Learning analytics in primary, secondary and higher education
Guest Editors: Markus Ebner and Martin Ebner

Number 2
93 Editorial boards
94 Small data as a conversation starter for learning analytics: exam results dashboard for first-year students in higher education
Tom Broos, Katrien Verbert, Greet Langie, Carolien Van Soom and Tinne De Laet
107 Evaluating emotion visualizations using AffectVis, an affect-aware dashboard for students
Leony Derick, Gayane Sedrakyan, Pedro J. Munoz-Merino, Carlos Delgado Kloos and Katrien Verbert
126 Entrepreneurship students distilled their learning experience through reflective learning log
Khar Kheng Yeoh
143 Learning analytics to improve writing skills for young children – an holistic approach
Nina Steinhauer, Michael Gros, Martin Ebner, Markus Ebner, Anneliese Hoppeitz, Mike Comann, Susanne Biermeier, Lena Burk, Konstanze Edtstadler, Sonja Gabriel, Martin Wolschky, Christian Aspalter and Susanne Martich
160 On predicting academic performance with process mining in learning analytics
Rahila Umer, Sae Siong, Anuradha Mathrani and Suradi Burdai
EDITORIAL ADVISORY BOARD

Dr Diana Bilimoria
Professor, Chair of the Department of Organizational Behavior, Weatherhead School of Management, Case Western Reserve University, Cleveland, OH, USA

Dr Marius Boboc
Professor of Education and Vice Provost for Academic Planning, Cleveland State University (CSU), Cleveland, OH, USA

Dr Curtis Bonk
Professor of Psychology and Technology, Instructional Systems Technology, Indiana University, IN, USA

Dr Peter Doolittle
Assistant Provost for Teaching and Learning, Executive Director for Instructional Development and Educational Research, Virginia Polytechnic Institute and State University, USA

Dr Martin Ebner
Professor, Head of the Department of Educational Technology, Graz University of Technology, Austria

Dr Sara Kelly
Associate Professor, Chair of the Department of Professional Studies, National University, USA

Tammy Lakes
USA Funds, Vice-President, Product Innovation and Management, USA

Dr Mimi Miyoungh Lee
Associate Professor, Curriculum and Instruction, University of Houston, USA

Dr Marilee Bresciani Ludvik
Professor, Department of Administration, Rehabilitation, and Postsecondary Education, San Diego State University, San Diego, CA, USA

Dr Hermann Maurer
Professor, Graz University of Technology, Member of European Academy of Sciences, Austria

Dr Rollin Nordgren
Professor, Chair, Department Educational Administration and School Counseling/Psychology, Sanford College of Education, National University, USA

Dr Peter Serdyukov
Professor, Department of Teacher Education, Sanford College of Education, National University, USA

Dr Todd Zakrajsek
Associate Professor, Department of Family Medicine; Associate Director, Fellowship Programs, University of North Carolina at Chapel Hill, USA

Dr Yong Zhao
Professor, School of Education, University of Kansas, USA

EDITORIAL REVIEW BOARD

Dr Shoba Bandi-Rao
CUNY, USA

Dr Marius Boboc
Cleveland State University, USA

Dr Tatia Codreanu (Wa)
French Institute, London, United Kingdom

Dr Chris Devers
Indiana Wesleyan University, USA

Dr Dee Fabry
National University, USA

Dr Ron Germaine
National University, USA

Dr Joachim Paul Hasebrook
Zeb Business School, Germany

Dr Jungwon Hur
Auburn University, USA

Dr Joyce Hwe Ling Koh
National Institute of Education, Singapore, USA

Dr Miguel Lara
Cal State Monterey Bay University, USA

Dr Dabae Lee
Houston State University, USA

Dr Mimi Lee
University of Houston, USA

Dr Tracey Messer
Case Western Reserve University, USA

Dr Jason Ohler
Full Sail University, USA

Dr Dilani Perea-Diltz
University of Houston-Spring, USA

Dr Latha Poonamallee
Michigan Tech University, USA

Dr Linda Purrington
Pepperdine University, USA

Dr Najia Sabir
Indiana University, USA

Dr Cynthia Schubert-Irastorza
National University, USA

Dr Maureen Spellman
Xavier University, Chicago, USA
Small data as a conversation starter for learning analytics
Exam results dashboard for first-year students in higher education

Tom Broos
Department of Computer Science, Katholieke Universiteit Leuven, Leuven, Belgium and
Department of Tutorial services, Faculty of Engineering Science, Katholieke Universiteit Leuven, Leuven, Belgium

Katrien Verbert
Department of Computer Science, Faculty of Engineering Science, Katholieke Universiteit Leuven, Leuven, Belgium

Greet Langie
Faculty of Engineering Technology, Katholieke Universiteit Leuven, Leuven, Belgium

Carolien Van Soom
Faculty of Science, Katholieke Universiteit Leuven, Leuven, Belgium, and

Tinne De Laet
Faculty of Engineering Science, Katholieke Universiteit Leuven, Leuven, Belgium

Abstract
Purpose – The purpose of this paper is to draw attention to the potential of “small data” to complement research in learning analytics (LA) and to share some of the insights learned from this approach.
Design/methodology/approach – This study demonstrates an approach inspired by design science research, making a dashboard available to n = 1,905 students in 11 study programs (used by n = 887) to learn how it is being used and to gather student feedback.
Findings – Students react positively to the LA dashboard, but usage and feedback differ depending on study success.
Research limitations/implications – More research is needed to explore the expectations of a high-performing student with regards to LA dashboards.
Originality/value – This publication demonstrates how a small data approach to LA contributes to building a better understanding.

Keywords Higher education, Case study, First-year students, Learning analytics, Dashboard, Small data

Paper type Research paper

Introduction
Learning analytics (LA) is a growing domain of research that is on the one hand enthusiastically exploring the potential of big data, but on the other hand provoking more and more critical considerations, for instance about ethics and privacy. Referring to maturity models in
information systems, we argue that LA research involving big data can be complemented by a more modest approach based on “small data” that is readily available within institutions. This approach makes it possible to learn about expectations regarding, and challenges for LA within the short term. We demonstrate this approach by presenting an evaluation study with $n = 1,905$ first-year students, offering them a dashboard for their exam results.

A plea for small data and modest LA.
LA is defined as “the measurement, collection, analysis and reporting of data about learners and their context, for purposes of understanding and optimizing learning and the environments in which it occurs.” (Siemens and Long, 2011). In contrast to the related concept of academic analytics, its primary subject is the learning process, while the latter has a broader scope. In this perspective, learners and faculty are the primary beneficiaries of LA, while administrators, governments, and authorities are exemplary for the target audience of academic analytics (Siemens and Long, 2011). Some insist on a further distinction, equating academic analytics to a concept parallel to business analytics, supporting operational, and financial (business) decision making, but in the specific context of higher education institutions (Van Barneveld et al., 2012).

In turn, the term analytics became a nominator for techniques applied when analyzing big data (Gandomi and Haider, 2015). Big data is different from “small” data due to its volume (terabytes, petabytes), its velocity (real-time or near real-time), and variety (including both structured and unstructured data) (Laney, 2001). It is exhaustive and aims to capture entire populations ($n = \text{all}$) (Kitchin, 2013), which further elucidates the distinction between big data analytics and traditional statistical analysis that regularly involves dealing with samples limited in size.

To some authors, big data analytics preludes “the end of theory,” making not just ex ante hypothesis unnecessary – if there is a pattern in the data, the algorithms will find it anyway – but even the need to theorize ex post about underlying causation or explanations (Prensky, 2009): “Forget taxonomy, ontology, and psychology. Who knows why people do what they do? The point is they do it, and we can track and measure it with unprecedented fidelity. With enough data, the numbers speak for themselves” (Anderson, 2008). It may come to no surprise that many scientists feel uncomfortable about this notion (Kitchin, 2013; Lazer et al., 2014) and that it may contribute to a distrust of LA from educational scientists. In his book “The Black Box Society,” Pasquale (2015) portrays a dystopian view on the growing use of big data to drive all kinds of decisions. He sketches algorithms that are unfair or unaccountable, proprietary, complex, and self-fulfilling.

Apart from the analytics, the collection of massive amounts of personal data is subject of debate in itself. One concern is, of course, the level of privacy. Because of the fine-grained character of big data, several examples of re-identification attacks demonstrated the difficulty of giving any guarantees about anonymization (Boyd and Crawford, 2012; Jensen, 2013). Ever increasing computational power may make data collection that is respecting today’s privacy requirements entirely retraceable to the personal level within a few years. Attempts to make the de-identification process more robust may diminish the utility of the data (Daries et al., 2014). Again, it may not come unexpected that many, not in the least data subjects themselves, such as teachers, students, and their parents, are reluctant to the use of big data in education. In 2014, the inBloom project fell victim to “exaggerated fears and a misunderstanding about the technology” (C.K.N., 2014). The project was conceived by American educators to answer a shared problem of managing student data in a secure, standardized, and efficient way, but its failure to convince concerned parents and privacy advocates resulted in a shutdown.

Within the European legal framework, the growing scope of data collection in education should be subjected to the proportionality principle (Berendt et al., 2017): the data should serve a legitimate aim within the broader view of the role of education in society; the use of...
data should be suitable to achieve the aim, as demonstrated by rigorously constructed scientific evidence; the data should be necessary, in the sense that similar results cannot be obtained with less; its collection should be reasonable, meaning that the objectives are balanced against the loss of privacy.

Privacy became an increasingly discussed topic in the domain of LA (e.g. Gašević et al., 2016; Hoel et al., 2017; Prinsloo and Slade, 2017), not only as an obstruction but as a constraint that needs to be considered early in the design. At the same time, efforts are being made to close the gap between “traditional” research in education and the use of big data (e.g. Gibson and Ifenthaler, 2017).

In their book “Competing on Analytics,” Davenport and Harris (2007) describe an analytics maturity model with five stages ranging from the “analytically impaired” organization that collects data but is generally incapable of deriving actionable insights from it, to “analytical competitors” that are entirely built around analytics. Maturity assessment models are common practice in management science and information systems. They provide a way to determine and describe the current state of organization within a specific scope and allow for goal-setting and gap analysis. Maturity can be evaluated from different perspectives (Mettler, 2011): process/structure, object/technology, and people/culture. Maturity or sophistication models describe a growth path, hence skipping maturity stages “seldom proves to be wise. There is valuable experience to be gained at each stage […]” (Wixom and Watson, 2012). Siemens et al. (2013) proposed an “LA Sophistication Model” that starts out from simple awareness and gradually shifts into the direction of transformation at the sector level.

This paper aims to complement the big data approach to LA, a promising but also challenging endeavor, with a more “modest” implementation in practice. A less esoteric approach facilitates an immediate involvement of students, practitioners, managers, and policymakers. Each of these stakeholders can offer useful input about their expectations, experiences, and suggestions for improvement. Apart from providing valuable input for further research, this approach contributes to a gradual introduction of LA, managing change in line with the principles addressed by maturity models. This is not to say that researchers should abandon big data in LA altogether, to the contrary. Big data has a long road ahead toward general acceptance in education. It will take time before questions about proof of value are sufficiently answered to justify the collection of vast amounts of personal data for LA. Meanwhile, there is a lot to be gained from bringing more conservative applications of LA into practice.

Involving students and study counselors
Designing LA interventions is not limited to technological aspects, but includes the entire educational context and the processes surrounding it (Wise, 2014; Gašević et al., 2015). An increasing number of authors are taking a critical position toward LA (e.g. Roberts et al., 2017), some addressing the lack of knowledge about the topic among students and the need to engage students in the decision making (Roberts et al., 2017). And indeed, a lot can be learned by engaging students and practitioners in focus groups to find out about their concerns and suggestions. In this paper, we discuss an alternative route inspired by design science research (DSR) in information systems – “learning through the act of building” (Kuechler and Vaishnavi, 2008). Many things can be learned by exposing a sizable number of students to iterative implementations of learning dashboards early in the process. This ensures that the basic concepts and subsequent development of LA within the institution become conceivable and provides handlebars for further discussion.

Several types of data about students are already being collected by higher education institutions. The data are being used to support research, management, and policy, but in several (most) cases not being fed back to students in an insightful way, if at all. Doing so
now in the context of LA provides an opportunity to start the conversation about LA in a non-threatening manner and a gentle way to guide the organization into the direction of an ongoing sophistication process.

**Background**

We have put the concept of starting with small data interventions into practice. In the next session, we present the MyScores dashboard that was offered to 1905 first-year students in 11 science and engineering (STEM) oriented study programs within the University of Leuven (KU Leuven) following shortly after the publication of their exam scores.

Within the different faculties at KU Leuven, study counselors are available to support first-year students in their transition from secondary to higher education. Much experience is present in these study counseling services. Study counselors are bridge persons; their role involves the reconciliation of interests from faculty and students and we consider them well-placed advisors when it comes to acceptance of LA interventions. By involving them actively in a co-creation process – for example, by asking them to contribute the accompanying texts and to evaluate the visualizations in the dashboard discussed below – some of the experience can be captured and a friendly communication line is opened to inform further research on LA, including studies that do have big data in scope.

**MyScores dashboard**

In this section and the next, we use the categories of Bodily and Verbert (2017) to introduce the MyGrades student dashboard (Figure 1) in a structured and comparable manner (see Table I for an overview).

**Intended goal.** The goal of the MyScores dashboard was to support students in the first year of higher education by providing them with the feedback on exam results, encouraging reflection and adding actionable recommendations for improvement. The dashboard was to demonstrate how simple, “small data” that is readily available within institutions can be used as a starting point to fuel the organizational learning process around LA and to collect the valuable feedback about the intervention.

**Information selection.** The MyScores dashboard visualizes exam results on the level of individual courses and the study program. These have been stored in computer systems for decades and results are being communicated to students in an electronic format for many years at KU Leuven, but the traditional “study progress file” only offers a limited and unprepossessing presentation.

**Needs assessment.** The first examination period in Flanders takes place in the middle of the academic year. On the course level, several reasons may lead to a certain degree of ambiguity when newcomers in higher education receive their first exam report. Students are not “graded on a curve,” meaning that scores are not relative to those of other students. Belgian universities use a common rating scale from 0 to 20, with 10 being the passing grade. In many study programs, it is common for distributions of scores to differ significantly from one course to another. Most secondary schools use different grading scales and even in relative terms, the comparability of results in secondary and higher education is limited. On the program level, students need to get acquainted with a crucial key performance indicator: the cumulative study efficiency (CSE). This is the ratio of credits a student passed throughout the years to the total number of credits taken within the study program. Most programs have a semester total of 30 credits and typical one-semester courses correspond to three to six credits. The CSE is an important indicator of progress and connected to institutional restrictions. Students who do not attain a CSE of at least 30 percent at the end of the first year are not allowed to continue the program; students with...
First-year students may find it difficult to properly assess their situation. In the presence of such uncertainty, social-comparison theory (Festinger, 1954) suggests that a CSE below 50 percent receive binding study advice and may eventually be refused enrollment to all programs if they fail to catch up.

Notes: The text, in Dutch, at the top explains the goal of the current tab (self-reflection), its contents and how to read the visualizations. The second paragraph reminds the student to the fact that the dashboard is only a partial representation, not accounting for individual circumstances. At the bottom of the dashboard, students are being invited to provide feedback on the dashboard.

Figure 1. Initial view of the “Scores per course” tab, before answering questions about expectations and satisfaction.
individuals will try to compare themselves to peers. The MyScores dashboard responds to this strategy, by offering students the option to compare their exam scores to those of other participants and to position their CSE in relation to other students in the same study program.

Additionally, the dashboard includes historic data that connects the first-year, first-semester CSE students from previous years to their success rate and time needed to finish the program. The objective is to help students to become aware of the meaning and impact of the progress indicator, beyond the simple number. Previously, only one of the participating faculties offered such guidance in a systematic manner, but without further personalization of the message.

**Visual design.** Most student dashboards represent information in a visual way, but it has been argued that including elaborate textual information is useful to provide additional guidance (Ramos-Soto et al., 2015). Previous experience within the institution (Broos et al., 2017) reinforced the idea that the provision of a flexible text parameterization system for study counselors is a good way to make use of their knowledge and experience and to foster their involvement in and acceptance of the intervention. Hence, in addition to data visualizations discussed below, the dashboard contains a considerable number of accompanying notes that are a composition of common parts, parts adapted to the study program and parts that are individualized based on the actual situation and grades of the student. To offer a better overview, the dashboard is composed out of five tabs:

1. **Introduction:** the first tab contains an introductory text, explaining the objectives, and contents of the other four tabs. It includes a message explaining that scores are only a partial view of who the students are and provide a direct link to the study counselor and/or support service assigned to the study program.

2. **Scores per course:** the second tab contains the exam results for each course the student is enrolled in. The initial view shows a numeric score on a 0-20 scale and prompts the student to select the applicable option for two statements: “This score is […] than what I expected after the exam” and “I feel […] about this result.” When both blanks are filled, a button to show the scores of peer students is unlocked. On pressing this button, a visualization containing the scores of all exam participants appears. Within this visualization, the student is positioned in one of four groups. A second button offers the option to position the individual score more precisely within the group, leaving it up to the students to choose whether they would like to see this information or not (see Figure 2). This sequence is repeated for each of the enlisted courses. The tab concludes with an invitation to reflect on what is shown above, some additional remarks to frame what is shown above and an open invitation to get in touch with a student counselor.

---

**Table I.**

<table>
<thead>
<tr>
<th>Category</th>
<th>Question</th>
</tr>
</thead>
<tbody>
<tr>
<td>Intended goal</td>
<td>What is the intended goal of the system?</td>
</tr>
<tr>
<td>Information selection</td>
<td>What types of data support your goal?</td>
</tr>
<tr>
<td>Needs assessment</td>
<td>What do students need? Does this need align with your goal?</td>
</tr>
<tr>
<td>Visual design</td>
<td>Why are you using the visual techniques or recommendations you have chosen?</td>
</tr>
<tr>
<td>Visualizations</td>
<td>What visual techniques will best represent your data?</td>
</tr>
<tr>
<td>Student use</td>
<td>How are students using the system? How often? Why?</td>
</tr>
<tr>
<td>Actual effects</td>
<td>What is the effect on student behavior/achievement?</td>
</tr>
<tr>
<td>Student perceptions</td>
<td>How do students perceive the reporting system?</td>
</tr>
<tr>
<td>Usability test</td>
<td>Is the system easy and intuitive to use? Usability</td>
</tr>
</tbody>
</table>

Small data as a conversation starter
Global progress: the middle tab informs about the CSE. The upper part (see Figure 3) summarizes the progress of the student on the program level. It contains a visualization to put the individual result in perspective, comparing it to other students within the same study program. As with the course level scores, the initial view is intentionally coarse. When preferred, a button allows the student to activate a more fine-grained positioning. The lower part (see Figure 4) of the "global" tab explains and visualizes how the CSE of students within the same study program in previous years relates the number of years they spent to obtain the bachelor diploma: three years (default), four years or five or more. Drop-out students are indicated as well. For the category that corresponds best to the student, this information is repeated in an accompanying text.

Tips: the fourth tab’s content is entirely textual and is divided into two parts. The first part contains four sections on how to process the information contained by the dashboard: “talk about it,” “realistic expectations,” “learning skills” – including a link to our previous dashboard on learning skills (Broos et al., 2017) – and other factors (e.g. personal situation). The second part contains actionable advice on how to set goals and tips on how to achieve them.

Regulations: the last tab provides a summary of the progress monitoring regulations with respect to the CSE (see above). As with most of the textual information in the dashboard, this information is also available online, but it is fragmented across several web pages following an organization oriented structure. The MyScores dashboard collects relevant information from different sources and offers an integrated and personalized view, adapted to the program and situation of the student, and focused on the specific context of the first exam results.

Visualization. The dashboard uses a consistent visualization method across different measures. Unit charts are used to assign student scores to groups. Every student is represented by one dot and attributed to one of the groups. In case of the course scores, the dots are color-coded according to their impact on the student: red for failed and
non-tolerable (0 to 7/20) or not completed, orange for failed but with a potentially tolerable score (8 or 9/20), green for passed with moderate score (10 to 13/20), and bright green for passed with high score (14 to 20/20). For the global process tab, the first visualization (see Figure 3) uses gray dots for current year’s students, combined into three groups (depending on the study program, e.g., 0-30, 30-70, and 70-100 percent for engineering science). The second visualization (see Figure 4) uses colored squares to represent percentage units of previous year’s students (rather than dots for single students). Here, colors correspond to the time these students required to obtain the bachelor degree: black for students that dropped out, red for those who needed five years or more, orange for students with one year of delay (four years), and green for students that obtained the degree within the default three years. For each of the visualizations, the interpretation of colors and markers (dot or square) is explained using inline elements within the text.
If students activate a more precise positioning of their own result (by clicking the corresponding button) the dots of all students with the same score become emphasized (unfilled and blinking a few times to catch attention).

Results and discussion

Student use

Out of 1,905 unique students that received an invitation by e-mail, 887 (47 percent) used the dashboard. There is a remarkable difference in click-through rate between study programs. Students in programs bioengineering, engineering science, and engineering science-architecture received the invitation one day after online publication of their results in the university’s portal. In total, 55, 68, and 60 percent of students in these programs clicked through to the dashboard, respectively. For students of the science faculty (chemistry, biology, biochemistry-biotechnology, geography, geology, mathematics, informatics and physics) a different procedure applies. Students need to collect results in person allowing study counselors to provide some extra information on the spot. These students received the dashboard invitation only after this procedure was completed, which may partly explain why the click-through rates were only between 22 percent and 40 percent for these programs.

There is a noteworthy difference in the click-through rate between CSE groups (see Figure 5). More than half (56.3 percent) of the students with a relatively good study progress – the high CSE group[1] – do click through from invitation to dashboard. Among students with an alarming low study progress – the low CSE group – only 34.8 percent are visiting the dashboard. The medium CSE group has a click-through rate of 45.9 percent. Possibly, this difference may be explained by students who have already given up and may have lost interest in further information. Additionally, it is possible that student characteristics that result in a lower interest in the dashboard also result in lower exam results. Although more research is required to explain the exact cause of the difference, the suggestion that student dashboards are more likely to appeal to “successful” students is already informative, especially when considering that many efforts in LA are focused on at-risk students and drop-out prevention. On the one hand, the suggestion is that additional effort is required to reach “weaker” students. On the other hand, dashboards should be adapted to provide appropriate feedback that is valuable to “successful” students, for example by informing them about honor programs or outlining the prospect of advanced academic training.

Actual effects

Verbert et al. (2013) presented a conceptual framework for evaluating LA applications that distinguishes four stages: awareness, reflection, sense-making, and impact. The awareness stage is limited to simply displaying data. Almost by definition, any dashboard would

![Figure 5. Click-through rate in relation to the study progress of the student](image-url)
achieve this phase. But the presentation of data is not the end goal. The reflection stage relates to how the data are used to provoke questions in the mind of the user. In the MyScores dashboard, we opted to steer students in this direction by requiring them to answer questions related to their self-assessment and satisfaction to unlock more detailed information. By placing individual results within context, the MyScores dashboard aims to provoke additional questions, which in case of students' endeavor to answer them, should lead to new insights in the sense-making phase.

The fourth phase of the framework, impact, requires the assessment of behavioral change. While the MyScores dashboard provides several starting points to guide the students in the direction, the assessment of impact would require a longitudinal study and was left out of scope here. To assess awareness, reflection, and sense-making, three perspectives were used: do students use the footholds offered by the dashboard, how do students perceive the dashboard, and does the evaluation of the dashboard depend on the situation (grades, progress) of the student?

**Student perceptions and usability**

Students were asked to respond to three statements on a 1 (−) to 5 (+) scale:

1. I find this information useful (usefulness);
2. I find this information clear (clearness); and
3. this information influences how satisfied I am with my results (reflection).

These statements were presented to them in a noticeable yellow area at the bottom of the dashboard (see bottom Figure 1). Results are summarized in the Marimekko charts, as shown in Figure 6. In total, 289 out of 887 (32.6 percent) dashboard users provided full feedback on all three statements. Usefulness was rated positively (response 4-5) by 87.7 percent of feedback-providing students within the high CSE group for their study program, 90.8 percent by the medium CSE group, and 90.3 percent by the low CSE group. Clearness was rated similarly: 88.2, 88.0, and 90.3 percent in the high, medium, and low CSE groups, respectively. Students within the medium CSE group seem to respond with a slightly lower score on average for both statements. When asked if the information provided by the dashboard has influences on how satisfied students are with their results, based on the idea that providing perspective may alter how students think and feel about their own results, the response is more moderate. In total, 48.4 percent of students in the high CSE group, 40.5 percent of students in the medium CSE group, and 26.7 percent of students in the low CSE group provide a positive response. A large share of students provides an answer somewhere in the middle (35.9, 45.7 and 42.2 percent for the three groups, respectively). Up to 31.1 percent of students in the low CSE group seem to indicate that the dashboard did not alter satisfaction with their results. In retrospect, the third
statement may have given rise to some misinterpretation: some students may have intended to indicate that the information presented influenced their satisfaction negatively, rather than not (see Figure 1, lower right).

Evaluation and conclusion
This paper suggests complementing big data for LA within higher education institutions by a more modest, practice-oriented approach leveraging existing small data and to start gathering additional insights by deploying dashboards to students. Subsequently, it presents an implementation of this approach by offering the MyScores dashboard to 1905 first-year students in 11 science and engineering-oriented study programs, reaching 887 active users, from which 289 provided feedback on three statements about the dashboard. Generally, students rate the usefulness and clearness of the MyScores dashboard positively, but the appreciation of how the dashboard changes their satisfaction about their results is less consistent and differs depending on the study success. Moreover, “weaker” students seem to access the dashboard less, which raises questions about how to reach these students, but also underline the potential of targeting “stronger” students with targeted information. The difference in dashboard access between “weaker” and “strong” students is consistent with earlier finding within a largely overlapping population of STEM students (Broos et al., 2017). Further research is required to explore if “weaker” students are less likely to use LA tools overall, or if the finding is specific for the type of dashboards or the process used to inform students about their availability. For a follow-up study, we would choose a more structured approach, based on DSR, using dashboard artifacts to further probe into how students accept and use them, complementing qualitative usage information and questionnaire feedback by in-depth qualitative research. Focus group discussions may help to build the understanding of the difference in usage and response between students with low, medium and high study efficiency (CSE).

The practical implementation of a dashboard targeting a sizable number of first-year students involved many contributors, including students, study counselors, university and faculty management, IT and legal services, etc. This allowed us to see LA at the organizational level in its full complexity and from many angles, some of them previously overlooked. Many of the insights gained throughout the process were not discussed in this paper.

Our study remained on the surface with respect to the lessons learned from the MyScores intervention and would ideally be complemented by a more longitudinal approach to try to capture if the dashboard leads to behavioral change (impact). For each of the participating STEM study programs we received confirmation for continued participation. Several additional study programs confirmed participation in an upcoming iteration, including programs in humanities, social and biomedical sciences, expanding the population both in size and diversity.

Note
1. The allocation of students to a high, medium or low CSE group is based on program-specific CSE cutoff scores. The analysis preserves the allocation as used on the dashboard.

References


Further reading


Corresponding author
Tom Broos can be contacted at: tom.broos@kuleuven.be

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
Evaluating emotion visualizations using AffectVis, an affect-aware dashboard for students

Leony Derick
Department of Telematic Engineering, Universidad Carlos III de Madrid, Madrid, Spain

Gayane Sedrakyan
Department of Computer Science, Katholieke Universiteit Leuven, Leuven, Belgium

Pedro J. Munoz-Merino and Carlos Delgado Kloos
Department of Telematic Engineering, Universidad Carlos III de Madrid, Madrid, Spain, and

Katrien Verbert
Department of Computer Science, Katholieke Universiteit Leuven, Leuven, Belgium

Abstract

Purpose – The purpose of this paper is to evaluate four visualizations that represent affective states of students.

Design/methodology/approach – An empirical-experimental study approach was used to assess the usability of affective state visualizations in a learning context. The first study was conducted with students who had knowledge of visualization techniques (n = 10). The insights from this pilot study were used to improve the interpretability and ease of use of the visualizations. The second study was conducted with the improved visualizations with students who had no or limited knowledge of visualization techniques (n = 105).

Findings – The results indicate that usability, measured by perceived usefulness and insight, is overall acceptable. However, the findings also suggest that interpretability of some visualizations, in terms of the capability to support emotional awareness, still needs to be improved. The level of students' awareness of their emotions during learning activities based on the visualization interpretation varied depending on previous knowledge of information visualization techniques. Awareness was found to be high for the most frequently experienced emotions and activities that were the most frustrating, but lower for more complex insights such as interpreting differences with peers. Furthermore, simpler visualizations resulted in better outcomes than more complex techniques.

Originality/value – Detection of affective states of students and visualizations of these states in computer-based learning environments have been proposed to support student awareness and improve learning. However, the evaluation of visualizations of these affective states with students to support awareness in real-life settings is an open issue.

Keywords Learning analytics, Interactive learning environments, Human-computer interface, Visualization evaluation

Paper type Case study

The current issue and full text archive of this journal is available on Emerald Insight at: www.emeraldinsight.com/2397-7604.htm
1. Introduction

Emotions are known to play an important role in learning (Kort et al., 2001; Trigwell et al., 2012). Emotions drive attention, which, in turn, drives learning and memory (Värlander, 2008). Emotions are often a more powerful determinant of our behavior than our brain’s logical and rational processes (Sylwester, 1994). Furthermore, emotions play an essential role in studies on attitudes and motivation (Pintrich, 2003; Meyer and Turner, 2002). Several studies found that students experience a rich diversity of both positive and negative emotions in academic settings (Pekrun et al., 2002). Prior research has highlighted the importance of supporting learner awareness of these emotions (Ashkanasy and Dasborough, 2003). Information on affective states can, for instance, help students (or stimulate interest) to reflect on the type of emotions they felt, the activities that generated certain emotions or their evolution over time. By analyzing such information, students can take a pro-active role in regulating their learning as well as taking decisions on their improvement needs during learning processes, based for instance on information from studies that relate learning outcomes with affective states (Baker et al., 2010).

Recent research shows increased interest in the automatic detection of emotions in various contexts. There are studies that propose different methodologies and detectors of emotions that also demonstrate practical applications in different learning contexts. Examples include emotion detection algorithms based on facial and gesture recognition (Burleson, 2006). Such algorithms are mostly based on human body signals, such as brainwaves captured with various sensors (Azcarraga et al., 2014). Several studies attempted to correlate such data with student actions in different learning environments such as Intelligent Tutoring Systems (Pardos et al., 2013), MOOCs (Leony et al., 2015), or course-specific environments (Leony et al., 2013a). Recent research has also shown interest in biofeedback based on analysis of multi-modal data collected from various wearable sensors during learning tasks. In some studies, information on emotions is processed based on manual input by students when interacting with a learning environment (Muñoz-Merino et al., 2014).

An important issue is that information about affective states, such as the type and intensity of the experienced emotion, should be presented in computer-based learning environments in an intuitive way to the different stakeholders, including teachers, students, and managers. Visualization techniques are one of the most used techniques to present such information in the context of so-called learning dashboards (Verbert et al., 2014). The goal is to support stakeholders to gain insight from these visualizations, i.e. to provide information that can be of utility to support awareness, reflection, and decision making (Verbert et al., 2013). There are some works that present visualizations of emotions in computer-based learning environments (Leony et al., 2013b). However, to our knowledge, no studies can be found that evaluate the capabilities of different visualization techniques to support awareness of emotions in learning environments. Also, most evaluations of learning visualizations are done by teachers (Verbert et al., 2014). Empirical studies focusing on the evaluation of visualizations by students that provide insight into the usability of these visualizations are largely lacking.

In this manuscript, we focus specifically on evaluating the usability of visualizations of affective states with students using well-known usability assessment constructs such as the perceived usefulness and insight. Usability, also, refers to the ease of use (Davis, 1989) of the visualizations. We define the perceived usefulness as the perception of students about the importance of each one of the visualizations for the learning process. Insight is defined as the extent to which students can interpret the presented visualizations in a correct way (North, 2006). This manuscript aims to present the first evaluations and experiences from the use of visualizations of affective states for students.
and attempts to identify future research directions in this domain. The research questions are the following:

RQ1. How usable are the visualizations that we have developed for students in terms of their perceived usefulness and ease of use?

RQ2. Which insights are supported by the affective visualizations for students regarding their capability to support awareness?

In this work, we assess the usability of different visualizations using different groups of students in higher education. Bachelor and Master level students from two different study programs at two universities participated in the studies. The pilot study was conducted with a group of students with a background in visualization techniques, whereas the second study was conducted with a group of students with little knowledge of visualization techniques. The insights from the first user study ($n=10$) were used to improve the visualizations. We used suggestions of these students to create additional visualizations. The second user study was conducted with the enhanced environment with a larger group of students ($n=105$).

The rest of the paper is organized as follows: Section 2 presents related work on visualizations of affective states in the context of learning dashboards. Section 3 presents our methodology. Section 4 describes the AffectVis dashboard, including four different visualizations of affective states. Sections 5 and 6 present two user studies conducted in the context of two different courses, detailing the participants, data collection, data analysis, and post-study interview results. Section 7 discusses the findings and limitations of the work. Finally, Section 8 concludes the work proposing some future research directions.

2. Background

2.1 Learning analytics

Different dimensions have to be considered when developing learning analytics applications (Greller and Drachsler, 2012), including internal limitations, external constraints, instruments, data, objectives, and stakeholders. The objectives can be twofold: reflection and prediction. In this study, we focus on students as stakeholders and reflection as an objective.

Instruments usually rely on either information retrieval or information visualization technologies, or a combination of both. Information retrieval intends to infer high-level information from the analysis of raw data. Examples of this high-level information can be student characteristics, such as learning behavior patterns, and future performance indicators (Muñoz-Merino et al., 2013). These indicators are then visualized to provide useful insights for teachers, students, and managers (Ruipérez-Valiente et al., 2014). Visualizations are known to support self-regulated actions of learners in online environments, as they can simulate social engagement and reflection appropriate for the context of the learner (Glahn, 2009).

Detection of affective states in educational settings has been explored previously by several researchers in the field (Baker et al., 2010; Jaques and Vicari, 2007; Burleson, 2006; Azcarraga et al., 2014; Pardos et al., 2013). Leony et al. (2013a) presented a concrete case of inference of emotions from interaction data in a programming environment. The approach consists of a set of Hidden Markov Models (HMMs). During a programming task, students were asked to provide information about their affective state. This information was used to train the HMMs that were later used to predict emotions. In another approach, Leony et al. (2015) used a rule-based model for each emotion of interest, contextualizing emotion detection in MOOCs. In this work, frustration is for instance understood to occur when students either frequently fail exercises or fail an exercise about a topic that they thought they had sufficient knowledge about.
In this paper, we will focus specifically on visualizing data about affective states of learners and evaluating the usability measured by usefulness and insight of visualizations. Several studies have explored the typology of affective states that occur during learning (DMello, 2012). The results of previous research in this domain indicate that the basic emotions identified by Friesen and Ekman (1978), such as anger, fear, sadness, joy, disgust, and surprise, typically do not play a significant role in learning (Kort et al., 2001). Several studies have also identified subsets of affective states that typically do play a significant role in learning, at least in the case of college students. Craig et al. (2004) for instance found evidence for a link between learning and the affective states of confusion, flow, and boredom. D’Mello et al. (2006) in addition found significant relationships for happiness (Eureka), confusion, and frustration, but not for boredom. In this manuscript, we make use of the most common subset of affective states based on prior research suggestions, namely: frustration, confusion, boredom, happiness, and motivation. We visualize information on these five affective states experienced by students during their learning activities in a learning dashboard for students. We present the results of two user studies that assess the usability of different visualization techniques for these affective states.

2.2 Learning dashboards
Dashboards are instruments intended to improve decision making by amplifying or directing cognition and capitalizing on human perceptual capabilities (Yigitbasioğlu and Velcu, 2012). In a learning context, dashboards aim to support learning process awareness, ultimately targeting regulation of learning (Sedrakyan, Järvelä and Kirschner, 2016; Sedrakyan et al., 2017; Sedrakyan, Malmberg, Verbert, Järvelä and Kirschner, 2018). In recent years, several dashboard applications have been developed to support learning or teaching. Such dashboards provide graphical representations of the current and historical state of a learner to support decision making (Few, 2006). Dashboards have been deployed to support learners or teachers, or both, and used in traditional face-to-face, group work, or online/blended learning (Verbert et al., 2014). Examples of dashboards that are used to support face-to-face teaching include Backstage (Pohl et al., 2012), Classroom Salon (Barr and Gunawardena, 2012), and Participation Tool (Janssen et al., 2007). The overall objective of these dashboards is to stimulate learner engagement during face-to-face sessions. Several dashboards also focus on group work and collaboration. TinkerLamp (Son, 2012) and Collaid (Maldonado et al., 2012) are some prominent examples. Most dashboards, however, focus on online or blended learning. Course signals (Arnold and Pistilli, 2012) is one of the more prominent examples in this category. The dashboard predicts and visualizes learning outcomes based on three data sources: grades in the course so far, time on task, and past performance. Most of our work is also part of this category.

In the context of learning dashboards, there are a few interesting observations that are relevant to the content of this paper:

- One observation is that usability evaluations have been conducted most often with teachers. Teachers were often asked to indicate how useful they think a dashboard would be for learners. Such a perceived usefulness evaluation was conducted for instance with both student inspector (Scheuer and Zinn, 2007) and LOCO-Analyst (Ali et al., 2012), both yielding positive results. Results of our evaluations with SAM (Govaerts et al., 2012) and StepUp! (Santos et al., 2013) indicate that the perceived usefulness is often higher for teachers than for students. In this paper, we focus specifically on evaluations with students. In contrast to earlier studies, which were often conducted with a relatively small number of participants, we present results of a case study with a relatively large number of students.
Also, most dashboards focus on visualizations of utilized resources, time, test results, and social interactions (Verbert et al., 2014). To the best of our knowledge, only a few dashboards have been presented that focus on the representation of student emotions. In a recent study, Ruiz et al. (2016) focus on the methodological aspects of developing dashboards that support emotion-related information. Only one study was conducted that evaluates the utility aspects of inclusion of emotion-related information into learning dashboards (GhasemAghaei et al., 2016). The focus of the study is the utility of such a dashboard for instructors. In our work, we focus on visualization of affective data for learners, motivated by the fact that such data have shown to be an important player in learning behavior regulation (Baker et al., 2010).

Finally, little is known about the effectiveness of different visualization techniques to give students insight into their learning-related data. Different visualizations have been proposed in earlier work, but to which extent these visualizations can be interpreted in correctly by students, and which techniques work better than others, both need further research.

3. General methodology

In this work, we followed the principles of Design Science in information systems research, which targets building and evaluating innovative artifacts to help understand and solve knowledge problems (Von Alan et al., 2004). Our artifact includes a learning dashboard that shows emotion-related data during a learning process. The goal of the dashboard is to support student awareness on their affective states during their learning activities.

We use visualization techniques to represent affective states in the context of a learning dashboard, motivated by the fact that interactive visualization techniques are known to support effective understanding of data, reasoning, and decision making (Keim, 2002). Several visualizations have been developed with the goal to allow students to obtain insight into affective information. We designed, implemented and deployed visualizations of a relevant subset of emotions based on prior studies, as explained in the previous section. These visualizations represent the intensity levels of emotions, learning activities during which the emotions were experienced by students, the evolution of emotions over time, as well as comparisons with data of peers.

An empirical, experimental study approach was used to assess the usability of such visualizations deployed in a dashboard (the design artifact). Two experiments have been conducted in two different universities with two different groups of students: students with and students without previous knowledge of information visualization techniques. During the experiments, students completed different learning tasks, further referred to as sessions. At the end of each session, students provided information about their emotions by answering a set of basic questions in an online survey, such as “indicate how frequently you felt motivated/happy/confused/frustrated/bored during this learning activity.” Based on the input of students, the learning dashboard generated affective visualizations. Context information about the students was also collected: students completed a questionnaire about their personal characteristics, such as gender, age, and previous knowledge on information visualization techniques.

The dashboard with different visualizations of affective states was deployed and provided to students. The following data measurements were used:

- a five-position Likert-type scale was used to score subjective judgments of students about the proposed affective visualization method, such as ease of use, perceived utility, and insight;
- the System Usability Scale (SUS) method (Brooke, 1996) has been used to measure the usability;
insight was measured using comparison between the actual data and student perceptions: exploratory correlation analyses have been performed to study the differences of students’ visualizations interpretations with the intended goal of the visualization; and

- subjective perceptions and insight have been, in addition, explored using a set of objective questions in a post-study interview.

To isolate the impact of pro-social behavior (Mitchell and Jolley, 2012), the anonymity of participants was ensured by not disclosing any identifiable information.

4. AffectVis: a visual learning dashboard of affective states and learning activities in projects

For our studies, we have developed four visualizations with the general objective of allowing learners to reflect on their affective states and their connection with specific learning activities. The visualizations are web based. Thus, the only tool needed to access them is a web browser with JavaScript capabilities.

Figure 1 shows the first visualization (radial visualization), which includes an improved version of the visualization presented by Leony, Parada, Muñoz-Merino, Pardo and Delgado Kloos (2013). The technique makes use of a set of polar bars to present the average frequency of each affective state experienced per each learning activity. Affective states of a

**Figure 1.**
The radial visualization is showing the frequency of each affective state for each activity

**Source:** Polar bars show the values for the active student, while the solid line shows the class average. Sedrakyan et al. (2017) (used with permission)
learner are differentiated through the color of the bar, while labels are used to represent the associated activity. The solid line shows the average value of the class for each emotion and activity.

The second visualization (timeline visualization) presents the evolution of time dedication of each student during the course, as well as the average time dedication and emotion evolutions of the whole class. The visualization represents the accumulated time dedication of students; when the student selects a point of time on the horizontal axis, the values of the vertical axis indicate the accumulated levels of time dedicated to learning activities during the course until that moment. In addition, the timeline visualizes the evolution of each emotion during the course. Figure 2 presents an example of the timeline visualization used in the scope of this work.

The third visualization is a heatmap visualization, in which columns represent time units, such as days, weeks, and months, and rows represent students. Each affective dimension is represented by a cell, while the frequency level of each emotion is represented through the intensity of the cell color (a more intense color represents a higher level of the emotion). A portion of this visualization is shown in Figure 3.

Lastly, we designed a scatterplot visualization. In this visualization, each affective dimension has a different scatterplot associated to it. The X-axis corresponds to the exact date and time when the emotion takes place, and the Y-axis presents the frequency value of the emotion. Bubble sizes represent the amount of work dedication, and bubble colors indicate whether the data point belongs to the active student (blue) or a peer. Figure 4 presents an example scatterplot for the emotion “confusion.”

In its current form, the visualizations in the AffectVis dashboard rely on the data collected through systematic surveys of students about the typology and intensity of their emotion per different learning activity. Further details are described in the next sections.

5. Pilot study with a small group of students with knowledge in information visualization: user study 1
The main purpose of this user study was to perform an initial exploratory analysis of the developed visualizations with a small number of students. The radial visualization and the timeline visualization were deployed and evaluated in this study. Based on feedback and input from students participating in this study, the visualizations were improved. Moreover, two additional visualizations were developed and deployed for the second, more elaborate, user study.

Source: Students can also see the time dedication average of the class (Sedrakyan et al., 2017) (used with permission)
The first study was conducted with Master level students at Vrije Universiteit Brussel in Belgium. The profile of students, having knowledge about visualizations, was beneficial for obtaining targeted feedback. This user study would also allow observing differences between students with knowledge on visualizations and students without such knowledge (the students of user study 2). As indicated above, the radial visualization and the timeline were evaluated in this pilot study. At this stage, the timeline visualization included the aggregated time dedication of the student to different learning activities and the average time dedication of his/her peers.
The user study was conducted in the context of the course project, which lasted five weeks, from late February to early April of 2014.

5.1 Demographics of participants in user study 1
This user study was conducted in the context of a course on information visualization at a graduate level (Master degree program). First, students received theoretical and practical sessions about different concepts and visualization techniques. Next, and as part of the evaluation of the course, students implemented and presented a project which included 12 types of learning activities: brainstorming, designing visualization, gathering data, parsing, filtering and mining data, getting started with the visualization library D3.js, implementing the visualization, implementing interaction in the visualization, reading resources, reading research papers, preparing questions, and preparing research presentations.

As participants were students registered for an information visualization course, they all had a relevant level of knowledge of principles and theories involved in the creation of visualizations. Thus, their feedback was highly relevant during the stage of early definition and development of the visualizations.

In total, 42 students were registered for the course. Out of these 42 students, ten students participated in the first user study.

5.2 Data collection in user study 1
This pilot study mainly served to identify usability issues of the radial and timeline visualizations and to collect feedback and input from students for additional visualizations. In this study, we first conducted ten think-aloud sessions (Lewis, 1982), with one student at a time. Each session was organized in three phases: filling out a survey to capture emotion-related data about their work during the project, conducting tasks with the two visualizations, such as identifying the most frequent emotion with the visualizations, and filling out an evaluation survey about the visualizations.

The survey that intended to capture data about students’ activities during the project used explicit questions about the students’ affective state for each type of activity conducted in the project. For each type of learning activity, students had to indicate how frequently they have experienced the five affective dimensions known to occur in learning scenarios: motivation, happiness, boredom, confusion, and frustration (DMello et al., 2007). Students were also asked to indicate the amount of time they dedicated to the project during each week. Afterwards, the two visualizations were presented, i.e. the radial visualization showing the frequency of emotions for each type of activity and the timeline visualization. Both visualizations used the data collected in the previous phase.

After completing the tasks, such as identifying the most frequent emotion and comparing this value to the class average, students filled out an evaluation survey, including questions about the usability (usefulness and insight) of the visualizations. The usability was subsequently measured with the SUS method. Students also rated the two visualizations in the range of “not useful at all” to “very useful.”

Students were also asked which other information could be of interest to represent through visualizations. They were asked to rate the utility of five types of information on a five-point Likert scale: types of used resources (e.g. forums, blogs or files), detailed information about one student, comparing actions between two students, detailed statistics of most used resources and information about content creation by students.

5.3 Data analysis results in user study 1
5.3.1 Perceived usefulness. The usability results obtained from the evaluation of the SUS questions resulted in 72.5 points on average, which can be assessed as a positive belief
(Bangor et al., 2008). The timeline was perceived as the most useful by the students (average score above 3.5 on a scale from 1 to 5). The radial visualization was perceived as useful (score above 3 on a scale from 1 to 5).

Students indicated that they are interested in detailed information of one student, comparison of students and information about content creation by students. The information related to the types of used resources and the most used resources were the least prioritized. In Figure 5, we present a set of box plots illustrating the priorities by students given to each option.

5.3.2 Post-study interview results in study 1. The analysis of students’ answers to the interview questions provided useful insights for improvement needs for the visualizations. The results revealed that students experienced difficulties in interpreting certain aspects of visualizations. For instance, some students found it difficult to identify the values on the radial bars that were used to visualize affective states per activity. This difficulty was found to be due to user interface related issues such as having adjacent bars with similar colors or a low-level contrast, making them not easily distinguishable in the chart. Furthermore, some students were not able to interpret the meaning of several visual components. For example, there were students who were not able to detect that the class average was represented by solid lines on the timeline visualization. In general, students expressed that they “liked the timeline and the comparison with the class average” and prefer its use in the future. The radial visualization was difficult to understand by some students (“it is hard to see the information of all students,” “the red color (of bars representing frustration) is too distracting” or “it is confusing that bars do not start from zero”), which suggested that the interpretability of the visualization needed to be further improved.

Some of the post-study responses provided creative input in the form of suggestions for further information needs of students. For instance, there was a suggestion to include the equivalent of “return over investment,” where the dedicated time would represent the investment and the obtained mark would represent the return. Another suggestion was the inclusion of task types to which the time was dedicated, such as lectures, homework, studying, and group work.

6. Extended study with a larger group of students without knowledge in information visualization: user study 2

The second study was conducted with bachelor-level students at the Eindhoven University of Technology in the Netherlands. For the second user study, we improved the two visualizations based on the findings and suggestions of students of the pilot study. To address the difficulties of interpretation, the contrast of colors and the visibility of elements in both visualizations were improved. Interactivity was added to clarify the details of the visualizations: the radial visualization was adjusted to show the value of each bar when the mouse cursor hovered over it. The timeline was adjusted to offer the option for hiding and showing data series, etc. In addition, for this user study, we implemented
two new visualizations with emphasis on individual and detailed information, as such information was indicated as relevant by students of the first user study. These visualizations are the heatmap and scatterplot visualization described in Section 4. The purpose of this study was to evaluate the usability of the four visualizations, namely, the improved versions of the two visualizations used in user study 1, and the two new visualizations.

6.1 Demographics of participants in user study 2
Overall, 105 first-year bachelor students enrolled in technological programs took part in user study 2. The majority of participants are male (96.6 percent males and 9.4 percent females). Participants are between 18 and 40 years.

6.2 Data collection in user study 2
The study was conducted in the context of the course human-technology interaction. At the beginning of the course, an introduction was provided about all the concepts and processes involved in the design of usable interfaces for technological artifacts. At the end of the semester, students completed a project about the design of a thermostat. The project duration was four weeks, from late April to early June in 2014. In the end, students presented their project to the teaching staff and their peers.

For this project, in collaboration with the instructors, we defined six types of learning activities: brainstorming, interface design, implementation, writing documentation, experiment with users, and writing installation instructions. During the project, students received an e-mail with a link to an online survey each week, as well as a link to a web application that showed the visualizations of their emotion-related data. To maintain the anonymity of information, students were asked to create a personal identifier with which they could access the visualizations. Every week, students completed the following tasks: filling out a survey about their emotions and learning activities during that week, exploring their data in relation to data of other students with the proposed interactive visualizations, and filling out a survey about their perceptions and judgments on the visualizations.

The survey included questions about particular activities. Students were asked to indicate how frequently they had experienced each affective state while performing each of the project activities and the time they dedicated to the project. Students were allowed to report activities for a week different than the current one. After the data were submitted, the student could use a web application to access the visualizations.

The data collected in user study 2 were used as follows:

- Radial visualization: values for each affective emotion and each activity were shown explicitly. The average for all students was computed and shown as a solid line.
- Timeline visualization: values for the student were plotted according to the week they were provided. An average for the class was also included.
- Heatmap visualization: the intensity value of each cell represents the average value for the corresponding emotion for the given week.
- Scatterplot visualization: for each affective state, the visualization plots the values (scores) for each student along the date and time of the survey submissions.

In this study, the students were asked to answer questions of an evaluation survey to assess the usability of the visualizations on a continuous basis. The usability was evaluated through SUS questions, while the usefulness was evaluated using five-point Likert scales that rank each one of the visualizations from “not useful at all” to “very useful.” In addition
to these questions, we also used questions to objectively assess the insight of the visualizations as follows:

- five-point Likert scale to indicate whether the student is much below, below, average, above or much above the class average for each emotion and time dedication;
- indicate the most frequent emotion experienced during the project;
- identify the activity that motivated students (the whole class) the most;
- identify the activity that frustrated the student the most; and
- identify the activity during which the student is most different from the rest.

In addition to weekly evaluations, a think-aloud session took place at the end of the course. The think-aloud session was conducted with batches of two to four groups, involving 6 to 12 participants in each session. At the beginning of each session, the four visualizations were briefly explained. Then the students completed the tasks. Afterwards, we asked students feedback and inquired about their interests. Table I shows the questions asked in this final survey.

Overall, we received 298 submissions from 95 students for the data gathering survey, with 91 percent of the responses from male students and 9 percent of females. Most of the submissions (78.5 percent) belonged to students 20 years old or younger, 17.4 percent between 21 and 25, 1.7 percent between 26 and 30, 0.7 percent between 31 and 35, and 1.7 percent between 36 and 40. The survey for weekly evaluations received 218 responses from 85 students, while only 52 students participated in the final evaluation.

6.3 Data analysis results of user study 2

6.3.1 Perceived usefulness. The average SUS score for the set of visualizations was 60.1. Figure 6 presents a boxplot for each week of the study. The average usability stays constant over the weeks.

<table>
<thead>
<tr>
<th>Question</th>
<th>Correct (%)</th>
<th>Categories</th>
</tr>
</thead>
<tbody>
<tr>
<td>Survey of the last week (n = 35)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>Identify your most frequent emotion during the project</td>
<td>48</td>
<td>5</td>
</tr>
<tr>
<td>Identify the activity that motivated students the most</td>
<td>23</td>
<td>6</td>
</tr>
<tr>
<td>Identify activity that frustrated you the most</td>
<td>69</td>
<td>6</td>
</tr>
<tr>
<td>Identify activity where you differ the most from peers</td>
<td>31</td>
<td>6</td>
</tr>
<tr>
<td>Final survey (n = 47)</td>
<td></td>
<td></td>
</tr>
<tr>
<td>How does your confusion evolve along time?</td>
<td>61</td>
<td>3</td>
</tr>
<tr>
<td>Identify the week when students were more frustrated?</td>
<td>34</td>
<td>4</td>
</tr>
<tr>
<td>Do you think your time dedication affects your boredom?</td>
<td>53</td>
<td>3</td>
</tr>
</tbody>
</table>

Table I. Percentage of correct responses to the objective questions in the survey

Figure 6. SUS scores obtained for all visualizations during the four weeks of the user study.
The obtained usability results were found to be lower than in the user study 1. The reason for that can potentially be attributed to the profile of the students. In contrast to the user study 1 participants specialized in visualizations, the students in user study 2 had little or no knowledge about information visualization techniques.

Figure 7 presents the perceived usefulness of each visualization on a five-point Likert scale. The median of the usefulness score was on average 3 for all visualizations. The average was slightly incremented for the timeline and the scatterplot visualization.

6.3.2 Analysis of insight. The differences in student perceptions of their affective states based on the visualizations and the actual values according to data were tested using the Pearson correlation. The analysis, with $n = 34$, resulted in the following findings: frustration ($r = 0.634, p = 0.000$), confusion ($r = 0.620, p = 0.000$), boredom ($r = 0.551, p = 0.000$), happiness ($r = 0.684, p = 0.000$), motivation ($r = 0.829, p = 0.000$), and time dedication ($r = 0.374, p = 0.040$).

For all the relationships, a significant correlation was found ($r > 0.5$), except the time dedication. This suggests that in general students were able to correctly interpret the provided visualizations. However, the results also suggest that there is room for improvement, specifically for the interpretability of visualizations. Ideally, the understanding by all the students would result in higher correlation coefficients closer to 1.

Table I presents the percentage of correct answers based on the students’ interpretation of visualizations. These values represent the number of times that the students correctly interpreted the visualizations for different aspects. The number of categories gives an idea of the number of possibilities a student can choose from. These categories are the number of options a student can select to answer a question and are presented in the third column of Table I. For question 1, “identify your most frequent emotion according to visualization,” the number of possible answers was limited to 5. We observed that the more options a student has, the more difficult it is for the student to give the right answer. Therefore, the percentages of correct answers should be interpreted taking into account the number of different categories.

Some of the questions included in these surveys also contained a certain level of difficulty that needs to be further discussed. For instance, the affective states with the highest values (motivation and happiness) have a difference of only 0.21. The mean standard deviation for all states is 0.27. As such, in some cases, a student would select an incorrect emotion as his/her “most frequently occurred emotion,” however, the value of such emotion was in fact very close to the highest one. The second question of the final survey presents a similar case. Students had to identify the week when they were more frustrated. However, the values for week 2, week 3, and week 4 were similar, with the value of week 4 being just marginally higher than the values of week 2 and week 3. If we had considered all of these three options as valid, 98 percent of the answers would have been correct.

6.3.3 Post-study interview results of user study 2. The analysis of the responses to the interview questions showed that the perceived of the visualization varied among students. Some of them preferred the radial visualization of affective states per learning activity:

I liked the states per activity the most. After that [I] will go the timeline, followed by the heatmap. Finally the scatterplots.
Other students considered the timeline as the most useful:

The timeline is the easiest to interpret, since it is in a form I am used to and since it doesn’t contain that much data at the same time, which the others do. Especially the heatmap and scatterplot are containing too much detailed and deviating information, which makes it hard to get an overview. The emotion per activity is okay, but also not readable very easily, because some colored areas are very small and it is not always clear which color is represented at what place of the grey line.

Others valued the combination of data and design used in more complex visualizations such as the heatmap:

It is difficult finding some meaningful values in the scatterplots. However the information in the heatmap is grouped together nicely. Timeline shows a nice overview of how affective states progressed as well.

The teaching staff also provided valuable feedback about potential improvements. The main suggestion was to allow the instructor to indicate an expected amount of time dedication. This would allow students to know whether they were dedicating less (or more) time than what the instructor was planning. The inclusion of expected time dedication would also allow the teaching staff to analyze whether the work load is being set appropriately for the current group of students.

7. Discussion

The results of our study provide useful insights for the usability of different visualizations for students for presenting emotion-related data. However, there are also several limitations that should be articulated. While we were able to assess the usability, measured by perceived usefulness and insight, with a relatively large number of students in user study 2, the limited number of students in user study 1 does not allow to draw strong conclusions from the survey results with this group as the suggestions from this study were mainly used as a basis to improve the visualization design for user study 2.

Second, data collection was performed manually in both user studies. Thus, the accuracy of affective states could be subject to subjective judgments of students. We should, however, mention that, although some studies rely on methods and techniques for capturing and analyzing emotion-related data of students in an automatic way (Leony et al., 2013a, 2015), in this paper we focus on the evaluation of usefulness of representing such data to students. Nevertheless, the acquisition process may also influence perceived usefulness and interpretation of data.

In general, usability results indicate that the visualizations are easy to use for students with knowledge of visualization techniques. A SUS score of 72.5 can be assessed as strongly positive beliefs (Bangor et al., 2008). Although the same results could not be confirmed by user study 2, the results were still found to be acceptable. The average SUS score of 60.1 in this user study still reflects a positive attitude. Since the students who participated in the second study had little or no knowledge of information visualization techniques, the relatively lower scores can be attributed to difficulties with using the visualizations, as can also be inferred from students’ answers.

In general, the results suggest that visualization techniques need to be designed with care: the difficulty of interpretation of more complex visualizations, such as the heatmap and radial visualization, may be a barrier for uptake by a general audience with no background in information visualization.

The results on perceived usefulness show that students perceive a simple timeline that represents time dedication and evolution of affective states over time as the most useful visualization. This visualization was rated higher regarding its usefulness than the visualization of affective states per activity in user study 1. In the second user study, with two other visualizations added, this visualization resulted in the most positive scores.
on average. In general, the findings suggest that a simpler technique results in a higher perceived usefulness.

Insight was measured by correlations between the actual data indications and student perceptions from visualizations. In addition, we measured how well students were able to interpret the visualizations with the use of objective questions. All the correlations were significantly high, with results higher than 0.5, but still less than 0.7. This suggests that, in the majority of cases, students were able to correctly interpret the provided visualizations.

In summary, the results indicate that visualizations of emotions can support awareness and reflection of student data, but they need to be designed with care to address the needs of students. Simpler techniques, such as timeline visualization, may result in higher positive perceptions than more complex techniques, such as heatmap or radial visualizations. The type of data to include in such visualizations constitutes a further line of research. While students in the user study 1 showed interest in more detailed data about individual students, the representation of such data remains a challenge. Evaluation results of user study 2 indicate that the visualizations that we selected to address their needs (heatmap, scatterplot with three dimensions) are difficult to interpret by users with no background in information visualization. Our future work will focus on exploring visualizations that can represent emotion-related data in a simple and intuitive way to enable use by a general audience.

8. Conclusion
The evaluation presented in this paper showed the potential of dashboards and visualizations to support students awareness of affective information linked to learning activities in an educational scenario. In general, students expressed that they “liked seeing their emotion-related information linked to learning activities and their comparison with their peers.”

The results of student evaluations suggest that usability of the proposed visualizations was acceptable, but that there is also room for improvement. In addition, the simpler techniques, such as the timeline visualization, so far offer the highest potential with respect to usability, measured by perceived usefulness and insight. There were differences between students with knowledge and those without knowledge about information visualization. SUS results were higher in user study 1. This suggests that the fact of having knowledge about visualizations might have an influence on the perceived usefulness and insight, and that student training might be necessary in some cases.

Future work includes the improvement and design of new versions of the presented visualizations. Initial modifications will be based on the feedback received during the interviews. Some of these improvements include simplification and adding interaction to ease the interpretation of data. In addition, other visualizations of affective information can be designed to be used by the instructor rather than directly by students.

The work can be ultimately expanded to support integration of this kind of dashboard with emotion detection systems. Applying automatic detection would also provide levels of each emotion in an objective way rather than from a personal perspective, which certainly would improve the validity of the information. Exploring how data that can originate from a multitude of sources and formats can be harvested, curated and fused (Sedrakyan, De Vocht, Alonso, Escalante, Orue-Echevarria and Mannens, 2018) will allow integrating multi-modal data from various wearable sensors and audio/video streams in real-time automated solutions.

In addition, the evaluation presented in this work is limited to perceived usefulness and insight, and does not provide any insight related to potential impact on learning improvements. Thus, expanding the dashboard visualizations with mechanisms to capture a broader scope of learning processes could be another future direction. For instance, process analytics driven approaches (Sedrakyan, 2016; Sedrakyan, De Weerdt and Snoeck, 2016;
Sedrakyan et al., 2014) targeting broader learning-related indicators (Glahn, 2009) will be a relevant future study. Exploring mechanisms for coupling visualizations with textual advice, such as cognitive feedback and behavioral feedforward (Sedrakyan, 2016; Sedrakyan and Snoeck, 2017; Sedrakyan, Jarvela and Kirschner, 2016) as well as a generalizing for different learning goals and tasks (e.g. solo/collaborative learning), is yet another possible direction for future work. Furthermore, not many studies can be found in the domain of feedback automation (Sedrakyan and Snoeck, 2016) thus requiring further research for solid methodologies and frameworks for delivering automated feedback that communicates emotion-related information. Finally, stricter experimental designs with controlling broader evaluation variables and constructs for user acceptance are needed to gain in-depth insights both for future scientific and practical implications.

References


Corresponding author
Gayane Sedrakyan can be contacted at: gsedrakyan@gmail.com
Entrepreneurship students distilled their learning experience through reflective learning log

Khar Kheng Yeoh
School of Business Management, College of Business, University Utara Malaysia, Changlun, Malaysia

Abstract

Purpose – This Scholarship of Teaching and Learning research is a part of the larger study grant to analyze written reflections through learning log among the third and final year students undertaking BPME 3073 Entrepreneurship module in University Utara Malaysia (UUM). The paper aims to discuss these issues.

Design/methodology/approach – The data collection techniques are researcher-directed textual data through reflective learning log, taken from 140 students from three classes. A thematic approach was utilized to present the reflections of the students and all data were recorded in a verbatim format.

Findings – The findings show that most students have never written a reflective log or essay in the formative assessment. As a consequence, they had difficulty in writing the reflection when being requested to do so. A total 75 (approximately 55 percent) of the reflective logs were identified as level 1 (from 1 to 5 percent) in which reflections were simply written in a descriptive manner, resulted in a balance of 61 learning logs being utilized for further analysis. The students’ reflections on their entrepreneurship experience systematically categorize into four different themes comprised of: the nature of entrepreneurship module, entrepreneurial characteristics, opportunity recognition, and creativity and innovation.

Research limitations/implications – As for the limitation of the study, it is important to note to not underestimate the challenges of introducing a grade assessment that most of them are not familiar with in their university academic journey. Students need guidance, assurance and confidence writing something that requires personal opinion, own thinking, sensitive and personal nature of narration. For most students as found out in this study, self-confessional writing is hard to come by (they dare not attempt it in the first place), only a handful appreciating the writing start with “I,” “me” as first person. More research in this study should be conducted across the university to gauge the response from the students to see if the result of this study is only applicable to this group of students or to this discipline of studies. The researchers would also like to recommend for future studies which take the form of a longitudinal study of similar kind to examine the problems and challenges with regards to promoting learning reflection at the undergraduate level.

Practical implications – Based on the result of the 61 students who had demonstrated an ability in reflective writing, it is suggested that perhaps the university should consider offering coursework that contains a component of reflective writing as part of the assessment. As such, if this is implemented, students of such ability like the one in this sample group would have been benefitted from such assessment which look at reflective ability (Greene, 2014) and which they were allowed to form a broader perspective in relation to the module undertaken. This in turns will foster the growth of reflective ability which is recognized as a learned behavior (Gustafson and Bennett, 1999). In addition, for the future exercise of this reflective learning log, the researcher opined that we should encourage our students to engage with another student (e.g. close friend) in a way that encourages talking with, questioning, or confronting, helped the reflective process by placing the learner in a safe environment in which self-revelation can take place. In addition, students were able to distance themselves from their actions, ideas and beliefs, by holding them up for scrutiny in the company of a peer with whom they are willing to take such risks (Hatton and Smith, 1995).

Originality/value – The results of this research have strongly suggested the need to urgently develop among the students the skills in writing reflectively as they go through the process of higher education which
is useful in molding their future professional and entrepreneurial behavior as when they entered the job market which requires a critical reasoning ability.

**Keywords** Entrepreneurship, Thematic analysis, Critical reflection, Action research, Critical thinking, Reflective learning log

**Paper type** Research paper

**Introduction**

Entrepreneurship has always been conceptualized as a process of discovering an opportunity, sourcing and manipulate resources, planning, and execution intelligence. An entrepreneur knows all the parts and knows how to match and fit all those resources together. The above narration signifies that a process is quite predictable, but in reality, the nature of entrepreneurial activities (to undertake) is not predictable at all. Greene (2014) proposed a portfolio of four complementary techniques for teaching entrepreneurship as a method, not as a process. They are businesses start-up, reality-grounded simulations and gamification, design-oriented learning, and reflective practice.

Each method requires learners to extend beyond the process-based paradigm of knowing, analyzing, and talking, instead positioning the learners to create, apply and act (Greene, 2014). In this Scholarship of Teaching and Learning (SoTL) study, the researcher adopts the reflective practice as a method for teaching entrepreneurship and the toolkit used is reflective learning log. Given its importance, reflective skills have now being recognized as an important proficiency among the professionals and being treated as a good source for critical thinking development, enhancing self-monitoring as well as leading to one having developed good reasoning skills. Furthermore, reflective exercise has also become a vital component for most professional degree programs, namely nursing, teaching as well as many business capstone courses that require final year students to recap and reflect before the end of their university’s tenure (Chalk and Hardbattle, 2007).

Reflection is a significant process by which knowledge is derived from experience. When reflecting, one considers an experience that has happened and tries to understand or explain it, which often leads to insight and deep learning – or ideas to test on new experiences. Reflection is particularly important for puzzling experiences, operating under conditions of high ambiguity, and problem solving. As a result, it should not be a surprise that reflection is a pivotal component of entrepreneurship education and also a way of practicing entrepreneurship.

In this study, the researchers aspire the students in BPME 3073 Entrepreneurship to develop a strong sense of awareness and appreciation of reflecting by looking at what they are doing/learning now in entrepreneurship module. This type of reflection is often referred to as “reflection-in-action” (Schön, 1987). Students who lack of this reflection-in-action are likely to make mistake of repeating the same dysfunctional behavior/attitude/decision in their future career/entrepreneurial endeavor, therefore impede their advancement in their respective fields. Students who possess this appreciation of reflection demonstrate it through reflective writing. Reflective writing is a metacognitive, “thinking about your thinking” process (Martinez, 2006). As a metacognitive process, the student is able to appreciate the deeper, underlying issues (Martinez, 2006), rather than accepting a superficial interpretation of the problem, which may present in a professional/entrepreneurial context (Mair, 2011).

This research was carried out among the undergraduate students of the third and final year, majoring in entrepreneur and business administration in University Utara Malaysia (UUM). As to assess the appreciation of reflection-in-action among our students of entrepreneurship, the students were requested to hand in two reflective learning logs which contain their reflection that will be used as part of the coursework assessment for the entrepreneurship module. Yinger and Clark (1981) believed that reflection results
written down are more powerful than reporting them orally. This form of writing a reflective journal has definitely being regarded as new to the students given that the norm in the coursework assessment of this module usually being quiz, mid-semester test, business plan and final exam.

This research was carried out with the two main aims as to:

1. investigate whether students at the tertiary level who are undergoing an entrepreneurship module is able to engage in a critical reflection in the process of learning; and

2. analyze the emerging themes on entrepreneurship from those students who are able to achieve Levels 2, 3 and 4 of Hatton and Smith’s (1995) framework of critical reflection.

Significance of the study
In this study, the researcher adopts the reflective practice as a method for teaching entrepreneurship and the data collection techniques are researcher-directed textual data through reflective learning log. Given its importance, reflective skills have now being recognized as an important proficiency among the professionals and being treated as a good source for critical thinking development, enhancing self-monitoring as well as leading to one having developed good reasoning skills. Furthermore, reflective exercise has also become a vital component for most professional degree programs namely nursing, teaching as well as many business capstone courses that require final year students to recap and reflect before the end of their university’s tenure (Chalk and Hardbattle, 2007). However, actually putting vague thoughts or feelings into a format that other people read is not a process that some people enjoy or find easy (Harvey and Knight, 1996) and such skills can be especially difficult to develop for Malaysia’s students which traditionally have adopted instructivist learning techniques.

Critical reflective learning
It is not easy to write our thinking process in a form of reflection but it should be practiced in learning, it helps to encapsulate learning in much more meaningful manner. The use of reflection journals can assist to document thinking process which in turns helps teacher to draw conclusion on the learning progress of the students. The entire process of writing per se may have somehow encouraged students to reflect on what they have been taught in class and thus facilitate learning further. According to Cowan (1998, p. 16), a student is said to be doing reflection when “[…] she notes that there is something different about the case that she is considering, in comparison with the examples she has encountered in class; and when she also identifies what the difference is, and what she should do about it.”

Dewey (1933) suggested that reflective allow one to be engaging in deep thinking in order to get an in-depth meaning of something, converting uncertainty into understanding which leads to action. This is in accordance to what Moon (1999) in which reflection entails mental activity that occurs in relation to the processing of complex ideas which are commonly found in the process of learning. Thus, so as to realize the benefits of reflection, students should be required to reflect and to write a reflective log or journal so that they can easily see the importance of the learning activities that have been carried out. As a result, reflection journal will reveal the thinking since the entire process of writing is actually a display of thinking (Luidens, 1997). Furthermore, idea clarifications and modifications happen due to the need to present knowledge in different form during the reflection writing process. As such it is expected that during this process, the learner will be able to develop new understanding and view the information in a different perspective (Yinger and Clark, 1981).
Methodology
The research methodology used in this study was classroom-based educational action research (Angelo, 1991; Elliott, 1991). Part of effective teaching is the ability to reflect on what is happening in the classroom, and to identify any differences in what was planned and what actually occurred. By conducting “systematic, intentional inquiry” within his/her own classroom, the instructor builds a better understanding of his/her own practice (Cochran-Smith and Lytle, 1993, p. 7).

At the beginning of the academic session of 2015/2016, all students taking the BPME 3073 Entrepreneurship module in UUM had been introduced to the concept of reflection writing through the reflective learning log (Moon, 1999) that will be assessed as part of their coursework. The idea of including the reflective learning log as part of the coursework mark contribution is to entice students to participate in the exercise given that naturally students tend to behave strategically when come to their class participations. Such behavior commonly leads to the occurrence of a phenomenon known as “Strategic Students” (Kneale, 1997). In this phenomenon, students would normally resist to participate in learning beyond formal assessment requirements; point to the necessity to link assessment and learning together. Moreover, excluding the reflective learning log from formal assessment can send a negative signal to the students about the significance of reflection within the module. Nevertheless, on the hindsight, it is important to note that it can be quite impossible to know as to how the effect would take place on the comments of the students given the allocation of marks associating to the learning log.

In this qualitative research, data were collected using the reflective learning log from 140 students of three classes in which these students were requested to carry out written reflections which are assessed for final grades (contributing a maximum of 10 percent of the total coursework marks). However, students have also been informed that they will only be rewarded a maximum of 5 percent should they write their reflection in a very descriptive or reporting manner. As such, the researcher considered those students who were able to obtain a score of more than 5 percent for the reflective log exercise would be considered as those who were able to achieve the learning outcome of the formative assessment (see Table AII). Hence, in this study, students were required to keep a record of the lessons that they have learned for duration of ten weeks and then write a reflection on the particular issues of the entrepreneurship taught in each lesson based on the syllabus. At the end of the study, four students failed to hand in the learning log resulted in only 136 learning logs were used for further analysis.

Ethical guidelines were strictly adhered to which included getting their consent and keeping their identity hidden so that they could express themselves freely (Marshall and Rossman, 1999). Finally, to protect the students’ privacy, we use pseudonyms in the place of real names to report our findings.

Critical reflection framework
Given that the learning log is a graded assessment, the benefits of doing reflection in the academia and at work, as well as the various types of reflective writing techniques had been informed to the students. For the purpose of grading, the author adopts the reflection framework as suggested by Hatton and Smith (1995) for the purpose of grading and subjectivity reduction. The framework contained four levels of reflections (Hatton and Smith, 1995) (which was made know to the students) as depicted below:

1. Descriptive writing (contains no evidence of reflection) – Level 1.
2. Descriptive reflection (a description of events with reflection from one perspective) – Level 2.
Dialogic reflection (some “stepping back” from events and recognition of alternative viewpoints) – Level 3.

Critical reflection (awareness that the same actions and events are viewed in a different way by different individuals) – Level 4.

In executing the study, the researcher gave an assurance to the students that their grades will not be penalized when they provide constructive criticism on the entrepreneurship module or on the taught concepts or even on the style of the lecture. In the same manner the students were also being reminded that their grades will not be affected positively as it contain purely praises for the said module. The students were told that evaluation will solely be based on the quality of reflection shown based on the Hatton and Smith’s (1995) framework.

A very basic scaffolding framework was provided for the reflective task which gave students three prompts which they could structure their reflection around; what have you learnt from the module so far? What are the topics in this module that give you many insights on becoming an entrepreneur? How will you be able to use the knowledge from this module in your professional career? The prompts were aimed at leading the students away from writing a descriptive reflection ordered by the topics taught on the module. Only 12 students utilized the framework, the majority completed a free style reflection. The fact that so few students used the framework may mean they found it difficult to use because it was designed to discourage descriptive answers and promote reflective thinking. The students may also have felt that they did not need guidance to help them structure their reflection or even that the mark given (10 percent) for the reflection did not warrant spending too much time writing it.

Analyses and discussions

Based on the mark sheets for reflective learning log (see Table AII), Table I tabulates the frequency of score percentage for graded reflective learning log.

All of the reflections in the findings section were taken from the students that have achieved the score of 6 percentage points to 9 percentage points from the total score of 10 percentage points based on Table AII. A total 75 (approximately 55 percent) of the reflective logs were identified as Level 1 (from 1 to 5 percent) in which reflections were simply written in a descriptive manner in the absence of reflections on the weekly lessons. Consequently these learning logs were removed from analysis; resulted in a balance of 61 learning logs being utilized for further analysis. Given that the numbers of learning logs (75 logs altogether) that have been discarded for further analysis were large (approximately 55 percent), it may provide a significant indicator that the task of doing reflection can be challenging due to the lack of ability in reflective thinking or it could also be an indication that the students were simply not interested in pursuing the assessment completely.

Notwithstanding the importance of reflective exercise to undergraduate students, the author found out that most of the 140 students had taken subject ranging from 18 to 30 subjects (Table AI), yet this is the first time they were required in this entrepreneurship module to write reflective learning log in their formative assessment. Many of the samples did not achieve the learning outcome of the reflective learning log objectives (see Table AII).

<table>
<thead>
<tr>
<th>Score percentage</th>
<th>Frequency</th>
</tr>
</thead>
<tbody>
<tr>
<td>1-5</td>
<td>75 (Level 1)</td>
</tr>
<tr>
<td>6</td>
<td>35 (Level 2)</td>
</tr>
<tr>
<td>7</td>
<td>17 (Level 3)</td>
</tr>
<tr>
<td>8</td>
<td>7 (Level 4)</td>
</tr>
<tr>
<td>9</td>
<td>2 (Level 4)</td>
</tr>
<tr>
<td>10</td>
<td>0</td>
</tr>
</tbody>
</table>

Table I.
All of the 140 students in this module had taken more than 18 subjects prior to this module. Some of the students, at the time of this study being carried out, had even registered/taken 30 subjects prior to this entrepreneurship module (see Table AI) and yet it was the first time they were being asked to write reflective learning log in their formative assessment. Absence of such experience had led to the students not being able to do reflection effectively which in turns having many of them merely able to engage in Level 1 of writing that is descriptive writing while others only able to reach the Level 2 of writing known as the descriptive reflection which normally look at events from a single point of view.

A thematic approach was utilized to present the reflections of the 61 reflective learning logs (in anonymity) and all data were recorded in a verbatim format. In analyzing the data, we used open coding (based on constant comparative methods) (Strauss and Corbin, 1998) to identify themes to represent repeated ideas that emerged across the reflections about their experiences in entrepreneurship module. We assigned labels to the reflective learning logs, as a means to identify patterns in the students’ reflection and discussion. We engaged in peer examination to ensure that our analysis was reliable and that our own biases and predispositions would not affect our interpretation of the data (Gray, 2004).

Discussions
The students’ reflections on their entrepreneurship’s experience could be grouped into four different reflection organizing themes comprised of:

1. the nature of entrepreneurship module;
2. entrepreneurial characteristics;
3. opportunity recognition; and
4. creativity and innovation.

Each of these themes is described by student reflections that are found to be fulfilling the criteria of reflective learning log (all of these reflections were taken from the students that have achieved the score of 6 percent points to 9 percent points from the total score of 10 percent points). Last but not least, the final sub-section of the results highlights the descriptive nature of reflective learning log (of those which do not fulfill the reflective log criteria as according to the Level 1 of the Hatton and Smith’s (1995) framework with the score percentage point was from 1 to 5 percent). This research analyses the students’ reflections to ascertain their views on the topics that the module covered and their reflective ability. A reflective learning log normally contains a record of one’s experiences, thoughts, feelings and reflections (Miller et al., 1994).

Entrepreneurship module
This theme looks at whether students were able to do reflection which is related to the entrepreneurship module as itself. This study found that there were 45 samples contained reflections that were related specifically to the nature of entrepreneurship from a total of 61 reflection logs. In addition, there were 30 students who found the entrepreneurship module interesting even though some of the students (six students) felt the module is not benefitting them at all given the opinions that the module contains topics that do not relate to one being an entrepreneur. Some people say – entrepreneurship cannot be taught in the classroom setting. Below were the excerpts from our students whose names had been disguised:

Wan Ting: In my own opinion, entrepreneurship is a subject that not needs study in university. It is because, entrepreneurship is the person who dislike study and lazy to work and my friend tell me that, nowadays, if u got degree, your salary also no enough to cover your life. So, she decide to stop hers study and go to find hers own goal. That time, I can’t understand. Why she wants to stop hers
Wan Ting was questioning the legitimacy of this module as she found this module was not useful for her. From her anecdotal example presented above, she opined that to be an entrepreneur, she does not need to study entrepreneurship as a subject because many successful entrepreneurs never set foot in the ivory tower, they just do it (she quote the budding entrepreneur dislike study). She even goes to the great length to argue that a degree was not sufficient to guarantee a successful life thereafter:

Samantha: I felt this subject “Entrepreneurship” was an unnecessary subject for me. From my last experienced, I felt bored and I’m taking this subject because I’m just wants to finish up my syllabus! For my perspectives, I think we totally don’t have to take this course. For me, to become a successful entrepreneur is not just study all these theories only. Even though we have memories and study all the theories, but it is still not guarantee us to become a successful entrepreneur. Instead for those entrepreneurs, even though they are not taking this subject but they will also become a successful entrepreneur. I don’t think so this subject can cultivate the entrepreneur, maybe can but I think is just a few only! I hope also can become a successful entrepreneur in one day! (sic).

As like Wan Ting, Samantha, also poured her heart out on this subject in the negative way (unnecessary subject for me). Samantha wants to be a successful entrepreneur but she did not think study the module per se with all the entrepreneurship’s concepts and theories will transform her to become one (it is still not guarantee us to become a successful entrepreneur). Like Wan Ting, Samantha also questioned the legitimacy of this module as they felt bored, wasting their time and to the very core, they just to complete their credit hours of this module to obtain a degree.

But what expressed above by Wan Ting and Samantha did not reflect the current trend in the world of business and academic. In recent years, we have seen an extraordinary proliferation of entrepreneurship and small business courses and programs in colleges and universities worldwide (Solomon, 2006). From its origins until today, entrepreneurship research and teaching has met many important milestones. It appears that such rocketing interest in entrepreneurship has shaped not only scholarly writing, but also initiated a “revolution” in entrepreneurship education at academic institutions. This “revolution” has also sparked an interesting scholarly discourse between Kuratko (2004, 2005) and Katz (2006, 2008) about the maturity and legitimacy of entrepreneurship as not only a field of study but also a field relevant to higher education.

In other words, the adoption of programs (i.e. entrepreneurship) makes the higher education institutions like UUM relevant and it signals that UUM graduates have the preparation they need to succeed in a competitive environment. This goes as far as hiring employees based on the legitimacy they provide to the institution (UUM) rather than their (graduates) actual capabilities. This should augur well for UUM graduates like Samantha and Wan Ting as it was compulsory for all bachelor of business undergraduates to take this module.

As with entrepreneurship’s theories as highlighted by Samantha, Fiet (2001) specifically addressed this issue when he observed the large number of adjuncts that teach entrepreneurship:

Students must learn theory in order to understand the future consequences of their entrepreneurial decisions. Thus, it seems improbable that adjuncts can effectively teach the course. Teaching theory is rarely the strength of adjuncts (Fiet, 2001, p. 9).

Adi: My first impression on this class is that it is actually quite interesting as I am uninterested on this class at first as I thought that this class will be a very hard one and involve a lot of reading. Indeed it does involved a lot of reading but it is interesting that I sometimes imagine myself in the situation or topic that I read. For example, franchising is a very famous topic related to entrepreneurship. So, it makes me imagine that someday I will have my own franchise and how
I will manage it. As a Malay myself, of course I would really like to be just like these successful entrepreneurs below and I would really want to see myself to be as good as them in the future (sic).

Kee Keat: Based on the knowledge that I had learned in the class, I have a clearer understanding about what is entrepreneurship and it change many of my misconception about entrepreneurship. It really inspires me. However, becoming an entrepreneur is not an easy task because they have to face many problems and make the best choice of every decision. It needs a good planning before we start our business (sic).

Despite an initial lukewarm interest in the module, a change in Adi’s behavior and the comments from Kee Keat can be seen among the students when they started to view how the module content can be useful for them. Most of the reflection of the students has similar comments as the one depicted by Adi in which, the expression of relishing certain topics of the entrepreneurship module is prevalent and how the module as a whole had changed their worldview of entrepreneurship forever. This module also stimulates Adi’s imagination as he envisions himself as a successful franchise entrepreneur in the future. He went on to search list of successful Malay entrepreneurs so that Adi can emulate and be good as them. As for Kee Keat, he managed to correct its misinterpretations on the concepts of entrepreneur and aspire to become one. He also has high awareness that to start a business is no easy feat can be done with good planning (it needs a good planning before we start our business).

Entrepreneurial characteristics

All 61 logs contained reflections that related to the entrepreneurial characteristics. In total, 35 students made specific reflections that learning entrepreneurial characteristics had demonstrated the importance of “taking calculated risk, ideation, control of our own future,” something they had not done previously or that the “module has made me rethink my personal beliefs and has made me consider to become entrepreneur.” Many students made specific reflections about their intention to use whatever they had learnt on the module to help them venture into their small businesses and achieve their “dream.” In a nutshell, this topic uncovers what make a successful entrepreneur like Tony Fernandez and Richard Branson “thick”:

Hafiz: From the lesson of this week, I could understand that entrepreneur is a person who is taking a calculated risk. This is meant that entrepreneur is not a gambler who is just trying luck without any effort. To become an entrepreneur, firstly we have to identify the opportunity that we see and use the opportunity wisely to enhance our life and also the community’s life. As example, Tony Fernandez who is the owner of Air Asia starts his business by making a loan to government. His idea is brilliant (sic).

The above account by Hafiz highlighted many intrinsic value of an entrepreneur, namely, calculated risk, opportunity identification, community’s life enhancement and ideation. There is clear evidence of dialogic reflection by citing Tony Fernandez from Air Asia. Hafiz is making an association between all the good values espoused by an entrepreneur and connects it with his icon, Air Asia’s Boss. Hafiz further draws the inspiration on how young people like him can learn so much from Tony as a successful entrepreneur:

Saidatul: I found out that even women also have a big potential to become a successful entrepreneur. The entrepreneur not only belongs to man. This video proves that women also have the entrepreneur characteristics that the men had. The women also have the strong passion and interest to become an entrepreneur and this passion doesn’t less than men had. In my opinion, everyone also can become a successful entrepreneur no matter the gender, age, and nation. It is only depends on the person themselves that they really want to be (sic).

Saidatul reflects emotionally for the above account. As a woman, Saidatul strongly opined that successful entrepreneur does not belong to man only; woman as shown in the video during the class can also become successful. In her frame of mind, woman equal with man in the endeavor of entrepreneurship. Clearly, in this account, self-questioning is evident (an “internal dialogue”
Kamariah: But the most interesting thing in this first chapter is that it makes me think more whether I should be an entrepreneur or not, and if I want to be an entrepreneur why? As I think all over again, I think that being an entrepreneur might be a good thing for me because I really want to have a full control of my own life, and give something to my family especially my parents, and lives like how I have dream before. But then when think about the challenges to be an entrepreneur nowadays, it kind of breaking me apart as I picture myself in a very difficult situation if I become an entrepreneur. Just thinking about the competition that I will faced makes me wondering how tough the life it is for those successful entrepreneur when they first started (sic).

Like Saidatul, Kamariah is self-questioning herself (whether I should be an entrepreneur or not?). Kamariah is deliberating between different views of her own behavior (different views of her own and others). She delves into the question “Why I want to become entrepreneur”? She not only starts to envision the benefits of becoming one successful entrepreneur, but also worry about the huge challenges that are going to be encountering before becoming one. Kamariah says “But then when think about the challenges to be an entrepreneur nowadays, it kind of breaking me apart as I picture myself in a very difficult situation.” This shows Kamariah is learning new skills in contemplating and decision making. In addition, it unveils the feelings of insecurity when she visualizes the difficult road ahead of a budding entrepreneur.

Tina: There are a few characteristics of an entrepreneur. Dare to take risk, hardworking, creative and innovative, open-minded, energetic, self-confident, and optimistic are some of the characteristics of the entrepreneur. Some of the people believe that entrepreneurs are born not create. However, it is not true. Most of the entrepreneurs are force to be an entrepreneur because of their condition. Sometimes, when people migrate to the other place, they will start a new business too because they have nothing at the new place. Beside, entrepreneur nowadays are born through knowledge. There are myths that say that “entrepreneur is those who do not do well in their academic.” However, I do not agree with it (sic).

Tina critically analyzed and opined that: “Some people believe that entrepreneurs are born not create. However, it is not true.” She recognizes how prior experience, thoughts (own and other’s) interact with the production of her own behavior. She gave her own opinion and explains critically that: “Most of the entrepreneurs are force to be an entrepreneur because of their condition.” She goes on to explain all the rationale of why entrepreneur is not born, but due to their predicaments. She also cited that immigrants tend to become entrepreneur in the host country because their lack of origin endowment. Again, self-questioning is evident in her reflection; deliberating between different views of her own behavior (different views of her own and others). She further says that: “entrepreneur nowadays are born through knowledge” and use this statement to debunk the myths that says: “entrepreneur is those who do not do well in their academic.”

Tina succinctly supported her argument with: “entrepreneurship has become a core subject for every university’s students,” such as in the Malaysian context, therefore, all Malaysian entrepreneurs are well educated in the future. This is a logical argument. There is clear evidence of dialogic reflection in the final writing of Tina. Ironically, Tina reflection on the last thought of well-educated entrepreneur stand in total contrast with the reflection of Wan Ting, which says: “Entrepreneur is the person who dislikes study.”

**Opportunity recognition**

As with entrepreneurial characteristics, all 61 logs contained reflections that related to the theme of opportunity recognition. Opportunity recognition is the essence of
entrepreneurship as entrepreneur is all about spotting and acting on opportunity. According to the definition that research works consider opportunity to be laying at the heart of the entrepreneurial process, an entrepreneur is an individual who is able to identify, evaluate and exploit opportunities (Shane, 2003; Venkataraman, 1997):

Akhmal: How we define opportunities? What if, when opportunities come, and we don’t know anything about it? After the class I try to search on the internet about the definition of opportunity. Then only I realize that, there’s no any clear definition about. Let’s say in one situation, there’s very less people selling fried banana at the area, then a person start a fried banana business without realizing these factors. Does it mean that the person didn’t recognize the chance and grab it? In my understanding, I don’t think the person grab the opportunity although he done a great job. To me, opportunity recognition more towards the person keep thinking about what kind of food has not been selling at his area, then he figure out that it’s fried banana! (sic).

Reflection by Akhmal depicts the student engaging in asking self-critical question about what is “opportunity.” That made him/her thinks very hard on this magic word and then synchronizes the meaning of “opportunity” to his/her own cognitive map.

Self-questioning is evident: “What if, when opportunities come, and we don’t know anything about it?” This critical question posed by Akhmal demonstrated that there are sense of “mulling about,” discourse with self and an exploration of the role of self in entrepreneurship module. Akhmal try curiously to find the meaning of “opportunity.” Even though Akhmal cannot find one to his satisfaction, but he try to make sense of the term with a good example. The example given showed Akhmal try hard to find closure on the term “opportunity”:

Qistina: When talking about opportunity, here is a company that using opportunity wisely – Khan Academy. When I visit his website www.khanacademy.com there are thousands of education resources which provide variety of subject such Mathematic, History, Healthcare, Medicine, and a lot more. When I heard that Khan Academy is a nonprofit organization, I truly curious about how he makes his money and get the financial resources. Then after make a research actually his revenue come from donation from Bill Gates and Google and also from advertising. Finally, I found out that Khan Academy becoming a powerful brand name in education it is because “they delivered value to people” (sic).

Qistina reflection signifies descriptive reflective. There is basically description of events on Khan Academy, but shows some evidence of deeper consideration in relatively descriptive language (e.g. I truly curious about how Khan Academy makes his money and gets the financial resources). Qistina linked her writing to the global organization like Khan Academy and she did a lot of research in order to complete her reflection. This is out of curiosity on Khan Academy. This is one of the benefits of reflective learning log as students reflect and engage actively in certain topics/themes, they tend to search and research extensively in order to close their “knowledge gap”:

Nurahimah: I once worked at a supermarket that sells goods worth RM2. I found many buyers purchase goods in our store because the price is cheaper than other stores. In uncertain economic conditions, high prices can temp customers to switch to a store that can offer low prices for them. This shows that the entrepreneur can use creative pricing strategy in uncertain economic condition as an opportunity to sell their product.

In the above account, Nurahimah try to connect her working experience in supermarket and the entrepreneurial concepts of creative pricing, opportunity and economic conditions. Nurahimah use her keen sense of observation to deduce the concepts of “opportunity” by making a connection between low price strategy and uncertain economic situation. Her reflection signifies opportunity never fall from the thin air but rather through skillful orchestration between strategy and environment:

Kee Keat: Opportunity is very important for an entrepreneur because opportunity is rather than just idea. If an entrepreneur has a good idea, but he or she do not has the chance to present or show to the other which mean that his or her idea will not be known by the public and hard to success.
They always need to observe and understand the trend of the business. In addition, they will not miss the opportunity which will change the future of the business. Beside, entrepreneur must be sensitive to the surrounding. A little change of the environment may lead to a huge change of the business especially the technology. Therefore, every entrepreneur must put effort in finding the opportunity and sensitive to the surrounding to ensure that their business can be success.

Kee Keat reflective writing is a metacognitive, “thinking about your thinking” process (Martinez, 2006) through “Opportunity is very important for an entrepreneur because opportunity is rather than just idea. If an entrepreneur has a good idea, but he or she does not have the chance to present or show to the other which mean that his or her idea will not be known by the public and hard to success” and “A little change of the environment may lead to a huge change of the business especially the technology.”

The above account requires Kee Keat to use analytical skills (i.e. higher order thinking skills), to differentiate between idea and opportunity rather than merely create a narrative. As a metacognitive process, Kee Keat is able to appreciate the deeper, underlying issues of entrepreneurship (Martinez, 2006), rather than accepting a superficial interpretation of the problem, which may present in an entrepreneurial context such as technological advancement (Mair, 2011).

Creativity and innovation
Only 45 reflections contained comments that related to the topic creativity and innovation. Many reflections cite that there are vast differences between creativity and innovation. Shane (2003) emphasized an entrepreneur’s creative role in innovation, with the observation that many founding teams use various forms of brainstorming to increase the number of new ideas, and thus enhance creativity as an important foundation for innovation. For many business students and the budding entrepreneur, innovation becomes the focal point of the reflections:

Wen Hui: Entrepreneurs need to create new ideas in business to produce new products not yet available in the market and that cannot be imitated by competitors. For example, Cirque de Soleil, which uses “Blue Ocean Strategy”, in a business where no entrepreneurs who are able to provide competition in the business. Cirque de Soleil has been doing business unique circus where it has employees who have high expertise and exceptional in doing interesting acrobatic actions. This business received overwhelming response from customers because it is different from the normal circus (sic).

Wen Hui quote world class circus like Cirque du Soleil, which uses “Blue Ocean Strategy” come to the fore of his/her reflection. She was able to connect creativity and innovation with Cirque du Soleil. Blue Ocean Strategy requires entrepreneur to create new ideas to produce new products not yet available in the market and cannot be imitated by the competitors. The uses of this unique example demonstrated the depth of propositional knowledge and the ability of Wen Hui to connect innovation and business bottom-line:

Zuliana: Today lecturer let us watch a video about woman entrepreneurs in Uganda. From this video I can see it is a very poor country. Yet, people did not give up their life; they are doing hard work for success, especially for woman entrepreneurs. A woman entrepreneur collects used straw from rubbish site, recycle the used straw and transform product like into lady handbag, shoes, and woman accessories. From here I see the woman entrepreneurs are very creative. When the machine used to flatten the straw breakdown, she even used her own teeth and knife to flatten the straw (sic).

Zuliana is touched by the heart-wrenching video showcased true-grit creative woman entrepreneurs in a poverty stricken country called Uganda. Through this video, Zuliana narrates what she observed from the hard lives of woman entrepreneur in Uganda. The description is succinct – just sufficient to raise the issues. Extraneous information is not added. It is not a story. The focus is on the attempt to reflect on the event and to learn from it. There is more of a sense of Zuliana standing back from the event in the video; in order to reflect better on her actions and in order to be more effectively critical.
Descriptive writing (non-existence of reflection)

In this SoTL, there were 75 samples of reflective learning logs that scored 5 percent and below 5 percent point (see Table AII). All of them achieved Level 1 (non-existence of reflection) in the reflective framework proposed by Hatton and Smith (1995). Descriptive writing is a description of events or literature reports. There is no discussion beyond description. The writing is considered not to show evidence of reflection. The followings are some of the accounts:

Asmida: Franchise is a semi-independent business that individual pays the fees and royalty for the trademark that will be used to sell the product and their service. This business is easier compare to the individual that want to open their own business. For example like Mc Donald which has franchise for all over the whole that many franchisees pay the fees to run their business (sic).

This account written by Asmida is descriptive and it contains little reflection. The account describes what is franchise, its characteristics and some of the benefits but all in the context of an account of the event. Generally one point is made at a time and ideas are not linked. It was wordy, tended to be meaningless and lacked sophisticated vocabulary. Overall, the writing was particularly colloquial in style.

Tina: An entrepreneur is someone who organizes, manages, and assumes the risks of a business or enterprise. An entrepreneur is an agent of change. Also, an entrepreneur who is makes a profit. The profit-and-loss system of capitalism helps to quickly sort through the many new resource combinations entrepreneurs discover (sic).

The above definition written by Tina was well versed but the entrepreneurial concepts are taken on without questioning them or considering them in depth. It is very descriptive. It could be a reasonably written account of an event that could serve as a basis on which reflection might start on an entrepreneur’s risk taking behavior and efficient resource allocation through the profit-and-loss system, though it hardly signals any material for reflection—other than the last few words stated “positive side of business failure”.

Hafiz: During the week, I study the new chapter about related to business networking. In my opinion, in this kind of business network should be established and it is important to keep the business continue business gaining strength and develop. There are several types of networking. In the class as well, lecturer explains the strategy to build a network that is, build, consolidation and retention. My comment about this thing is my very interest to know more about this topic (sic).

The account is written only from one point of view—that of Hafiz. Its narration is both surface and superficial without any substance content for critical reflection in the context of business networking. The idea is not relevant or focused, such as “My comment about this thing is my very interest to know more about this topic.” There were also significant problems with grammar, characterized by multiple errors in the original transcript of the Hafiz. Apparently, there is a lack of checking in the process of preparing this reflective learning log.

Ahmad: Lecturer continue new chapter that getting funding or financing. In this chapter lecturer explains why most new ventures need financing or funding. In addition there is also help to find alternative financial sources such as personal fund, equity capital, debt financing and creative sources. Lecturers are also explains about bootstrapping, other word is thrifty and there are 9 example of bootstrapping methods (sic).

Instead of using first party address, Ahmad tends to use third party address such as lecturer in writing his reflection. Critical reflection very much focused on the individual discourse with self and an exploration of the role of self in events and actions. The above narration is very much descriptive in nature and hardly contains any significant materials for reflection. The writing reflects a mere regurgitation of learning materials from the textbooks and the lecturer’s power point slides. There was a lack of sophistication in the writing.
The reflection by the above four sampled students in this study indicated the reflective learning log were descriptive in nature. The writings from the reflection learning log were a mere regurgitation from the power point slides and textbooks given in lectures throughout the 14-week class. They were written in reporting style, all in the context of an account of the event and more like registering everything happen in those particular week. The reflective learning log failed to address things like thoughts, feelings, how well (or bad) it went, what the students had learnt, and what he/she will do differently next time (Miller et al., 1994). Based on the frequency of score for graded reflective learning log, 75 students merely self-reported the contents of the lectures throughout the semester that have no reflection at all (some of their reflection were highlighted in the reflections of students 16-20). Another 35 students produce the works that contain descriptive reflection from one perspective. Only 26 students were able to reflect critically at Level 4 (Hatton and Smith, 1995) in certain themes in entrepreneurship module.

**Implications and limitations**

This study highlighted that reflective writing was an unfamiliar genre to many business students and needed to be explicitly taught in the respective subject. The results of this study have strongly suggested an urgent need for the development of reflective writing skills among the students during their higher education studies so as to assist in the professional and entrepreneurial behavior development as they entered the job market. In addition, there is a necessity to equip the undergraduate with basic building blocks needed to mold them as reflective practitioners. The reflective framework provided by Hatton and Smith (1995) proved to be useful in categorizing the written reflections and making the grading of the reflective learning log as less subjective. The weightage given to this graded assessment was 10 percent and it is considered substantial in grade. While the result of the study may not provide strong evidence on the ability of the students in reflective writing, it will be strange should inability exist as the outcome of attitudinal problem of willingly forgo a set of marks by not writing reflectively in which may warrant further investigation.

Based on the result of the 61 students who had demonstrated an ability in reflective writing, it is suggested that perhaps the university should consider offering coursework that contains a component of reflective writing as part of the assessment. As such, if this is implemented, students of such ability like the one in this sample group would have been benefitted from such assessment which look at reflective ability (Greene, 2014) and which they were allowed to form a broader perspective in relation to the module undertaken. This in turns will foster the growth of reflective ability which is recognized as a learned behavior (Gustafson and Bennett, 1999). In addition, for the future exercise of this reflective learning log, the researcher opined that we should encourage our students to engage with another student (e.g. close friend) in a way that encourages talking with, questioning, or confronting, helped the reflective process by placing the learner in a safe environment in which self-revelation can take place. In addition, students were able to distance themselves from their actions, ideas and beliefs, by holding them up for scrutiny in the company of a peer with whom they are willing to take such risks (Hatton and Smith, 1995).

The researcher found that more than half of the students were only able to reach Level 1 of reflection as suggested by Hatton and Smith (1995), the idea of giving additional specific guidelines will lead the students to certain stereotypes of writing reflective learning log (Stamper, 1996) and that may prove counterproductive. As such Holland (2013) suggested that it is only necessary for reflective writing skills being developed with the reflective thinking skills so as to assist in the ability of the students to write a good reflection.

Results which show the mark sheets for reflective learning log (see Table AII) indicate that most of the students need to be given greater opportunities for reflective writing skills development throughout their tertiary learning. This study suggested the importance of
having a blend of reflective writing exercise in most of the subjects at the undergraduate degree so as to facilitate not only dispersing knowledge on their module professional aspects but also create pools of reflective practitioners in the future.

As for the limitation of the study, it is important to not to underestimate the challenges of introducing a grade assessment that most of them are not familiar with in their university academic journey. Students need guidance, assurance and confidence writing something that require personal opinion, own thinking, sensitive and personal nature of narration. For most students as found out in this study, self-confessional writing is hard to come by (they dare not attempt it in the first place), only a handful appreciating the writing start with “I,” “me” as first person. More research in this study should be conducted across the university to gauge the response from the students to see if the result of this study is only applicable to this group of students or to this discipline of studies. The researchers would also like to recommend for future studies which take the form of a longitudinal study of similar kind to examine the problems and challenges with regards to promoting learning reflection at the undergraduate level.

Conclusions
By maintaining a learning log for the past 14 weeks in this entrepreneurship module, students can record and comment their every week entrepreneurship lessons in a profound manner. They can self-track their perspectives in their first class in this module and monitor their evolved entrepreneurship knowledge and perspectives as the lecture progress until the end of the semester. In the process, they immerse in self-discovery in the world of entrepreneurship and self-tracking personal growth and cognitive development. Reflective learning log enables students to clearly identify their own thinking in entrepreneurship, the risks involved, the myths of becoming an entrepreneur, opportunity and execution intelligence – a method that promotes deep learning as depicted in the student’s comments in the findings section in this study. Research has identified that reflection can help people to change. Although most of the students registering for this module are young undergraduate, and have limited experience to draw from to reflect upon certain issues (this is the findings from this study). But despite these barriers, reflective learning log makes entrepreneurship learning congruent with the suggestion to teach entrepreneurship as a method rather than as a process. Reflective learning log provides students with a profound insight into their own world of academic and; hopefully in future, their professional growth as reflection become a habit for my student.

References
Dewey, J. (1933), How We Think?, D.C. Heath and CO, Boston, MA.
Greene, P.G. (2014), Four Approaches to Teaching Entrepreneurship as a Method, Babson College, MA.


Corresponding author
Kharkheng Yeoh can be contacted at: kharkheng@uum.edu.my
## Appendix 1

<table>
<thead>
<tr>
<th>Matric no.</th>
<th>Total subjects taken (subject with reflective learning log)</th>
<th>Matric no.</th>
<th>Total subjects taken (subject with reflective learning log)</th>
<th>Matric no.</th>
<th>Total subjects taken (subject with reflective learning log)</th>
<th>Matric no.</th>
<th>Total subjects taken (subject with reflective learning log)</th>
</tr>
</thead>
<tbody>
<tr>
<td>219,792</td>
<td>30 (1)</td>
<td>227,437</td>
<td>17 (1)</td>
<td>225,447</td>
<td>17 (1)</td>
<td>227,810</td>
<td>20 (1)</td>
</tr>
<tr>
<td>219,868</td>
<td>25 (2)</td>
<td>227,440</td>
<td>21 (1)</td>
<td>225,513</td>
<td>19 (1)</td>
<td>227,976</td>
<td>19 (1)</td>
</tr>
<tr>
<td>220,372</td>
<td>28 (2)</td>
<td>227,576</td>
<td>21 (2)</td>
<td>225,628</td>
<td>18 (1)</td>
<td>228,031</td>
<td>20 (1)</td>
</tr>
<tr>
<td>220,488</td>
<td>30 (2)</td>
<td>227,584</td>
<td>21 (1)</td>
<td>225,707</td>
<td>17 (1)</td>
<td>228,050</td>
<td>18 (1)</td>
</tr>
<tr>
<td>220,540</td>
<td>29 (1)</td>
<td>227,875</td>
<td>19 (1)</td>
<td>225,740</td>
<td>17 (1)</td>
<td>228,068</td>
<td>18 (1)</td>
</tr>
<tr>
<td>221,390</td>
<td>29 (1)</td>
<td>228,011</td>
<td>19 (1)</td>
<td>225,837</td>
<td>19 (1)</td>
<td>228,346</td>
<td>19 (1)</td>
</tr>
<tr>
<td>221,573</td>
<td>29 (2)</td>
<td>228,018</td>
<td>19 (2)</td>
<td>225,949</td>
<td>19 (1)</td>
<td>228,348</td>
<td>19 (1)</td>
</tr>
<tr>
<td>221,577</td>
<td>29 (1)</td>
<td>228,064</td>
<td>17 (1)</td>
<td>225,955</td>
<td>18 (1)</td>
<td>228,492</td>
<td>16 (1)</td>
</tr>
<tr>
<td>221,632</td>
<td>29 (1)</td>
<td>228,382</td>
<td>19 (1)</td>
<td>225,986</td>
<td>19 (1)</td>
<td>228,504</td>
<td>17 (1)</td>
</tr>
<tr>
<td>221,812</td>
<td>28 (1)</td>
<td>228,405</td>
<td>19 (2)</td>
<td>226,036</td>
<td>18 (1)</td>
<td>229,131</td>
<td>19 (1)</td>
</tr>
<tr>
<td>221,822</td>
<td>29 (1)</td>
<td>228,416</td>
<td>18 (1)</td>
<td>226,058</td>
<td>17 (1)</td>
<td>229,135</td>
<td>18 (1)</td>
</tr>
<tr>
<td>221,862</td>
<td>29 (1)</td>
<td>228,563</td>
<td>14 (1)</td>
<td>226,100</td>
<td>18 (1)</td>
<td>216,678</td>
<td>33 (1)</td>
</tr>
<tr>
<td>222,068</td>
<td>30 (1)</td>
<td>228,592</td>
<td>15 (1)</td>
<td>226,211</td>
<td>18 (1)</td>
<td>222,146</td>
<td>30 (1)</td>
</tr>
<tr>
<td>222,071</td>
<td>29 (1)</td>
<td>229,128</td>
<td>19 (1)</td>
<td>226,309</td>
<td>18 (1)</td>
<td>223,193</td>
<td>29 (1)</td>
</tr>
<tr>
<td>222,083</td>
<td>29 (1)</td>
<td>230,227</td>
<td>12 (1)</td>
<td>226,633</td>
<td>18 (1)</td>
<td>223,206</td>
<td>31 (1)</td>
</tr>
<tr>
<td>222,302</td>
<td>33 (1)</td>
<td>230,276</td>
<td>12 (1)</td>
<td>226,689</td>
<td>18 (1)</td>
<td>223,214</td>
<td>30 (1)</td>
</tr>
<tr>
<td>222,342</td>
<td>28 (1)</td>
<td>219,701</td>
<td>30 (1)</td>
<td>226,857</td>
<td>18 (1)</td>
<td>223,244</td>
<td>29 (1)</td>
</tr>
<tr>
<td>224,135</td>
<td>23 (1)</td>
<td>220,781</td>
<td>30 (1)</td>
<td>227,145</td>
<td>18 (1)</td>
<td>223,305</td>
<td>29 (1)</td>
</tr>
<tr>
<td>225,622</td>
<td>18 (1)</td>
<td>220,807</td>
<td>29 (1)</td>
<td>227,160</td>
<td>18 (1)</td>
<td>225,727</td>
<td>29 (1)</td>
</tr>
<tr>
<td>225,976</td>
<td>18 (1)</td>
<td>221,298</td>
<td>28 (1)</td>
<td>227,167</td>
<td>17 (1)</td>
<td>225,776</td>
<td>18 (1)</td>
</tr>
<tr>
<td>226,101</td>
<td>21 (1)</td>
<td>221,318</td>
<td>28 (1)</td>
<td>227,272</td>
<td>17 (1)</td>
<td>225,781</td>
<td>18 (1)</td>
</tr>
<tr>
<td>226,168</td>
<td>19 (1)</td>
<td>221,447</td>
<td>30 (1)</td>
<td>227,315</td>
<td>19 (1)</td>
<td>226,106</td>
<td>18 (1)</td>
</tr>
<tr>
<td>226,245</td>
<td>18 (1)</td>
<td>222,041</td>
<td>29 (1)</td>
<td>227,492</td>
<td>19 (1)</td>
<td>226,176</td>
<td>18 (1)</td>
</tr>
<tr>
<td>226,281</td>
<td>18 (1)</td>
<td>222,093</td>
<td>29 (1)</td>
<td>227,599</td>
<td>17 (1)</td>
<td>226,184</td>
<td>18 (1)</td>
</tr>
<tr>
<td>226,296</td>
<td>18 (1)</td>
<td>223,304</td>
<td>18 (1)</td>
<td>227,606</td>
<td>18 (1)</td>
<td>226,297</td>
<td>18 (1)</td>
</tr>
<tr>
<td>226,566</td>
<td>18 (1)</td>
<td>225,280</td>
<td>18 (1)</td>
<td>227,649</td>
<td>18 (1)</td>
<td>226,321</td>
<td>18 (1)</td>
</tr>
<tr>
<td>226,658</td>
<td>17 (1)</td>
<td>225,239</td>
<td>17 (1)</td>
<td>227,663</td>
<td>18 (1)</td>
<td>226,323</td>
<td>17 (1)</td>
</tr>
<tr>
<td>227,379</td>
<td>18 (1)</td>
<td>225,378</td>
<td>17 (1)</td>
<td>227,685</td>
<td>18 (1)</td>
<td>226,328</td>
<td>20 (1)</td>
</tr>
<tr>
<td>228,426</td>
<td>18 (1)</td>
<td>226,261</td>
<td>18 (1)</td>
<td>226,704</td>
<td>18 (1)</td>
<td>227,525</td>
<td>18 (1)</td>
</tr>
<tr>
<td>226,430</td>
<td>18 (1)</td>
<td>226,608</td>
<td>18 (1)</td>
<td>226,718</td>
<td>18 (1)</td>
<td>227,569</td>
<td>19 (1)</td>
</tr>
<tr>
<td>226,438</td>
<td>18 (1)</td>
<td>226,616</td>
<td>18 (1)</td>
<td>226,764</td>
<td>18 (1)</td>
<td>227,868</td>
<td>18 (1)</td>
</tr>
<tr>
<td>226,456</td>
<td>17 (1)</td>
<td>226,688</td>
<td>18 (1)</td>
<td>227,075</td>
<td>18 (1)</td>
<td>227,901</td>
<td>18 (1)</td>
</tr>
<tr>
<td>227,919</td>
<td>19 (1)</td>
<td>228,301</td>
<td>18 (1)</td>
<td>227,390</td>
<td>18 (1)</td>
<td>229,129</td>
<td>18 (1)</td>
</tr>
<tr>
<td>227,982</td>
<td>18 (1)</td>
<td>228,303</td>
<td>18 (1)</td>
<td>227,445</td>
<td>17 (1)</td>
<td>230,183</td>
<td>12 (1)</td>
</tr>
<tr>
<td>227,983</td>
<td>18 (1)</td>
<td>228,340</td>
<td>18 (1)</td>
<td>228,273</td>
<td>19 (1)</td>
<td>228,350</td>
<td>18 (1)</td>
</tr>
</tbody>
</table>

**Note:** $n = 140$ students

**Table AI.** Records on the total subjects taken and total subjects with reflective learning log

Entrepreneurship students
<table>
<thead>
<tr>
<th>Matric no.</th>
<th>Reflective learning log (10%)</th>
<th>Matric no.</th>
<th>Reflective learning log (10%)</th>
<th>Matric no.</th>
<th>Reflective learning log (10%)</th>
<th>Matric no.</th>
<th>Reflective learning log (10%)</th>
<th>Matric no.</th>
<th>Reflective learning log (10%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>219,792</td>
<td>6</td>
<td>227,437</td>
<td>5</td>
<td>225,447</td>
<td>5</td>
<td>227,810</td>
<td>6</td>
<td>226,426</td>
<td>5</td>
</tr>
<tr>
<td>219,868</td>
<td>6</td>
<td>227,440</td>
<td>5</td>
<td>225,513</td>
<td>5</td>
<td>227,976</td>
<td>6</td>
<td>226,430</td>
<td>5</td>
</tr>
<tr>
<td>220,372</td>
<td>6</td>
<td>227,576</td>
<td>8</td>
<td>225,628</td>
<td>5</td>
<td>228,031</td>
<td>6</td>
<td>226,438</td>
<td>5</td>
</tr>
<tr>
<td>220,488</td>
<td>6</td>
<td>227,584</td>
<td>6</td>
<td>225,707</td>
<td>5</td>
<td>228,050</td>
<td>5</td>
<td>226,456</td>
<td>5</td>
</tr>
<tr>
<td>220,540</td>
<td>5</td>
<td>227,875</td>
<td>5</td>
<td>225,740</td>
<td>5</td>
<td>228,068</td>
<td>5</td>
<td>226,591</td>
<td>5</td>
</tr>
<tr>
<td>221,390</td>
<td>-</td>
<td>228,011</td>
<td>6</td>
<td>225,837</td>
<td>4</td>
<td>228,346</td>
<td>4</td>
<td>226,608</td>
<td>6</td>
</tr>
<tr>
<td>221,573</td>
<td>6</td>
<td>228,018</td>
<td>5</td>
<td>225,949</td>
<td>5</td>
<td>228,348</td>
<td>4</td>
<td>226,616</td>
<td>5</td>
</tr>
<tr>
<td>221,577</td>
<td>-</td>
<td>228,064</td>
<td>5</td>
<td>225,955</td>
<td>5</td>
<td>228,492</td>
<td>5</td>
<td>226,688</td>
<td>7</td>
</tr>
<tr>
<td>221,632</td>
<td>-</td>
<td>228,382</td>
<td>5</td>
<td>225,986</td>
<td>4</td>
<td>228,504</td>
<td>5</td>
<td>226,704</td>
<td>5</td>
</tr>
<tr>
<td>221,812</td>
<td>6</td>
<td>228,405</td>
<td>5</td>
<td>226,036</td>
<td>4</td>
<td>229,135</td>
<td>6</td>
<td>226,718</td>
<td>5</td>
</tr>
<tr>
<td>221,822</td>
<td>6</td>
<td>228,416</td>
<td>7</td>
<td>226,058</td>
<td>5</td>
<td>229,135</td>
<td>6</td>
<td>226,764</td>
<td>-</td>
</tr>
<tr>
<td>221,862</td>
<td>7</td>
<td>228,563</td>
<td>5</td>
<td>226,100</td>
<td>5</td>
<td>216,678</td>
<td>1</td>
<td>227,075</td>
<td>5</td>
</tr>
<tr>
<td>222,068</td>
<td>6</td>
<td>228,592</td>
<td>5</td>
<td>226,211</td>
<td>6</td>
<td>222,146</td>
<td>5</td>
<td>227,290</td>
<td>5</td>
</tr>
<tr>
<td>222,071</td>
<td>5</td>
<td>229,128</td>
<td>6</td>
<td>226,309</td>
<td>5</td>
<td>223,193</td>
<td>7</td>
<td>227,445</td>
<td>5</td>
</tr>
<tr>
<td>222,083</td>
<td>6</td>
<td>230,227</td>
<td>5</td>
<td>226,633</td>
<td>6</td>
<td>223,206</td>
<td>5</td>
<td>227,525</td>
<td>6</td>
</tr>
<tr>
<td>222,302</td>
<td>5</td>
<td>230,276</td>
<td>8</td>
<td>226,688</td>
<td>4</td>
<td>223,214</td>
<td>5</td>
<td>227,609</td>
<td>6</td>
</tr>
<tr>
<td>222,342</td>
<td>6</td>
<td>219,701</td>
<td>5</td>
<td>226,857</td>
<td>6</td>
<td>223,244</td>
<td>7</td>
<td>227,868</td>
<td>5</td>
</tr>
<tr>
<td>224,135</td>
<td>5</td>
<td>220,781</td>
<td>6</td>
<td>227,145</td>
<td>8</td>
<td>223,305</td>
<td>6</td>
<td>227,901</td>
<td>5</td>
</tr>
<tr>
<td>225,622</td>
<td>3</td>
<td>230,807</td>
<td>5</td>
<td>227,160</td>
<td>4</td>
<td>225,727</td>
<td>7</td>
<td>227,919</td>
<td>6</td>
</tr>
<tr>
<td>225,976</td>
<td>5</td>
<td>221,208</td>
<td>5</td>
<td>227,167</td>
<td>5</td>
<td>225,776</td>
<td>6</td>
<td>227,982</td>
<td>5</td>
</tr>
<tr>
<td>226,101</td>
<td>5</td>
<td>221,318</td>
<td>7</td>
<td>227,272</td>
<td>8</td>
<td>225,781</td>
<td>5</td>
<td>227,983</td>
<td>6</td>
</tr>
<tr>
<td>226,168</td>
<td>8</td>
<td>221,447</td>
<td>6</td>
<td>227,315</td>
<td>4</td>
<td>226,106</td>
<td>7</td>
<td>228,273</td>
<td>7</td>
</tr>
<tr>
<td>226,245</td>
<td>7</td>
<td>222,041</td>
<td>5</td>
<td>227,492</td>
<td>5</td>
<td>226,176</td>
<td>7</td>
<td>228,301</td>
<td>6</td>
</tr>
<tr>
<td>226,281</td>
<td>7</td>
<td>222,093</td>
<td>6</td>
<td>227,599</td>
<td>7</td>
<td>226,184</td>
<td>5</td>
<td>228,303</td>
<td>9</td>
</tr>
<tr>
<td>226,295</td>
<td>6</td>
<td>223,304</td>
<td>6</td>
<td>227,606</td>
<td>6</td>
<td>226,287</td>
<td>5</td>
<td>228,340</td>
<td>9</td>
</tr>
<tr>
<td>226,566</td>
<td>6</td>
<td>225,280</td>
<td>5</td>
<td>227,649</td>
<td>5</td>
<td>226,321</td>
<td>7</td>
<td>228,350</td>
<td>5</td>
</tr>
<tr>
<td>226,958</td>
<td>6</td>
<td>225,329</td>
<td>7</td>
<td>227,663</td>
<td>5</td>
<td>226,323</td>
<td>5</td>
<td>229,129</td>
<td>5</td>
</tr>
<tr>
<td>227,379</td>
<td>8</td>
<td>225,378</td>
<td>8</td>
<td>227,885</td>
<td>5</td>
<td>226,328</td>
<td>7</td>
<td>230,183</td>
<td>7</td>
</tr>
</tbody>
</table>

Note: $n = 136$
Learning analytics to improve writing skills for young children – an holistic approach

Nina Steinhauer and Michael Gros
Landesinstitut für Pädagogik und Medien, Saarbrücken, Germany

Martin Ebner
Department of Educational Technology, Graz University of Technology, Graz, Austria

Markus Ebner
Institute for Interactive Systems and Data Science, Faculty of Computer Science and Biomedical Engineering, Technische Universität Graz, Graz, Austria

Anneliese Huppertz and Mike Cormann
School of Raeren, Raeren, Belgium

Susanne Biermeier and Lena Burk
Albert-Weisgerber School, St Ingbert, Germany

Konstanze Edtstadler
Institute of Professionalisation in Early Childhood and Primary Teacher Education, University College of Teacher Education Steiermark, Graz, Austria, and

Sonja Gabriel, Martina Wintschnig, Christian Aspalter and Susanne Martich
University College of Teacher Education Vienna/Krems, Vienna, Austria

Abstract
Purpose – Due to the important role of orthography in society, the project called IDeRBlog presented in this paper created a web-based tool to motivate pupils to write text as well as to read and to comment on texts written by fellow students. In addition, IDeRBlog aims to improve student’s German orthography skills and supports teachers and parents with training materials for their students. The paper aims to discuss these issues.

Design/methodology/approach – With the aid of learning analytics, the submitted text is analyzed and special feedback is given to the students so that they can try to correct the misspelled words themselves. The teachers as well as the parents are benefiting from the analysis and exercises suggested by the system.

Findings – A recent study showed the efficiency of the system in form of an improvement of the students’ orthographic skills. Over a period of four months 70 percent of the students achieved a significant reduction of their spelling mistakes.

Originality/value – IDeRBlog is an innovative approach to improving orthography skills combining blogging and new media with writing and practice.

Keywords Technology enhanced learning, Learning analytics, German orthography,

German spelling acquisition

Paper type Research paper
Introduction

IDeRBlog enables students aged eight and above to write texts about a favorable topic on a computer, smartphone, or tablet connected to the web. Besides enhancing their media literacy by taking advantage of innovative ways of writing, students also improve their orthography skills. This is due to the fact that underlying the website is an "intelligent dictionary" which was developed within the frame of the project (Edtstadler et al., 2015). The dictionary analyses the spelling mistakes made by students. It provides not only feedback of how to correct mistakes, but also offers students and teachers alike an evaluation of the analysis of their spelling errors. On basis of this analysis students are directed to individual online or print exercises and training courses, in order to work on their difficulties. At the same time, students can publish their texts in a blog which can either be accessed by all members of the class, or the school users of the IDeRBlog-platform. Consequently, students can read the texts of fellow students and have the possibility to comment on them.

The aim of the website is to intrinsically motivate pupils to write texts as well as to read and to comment on texts written by fellow students and continuously considering the improvement of German orthography. The skill of writing orthographically correct is widely regarded as the basis of successful participation in school, later on in work life and everyday life within society (Grünke and Weber, 2015, p. 176). Due to Edtstadler et al. (2015, p. 6) the explanation for the significance of the orthography may be the following: “In contrast to other areas of language learning, there is hardly space to argue about the correct or incorrect spelling of a word. This orthographical stiffness can probably serve as an explanation for its importance.” Consequently, it is extremely important to detect possible spelling difficulties, in order to help students by giving individual support and tools to improve orthography skills autonomously (Ebner et al., 2015, p. 118).

The ongoing digitalization of society and everyday life changes the way of writing fundamentally. While students write with pen and paper in school they type in words, sentences and texts into smartphones, and computers when it comes to their leisure time. The use of social networking sites, communication tools, and apps is continuously increasing (Nagler et al., 2016).

Media literacy becomes more and more important. It is not only the required ability to write using new media, but also the question of suitability of statements and media for a publication on the internet that arises. Privacy and the so called “right to forget” are important issues as well. All of these aspects of orthography, writing, media literacy, and motivation are addressed by IDeRBlog. IDeRBlog is an abbreviation for “Individuell Differenziert Richtig Schreiben mit Blogs” (Individually DifferEntiated correctly wRiting with Blogs). The major aim of IDeRBlog was the development of a website that combines free writing opportunities with the improvement of orthography skills.

Registration and privacy policy

In order to use IDeRBlog students, teachers and their corresponding schools must be registered on the website. In this context, necessary information is the name of the teacher, the e-mail address, and the name of the school. Classes as well as students can be registered by teachers, too. The teacher can choose to register the students anonymously, e.g. with different names to ensure privacy. Due to privacy policies of the website, one cannot retrieve or even save this data. Privacy policies of Austria and the European Union apply to the website as well as the management and storage of the data.

Theoretical background

Learning analytics (LA)

LA is “the measurement, collection, analysis and reporting of data about learners and their contexts, for purposes of understanding and optimizing learning and the environments in
which it occurs” (Siemens and Long, 2011). This process of analysis has different steps: capture, report, predict, and refine (Campbell et al., 2007). These four steps can be used for a closed iterative LA-cycle, according to Clow (2012). Further, stakeholders can be added to the cycle, stating their mission and visions, according to Khalil and Ebner (2016). Interactions within platforms can be captured for later analysis to gain a deeper understanding of the learning process (Khalil and Ebner, 2015). LA allows to use these information to perceive learning issues and makes it possible for teachers to actively intervene (Siemens and Long, 2011; Greller and Drachsler, 2012; Greller et al., 2014). Appropriate visualizations can be used to provide insights for teachers, students, parents and academic personnel as well (Ruiprez-Valiente et al., 2015). This feedback has to be presented in a simple and informative way to reach out to all stakeholders involved in the process (Baker et al., 2012; Neuhold, 2013; Leitner et al., 2017).

Intelligent dictionary

The project’s main objective is to create more than a simple spell checker by developing the so called “intelligent dictionary.” The main idea is that the student gets a hint for correcting spelling mistake by providing specific feedback instead of only flagging a mistake and/or offering the correctly spelled word. In order to accomplish this task (for details see Edtstadler et al., 2015), the misspelled words are categorized into different orthographic areas, which are assigned to phenomena and are linked with corresponding feedback for giving hints for correcting mistakes. This way a modern didactic approach to learning and teaching German orthography is implemented, which focus on cognitive clarity (Ebner et al., 2017). In order to retrieve a manageable amount of categories for the teachers’ qualitative analysis, the phenomena are merged into categories, for which additional orthographic exercises are available on the platform.

Learning theory constructivism

The IDeRBlog website is mainly based on the principles of constructivism as a learning theory. Within this theory, knowledge is defined as an individual construct of the human mind rather than a copy of reality (Reimann and Mandl, 2006, p. 626). This construction is characterized by self-organization and active autonomous learning of the individual (Reimann, 2005, p. 155). These two aspects are especially encouraged while working with IDeRBlog. This is due to the fact that students manage their account as well as their texts on their own and work autonomously.

The design and realization of the website was mainly influenced by the concept of situated learning (Gros et al., 2015, p. 22). In the context of learning and teaching, situated learning evolved within the situated cognition movement. The movement emphasized constructivist elements of learning and prioritized constructivist aspects of pedagogy. From this point of view, learning always occurs within a social context and is characterized by the individual experiences of the learner. On basis of these experiences, knowledge is actively constructed by the learner (Reimann and Mandl, 2006, p. 627). In the process of learning, various tools and aids are used in order to complete an exercise or solve a specific problem (Fischer et al., 2009, p. 754).

Students working with the IDeRBlog website using different tools and aids in order to improve their German orthography skills. First of all, students receive feedback from the “intelligent dictionary” giving them advice on how to correct their spelling errors. Second, the teacher supports the students if they cannot draw the right conclusions based on the feedback. Third, in a later stage of the working process the discussion of the texts within the blog realized by the commentary function enables students to support each other. Especially, advice on writing skills beyond orthography skills, for example, choice of words, grammar or narrative style can be discussed within the blog.
The primary aim with constructivism is the flexible application of newly acquired knowledge. Besides that, the skills to organize and to solve problems autonomously are prioritized. Last but not least, the creation of (cognitive) strategies is defined as a major aim in a constructivist learning experience (Reinmann and Mandl, 2006, p. 627). Hence, strategies learned and internalized with the help of the feedback by the “intelligent dictionary” ought to be applied to different words of the same orthographic phenomena. Therefore, students should be able to correct and to prevent spelling mistakes with the help of those strategies developed.

In the context of constructivism, a formative evaluation of learning outcomes is considered as being rather significant. The focus is on the development of the individual learning process of each student. This is observed and evaluated independently from the performance of other students or the achievement of strictly defined learning goals (Reinmann and Mandl, 2006, p. 628). Consequently, continuous feedback concerning the individual learning process plays a major role. It serves as a point of reference, not only for the teacher but at the same time for the student (Reinmann and Mandl, 2006, p. 628). Furthermore, the feedback allows the student to review and control his learning progress (Fischer et al., 2009, p. 758). This aspect is realized within IDeRBlog by the continuous support of the learner through the feedback of the “intelligent dictionary” as well as the analysis of the spelling errors.

The IDeRBlog website can be considered as an example of a moderate constructivism (Reinmann and Mandl, 2006, p. 638). Moderate basically means that the assumption that all knowledge is exclusively constructed by the learner cannot entirely be hold up. There needs to be at least a small foundation of knowledge upon which the individual can build and acquire as well as construct new knowledge (Reinmann and Mandl, 2006, p. 638). Hence, new knowledge is based and constructed upon former knowledge and experiences of the learner. As a result, a moderate constructivism cannot completely refrain from instruction. In this context, the acquisition of new knowledge is rather an interplay between construction and instruction (Reinmann and Mandl, 2006, p. 639). The already existing knowledge is the learner’s ability to allocate sounds to corresponding letters correctly (Grünke and Weber, 2015, p. 178) and a basic knowledge of German orthography. This knowledge is enhanced during the work with IDeRBlog due to the fact that the text production is based upon an elementary understanding of the German orthography (Grünke and Weber, 2015, p. 177).

Research design
The development of the prototype (Holzinger et al., 2005) is based on four steps according to Alavi (1984) and Larson (1986): identifying basic requirements, development of a working prototype, implementation and usage (field study), and revision. Concerning the “intelligent dictionary” in a first step the orthographic areas, categories, and phenomena were constructed in a way to fulfill the scientific, practical, and technical requirements (Edtstadler, 2015) as well as other aspects of a qualitative analysis of orthographic mistakes (Edtstadler, 2015). In a next step, the words were selected and the orthographic mistakes were assigned to the system and the “intelligent dictionary” was implemented to the parallel programmed web-based platform. We collected the requirements and implemented a prototype (Ebner et al., 2016) with pupils as co-designer (Johnson et al., 2014) and focus on LA. The evaluation was conducted with our partner schools (Ebner et al., 2017).

Pedagogical concept of IDeRBlog
Writing, as well as reading and calculating, is one of the key competences of a society. All three of these competences build the foundation of a successful participation in everyday school life and in the later work and social life.
The competence of writing contains several fields of competences (Bremerich-Vos, 2011), such as the ability of writing (e.g. handwriting or using the computer), spelling (e.g. considering the orthography), and composing texts (e.g. planning, structuring and revising texts).

These aspects are taken into account within the IDeRBlog project and brought together by using digital technologies and are enhanced by reading as an additional area of competence. Furthermore, several aspects of the fourfold model for teaching German in primary school (Brinkmann, 2015) are considered by using the platform. In the writing area of IDeRBlog students apply their existing knowledge of writing while composing their texts. Furthermore, working with the “intelligent dictionary” contains a systematic approach to German orthography and enables students to practice their orthographic skills with the goal of writing correctly. Reading skills are improved by perusal of blog entries written by fellow students.

A necessary requirement for using the IDeRBlog-platform is according to Ebner et al. (2017) “that children have acquired the alphabetic principle of German orthography. This means that the children should apply at least the basic correspondences between phonemes and graphemes. The deeper understanding of other strategies, e.g. the morphological strategy, is supported by the intelligent dictionary.”

Due to the fact that IDeRBlog allows and encourages the publication of written texts it addresses the motivational aspect of writing. One can presume that the relevance of the written text is elevated by the opportunity to publish it on the blog of the class or school. Consequently, this spurs students into being motivated to write even more texts having a positive impact on the author’s orthography skills as the skill of orthographically correct writing can only be acquired and further improved by writing (Augst and Dehn, 2015). A crucial part of text writing is revising concerning two aspects: One aspect refers to the mode of expression and the correct application of grammar. This review of the text can be done in different ways. Students can check their texts on their own or work with a fellow student in a so-called writing conference in order to improve the style of the text. The second aspect refers to orthography and is strongly linked to text writing as the motivation of “wanting to write more” plays a decisive role in acquiring German orthography, which is supported by the “intelligent dictionary” of the IDeRBlog-platform. This digital tool highlights orthographic mistakes in two different background colors in order to direct the student’s attention toward them. In case the mistake is categorized and thus, a feedback is available, the students apply different strategies in order to spell words correctly with the help of the hint for correcting the mistake provided by the “intelligent dictionary.” These strategies are learned beforehand – either by the teacher or by the training course of the platform – and systematically applied repeated by the “intelligent dictionary.” This way, the strategies are (hopefully) memorized and also applied in different contexts.

In research and practice, a wide range of didactic approaches for teaching German orthography exists (for an overview: Brinkmann, 2015). They range from rather linguistically orientated approaches, taking into account the theory of German orthography with its principles and rules to quite simplified views. The so-called “FRESCH” method according to Renk and Brezing (2015), for example, is based less on orthographic rules but rather on a small set of strategies. In view of this approach, students apply a small number of different strategies in order to deduce the correct spelling of words. IDeRBlog takes both of these approaches into account. The feedback of the “intelligent dictionary” can be configured according to the approach already used in class. Hence, the use of IDeRBlog does not necessarily go along with a break in teaching strategies.

All of the pedagogical approaches are based on the assumption that orthography and the correct spelling of words need to be trained and repeated regularly. IDeRBlog addresses this aspect with an integrated training database. Students are able to access more than
500 exercises in order to improve their orthographic skills without registration and free of charge. The database includes online and print exercises, which focus on the training of German orthography in isolation, allowing a balanced training with analogue and digital media alike. Within the training area, students can choose between various exercises belonging to different orthographic categories. These exercises can be worked on independently in order to improve one’s orthography skills. Especially digital exercises have the benefit of providing students with a direct feedback if they make a spelling error which is not the case with analogue worksheets or textbooks. This blending of different teaching approaches makes IDeRBlog diverse (Steinhauer, 2017).

In addition to “analogue” pedagogical approaches to writing, new approaches influenced by writing within digital media are considered and pursued. These new approaches are consistent with those being depicted and called for by the strategy of the Kultusministerkonferenz (conference of ministers of culture) in their paper “Bildung in der digitalen Welt” (education in a digital world) (Klingenberg, 2017). Furthermore, the normative writing of texts with the help of a keyboard is trained while working with IDeRBlog. As a result, the implementation and application of text processing programs and related digital writing situations is initiated.

**Platform concept**

The diagram shown in Figure 1 illustrates the workflow on the IDeRBlog website with the student’s area being the core of the platform. Most of the other functions evolve around it, emphasising on the student and his or her learning process in the central position.

**Figure 1.**
Workflow of the IDeRBlog website

Source: Ebner et al. (2017)
Student’s area

After the login students are directed to the student’s area (1). This is the starting point for all activities how students can engage with IDeRBlog. The core activities consist of the writing area, the access to texts already written by the student, the blog and the evaluation of the analysis.

Writing area

In the writing area (2) the student types their text into an editor, as shown in Figure 2. Having finished the text, the student can have their texts spell checked by the “intelligent dictionary” (3). Spelling mistakes are highlighted and feedback for an autonomous correction is offered. The student then corrects their spelling mistakes and if necessary has the text spell checked by the “intelligent dictionary” once more. At any time of the writing process, there is always the possibility of saving the text for later in case the student prefers to continue their work at some other time.

If the student has finished the text, it can be handed in. Consequently, the teacher receives the pupil’s text (4) and can proofread it. If necessary, further alterations and correction can be performed by the teacher. Additionally, the teacher can write a personal feedback to the student. The teacher then decides whether the text needs to be rewritten or if it is well done and suitable for publication on the blog. In this case, suitability is not only based on spelling mistake or grammar, but also on the general content of the text and its appropriateness for publication (5).

Last but not least, the student gets the text back and decides whether the text should finally be published in the blog or not (6).

Evaluation and exercises

Parallel to the writing process the intelligent dictionary analyses and logs all of the spelling mistakes and attempts of corrections made. After having written a certain amount of words and having made a number of spelling errors the student receives an evaluation in terms of orthography skills. Based on the analysis, the student get offered various print and online

Figure 2.
Text writing area
exercises. The teacher receives an evaluation of the orthography skills of the whole class as well as of each individual student, too. The process of analyzing and evaluating will be discussed in more detail later on in this paper.

**Blogging**

The fact that the text is finally published by the student is regarded as a part of media literacy education. The student needs to decide whether the text is suitable for a wider audience or not. Further, the student as the author of the text, chooses to publish or to retain his work.

In this context, it must be stated that each blogging entry only shows the nickname of the user as well as date and time of publication. The real name of the student cannot be inferred from it, not even in a public blog. Yet, within a class – as it is often on the internet – it is rather easy especially for the teacher to identify to which student a blog entry belongs to.

Independent from the configurations made by the teacher beforehand the student’s text is automatically published in a personal blog accessible only by the student. Whether the entries of the students are published within the class, the school or on the internet depends on the teacher’s choice of configurations.

**Commenting and reading**

On the IDeRBlog website, pupils are not only able to write and publish their texts, they also get the opportunity to read and comment on published texts of fellow students, as shown in Figure 3. These comments can only be seen by those people who are also able to see the

---

**Figure 3.** Area for Blog entries and comments
blog itself. For instance, a class blog cannot be accessed and commented on by the public or someone belonging to another class or school.

In general, it is the teacher’s choice to allow comments or not. Furthermore, the teacher also decides whether comments can be published directly or whether the teacher has to check these beforehand. Allowing comments to be made without censorship at the teacher, creates a lot of opportunities to improve the media literacy of students. If some inappropriate comments are written, these can be discussed in class. Guidelines for an appropriate behavior and use of language on the internet can be developed starting with a minor mistake of a student having, for example, written an offensive comment. Furthermore, differences between analog and digital communication can be discussed and elaborated. In a worst-case scenario, inappropriate comments can also be deleted by teachers.

Training database
In addition to writing, blogging, and reading pupils can improve their orthography skills by making use of the exercise database on the IDeRBlog website. Consisting of around 500 print and online exercises, the database can be accessed without registration. The database consists of worksheets and various online exercises created by the IDeRBlog project team. The other exercises derive from extensive research on the internet and are integrated as links to the website. These collected exercises were checked at least twice by the project team according to established criteria with regard to the exercises being correct, accessible and suitable for the target group of younger students.

The exact same pool of exercises can be retrieved via the “Erwachsenenseite” (a section of IDeRBlog created for teachers and parents) or via the “Kinderwelt” (starting page for students). Concerning the training database, these two only differ in the order and display of the individual exercises.

Accessing the exercises via the “Kinderwelt” the structure is based on the “Freiburger Rechtschreibschule (FRESCH).” FRESCH is an approach to improve orthographic skills by relying less on formal rules and more on spelling error prevention as well as avoidance. In order to put this concept into action, different strategies of deducting the correct spelling are learned, memorized, and applied (Renk and Brezing, 2015).

Students can decide whether they prefer to improve their orthography skills with the help of online or print exercises. In a next step, they chose between six different FRESCH categories. These derive from the different FRESCH strategies and encompass:

1. “Ableiten” (students deduct for example that words are written with “ä” in the plural form due to the fact that they derive from a word written with a in the singular form).
2. “Groß und Klein” (a set of questions is applied in order to decide whether a word begins with a capital letter or not).
3. “Merkwörter” (words that have to be learnt by heart).
4. “Rhythmisches Verlängern” (students form the plural form of words in order to decide on the final consonant being an d or t, g or k, p or b which are often homophones in the singular form of the word).
5. “Schwingen und Schreiben” (students divide words into single syllables in order to hear all of the word’s letters and write it correctly).
6. “Zusammen/Getrennt” (students decide whether two or more words build a compound or not).

Then the exercises involving the chosen FRESCH strategy are displayed in Figure 4. The title of the exercise as well as the FRESCH category and a screenshot are provided likewise. A click on the link below the exercise causes a new window to open and thereby
leads the student directly to the exercise. Having finished the exercise the student closes the window and automatically finds himself or herself back at the database of exercises and can choose his or her next tasks.

In contrast to that, exercises in the “Erwachsenenwelt” are structured based on the theory of German orthography. First, the teacher chooses between the morphological, phonological, syntactical or lexical level. Then the exercises belonging to the chosen category are displayed. Information about each exercise is given in the title and is enriched by a screenshot. Furthermore, teachers and parents receive additional information about each exercise before being directed to the exercise itself. Short profiles of the exercises are presented containing information about the author and topic of the exercise. In addition, the amount of words as well as miscellaneous information about the task is given. Consequently, the teacher can easily find an exercise suitable for the individual needs and skills of the student. Due to the numbering of the exercises the teacher is able to tell the student which exercise to search for and complete.

Both areas of the website, the “Kinderwelt” as well as the “Erwachsenenseite”, provide courses on various orthographic phenomena. These courses can be completed online and consist of two explanations of the orthographic phenomenon, online as well as print exercises and links to additional exercises of the chosen linguistic or FRESCH category.

In the context of the training database the IDeRBlog website also offers the possibility of searching for exercises in the print as well as in the online section of the database. The applicable search parameters are listed below the search field and are in accordance with the children’s as well as the teacher’s version of the database.

**Supporting teacher and parents**

IDeRBlog aims at a wide-ranging support for teachers implementing IDeRBlog in their school and lessons. These can be found on the “Erwachsenenseite.” Besides the materials offered within the frame of the training database, several tutorials are provided.
These are short videos dealing with different IDeRBlog topics and explaining various functions of the website.

Furthermore, manuals depicting almost all of the website’s functions were drawn up. One version was specifically written for students aged 8 and above while a second more detailed one was written having teachers and parents using IDeRBlog in mind. Dates for differentiated workshops and online seminars introducing IDeRBlog are regularly published on the website.

IDeRBlog also offers information and digital presentations with regard to independently organized workshops or parents-teacher conference. These materials contain first of all general information about IDeRBlog. In addition, different versions of the presentations were created containing references to curricula in Germany, Austria, and the German-speaking community of Belgium. In this context, a consent form for parents allowing their child to work with IDeRBlog in school is provided, although, this mainly concerns the work with a public blog.

The IDeRBlog website itself as well as all of the provided materials are licensed under a Creative Commons License (CC-BY) and serve as an open educational resource (Schön and Ebner, 2017). Hence, these materials can easily be adapted and distributed according to individual needs.

LA approaches

IDeRBlog pursues an approach to data mining and LA as defined by Siemens and Long (2011). The LA method chosen for the IDeRBlog website is a content analysis. It analyses the text typed in by students with a focus on spelling errors. In doing so a vast amount of data is collected involving, for example, the duration of the text production as well as the number of attempts of correcting the text.

Within the pursued LA approach the student is provided with the collected data involving their learning process in order to review and to improve their performance. Underlying this procedure is a circular workflow shown in Figure 5.

With each text, a student types into IDeRBlog the circular workflow repeats itself. In the following, the individual stages of the circular flow are depicted and related to the work performed by the student on the platform.

*Input collection of data*

Having logged into the writing area of the website by typing in their nickname and password, students write their text (see Figure 6). The writing of the text is independent of...
time and place. The website is permanently available. Students cannot only check their text for spelling errors, they can also safe the text they are working on for later. The orthographic analysis of the written text starts with a click on “Jetzt Text kontrollieren” (German for: “check your text now”).

Analysis of the data and direct feedback
All of the spelling errors made by the student are displayed. There are two dictionaries underlying the employed spell checker system (see Figure 7). Words marked yellow derive from an open dictionary. Those words marked red are part of the intelligent dictionary and prioritized over the open dictionary for feedback.

In this context, a direct feedback which allows students to correct their spelling errors autonomously is prioritized. In this phase, students analyze and reflect on their writing skills and spelling errors made. They gain routine and a feeling for the German language by reflecting on and remembering the system of orthographic rules. Hence, students improve their orthographic skills.

Boundaries of the employed dictionaries
In the process of spell checking, the entire spelling mistake cannot be detected. The dictionaries are not able to consider the semantic context in which a word is written. Consequently, words that are clearly misspelled in the given context yet would be correctly spelled in a different context cannot be highlighted as spelling errors. As shown in the screenshot above, one can gather from the semantic context that the student intended to write “heute” meaning today whereas he or she spelled it as “häute” meaning skins. The spelling mistake clearly derives from the fact that those words are homophones in German.

The words that are highlighted with a yellow background color solely tell the student that this is a potential spelling error.

In contrast to that, words highlighted with a red background color give the student a detailed feedback in form of a differentiated advice depending on the linguistic phenomena the spelling error belongs to, as shown in Figure 8. This is due to the fact that the spelling errors highlighted in red are part of the “intelligent dictionary” programmed by IDeRBlog. The detailed feedback ought to support the students to write the words orthographically correct.
There are two different feedback systems for the students highlighting the spelling mistakes in red. On the one hand, there is the feedback according to the orthography strategies of FRESCH which is depicted in Figure 8. On the other hand, there is a more general feedback solely based on the general rules of German orthography instead of specific strategies (see Figure 8). Which of the two options for the feedback is eventually applied can be decided by the teacher beforehand in the class’ configurations menu.

Independent from the feedback chosen by the teacher all of the misspelled words are recorded and matched to one of the categories of German spelling errors. These categories form the basis for the analysis and evaluation of the student’s texts. The evaluation is only statistically significant after a certain amount of words (500) has been written and a certain amount of spelling errors (60) was made. Hence, the evaluation can only be displayed to the teacher and the student, if these criteria are fulfilled.

Interpretation of data
Having reached the abovementioned minimum number of words and spelling errors students receive an evaluation of their orthographic skills. The five categories in which most
of the spelling mistakes were made are displayed in a pie chart, as shown in Figure 9. Different categories are represented by different colors allowing the student to grasp the main areas of spelling mistakes intuitively. Teachers are provided with the same evaluation for each student. Additionally, teachers receive an evaluation of the orthographic skills of the whole class which is displayed in form of a similar pie chart.

These two evaluations allow the teacher to focus on the main orthographic difficulties of their class at any given time throughout the lesson. When teaching the whole class teachers can base their methods on their knowledge of the difficulties in orthography of the entire class. Working individually during a lesson each student is able to work on his or her struggles with orthography. On basis of the evaluation, remedial teaching can be planned and executed effectively, for example, by constructing groups of students experiencing the same difficulties in the same areas of German orthography.

**Reaction training materials**

Below the depiction of the evaluation links to individual training materials on the basis of the analysis is provided. These materials consist of online and print exercises alike. Besides, online courses on individual phenomena of German orthography can be accessed by the student offering a rather intensive training including explanations and various exercises. This is the last stage of the circular flow encompassing the entry of the text, the text analysis and correction, the evaluation and the follow-up exercises. In conclusion, a repeated looping of the circular flow consequently leads to a continuous increase in the orthographic performance of the student.

**Conclusion**

IDeRBlog offers students a unique blending of blogging and various opportunities to improve their orthographic skills. While the writing and working with the “intelligent dictionary” can be considered as the core function of the website it is enriched by a vast amount of print and online exercises as well as courses on various orthographic phenomena. From the perspective of LA IDeRBlog offers an exclusive insight into the acquisition and improvement of orthographic skills. Figure 10 shows, the holistic approach and the integration of LA in the whole process. Individual evaluations are provided to each text as well as recommendations for exercises. In comparison to studies on orthography, IDeRBlog benefits from a large number of users and hence it has the advantage of exceptionally huge sample of student’s texts and spelling mistakes. Analyzing and evaluating the data are assumed to uncover new aspects of the learning process.
References


Neuhold, B. (2013), Learning Analytics-Mathematik Lernen neu gedacht, BoD – Books on Demand, Norderstedt.

Further reading


Corresponding author

Nina Steinhauer can be contacted at: nina.steinhauer@web.de
On predicting academic performance with process mining in learning analytics

Rahila Umer
Institute of Natural and Mathematical Sciences, College of Sciences, Massey University, Auckland, New Zealand
Teo Susnjak and Anuradha Mathrani
Massey University, Auckland, New Zealand, and
Suriadi Suriadi
Faculty of Science and Engineering, Queensland University of Technology, Brisbane, Australia

Abstract

Purpose – The purpose of this paper is to propose a process mining approach to help in making early predictions to improve students’ learning experience in massive open online courses (MOOCs). It investigates the impact of various machine learning techniques in combination with process mining features to measure effectiveness of these techniques.

Design/methodology/approach – Student’s data (e.g. assessment grades, demographic information) and weekly interaction data based on event logs (e.g. video lecture interaction, solution submission time, time spent weekly) have guided this design. This study evaluates four machine learning classification techniques used in the literature (logistic regression (LR), Naïve Bayes (NB), random forest (RF) and K-nearest neighbor) to monitor weekly progression of students’ performance and to predict their overall performance outcome. Two data sets – one, with traditional features and second, with features obtained from process conformance testing – have been used.

Findings – The results show that techniques used in the study are able to make predictions on the performance of students. Overall accuracy (F1-score, area under curve) of machine learning techniques can be improved by integrating process mining features with standard features. Specifically, the use of LR and NB classifiers outperforms other techniques in a statistical significant way.

Practical implications – Although MOOCs provide a platform for learning in highly scalable and flexible manner, they are prone to early dropout and low completion rate. This study outlines a data-driven approach to improve students’ learning experience and decrease the dropout rate.

Social implications – Early predictions based on individual’s participation can help educators provide support to students who are struggling in the course.

Originality/value – This study outlines the innovative use of process mining techniques in education data mining to help educators gather data-driven insight on student performances in the enrolled courses.

Keywords Prediction, MOOCs, Machine learning, Learning analytics, Process mining, Education data mining

Paper type Research paper

1. Introduction

Massive open online courses (MOOCs) have become very popular among student communities, since it provides them an opportunity to register for courses offered by prestigious universities around the world. MOOCs provide a learning environment which attracts large number of learners having different goals and motivations. Coursera, edX
and Udacity are the three pioneers of MOOCs platform which are then closely followed by several around the world like Miriada and Spanish MOOC in Spain, Khan Academy in North America, Iversity in Germany, FutureLearn in England, Open2Study in Australia, Fun in France, Veduca in Brazil, Schoo in Japan and xuetangX in China.

MOOC environment has revolutionized education by centralizing global resources and restructuring the learning environment to bring it closer to students reach. Unlike traditional higher education learning environment, MOOCs provide an open access to courses to anyone with access to the internet. It provides free and open access to high-quality advanced courses comprising of video lectures, reading materials, quizzes, problem sets and forums for productive discussions to foster learning process and develop learning communities. The use of technology in teaching like forum, blogs (Ebner et al., 2010), wiki or educational software has improved the learning process. The increased availability of recorded data from such environments has provided an opportunity to closely investigate student learning behaviors and work toward improving their learning process.

Although there are massive enrollments in courses offered via MOOCs, the completion rate and retaining of persistent students are rather low, often less than 20 percent (Kizilcec et al., 2013). One of the criticisms in MOOCs is the low retention rate of the students which is heavily criticized. Therefore, predicting the likelihood of dropout is necessary, so that steps can be taken to retrain students by encouraging them in their learning activities.

Learning analytics (LA) has recently emerged as a new research lens that focuses on computational techniques to inform on students’ practices. Every online interaction by students like click, page visited or video viewed while pursuing the course is recorded in the log history (Clow, 2013a). How can we get insights from the log history data so as to make pedagogical interventions to support student learning during the course? Campbell et al. (2007) identified five steps, namely, capture, report, predict, act and refine, as the central theme in LA. Once we have captured data that report students’ interactions with the course, analysts works toward making predictions for pedagogical intervention, which is gradually refined.

Several works (Marquez-Vera et al., 2013; Ye and Biswas, 2014; Bayer et al., 2012; Manhães et al., 2014; Martinho et al., 2013; Simon et al., 2006; Watson et al., 2013) have suggested EDM techniques as the way forward to help in predictions of academic failure among students. In this study, the focus is on predicting students’ performance through the traces they leave while pursuing a course. The aim is to apply data mining/machine learning algorithms to students’ data, as students are progressing through a course, in order to predict which students are at risk of not satisfying course requirements, or are rather likely to fail. The identification of such students would then enable educators to carry out various forms of early intervention or provide additional and more tailored support as mitigation measures.

The study is guided by the following research questions:

**RQ1.** Which machine learning algorithms are effective at predicting students, who are at risk of failure on MOOCs data set?

**RQ2.** Is the integration of process mining features able to increase the effectiveness of the machine learning algorithms for the MOOCs problem?

The significance of our study is the integration of process mining to extend the features. Extended features are obtained as a result of process conformance testing. EDM techniques are then evaluated on two kinds of data sets, one with process mining features and other without. We used some of the widely used (Wu et al., 2008) classifiers, namely, random forest (RF) (Breiman, 2001), logistic regression (LR), Naïve Bayes (Cortes and Vapnik, 1995) and K-nearest neighbor (KNN) (Hechenbichler and Schliep, 2004), to answer the posed questions.
This section has laid the foundation of our research inquiry. The remainder of the paper is organized as follows. In Section 2, we present some of the related work in similar field. Section 3 discusses data sources and the limitations of data set used in the study. Section 4 describes the methods applied in our experiments. In Section 5, we present answers to the research questions and discuss experimental results. Finally, in Section 6, we make conclusion.

2. Related work

Several studies have reported and provided promising results in prediction of students who are likely to fail in a given course. In most of these studies, the data used for prediction consist of non-academic information; all of which require extra effort to collect. Our study is the first of its kind which has used process mining to enhance existing features identified in the literature.

Khobragade (2015) proposed an approach where they have predicted the students’ academic failure using decision tree, Naive Bayes and using classifiers that are based on induction rule and decision tree. Data used for classifications involved social, academic and background information of the student. These data have been collected through surveys. A total of 11 features were used for prediction after applying the feature selection algorithm. Classifiers have then been evaluated based on accuracy. Naive Bayes provided the best accuracy of above 87 percent. However, data were gathered through surveys which are time consuming and also involved methods that make overall prediction and do not consider early prediction.

Marquez-Vera et al. (2013) evaluated white-box classifiers. They used induction rules and decision tree for predicting academic failures for students in middle or secondary school. Detailed information of students’ social background and academic information were used. Again the data collection process used here was very extensive and time consuming, since non-academic information was collected through surveys. The impact of different data pre-processing approaches was also analyzed on classification accuracy. The proposed methods show promising results for making prediction of overall academic performance of students.

Costa et al. (2017) presented a comparative study of EDM techniques to predict those students who are likely to fail in a programming course. The significance of this study is that these techniques used could predict the students’ performance at early stages so that some intervention strategy could be made to help students. This study also analyzed the impact of data pre-processing methods and algorithms fine-tuning tasks on prediction results. The study showed that support vector machine outperformed other techniques in a statistical significant way, and data pre-processing and algorithm fine-tuning tasks can improve accuracy.

The work proposed by Ahmad et al. (2015) presents an approach where EDM techniques are used to predict academic performance of first-year students in a computer science course. EDM techniques used are decision tree, Naive Bayes and rule-based classification. The data used during the course of study include demographic data, previous academic records and other family-related information. Rule-based classifiers outperformed other methods and provided prediction accuracy of 71 percent.

Yukselturk et al. (2014) predicted students’ dropout in an online course using EDM classifiers: Naive Bayes, decision tree, KNN and neural network. The data used for prediction consisted of demographic information, online technology self-efficacy scale, readiness for online learning questionnaire, locus of control scale and prior knowledge questionnaire. A total of ten features were used for predicting class label (dropout/not). The maximum prediction accuracy was obtained by KNN (87 percent). Data were collected through surveys and also did not provide prediction at early stages of course.
Another research, conducted by Boongoen (2017) in a Thai university, used a link-based cluster ensemble method as a data transformation framework for prediction. The research has compared several state-of-the-art dimensionality reduction techniques.

Ye and Biswas (2014) used and extended standard features for MOOC analysis with higher granularity to make more accurate predictions for dropout and performance. Analysis was made using data collected from video lectures, weekly quizzes and peer assessments from the ten-week course. Standard features were extended using some detailed temporal features like when some assessment was started during the week, or when the first lecture was viewed.

The findings compared with existing studies showed that these features improved the prediction accuracy. The time when a student starts the peer assessment assignment was found to be a good predictor. Once the peer assessment score was available, the prediction performance improved. Analysis shows that the students who watched video and did not take quizzes were the ones who mostly dropped out. Overall results show that more precise temporal features and more quantitative information improved early prediction accuracies and false alarm rates as compared to using only assessment score features.

Bydzovska (2016) proposed an approach to predict the students’ performance using course characteristics and previous grades. Two different approaches were used. In the first approach, classification and regression were used to predict performance using academic-related data and data about student’s social behavior. The findings were significant with the small number of students. In the second approach, collaborative filtering techniques were used to predict the student’s performance based on similarity of achievements. Classification algorithms, namely, support vector machine, decision tree, part, IBL, RF, Naive Bayes and rule-based classifier, were used, where support vector machine produced best predictions which were further improved by integrating social behavior data.

Cambruzzi et al. (2015) used learning analytics to predict dropout rates in distance education. They developed a system that predicts dropout and supports to integrate pedagogical interventions and textual analysis to reverse identified dropout tendencies. The system was able to predict dropout with 87 percent precision, later the dropout was decreased by 11 percent by implementing specific pedagogical actions.

3. Data sources
In this study, we analyzed data obtained from Coursera for course “Principles of Economics” offered in Summer 2014. The data set consisted of assessments grades, solution submission time, video lecture interaction log, participant’s demographic information, time spent weekly and final grades. The course was designed as an eight-week introduction to the study of economics. The total number of students was more than 3,000; however, we only included data of students who were registered at the time the course started (i.e. on June 24, 2014) and whose final score was not missing. We extracted data of total 167 students, out of which 40 students passed the course while rest had failed. Students with scores greater than 0.5 were considered passed. The final data set obtained was thus imbalanced in regard to the final grade distribution (Figure 1).

3.1 Data set 1 – standard features
The data set comprises of features like demographics, assessment grades, time spent on activities, video watch activities, etc. This has been characterized as “standard” features shown in (Table I).

3.2 Data set 2 – process mining features
A data set has been generated next using logs of weekly activities during the course. It includes features that reflect the differences in the behavior of students with respect to the
behavior of top performing students in the course. These measures have been obtained as a result of process conformance testing (Van der Aalst et al., 2012).

In the process conformance testing, given a normative model $M$ and an event log $L$, difference between the process behavior and $L$ can be explained. Conformance checking was performed using the model representing top student’s weekly activities and log of other student’s weekly activities. The log was replayed using the model to establish a precise relationship between event and model elements and to analyze the deviation of students from modeled behavior. Output of conformance testing is a fitness score that is assigned to each student (case). The fitness scores, obtained based on weekly logs of top performing students and other students, were used as features and integrated with standard features.

The following steps have been performed to prepare data set with process mining features.

3.2.1 Step 1. Using the inductive miner method in ProM (Van der Aalst et al., 2009), we generated a process model using activity logs of top performing students having grade more than 90 percent. The result of this step is a process model shown in Figure 2.

3.2.2 Step 2. By using “Replay a log on Petri net for performance/conformance analysis” method in Prom, the log model alignment was generated, as shown in Figure 3. Inputs to this method were process model of top performing students and log of activities of other students (Figure 4).

3.2.3 Step 3. The log model alignment generated and exported it in the comma separated value format. We extracted fitness scores for each student and integrated them with the standard features in data set 1. This helped characterize the fitness scores as process mining features in data set 2. Same process was repeated and data sets were created using weekly logs.
3.3 Limitations of data sets

MOOCs environments are different from traditional learning setups, which makes it challenging to analyze the data. Large amount of missing data, multiple number of attempts for assignment submission, multiple time registration and higher rate of dropout are some of the major challenges faced during the analysis of MOOCs data.
In this study, a total of 167 students’ data have been used, of which the majority of students belonged to one class and also final data set was imbalanced. It has been found that in a typical MOOCs setting, students are active in the early weeks and become inactive or withdraw from the course in later weeks which result in a large number of missing values.

Figure 5 shows the average number of quizzes attempted by students during each week. Students attempted most of the quizzes in week-1 only. Figure 6 shows the average number of lectures watched during the course. It is evident that last three weeks were the most inactive weeks. The majority of the students either did not watch lectures or withdrew from the course. These indicate unequal patterns of participation. Figure 7 shows average time spent during the course. The graph shows week-7 to be the least active week. However, online engagement traces do not necessarily reflect all activities related to learning process. Due to this reason, measurement of success and participation in MOOC environment must be reconsidered (Clow, 2013b; Bergner et al., 2015).

This study recognizes these limitations in the data sets; however, this subset (of 167 students) is representative of student participation and completion from a real-world MOOC environment. The data sets have been used to mainly demonstrate the effectiveness of data mining techniques using process mining features.

---

**Figure 5.** Average number of quizzes attempted (failed or passed) by students during the course.

**Figure 6.** Average number of lectures watched by students during the course.

**Figure 7.** Average number of time (seconds) spent during the course.
4. Experimental design

The aim of this study is to compare the effectiveness of existing popular machine learning algorithms for early identification of students, who are likely to fail and to investigate the effect of process mining features in the performance of the techniques.

4.1 Classification methods

We used classification methods that have been utilized in the field of education domain and are suitable for imbalanced data set. Machine learning algorithms used in the experiment are as follows: Naive Bayes, RF, LR and KNN. In the following subsections, the classification methods used are briefly explained. Table II shows the parameters used for classifiers.

4.1.1 LR. LR is a parametric method, used in classifications wherein a sigmoid function is estimated based on the training data. This method is based upon the assumption that the probability of event occurring follows a logistic distribution. The distribution is defined as follows:

$$P(\text{outcome} = \text{Pass}/X) = \frac{1}{1 + e^{-X^T\beta}}$$

where $X^T\beta = \beta_0 + \beta_1x_1 + \beta_2x_2 + \cdots + \beta_nx_n$ and $X$ is a set of measurements, $X = [x_1, x_2, \ldots, x_n]$.

Using this function, input space is partitioned into two regions. New instances are classified to the region they belong. The distribution is in the shape of an “S,” which indicates that difference at the extreme ends will not effect much, as compared to the difference around center. LR is bounded by 0 and 1 to represent the probabilities. The upper and lower portions of “S” represent high probabilities and low probabilities of the same even, respectively.

This approach is widely used in the literature to predict the retention with high accuracy (Lin and Reid, 2009; Mertes and Hoover, 2014; Veenstra et al., 2009; Dunn and Mulvenon, 2009).

4.1.2 Naive Bayes. Naive Bayes is a simple supervised method, which is a special form of discriminant analysis. It is based on the Bayes theorem (Benbassat, 1990), returns probability of prediction using the evidence derived from observed data. This method relies on two assumptions: all attributes are conditionally independent and contribute equally to the final outcome of the class and that there are no hidden attributes that can affect the process of prediction. The Naive Bayes classifier assigns to each instance the class value with the highest conditional probability. This method is extensively used since 1950s and is a very popular method especially in the domain of text mining. Naive Bayes performs surprisingly well in cases where even the attributes are not independent. This method is used in several studies in the domain of education data mining (Pittman, 2008; Zhang et al., 2010; Khobragade, 2015; Nandeshwar et al., 2011; Mashiloane and Mchunu, 2013; Sharma and Mavani, 2011; Costa et al., 2017; Ahmad et al., 2015).

4.1.3 RF. RF uses a standard machine learning technique called a “decision tree.” Decision trees build a classification model by a recursive binary partition of a labeled data

<table>
<thead>
<tr>
<th>Classifier</th>
<th>Training setting</th>
<th>Implementation source</th>
</tr>
</thead>
<tbody>
<tr>
<td>KNN</td>
<td>$K = 3$</td>
<td>Scikit-learn (Pedregosa et al., 2011)</td>
</tr>
<tr>
<td>Random forest</td>
<td>Estimator = 10</td>
<td>Scikit-learn (Pedregosa et al., 2011)</td>
</tr>
<tr>
<td>Naive Bayes</td>
<td>Gaussian default setting</td>
<td>Scikit-learn (Pedregosa et al., 2011)</td>
</tr>
<tr>
<td>Logistic regression</td>
<td>default settings</td>
<td>Scikit-learn (Pedregosa et al., 2011)</td>
</tr>
</tbody>
</table>

Table II. Machine learning parameters
set into increasingly homogeneous nodes. Homogeneity is measured by the Gini index, which is defined as:

\[ G = \sum k P(k)(1 - \beta(k)) \]

where \( P(k) \) is the proportion of observations in the \( k \)th class.

At each step, an optimization is carried out to select, in each node, the feature and the numeric threshold or group of values if the variable is categorical that would produce the lowest \( G \) value if used to divide the node. This process continues until it is not possible to reduce the Gini index in any node. The final output is a classification tree with completely homogeneous lower nodes. However, this is not always the case, and the predominant class is used to label the node, the other cases being classification errors. On the basis of these errors, the tree is pruned to allow a higher generalization capacity. Small modification in a data set affects the results of classification in a case of single tree. However, this limitation could be overcome using ensemble learning techniques to obtain a better performance. Bootstrapped sample of the available instances is used to generate unpruned trees in a large number (500-2,000). In order to add randomness and to decrease correlation, each node division is carried out with a randomized subset of the predictors. In ensemble techniques, correlation is not a desirable property because the different results make sense to the voting system. New instance is classified to the class it belongs, based on the aggregate number of votes given by multiple trees. This method is widely used in the prediction of student’s performance (Bydžovská, 2016; Mashiloane and Mchunu, 2013; Marquez-Vera et al., 2013).

4.1.4 KNN. KNN (Hechenbichler and Schliep, 2004) is a classification method that estimates the class for every new instance using the \( k \)-closest instances, by calculating a distance metric, from the training set. Class probabilities for the new instance are estimated as the proportion of training set neighbors in each class. Ties are broken randomly or by including the \( k + 1 \) closest neighbor in the calculation. \( K \) is the number of neighbors, an important parameter to be considered when using this method. A small value leads to an increase in the probability of over-fitting, while too large a value causes a high-bias classification. This simple algorithm has been successful in a large number of classification problems (Gray et al., 2014; Minaei-Bidgoli et al., 2003; Mayilvaganan and Kalpanadevi, 2014; Yukselturk et al., 2014).

4.2 Evaluation measures
In order to compare the performance of each classifier, F1-score and area under curve (AUC) were used. Due to the imbalanced nature of data set, overall accuracy might be misleading.

4.3 F1-score
F1-score is widely used in binary classification problems. F1 score is the harmonic mean between Precision and Recall:

\[
\text{Precision} = \frac{TP}{TP + FP}
\]

\[
\text{Recall} = \frac{TP}{TP + FN}
\]

where TP is the number of positive instances correctly classified as positive; FP is the number of negative instances incorrectly classified as positive; and FN is the number of
positive instances incorrectly classified as negative:

\[ F1 - score = \frac{2 \times \text{precision} \times \text{recall}}{\text{precision} + \text{recall}} \]

### 4.4 AUC

A receiver operating characteristic curve is a way to compare diagnostic tests. It is a plot of the true positive rate against the false positive rate. The AUC is a number between 0 and 1:

\[ \text{False positive rate} = \frac{FP}{(FP+TN)} \]

\[ \text{True positive rate} = \frac{TP}{(TP+FN)} \]

### 4.5 Training procedure

To estimate the generalization capability of the model to future data set, ten-fold cross-validation technique was used. This technique splits the original data set into ten subsets of equal size, preserving the original ratio of minority and majority class instances. One subset is left for the validation and rest are used for training the model. This process is repeated ten times using different subsets for training and validation each time. In the end, average results across each iteration are computed. Performance of classification methods described in Section 4.1 was evaluated next. These methods are used for the prediction of student’s final outcome of the course as Pass or Fail on two data sets that are discussed in Section 3. Prediction was based on the learners’ demographics and dynamic data of the previous week. Each data set was divided on weekly basis. After each week, prediction was made based on the available data of current and previous weeks. In definition, week-3 data set means that it consists of all available data till week-3 which includes week-2 and week-1 data as well. We assume that prediction accuracy improves as more data become available in upcoming weeks. For instance, prediction after week-4 means that we used all available data till week-4, which might include scores of assessments and quizzes, which are part of final score and ultimately improve accuracy. Prediction accuracy at early stages is important so that timely interventions can be made to help students.

### 5. Experimental results

This section first presents the experimental results which are discussed next. Table III displays the result of data set 1 which includes features like demographic and grading scores.

<table>
<thead>
<tr>
<th>Data set</th>
<th>LR</th>
<th>RF</th>
<th>NB</th>
<th>KNN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week-1</td>
<td>0.77</td>
<td>0.712</td>
<td>0.788</td>
<td>0.724</td>
</tr>
<tr>
<td>Week-2</td>
<td>0.879</td>
<td>0.866</td>
<td>0.89</td>
<td>0.848</td>
</tr>
<tr>
<td>Week-3</td>
<td>0.794</td>
<td>0.803</td>
<td>0.618</td>
<td>0.678</td>
</tr>
<tr>
<td>Week-4</td>
<td>0.8</td>
<td>0.799</td>
<td>0.715</td>
<td>0.722</td>
</tr>
<tr>
<td>Week-5</td>
<td>0.836</td>
<td>0.858</td>
<td>0.764</td>
<td>0.75</td>
</tr>
<tr>
<td>Week-6</td>
<td>0.836</td>
<td>0.84</td>
<td>0.743</td>
<td>0.747</td>
</tr>
<tr>
<td>Week-7</td>
<td>0.85</td>
<td>0.842</td>
<td>0.503</td>
<td>0.75</td>
</tr>
<tr>
<td>Week-8</td>
<td>0.841</td>
<td>0.866</td>
<td>0.536</td>
<td>0.801</td>
</tr>
<tr>
<td>Rank(mean)</td>
<td>1.75</td>
<td>1.875</td>
<td>3.125</td>
<td>3.25</td>
</tr>
</tbody>
</table>

**Table III.** Comparative results of the effectiveness of machine learning algorithms on the data set using standard features and mean ranks of classifiers from highest (1) to lowest (N)
Table IV displays the result of data set 2 which enriches the standard features with process mining features. Next, we answer the research questions in the light of our results of analysis:

**RQ1.** Which machine learning algorithms are effective at predicting students, who are at risk of failure on MOOCs data set?

In order to answer this question, prediction was performed using four machine learning techniques on two data sets. Table III shows the results of effectiveness of machine learning algorithms using data set 1 (using standard features only) to predict students who are likely to fail. The results show that maximum F1-score obtained is 0.78 by Naive Bayes classifier after week-1. For week-2, F1-score improved to 0.89 by Naive Bayes classifier. After week-2, F1-score of all classifiers drops. In MOOC environment, it is normal that students are active in first week and also the assessments are easy to score high compared to the later weeks. After week-4, we observe continuous growth in F1-score for almost all classifiers. Maximum score achieved after week-8 is 0.86 by RF. Different classifiers performed differently for each week data, but overall RF and LR performed better than rest of the techniques. The performance of the models looks promising, till the mid of the course (after week-4) F1-score reaches to 80 percent by LR.

Table IV shows the results of classifiers using process mining features. Using process mining features, F1-score increases for almost all weeks. After week-5 and week-6, F1-score drops but still maximum score is 0.87 by Naive Bayes. The results show that overall all classifiers performed well in prediction; however, the Naive Bayes method outperforms all methods by scoring maximum accuracy of 0.89 after week-8.

In order to measure the significance of above findings, we used the Friedman's test methodology for comparison of multiple classifiers over multiple data sets. The Friedman's test is a non-parametric test used to compare observations repeated on same subjects. Chi-square with k-1 degree of freedom is the test statistic for the Friedman's test, where k is the number of repeated measures. When the p-value is small (p < 0.05), null hypothesis is rejected. The goal of this test is to see that there is any significance difference among the performance of machine learning techniques in our experiment. Null hypothesis of our study is “There is no difference among the performance of multiple classifiers.”

After applying the Friedman's test, p-values obtained are 0.02 and 0.001 for data set 1 and data set 2, respectively. As p-values are less than 0.05, null hypothesis is rejected. We conclude that there is significance difference between the performance of classifiers.

The calculation of mean ranks of classifiers (from highest to lowest) shows that the LR and Naive Bayes scored highest ranks for data set 1 and data set 2, respectively, and thus outperformed other classifiers on these data sets.

<table>
<thead>
<tr>
<th>Table IV. Comparative results of the effectiveness of machine learning algorithms on the data set using process mining features and mean ranks of classifiers from highest (1) to lowest (N)</th>
<th>Data set</th>
<th>LR</th>
<th>RF</th>
<th>NB</th>
<th>K.NN</th>
</tr>
</thead>
<tbody>
<tr>
<td>Week-1</td>
<td>0.831</td>
<td>0.817</td>
<td>0.829</td>
<td>0.816</td>
<td></td>
</tr>
<tr>
<td>Week-2</td>
<td>0.831</td>
<td>0.833</td>
<td>0.861</td>
<td>0.796</td>
<td></td>
</tr>
<tr>
<td>Week-3</td>
<td>0.842</td>
<td>0.852</td>
<td>0.872</td>
<td>0.808</td>
<td></td>
</tr>
<tr>
<td>Week-4</td>
<td>0.87</td>
<td>0.845</td>
<td>0.871</td>
<td>0.825</td>
<td></td>
</tr>
<tr>
<td>Week-5</td>
<td>0.878</td>
<td>0.892</td>
<td>0.878</td>
<td>0.844</td>
<td></td>
</tr>
<tr>
<td>Week-6</td>
<td>0.854</td>
<td>0.889</td>
<td>0.88</td>
<td>0.835</td>
<td></td>
</tr>
<tr>
<td>Week-7</td>
<td>0.865</td>
<td>0.868</td>
<td>0.879</td>
<td>0.828</td>
<td></td>
</tr>
<tr>
<td>Week-8</td>
<td>0.879</td>
<td>0.886</td>
<td>0.89</td>
<td>0.848</td>
<td></td>
</tr>
<tr>
<td>Rank(mean)</td>
<td>2.56</td>
<td>2.0</td>
<td>1.43</td>
<td>4</td>
<td></td>
</tr>
</tbody>
</table>
Figures 8-11 show the performance of classifiers when compared with second metric, i.e. AUC. The results are almost similar like in case of F1-score. All classifiers performed better with process mining features than standard features alone:

**RQ2.** Is the integration of process mining features able to increase the effectiveness of the machine learning algorithm for the MOOCs problem domain?
In order to answer this question, we performed same prediction experiment on data set 2, which consists of features used in data set 1 and additional features obtained in the result of conformance testing. Table IV displays the F1-score obtained after evaluating machine learning algorithms on data set 2. F1-score improves after each week as expected. Figures 8-11 show the comparative results of the effectiveness of machine learning algorithms when applied on data set 1 with standard features and data set 2 which contains process mining features as additional features to the standard features. The results showed that for all weeks, F1-score improved using process mining features except for week-2 for some methods. In order to measure the significance of these results, paired \( t \)-test was applied on the results. Following \( p \)-values were obtained as a result of \( t \)-test: \( p \)-value (LR) = 0.052; \( p \)-value (RF) = 0.02; \( p \)-value (KNN) = 0.007; and \( p \)-value (Naive Bayes) = 0.01. According to Gorunescu (2011), in order to present a significance difference, \( p \)-value should be normally less than 0.05. Based on this discussion, we conclude that all classifiers present a statistically significant improvement in F1-score when process mining features were integrated with standard features, except LR.

5.1 Feature importance
In order to identify the relevant predictor variables, we used variable importance measure produced by the RF classifiers. We trained an RF model with 10,000 trees on data set 2 and rank the ten features by their respective importance measures. We chose data set 2, as it comprises of both standard features and process mining features. We wanted to investigate which features are more informative to the target variable. Table V shows the top ten important features for each week. The results show that features that measure time spent weekly and video watch activities were the most important features for all weeks. Second most important features are related to process mining.

6. Discussion and conclusion
Educational data mining provides an insight from educational data. However, most of the EDM studies used traditional data mining techniques. This work describes a possibility to integrate process mining approaches in order to achieve high prediction accuracy. The use of features obtained from process mining approach for the purpose of prediction of students’ performance is novel.

We took Coursera MOOC as a case study with the focus on predicting student’s performance through the traces they leave while pursuing a course. Data mining/machine learning algorithms were applied to weekly generated student data, as students are progressing through a course, in order to predict which students are at risk of not satisfying course requirements, or are rather likely to fail.

This study conducted a comparative analysis of four techniques (LR, RF, Naive Bayes and KNN). These techniques were evaluated on using two data sets, one with standard features used in the literature and second with features obtained from process conformance testing. The impact of process mining features has been analyzed on the effectiveness of mentioned techniques. The results show that techniques used in the study are able to predict the performance of students at early stage. By integrating process mining features with traditional features, effectiveness of the some techniques has improved. LR and Naive Bayes classifiers outperform other techniques in a statistical significant way for data set 1 and data set 2, respectively. We also measured the importance of features using RF classifier. The results show that process mining features were among top ten important features for all data sets; however, features related to time spent weekly and video watch activities were most important features among all.

The significance of our study is the use of process mining to enrich the features, and the results show that overall performance is statistically significantly improved using
<table>
<thead>
<tr>
<th>Rank</th>
<th>Week-1</th>
<th>Week-2</th>
<th>Week-3</th>
<th>Week-4</th>
<th>Week-5</th>
<th>Week-6</th>
<th>Week-7</th>
<th>Week-8</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>LecLagw1</td>
<td>Vid-Act-w2</td>
<td>Timespentw3</td>
<td>VidActW4</td>
<td>VidActw4</td>
<td>VidActw4</td>
<td>VidActw4</td>
<td>VidActw4</td>
</tr>
<tr>
<td>2</td>
<td>VidActw1</td>
<td>VidActw1</td>
<td>VidActw1</td>
<td>Timespentw3</td>
<td>VidActw5</td>
<td>VidActw5</td>
<td>VidActw5</td>
<td>VidActw5</td>
</tr>
<tr>
<td>3</td>
<td>Timespentw1</td>
<td>LecLagw1</td>
<td>VidActw1</td>
<td>Timespentw4</td>
<td>Timespentw3</td>
<td>Timespentw3</td>
<td>Timespentw3</td>
<td>Timespentw8</td>
</tr>
<tr>
<td>4</td>
<td>Tracefitness</td>
<td>Timespentw1</td>
<td>Quizattempw3</td>
<td>VidActw2</td>
<td>Timespentw4</td>
<td>Timespentw6</td>
<td>Timespentw6</td>
<td>Timespentw3</td>
</tr>
<tr>
<td>5</td>
<td>Movemodelfit</td>
<td>Timespentw2</td>
<td>LecLagw1</td>
<td>VidAct-w1</td>
<td>Queue-state</td>
<td>Timespentw4</td>
<td>Timespentw7</td>
<td>Timespentw6</td>
</tr>
<tr>
<td>6</td>
<td>Movemodelfit</td>
<td>Queuestate</td>
<td>Timespentw2</td>
<td>Tracefitness</td>
<td>VidActw2</td>
<td>Timespentw4</td>
<td>Timespentw4</td>
<td>Timespentw7</td>
</tr>
<tr>
<td>7</td>
<td>Queuestate</td>
<td>Movemodelfit</td>
<td>Queuestate</td>
<td>Movemodelfit</td>
<td>Movemodelfit</td>
<td>Tracefitness</td>
<td>Queuestate</td>
<td>Timespentw4</td>
</tr>
<tr>
<td>8</td>
<td>Rawfitcost</td>
<td>Movemodelfit</td>
<td>Tracefitness</td>
<td>Movemodelfit</td>
<td>Tracefitness</td>
<td>Tracefitness</td>
<td>Queuestate</td>
<td>Timespentw4</td>
</tr>
<tr>
<td>9</td>
<td>Quizattempw1</td>
<td>Tracefitness</td>
<td>Movemodelfit</td>
<td>Queuestate</td>
<td>Movemodelfit</td>
<td>Movemodelfit</td>
<td>Movemodelfit</td>
<td>Movemodelfit</td>
</tr>
<tr>
<td>10</td>
<td>Tracelength</td>
<td>Quizlagw2</td>
<td>Movemodelfit</td>
<td>Timespentw2</td>
<td>VidActw1</td>
<td>Movemodelfit</td>
<td>Movemodelfit</td>
<td>Tracefitness</td>
</tr>
</tbody>
</table>

Table V. Feature importance by random forest classifier for data set 2.

Process mining in learning analytics.
process mining features. The limitation of this study is the missing values and the small size of the data. This study recognizes these limitations in the data sets; however, this subset (of 167 students) is representative of student participation and completion from a real-world MOOC environment. The data sets have been used to mainly demonstrate the effectiveness of data mining techniques using process mining features.

References


Corresponding author
Rahila Umer can be contacted at: r.umer@massey.ac.nz

For instructions on how to order reprints of this article, please visit our website: www.emeraldgrouppublishing.com/licensing/reprints.htm
Or contact us for further details: permissions@emeraldinsight.com
Learning analytics in primary, secondary and higher education
Guest Editors: Markus Ebner and Martin Ebner

Number 2
93 Editorial boards
94 Small data as a conversation starter for learning analytics: exam results dashboard for first-year students in higher education
Tom Bros, Katrien Verbert, Greet Lamana, Caroline Van Boon and Tinne De Loor
107 Evaluating emotion visualizations using AffectVis, an affect-aware dashboard for students
Leony Derick, Gayane Sedrakyan, Pedro J. Munoz-Merino, Carlos Delgado Kloos and Katrien Verbert
126 Entrepreneurship students distilled their learning experience through reflective learning log
Khar Kheng Yeoh
143 Learning analytics to improve writing skills for young children – an holistic approach
Nina Steinhauer, Michael Gros, Martin Ebner, Markus Ebner, Annelies Huppertz, Mike Comann, Susanne Bremikler, Lea Burk, Konstanze Edtstadler, Sonya Gacek, Martin Wachter, Christian Aspalter and Susanne Martich
160 On predicting academic performance with process mining in learning analytics
Rahila Umer, Jee Suoje, Anuradha Mathiran and Suradi Buniadi