidence Interval of the Difference Lower	Upper			
Pair 1	Self efficacy 1 Self efficacy 2	.08333	.16997	.06009	-.05876	.22543	1.387	7	.208

a. group = end

Appendix C.3. One-way ANOVA test

ANOVA

		Sum of Squares	df	Mean Square	F	Sig.
Self efficacy 1	Between Groups	1.187	3	.396	.837	.482
	Within Groups	18.910	40	.473		
	Total	20.097	43			
Self efficacy 2	Between Groups	10.029	3	3.343	6.964	.001
	Within Groups	18.721	39	.480		
	Total	28.750	42			

Post Hoc Tests

Multiple Comparisons

Tukey HSD

Dependent Variab	(I) group	(J) group	Mean Difference (I-J)	Std. Error	Sig.	95% Confidence Interval Lower Bound	Upper Bound
Self efficacy 1	beginning	beginning and er	-.26901	.27823	.769	-1.0148	.4768
		middle	-.03289	.28979	.999	-.8096	.7439
		end	-.40789	.28979	.502	-1.1846	.3689
	beginning and en	beginning	.26901	.27823	.769	-.4768	1.0148
		middle	.23611	.33410	.894	-.6594	1.1316
		end	-.13889	.33410	.975	-1.0344	.7566
	middle	beginning	.03289	.28979	.999	-.7439	.8096
		beginning and er	-.23611	.33410	.894	-1.1316	.6594
		end	-.37500	.34379	.697	-1.2965	.5465
	end	beginning	.40789	.28979	.502	-.3689	1.1846
		beginning and er	.13889	.33410	.975	-.7566	1.0344
		middle	.37500	.34379	.697	-.5465	1.2965
Self efficacy 2	beginning	beginning and er	-1.13158*	.29201	.002	-1.9151	-.3480
		middle	-.78158	.29201	.051	-1.5651	.0020
		end	-.93158*	.29201	.014	-1.7151	-.1480
	beginning and en	beginning	1.13158*	.29201	.002	.3480	1.9151
		middle	.35000	.34642	.744	-.5796	1.2796
		end	.20000	.34642	.938	-.7296	1.1296
	middle	beginning	.78158	.29201	.051	-.0020	1.5651
		beginning and er	-.35000	.34642	.744	-1.2796	.5796
		end	-.15000	.34642	.972	-1.0796	.7796
	end	beginning	.93158*	.29201	.014	.1480	1.7151
		beginning and er	-.20000	.34642	.938	-1.1296	.7296
		middle	.15000	.34642	.972	-.7796	1.0796

*. The mean difference is significant at the .05 level.

Homogeneous Subsets

Self efficacy 1

Tukey HSD[a,b]

group	N	Subset for alpha = .05
		1
beginning	19	4.1754
middle	8	4.2083
beginning and end	9	4.4444
end	8	4.5833
Sig.		.566

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 9.668.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Self efficacy 2

Tukey HSD[a,b]

group	N	Subset for alpha = .05	
		1	2
beginning	19	3.5684	
middle	8	4.3500	4.3500
end	8		4.5000
beginning and end	8		4.7000
Sig.		.086	.696

Means for groups in homogeneous subsets are displayed.

a. Uses Harmonic Mean Sample Size = 9.354.

b. The group sizes are unequal. The harmonic mean of the group sizes is used. Type I error levels are not guaranteed.

Appendix D. List of published papers

Cocea, M. (2006a). Assessment of motivation in online learning environments. In: V. Wade, H. Ashman, B. Smith (Eds.) *Adaptive Hypermedia and Adaptive Web-Based Systems, 4th International Conference AH 2006 Proceedings*, Springer-Verlag, 414-418.

Cocea, M. (2006b). Extendibility of Educational Systems to Include a Learner-Adaptive Motivational Module. In Radu Vasile, Risto Kimari, Diana Andone (Eds.). *The 12th NETTIES (Networking Entities) International Conference: The Future of E:Advanced Educational Technologies for a Future e-Europe*, Ed. Orizonturi Universitare, Timisoara, 195-198.

Cocea, M., Weibelzahl, S. (2006a). Can Log Files Analysis Estimate Learner's Level of Motivation? In Schaaf, M. and Althoff, K-D. *LWA 2006. Lernen-Wissensentdeckung – Adaptivität.* Universität Hildesheim, 32-35.

Cocea, M., Weibelzahl, S. (2006b). Motivation – Included or Excluded From e-Learning. In Kinshuk, Sampson, D. G., Spector J.M., Isaias, P., *Cognition and Exploratory Learning in Digital Age, CELDA 2006 Proceedings*, 435-437.

Cocea, M., Weibelzahl, S. (2007a). Eliciting motivation knowledge from log files towards motivation diagnosis for Adaptive Systems. In Conati, C., McCoy, K., Paliouras, G. (eds.) *User Modelling 2007. Proceedings of 11th International Conference, UM2007, Lecture Notes in Artificial Intelligence (LNAI)*, vol. 4511, 197-206, Springer, Berlin.

Cocea, M. (2007). Learning Engagement: What Actions of Learners Could Best Predict It? *In Proceedings of AIED 2007, 13th International Conference on Artificial Intelligence in Education* (Young Researcher's Track; online proceedings).

Cocea, M., Weibelzahl, S. (2007b). Cross-System Validation of Engagement Prediction from Log Files. In Duval, E., Klamma, R. and Wolpers, M. (Eds.) *Creating New Learning Experiences on a Global Scale, Second European Conference on Technology Enhanced Learning, EC-TEL 2007, Crete, Greece, September 2007 Proceedings, Lecture Notes in Computer Science (LNCS)*, vol. 4753, 14-25, Springer Berlin/ Heidelberg.