Karl LEDERMÜLLER¹ & Irmgard FALLMANN (Vienna)

Predicting learning success in online learning environments: Self-regulated learning, prior knowledge and repetition

Abstract

The emergence of new trends sometimes carries the risk that established, well-proven concepts rooted in other disciplines are not properly integrated into new approaches. As Learning Analytics seems to be evolving into a highly multidisciplinary field, we would like to demonstrate the importance of embedding classic theories and concepts into a Learning Analytics, system-data-driven setting. Our results confirm that classical factors that are operationalized with the help of system-generated data outperform more recent survey-based models. Therefore, we want to stress the point that system-generated data should not be left behind in the quickly evolving field of Learning Analytics.

Keywords

Repetition, memory, prior knowledge, self-regulated learning, learning effectiveness

¹ email: karl.ledermueller@wu.ac.at

Scientific Contribution
1 Introduction

One of the major aims of Learning Analytics is to investigate learning effectiveness in online learning environments. Questions related to this field have been discussed for a rather long time. However, the focus on the learner (UĞUR et al., 2009), educational massification (and the related demand for highly scalable, online-supported learning environments such as MOOCs) and trends such as lifelong learning generate a wide range of learning-related data that can be used to find different ways to meet challenges related to learning effectiveness.

Various disciplines (e.g. educational psychology) have been investigating effects on learning effectiveness for the last 100 years. Two very influential factors in educational psychology are prior knowledge and time invested in the learning process. In addition, a wide range of learning strategies is particularly important in online learning environments with a high degree of self-regulation. Self-regulated learning (SRL) has become a major topic in education and is widely discussed by researchers as well as by teachers and educators. Distance education or blended learning scenarios are of great help in reaching a huge number of students, independent of age and location. This fact becomes increasingly important when combined with lifelong learning and educational massification supported by the increasing use of technology (see e.g. BRYANT et al., 2005). Furthermore, the use of technology in education offers a more flexible way of learning and takes different learning styles into account. In accordance with constructivism, learning is described as an active, constructive, emotional and self-regulated process (KOPP & MANDL, 2009). Knowledge transfer is no longer primarily triggered by professors, but rather the students must actively acquire knowledge. Since the students themselves are responsible for a successful learning process, skills in self-regulation are an essential prerequisite.

In our research, we focused on the students’ self-regulated learning skills in the course “Accounting and Management Controll II (AMCII)”, which is supported by a wide range of online materials. We explored whether the online exercises had any influence on achievement. The influence of the learning strategies was controlled.
by classical factors (see chapter 2.4.) that predict learning effectiveness (e.g. prior knowledge, learning time invested), derived from system-generated data related to Learning Analytics. In chapter two, we briefly explain the widely investigated effect on learning time and prior knowledge and then go on to provide the theoretical background for self-regulation by explaining Zimmerman’s phase model of self-regulation. Chapter three explains the research rationale, followed by an explanation of the research design in chapter four. Chapter 5 then presents the results, and chapter 6 offers a conclusion.

The following publication can be seen as a partial outcome of a research project focusing on learning effectiveness within an unstructured learning environment (see also: FALLMANN & LEDERMÜLLER, 2016).

2 Theoretical background

2.1 Learning Analytics and educational data mining

Both technology-enhanced learning and the general digitalization of education generate a growing amount of data, which provides information on the educational process at different levels. Whereas administrative databases contain information such as enrolment data, grades, pre-university data (e.g. school grades) and demographics, learning management systems store more fine-grained data. Logfiles provide information about the usage of learning resources such as videos or ebooks, or participation in forum discussions (CALDERS & PECHENIZKIY, 2012). The research disciplines of educational data mining and Learning Analytics make use of the available data in order to support learning and improve educational systems.

Learning Analytics uses learner-generated data and combines them with an analysis model to predict student progress and performance. The acquired information is used to adapt the e-learning environment to support and improve individual learning.
SIEMENS (2010) goes a step further and argues for the concepts behind educational data mining. He suggests that Learning Analytics can not only be used to improve existing educational systems, but can also trigger a modification of the system as a whole. Learning Analytics can help universities to identify difficulties with learner performance and consequently adapt their programmes. Since a holistic approach is essential, it is necessary to adapt not only the e-learning environment, but also aspects related to the curriculum and pedagogical factors (SIEMENS, 2010).

Both educational data mining and Learning Analytics are emerging multidisciplinary research areas which have the potential to find solutions to improve learning. When new trends develop, there is a risk that established, well-proven concepts rooted in other disciplines are not properly integrated into new approaches. Since Learning Analytics seems to be developing into a highly multidisciplinary field, we would like to show the importance of classical theories and concepts. Effects that easily can be measured with the help of system-generated data drawn from e-learning environments could be theoretically grounded in basic concepts from other disciplines.

2.2 Learning, repetition and memory

It is widely accepted that memory and repetition play a major role in learning effectiveness. EBBINGHAUS (1885) has shown that repetition of content leads to increased memorization rates of content functionally described in the learning/forgetting curve. ANDERSON (2000) describes the historical development of learning and memory. Memory is not only connected with repetition and time between repetition cycles, but is also highly interlinked with a range of different factors. CRAIK & LOCKHART (1972), for example, introduce the idea of depth of processing, which strongly is strongly related to the idea of prior knowledge (see chapter below). The depth of processing and embedding into prior knowledge in turn highly correlate with the memorization of content. WICKELGREN (1981) and LIEBERMAN (2011) give comprehensive overviews of different theories and influencing factors related to learning and memory.
The main didactical tool of Accounting and Management Control II (AMCII) was learning from ‘worked examples’, a teaching (and learning) method specifically found in areas such as mathematics, physics, statistics and computer programming. ATKINSON et al. (2000) describe worked examples as follows: “they typically include a problem statement and a procedure for solving the problem.” By working through worked examples, students should construct a schema which can help them to solve similar problems. According to SWELLER et al (1998), knowledge is stored in long-term memory in the form of schemata. The acquisition of schemata is an active, constructive process. Older findings from research on worked example showed that practice is one of the most important predictors for the acquisition of skills and schemata (ATKINSON et al., 2000). VAN ENGEN (1959), a mathematics education professor, declared, “the best way to teach children how to solve problems is to give them lots of problems to solve.” We measured our study time variable by the logarithm of the number of quizzes students solved in the e-learning environment of AMCII. A quiz consists of a problem statement, five alternative answers and an automated feedback, which illustrates a correct way of solving the problem. There are approximately 450 quizzes available.

2.3 Prior knowledge

There are other factors beyond having a technical learning process that considers time issues that influence learning effectiveness. AUSUBEL (1978, p. 235) states that prior knowledge has a huge impact on learning effectiveness. “If I had to reduce all of educational psychology to just one principle, I would say this: the most important single factor influencing learning is what the learner already knows. Ascertain this and teach him accordingly”. According to the cognitive load theory, a successful learning process connects new information to existing knowledge. If there is no relevant prior knowledge available, the learner has to search randomly for a solution. This is primarily done via trial and error, which imposes an unnecessary cognitive load (SWELLER, 1998). Therefore, sufficient prior knowledge maximizes the effectiveness of any learning environment. SONG, KALET & PLASS (2015) have demonstrated a direct positive effect of prior knowledge on learning.
outcomes in a complex multimedia learning environment. Hailikari, Nevgi and Lindblom-Ylanne (2007) developed a model of prior knowledge that differentiates between declarative knowledge (knowledge of facts and meanings) and procedural knowledge (integration and application of knowledge). A survey with 115 pharmacy students showed a significant influence of prior knowledge on student achievement, with procedural prior knowledge being particularly important (HAILIKARI, KATAJAVUORI & LINDBLOM-YLANNE; 2008).

2.4 Theories on self-regulated learning

Learning effectiveness – especially in an online setting that offers high degrees of freedom in terms of learning structure – is highly affected by the self-regulation of the learners. According to BOEKAERTS (1999), “Self-regulation means being able to develop knowledge, skills and attitudes which can be transferred from one learning context to another and from learning situations to a leisure and work context”. The ability to self-regulate one’s own learning process is a key factor for successful learning. Many researchers have developed various models on SRL. Although the structures of these models differ, most of them are based on three basic schools: (1) research on learning styles, (2) research on metacognition and regulation styles, and (3) theories of the self, including goal-directed behaviour.

The three-layered model of self-regulated learning (BOEKAERTS, 1999) (see figure 1) integrates these three dimensions and emphasizes the interaction between the layers.
Figure 1: The three-layered model of self-regulated learning (BOEKAERTS, 1999)

ZIMMERMAN (2000) follows a similar approach. He describes self-regulation as “self-generated thoughts, feelings and actions that are planned and cyclically adapted to the attainment of personal goals”. Self-regulation is described as a process of adaptation consisting of three cyclical phases: (1) forethought, (2) performance/volitional control und (3) self-reflection (see table 1).
Table 1: Phase structure and sub-processes of self-regulation (ZIMMERMAN, 2000)

<table>
<thead>
<tr>
<th>FORETHOUGHT</th>
<th>PERFORMANCE/ VOLITIONAL CONTROL</th>
<th>SELF-REFLECTION</th>
</tr>
</thead>
<tbody>
<tr>
<td>Task analysis</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Goal setting</td>
<td><em>Self-control</em></td>
<td><em>Self-judgment</em></td>
</tr>
<tr>
<td>Strategic planning</td>
<td><em>Self-instruction</em></td>
<td><em>Self-evaluation</em></td>
</tr>
<tr>
<td></td>
<td><em>Imagery</em></td>
<td><em>Causal attribution</em></td>
</tr>
<tr>
<td></td>
<td><em>Attention focusing</em></td>
<td></td>
</tr>
<tr>
<td></td>
<td><em>Task strategies</em></td>
<td></td>
</tr>
<tr>
<td>Self-motivation beliefs</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Self-efficacy</td>
<td><em>Self-observation</em></td>
<td><em>Self-reaction</em></td>
</tr>
<tr>
<td>Outcome expectations</td>
<td><em>Self-recording</em></td>
<td><em>Self-satisfaction/ affect</em></td>
</tr>
<tr>
<td>Intrinsic interest/value</td>
<td><em>Self-experimentation</em></td>
<td></td>
</tr>
<tr>
<td>Goal orientation</td>
<td></td>
<td>Adaptive-defensive</td>
</tr>
</tbody>
</table>

In the first phase (*forethought*), students plan their learning process. Besides the planning of learning activities and goal setting, self-motivational beliefs (e.g. self-efficacy, outcome expectations, intrinsic interest, and goal orientation) play a major role. A learner who is highly self-regulated will define hierarchically organized goals and will be highly motivated to reach them. Thereby, self-efficacy is essential, since the more students believe they will be able to solve the task, the better they will perform, as they will invest more effort into reaching their goals.

The second phase (*performance/volitional control*) describes the learning process itself. The self-control phase concentrates on the application of cognitive strategies,
such as memorizing, critical thinking, and elaboration. Self-observation describes one’s ability to track one’s own performance and devise corrective measures, if necessary.

In the third phase (*self-reflection*), the learning process as a whole is evaluated. In the self-judgment phase, students compare their results with a standard goal and search for explanations for deviations. The self-reaction phase describes the ability to deal with disappointing results and to search for necessary adaptations in the learning process. This phase directly influences the self-motivational belief phase, which shows the cyclical nature of the self-regulation process. Students with high self-efficacy, for instance, are more likely to attribute bad marks to insufficient effort or an inadequate strategy. In the next learning process, they will try to act differently, but they remain confident in their ability to reach the goal. In contrast, students who are self-doubters interpret poor results as a confirmation of their limited abilities and therefore do not believe that they will reach their goals in the next learning process either.

### 3 Research objectives

Generally speaking, there is an observable trend towards learning environments that (theoretically) react with some degree of flexibility to the learners’ usage patterns or learner-related factors (e.g. sociodemographic variables). Learning Analytics must therefore develop a wide range of solutions on both the theoretical and conceptual levels.

This study addresses the predictive power of classical descriptive parameters in a high-dimensional model. We explored whether self-regulated learning skills influence the results on the final exam of the “Accounting and Management Control II (AMCII)” course at the Vienna University of Economics and Business (WU). To compare this model with classical concepts operationalized by system-generated data, we included the concepts of repetition (number of quizzes solved) and memory (learning time) and prior knowledge.
4 Method

4.1 Research setting

The research participants in this study were selected from the “Accounting and Management Control II“ (AMCII) course, which is compulsory for all bachelor students at the Vienna University of Economics and Business (WU). Each term, approximately 1,000 students are registered for the course. Students are assessed at the end of the term via a highly standardized multiple choice exam. Our survey is based on the exam results from the year 2010.

The main instructional devices are worked-examples. In a large lecture format, lecturers present problem statements and a procedure for solving them. There are 11 face-to-face units offered across the whole term. Attendance at the lectures is voluntarily. Additionally, students can use the e-learning environment Learn@WU, which enables distance learning. It includes course information, quizzes with sample solutions, online test exams, additional downloads, lecture videos and a glossary. Students are supposed to gain a theoretical background through studying printed reading material and applying their knowledge by solving the online quizzes. If they have questions, they can either search for explanations in the lecture videos or communicate with an expert or their peers via a moderated discussion forum. The e-learning environment is used very frequently, especially right before an exam. A high percentage of students prepare themselves only via the e-learning environment, which requires strong self-regulated learning skills.

4.2 Research participants

In order to participate in the exam, students are required to register via an online tool. The population for this survey consisted of 801 students who registered for the exam at the end of the term, with 110 individuals participating in the study (response rate = 14%). The sample consisted of 62 females and 48 males, who varied in age between 19 and 49, with a mean age of 23 years (standard devia-
tion=4.75, median=21). All students are familiar with the e-learning platform Learn@WU, as they had to pass several other e-supported courses before being allowed to write the AMCII exam. Sixty-four individuals (58%) stated that they attended class on a regular basis, while 46 (42%) prepared for the exam exclusively via self-study.

4.3 Measures

This section discusses the measurement of the variables used in our model.

*Self-regulated-learning* was surveyed via an online questionnaire. Students had to express a level of agreement with statements concerning their learning behaviour on a 5-point Likert scale. The items were summarised into scales, which reflect Zimmerman’s phases of self-regulation. The scales were derived from various sources (see table 2).

The main instrument was the inventory for acquisition of learning strategies in tertiary education (LIST) developed by WILD et al. (1994). The questionnaire included 11 scales divided into three categories: cognitive strategies, metacognitive strategies and resource management strategies. The scales were developed based on the Motivated Strategies for Learning Questionaire (MSLQ) by PINTRICH et al. (1991). In this study, the focus was solely on the scales concerning self-regulated learning in accordance with Zimmerman’s model.

Zimmerman’s self-reflection phase was measured by two scales developed by WOSNITZA (2002). Self-motivational beliefs were measured with scales developed by WAGNER et al. (2010).

Table 2 shows the scales used in the survey:
Table 2: Listing of scales

<table>
<thead>
<tr>
<th>Phase of self-regulated learning (Zimmerman)</th>
<th>Scale</th>
<th>Nr. of items</th>
<th>Source</th>
</tr>
</thead>
<tbody>
<tr>
<td>FORETHOUGHT</td>
<td>interest</td>
<td>3</td>
<td>Wagner et.al</td>
</tr>
<tr>
<td></td>
<td>self-efficacy</td>
<td>2</td>
<td>Wagner et.al</td>
</tr>
<tr>
<td>PERFORMANCE</td>
<td>repetition strategy</td>
<td>2</td>
<td>Wild</td>
</tr>
<tr>
<td></td>
<td>organisation</td>
<td>4</td>
<td>Wild</td>
</tr>
<tr>
<td></td>
<td>time management</td>
<td>3</td>
<td>Wild</td>
</tr>
<tr>
<td></td>
<td>elaboration</td>
<td>4</td>
<td>Wild</td>
</tr>
<tr>
<td></td>
<td>critical thinking</td>
<td>3</td>
<td>Wild</td>
</tr>
<tr>
<td></td>
<td>peer learning</td>
<td>4</td>
<td>Wild</td>
</tr>
<tr>
<td>SELF-REFLECTION</td>
<td>helplessness</td>
<td>3</td>
<td>Wagner et.al</td>
</tr>
<tr>
<td></td>
<td>self-reaction</td>
<td>3</td>
<td>Wosnitza</td>
</tr>
</tbody>
</table>

The study time/repetition variable was measured by analysing the log files of the e-learning platform. Any time throughout the term when a student accesses learning materials on the platform, this access was logged. Therefore, study time was approximated by the number of clicks in the environment. The main learning activity in AMCII consists of solving problem statements. The more examples students work on, the better their ability to apply theoretical knowledge should be. We used the logarithm of the number of solved examples to account for the “saturation” of
learning efforts. All studies on learning curves show a (more or less) logarithmic function (EBBINGHAUS, 1885). NETTEKOVEN & LEDERMUELLER (2011) also showed that additional learning effort in an e-learning environment has a certain saturation effect. In other words, solving the thousandth example has less learning effect than solving the hundredth.

Before students can attend AMCII, they have to pass Accounting and Management I (AMCI). Similar to AMCII, students have to pass a multiple-choice exam in AMCI. As a variable for prior knowledge, we used the test scores achieved in AMCI.

The main goal of our study was to measure the effect of the variables described above on the learning achievement. Our dependent variable was measured by the test scores of the AMCII final exam.

4.4 Analysis

In our study, we used structure equation modelling\(^2\) in order to predict the test scores of the final exam. In contrast to classical regression models, structural equation models integrate dependencies between (latent) variables, which have to be taken into consideration in modelling human behaviour. The model includes self-regulated learning skills, learning time and prior knowledge.

Calculations were established with the open-source statistical environment R (R DEVELOPMENT CORE TEAM, 2011). For structural equation modelling and visualization, we used the R packages lavaan (ROSSEL, 2012), semTools (SEM-TOOLS CONTRIBUTORS, 2016) and semPlots (EPSKAMP, 2013).

We formulated our structural equation model in the following form:

The following equations describe the formulation of the latent variables related to the self-regulated learning concept...

elaboration =~ elaboration02 + elaboration03 + elaboration04 + elaboration05
helplessness =~ helplessness01 + helplessness02 + helplessness03
interest =~ interest01 + interest02 + interest03
self-efficacy =~ self-efficacy01 + self-efficacy02
critical thinking =~ critical thinking01 + critical thinking02 + critical thinking03
peer learning =~ peer learn01 + peer learn02 + peer learn03 + peer learn04
repStrat =~ repStrat01 + repStrat03
organisation =~ organisation01 + organisation02 + organisation03 + organisation04
self-reaction =~ self-reaction01 + self-reaction02 + self-reaction03
time management =~ time manag01 + time manag02 + time manag03

...which were regressed against the scores achieved on the final exam.

scores on final exam ~ elaboration
scores on final exam ~ helplessness
scores on final exam ~ interest
scores on final exam ~ self-efficacy
scores on final exam ~ critical thinking
scores on final exam ~ peer learning
scores on final exam ~ repStrat
scores on final exam ~ organisation
scores on final exam ~ self-reaction
scores on final exam ~ time management

# to model prior knowledge, we regressed the score of AMC I against the score of AMC II. To include the concept of repetition, we included the logarithm of the quantity of solved exercises on WU’s online platform.
scores on final exam ~ study time (number of quizzes solved)
scores on final exam ~ scores AMCI

5 Results

Table 3 shows the result of the regressions underlying the structural equation model. The Comparative Fit Index (CFI) and the Tucker-Lewis Index (TLI) are slightly below 0.9, which is due to the size of the variables in the model. Linear regression of the data with aggregated factors (FALLMANN & LEDERMÜLLER, 2016) shows similar results in a linear regression model, with an $R^2$ of 0.3496.

As table 3 indicates, only the concepts of ‘repetition’ (number of quizzes solved) and ‘prior knowledge’ (which were based on system-generated data and have a very long tradition in educational psychology) show strong significant effects on the score received on the final exam. Self-regulated learning strategies seem to have no significant influence on test scores (when including repetition and prior knowledge in the model). The variables ‘helplessness’ and ‘self-reaction’ show an effect with a statistical significance between 0.1 and 0.15, whereby it could be argued that higher sample sizes could lead to significant effects.
Repetition, as measured by the number of questions answered in WU’s online environment, has a significant effect on learning effectiveness. Our findings confirm VAN ENGEN’s statement (1959): “the more problems students solve, the better their test scores.”

Prior knowledge, as was measured by the test scores from AMCI, also shows a significant influence in our model. Each one-point increase in the AMCI test score predicted a 0.568 increase in the test scores of AMCII. The teaching method, assessment and topics of AMCI are quite similar to those of AMCII. Students with high achievement in AMCI will have developed helpful schemata for problem solving, which also helps them succeed in AMCII. The importance of prior knowledge can also be detected in ZIMMERMAN’s cyclical model (2000). Feedback from prior performance is essential in the cyclical process of self-regulation, as it helps to make adjustments during current learning efforts. If learning strategies led to high achievements in AMCI, students may stick to them when preparing for AMCII and succeeded there as well. Furthermore, positive results in AMCI could be quite motivating for preparation in AMCII.

Table 3: Regression output of the SEM

| Regression                      | Estimate | Std.Err | Z-value | P(>|z|) |
|---------------------------------|----------|---------|---------|---------|
| Scores on final exam            |          |         |         |         |
| elaboration                     | -4.870   | 17.252  | -0.282  | 0.778   |
| helplessness                    | -9.426   | 6.477   | -1.455  | 0.146   |
| interest                        | -0.892   | 4.408   | -0.202  | 0.840   |
| self-efficacy                   | 9.325    | 8.228   | 1.133   | 0.257   |
| critical thinking               | 8.404    | 21.648  | 0.388   | 0.698   |
| peer learning                   | -1.326   | 3.077   | -0.431  | 0.667   |
| repStrat                        | 3.607    | 4.917   | 0.734   | 0.463   |
| organisation                    | 5.275    | 5.885   | 0.896   | 0.370   |
| self-reaction                   | -34.200  | 22.676  | -1.508  | 0.131   |
| time management                 | -0.334   | 5.261   | -0.064  | 0.949   |
| study time (clicks)             | 0.020    | 0.005   | 4.092   | 0.000   |
| prior knowledge (AMCI)          | 0.568    | 0.179   | 3.169   | 0.002   |
6 Conclusions and limitations

Our study shows that Learning Analytics can help to predict student learning effectiveness. Two factors representing repetition and prior knowledge had a strong impact on predicting individual learner success. These indicators are not only theoretically backed by more than a century of educational research, but are also easily (re)producible within learning environments and Learning Analytics questions.

To sum up, after prior knowledge, time spent and repetition effort in solving problems in the e-learning environment was the strongest predictor for good results on the final exam. The examples offered in the e-learning environment seem to support students in developing cognitive schemata to solve exam problems successfully. We were also able to show that prior knowledge significantly influences test scores, while self-regulated learning ability shows no significant influence on achievement. Although our model included different learning strategies, but none of them showed significant results.

We tried to find a model which describes the learning process and considers both classical factors and self-regulated learning ability. As learning is a very complex process, there may be many more system-generated variables (e.g. learning sequences, short- and long-term learning strategies and other factors) which influence achievement.

The participation in the survey was voluntary. In future research, we will focus on a higher amount of system-embedded data to increase the sample size, and we will investigate the effects in other classes. Furthermore, we will consider expanding our model with additional meaningful data, in order to explain the complex learning process in more detail.
7 References


Calders, T., & Pechenizkiy, M. (2012). Introduction to The Special Section on Educational Data Mining. ACM SIGKDD Explorations Newsletter.


Authors

Dr. Karl LEDERMUELLER  ||  WU Vienna, Evaluation and Quality Enhancement (Head)  ||  Welthandelsplatz 1, A-1020 Vienna

www.wu.ac.at
karl.ledermueller@wu.ac.at


www fh-wien.ac.at
irmgard.fallmann@fh-wien.ac.at