A Quantitative Evaluation of LOCO-Analyst: A tool for Raising Educators' Awareness in Online Learning Environments

Abstract—LOCO-Analyst is a Semantic Web technologies-based tool that provides educators with feedback about relevant aspects of the learning processes in online learning environments. The developed types of qualitative and quantitative feedback about students' performance presented in LOCO-Analyst should help educators in adapting their instructional strategy. The feedback is based on the analysis of user traces in online learning environments structured and integrated using the Learning Object Context Ontology (LOCO) framework. So far, two versions of the tool have been developed in 2006 and 2009. The second version differs from the first one in employing more effective visualization techniques for representing the different types of feedback for educators; it also offers some new kinds of feedback that are built on top of Social Web principles. In order to empower educators with most suitable types of feedback, we aimed at evaluating the effectiveness of the LOCO-Analyst's feedback mechanisms and discovering whether the changes in the Graphical User Interface (GUI) of the new version of the tool make feedback more intuitive and easy to understand for the users. Therefore, we have designed an experimental study to evaluate LOCO-Analyst. The results show that the users are highly satisfied with both the feedback set and the changes in the tool that make its GUI more effective than the previous version. Finally, exploratory analysis of the data revealed interesting relations between different variables characterizing the participants' perceived value of the tool's features. The results of the study will be used to improve the LOCO-Analyst tool to make it more effective and helpful for educators. We also believe that these results could help researchers and developers who work in the e-learning domain to provide more effective tools for educators.

1 INTRODUCTION

Today's Learning Content Management Systems (LCMS) provide substantial support for online teaching. Educators of online courses have many well-designed tools available for preparing course content and activities, as well as structuring and organizing their courses according to their preferred instructional approach. Moreover, they can interact with students and coordinate their activities using online communication tools. However, when it comes to personalization of the learning process, the support offered by LCMSs is rather limited [Willging, 2005] [Dawson et al, 2008].

Nowadays, educators need to constantly adapt their on-line courses to assure high performance and learning efficiency of their students [Gasevic et al, 2007]. To be effective, this adaptation requires educators' awareness of the students' activities in an online learning environment, the difficulties they are experiencing, their ability to comprehend and follow the course content and the like. For this reason, educators need comprehensive, but at the same time informative and reliable feedback about their online courses. To be comprehensive, the feedback has to be based on semantically interlinked (integrated) data about all major elements of a learning process: learning activities (e.g. reading and discussing), learning content, learning outcomes and, students [Jovanovic et al, 2007]. To be informative, the feedback should be such that an educator can quickly and easily get an overall insight into a certain aspect of the learning process. However, today's LCMSs offer limited support for such insights, such as superficial feedback including simple statistics about the technology the students have used or low level data about their interaction with learning content (e.g. page view). In addition, traditional course evaluation done at the end of a semester often does not provide reliable feedback, as it has been shown that most students enrolled in online courses do not complete standard course evaluation forms, and that their response rate is far lower than for students attending conventional classroom-based courses.

Having recognized the above stated issues of online education, in our previous research, we have identified the kinds of feedback that would be highly appreciated by online educators and subsequently developed a tool named LOCO-Analyst that is capable of generating the feedback needed by educators [Jovanovic et al, 2008]. The tool provides feedback on diverse levels of content granularity and feedback about different types of learning content (e.g., lessons and tests). Furthermore, LOCO-Analyst provides educators with rich feedback about each individual student - the student's interactions with the learning content as well as interactions with other students (Table 1).

In 2006, we conducted the initial evaluation of the LOCO-Analyst tool [Jovanovic et al, 2007]. The general impression of the study participants was rather positive, but equally important for us were the participants' comments and suggestions for further improvement of the tool. The great majority of them suggested improvements in the representation of feedback (primarily through visualizations); there were also a few suggestions for improving the existing and adding some new kinds of feedback.

Accordingly, we developed a new version of the tool by adding new kinds of feedback and enhancing presentation features (visualization) as outlined in Sections 2.1-2.2. In order to verify whether this new version of the tool would provide users with better user experience and whether it would better address their feedback requirements, in 2009, we organized a new evaluation study. Another goal of this study was to explore new research challenges when it comes to providing educators with feedback about the learning process within an online learning environment.

Whereas in the first (2006) study we did just some basic statistical analysis (descriptive statistics) of the collected data, within the second (2009) study, we conducted a comprehensive statistical analysis aiming to leverage the collected data to answer our research questions. Specifically, we wanted to answer the following questions:

- First, does the feedback offered by the tool provide educators with sufficient awareness of students’ online learning activities and their performance?
- Second, how intuitive is the graphical user interface (GUI) of the tool for educators and whether the changes done in the tool’s GUI make the tool easier to understand for educators?
- Finally, is there any correlation between the users’ characteristics and their perception of different types of feedback provided by the tool, as well as different features that the tool offers?

Generally, the results of the study have shown that the users were satisfied with the feedback offered by the tool. In addition, the changes that were done in the representation of the feedback and the tool’s GUI have made the GUI more effective when compared with the previous version. Therefore, we were able to conclude that the new version of the tool provides more effective and helpful feedback for educators. Furthermore, we performed an exploratory analysis (correlations) on the collected data and identified some interesting relations between the participants’ characteristics and the tool’s features. We believe that these findings could help researchers and developers who work in the e-learning domain to provide more effective tools for educators.

The rest of the paper is organized as follows: the next section gives a brief overview of LOCO-Analyst and explains the features that were introduced in the new version of the tool. Section 3 presents the quantitative evaluation of LOCO-Analyst in detail: first, it introduces and gives rationale for our research questions; subsequently, the research method (i.e. the study design) is explained. Descriptive and inferential results are given in Section 4 as well as their interpretation and discussion. Threats to validity of our experimentation are discussed in Section 5. After presenting the related work (Section 6), we conclude the paper and outline the future work in Section 7.

2 LOCO-Analyst

LOCO-Analyst is an educational tool aimed at providing educators with feedback on the relevant aspects of the learning process taking place in a web-based learning environment, and thus it helps them improve the content and the structure of their web-based courses. It provides educators with feedback regarding:

- the kinds of activities their students performed and/or took part in during the learning process,
- the usage and comprehensibility of the learning content they had prepared and deployed in the LCMS,
- contextualized social interactions among students (i.e., social networking) in a virtual learning environment.

Table 1 shows feedback types provided by tool for the educators.

The generation of feedback in LOCO-Analyst is based on the notion of Learning Object Context, which is about a student (or a group of students) interacting with a learning content by performing a certain activity (e.g., reading, quizzing, chatting) with a particular purpose in mind. The purpose of learning object context is to facilitate abstraction of relevant concepts from user-tracking data of various e-learning systems and tools. Despite differences in the format of the tracking data provided by various LCMSs, there are conceptual commonalities in their content and structure (e.g., history of pages visited, students’ marks on quizzes, and messages posted in online forums). These commonalities are captured in the form of learning object context data, formalized through LOCO (Learning Object Context Ontologies) ontologies [Jovanovic et al, 2007], and used for feedback generation in LOCO-Analyst. By grounding its feedback generation on the formalized learning object context model, LOCO-Analyst achieves its independence from any specific LCMS.

LOCO-Analyst is implemented as an extension of Reload Content Packaging Editor3, an open-source tool for creating courses compliant with the IMS Content Packaging4 specification. By extending this tool with the feedback provisioning functionalities, we have ensured that educators effectively use the same tool for creating e-learning courses, receiving and viewing automatically generated feedback about their use, and modifying the courses, accordingly.

Detailed explanation of the tool’s features, its architecture and semantic technologies it uses for feedback provisioning is given in [Jovanovic et al, 2008]. In what follows, we present the new features that were introduced in the second version of the tool.

2.1 Improved visualization

As previously stated, the majority of improvements in the second version of the tool were related to the tool’s user interface and the way it communicates feedback to its end users (i.e., educators). Since we are not able to present all the improvements in the feedback presentation, due to the space limit, we have chosen to present visualizations

---

2 In the scope of the LOCO-Analyst tool, users are educators who might of different role in the educational process such as instructors, facilitators, mentors, teachers, tutors, teaching assistants, or content authors.

3 http://www.reload.ac.uk/editor.html

4 http://www.imsglobal.org/content/packaging/
that are part of the feedback about learning activities of individual students. The rationale for this decision lies in the fact that for that specific kind of feedback, in the first (2006) evaluation study we received the highest number of comments suggesting better visual representation.

Table 1. LOCO-Analyst’s feedback and type of information they provide.

<table>
<thead>
<tr>
<th>Feedback</th>
<th>Provided Information</th>
</tr>
</thead>
<tbody>
<tr>
<td>F1 Individual lesson</td>
<td>Basic statistical indicators related to the lesson (re-)visits and dwell time; estimated difficulty level of the lesson and how it compares w.r.t. other related lessons; students' messages discussing the topics of the lesson; tags attached to the lesson by students.</td>
</tr>
<tr>
<td>F2 A group of (related) lessons</td>
<td>The same kind of data as for an individual lesson but on a more coarse grained level, i.e., the level of a lessons (topically-related) group.</td>
</tr>
<tr>
<td>F3 A learning module as a whole</td>
<td>The same as for the previous two but on an even higher generality level, referring to a learning module as a whole.</td>
</tr>
<tr>
<td>F4 Students' performance on a quiz</td>
<td>Overview of the quiz (test) results, including, besides the basic statistics also the list of the most difficult question and comparison with students performance on quizzes in other learning modules (both on individual and group level).</td>
</tr>
<tr>
<td>F5 Student's activities in discussion forums and chat rooms</td>
<td>Sent/received messages across different forums and chats rooms; an insight into the social network established through the students' online communications (Fig. 2), e.g., the most frequent communication partners.</td>
</tr>
<tr>
<td>F6 Student's interaction with the learning content (lessons)</td>
<td>Visual representation of the student's engagement with the different lessons of different learning modules; one view is given in Fig. 1.</td>
</tr>
<tr>
<td>F7 Student's comprehension of the studied topics (based on his/her annotations)</td>
<td>Different kinds of annotations the student used to annotate the lessons of the learning module, including tags (presented within the tag cloud of the entire class), notes/comments, and highlights (Fig. 3).</td>
</tr>
</tbody>
</table>

Fig. 1 A screenshot presenting the feedback regarding a student’s interactions with the learning content.

For presenting educators with feedback about a selected student, LOCO-Analyst uses a dialog-box (Fig. 1) implemented through four tab panels, each one presenting a specific kind of information that LOCO-Analyst possesses about the student:

- ‘Forums’ and ‘Chats’ panels inform an educator about the student's online interactions (in discussion forums and chat rooms, respectively) with

5 Learning module is considered here as a relatively independent unit of a course which focuses on one specific subject area taught within the course (often correspond to one chapter of the course's textbook).
other participants in the learning process;
- The ‘Learning’ panel is intended for presenting information regarding the student’s interaction with the learning content;
- The ‘Annotations’ panel provides feedback based on the student’s annotations (notes, comments, tags) and will be explained in the next (2.2) subsection.

Here, we present the ‘Learning’ and ‘Chats’ (Fig. 1 and Fig. 2) panels as they contain the most interesting visualizations.

Fig.2 presents feedback about a student’s interactions with his fellow-students through chat rooms. On the left hand side of the screen, there is a table presenting the student’s activity in chat rooms in terms of exchanged messages. The student from our example participated in only one chat room – the one devoted to asking for help and exchanging tips and hints. Below this table, a list of the student’s frequent chatmates is given. For each chatmate, the list entry states the number of chats the two of them participated in together, as well as the number of chats in which they were the only participants (dubbed as ‘one-to-one chats’). The largest part of the screen is taken by an interactive graph, which visualizes the students’ social network established through the chatting activity. The pink-highlighted node in the centre of the graph represents the student for whom the feedback is presented; the yellow coloured nodes represent all his chatmates; and blue coloured nodes are all other members of the online learning community with whom the given student did not interact online. The graph is interactive, so by making a double-click on any node (i.e., by selecting any student), that node gains the focus (i.e., becomes the central node) and its neighbours (i.e., the chatmates of the newly selected student) are made visible.

2.2 The New Kind of Feedback

Aiming to extend the feedback set offered by LOCO-Analyst, we were investigating how students’ collaborative tags (i.e., folksonomy) can be leveraged for providing educators with enhanced feedback about students’ comprehension of the course content. The result was a new kind of feedback (introduced in the second version of the tool), which makes use of the data about students’ annotations of the course content (including their collaborative tags) to inform educators about their comprehension of that content. The assumption is that notes and tags that students use for annotating the course content reflect their perception (or even comprehension) of the content.

To present an educator with this new kind of feedback, we have extended the dialog-box that LOCO-Analyst uses for displaying feedback about one particular student (Section 2.1) with the “Annotations” panel (Fig.3).

![Fig. 2 A screenshot showing feedback about a learner’s chatting activity](image)

On the left hand side of the panel, there is a tag cloud presenting tags that all students used for annotating the course content. We make a visual distinction between the tags that the selected student used, and those that other students used but the selected one did not. This distinction is visualized by making active and painted in blue only those tags that the chosen student used (i.e., mouse pointer turns into a hand indicating clickable tag), whereas other tags are not clickable and are painted in grey. This allows the educator to identify to what extent the student’s perception of the course content overlaps with that of his/her fellow students. After the instructor selects
one of the student’s tags from the tag cloud, the course content annotated with that tag is presented in a tree-like structure (the upper part on the right). The tree root represents the course, branches are lessons annotated with the selected tag and tree leaves are parts of the lesson’s content annotated with the selected tag. After the instructor selects one annotation (i.e., a tree leaf), the part of the lesson forming its ‘context’ is presented in the Annotation Preview panel and the student’s notes related to that annotation (if available) are listed in the Notes Preview panel (lower right part of the screen).

3 Quantitative Evaluation of LOCO-Analyst

3.1 Research Questions and Hypotheses

In our evaluation study we targeted the following six main research questions:

- **RQ1**: Do the educators perceive the feedback offered by the LOCO-Analyst tool as useful?
- **RQ2**: Have the changes done in the user interface of the tool (cf. Sections 2.1 and 2.2) resulted in a more intuitive and easy to use interface?
- **RQ3**: Does the educator’s role (e.g., instructor, teaching assistant, or research student) significantly influence his/her perception of the LOCO-Analyst’s feedback on learning content?
- **RQ4**: Does the educator’s role (e.g., instructor, teaching assistant, or research student) significantly influence his/her perception of the LOCO-Analyst’s feedback on particular student’s activities in an online learning environment?
- **RQ5**: Does the educator’s experience with related tools (e.g., Reload Content Packaging Editor) significantly influence his/her perception of the LOCO-Analyst’s GUI and the feedback it provides?
- **RQ6**: Are there some additional variables (besides the educators’ role and experience with related tools) that significantly influence the users’ perception of the LOCO-Analyst’s GUI and the feedback it provides?

**RQ1 Rationale**: As mentioned in the introduction section, in order to assist educators with the adaptation of their courses, we aim to provide them with an information-rich feedback about students’ activities and performance. Accordingly, LOCO-Analyst was designed to offer various kinds of feedback related to different elements of the learning process (cf. Section 2). We wanted to explore how informative this feedback is for educators and to what extent it can help them in their teaching practices. This was to be done by evaluating the participants’ answers to the questions that addressed the helpfulness of each kind of feedback offered by the tool, as well as the questions about the usefulness of the tool for improving the course content/instruction. Moreover, through exploration of this research question, we expected to be able to identify some weaknesses of the feedback the tool provides (e.g., not containing all relevant information or being irrelevant) and analyze them so that we can improve the tool. In order to answer RQ1, we performed a descriptive analysis (including, e.g., mean and standard deviation) of responses obtained from the study participants, as will be presented in Section 4.1.

![Figure 3](image-url) A screenshot showing feedback based on students annotation of course content

**RQ2 Rationale**: Based on the feedback that we received in the initial (2006) evaluation of LOCO-Analyst, the tool was changed (cf. Section 2). In a nutshell, we introduced different kinds of visualizations (charts and
graphs) for feedback presentation and added a new kind of feedback based on the students’ annotations of the course content. Having done these changes, we aimed at examining if the new version of the tool is perceived as more intuitive and its feedback is more comprehensible to the users. Specifically, we focused RQ2 on exploring whether the changes in the Graphical User Interface (GUI) of the tool have improved its intuitiveness and decreased the users’ perception of being overburden with information (present in the previous version of the tool). The results were supposed to help us identify the weakness and areas for improvement of the tool’s GUI so that it becomes more convenient for the users. In order to analyze results, we used statistical tests that checked for any significant difference in responses from the 2006 study participants and the 2009 study participants (cf. Section 4.2).

**RQ3-6 Rationale:** These research questions (RQ3-6) are related to our ongoing efforts to improve the existing and introduce new kinds of feedback. These efforts are partially directed towards the provision of feedback customized to the specific requirements of the course and/or user (e.g., the educator’s specific instructional strategy). To achieve that, we need to identify the relations between the different kinds of feedback the tool currently provides and how they are perceived by (different kinds of) potential users. Having identified those relations, we will be able to improve and customize the feedback based on the users’ preferences and/or course subject. Hence, the RQ3-6 research questions are investigated through an exploratory study of the users’ responses to find interesting relations which could be used for further improvement of the feedback the tool offers.

### 3.2 Study Design

#### 3.2.1 Design

As explained in Section 2, two versions of the tool were developed in years 2006 and 2009, respectively. Accordingly, we conducted two evaluation studies in the corresponding years. In both studies, we aimed to collect and explore the participants’ perceptions of our LOCO-Analyst tool. The participants were first presented with the main functionality and features of the tool; then, they were asked to try the tool on their own; and finally, they were asked to fill in the supplied questionnaire. Once the data were collected, we used quantitative methods for data analysis.

#### 3.2.2 Participants

For the 2006 study, we recruited 18 participants from Simon Fraser University (SFU), University of Saskatchewan (USASK), and University of Belgrade (UB). In the subsequent, 2009 study, 22 participants were recruited from SFU, UB, Athabasca University (AU), and a private Canada-based company developing and offering technology and content for professional training. In both studies, the participants were also asked to express their role in online education. We distinguished between the following three roles:

- **instructor** – a person who had independently instructed at least one entire course;
- **teaching assistant** – a person with teaching assistant experience only;
- **research student/practitioner** – a person who did research related to online education, or practiced online education in industry through software and content development and delivery.

The average experience of the participants in their role was 3.09 years (SD=2.2) and 6.45 years (SD = 5.58) in the first and second study, respectively. Seven participants from the first (2006) study, took part in the second (2009) study. The majority of participants in both studies came from Computer Science or Information Systems background: 16 out of 18 participants in 2006, and 21 out of 22 participants in 2009. In both studies, all the participants agreed to take part in the experiments, and received neither financial nor non-financial credits for their participation. All participants who responded to our invitation to take part in the study, successfully completed all the tasks of the study.

**3.2.3 Materials**

In both studies (2006 and 2006), questionnaires were used. Generally, the two questionnaires had the same structure with their questions grouped in the following main categories:

1. **General and conclusion questions**: the introductory part of the questionnaires contained general questions about the participants’ role (e.g., instructor, teaching assistant) and their experience related to the chosen role. In the concluding part, we asked for participants’ overall opinion about the feedback offered by the tool, their willingness to use the tool in their future work, and suggestions for improvement of the tool.

2. **Questions about the tool’s functionality**: The questions in this category were aimed at gathering the participants’ opinion about different features and functionality provided by tool. These questions were divided into two subcategories: questions regarding the feedback about the learning content (i.e. single lessons, composite lessons, learning module, quiz), and questions regarding students’ learning activities (i.e. activities in discussion forums and chat rooms, usage and annotation of learning content).

3. **Graphical User Interface (GUI) related questions**: questions within this category were aimed at eliciting the participants’ opinion about different aspects of the tool’s GUI, like intuitiveness and information load, as well as how it compares to some related tools.

The questionnaires used in these two studies were not exactly the same – the questionnaire used in the second (2009) study could be considered as an extension and improvement of the one used in the initial (2006) evaluation study. More precisely, there were two major differences between the two questionnaires. First, a new category of

---

6 In fact, we did both quantitative and qualitative analyses, but in this paper we focus exclusively on former.
questions were used for testing LOCO-Analyst’s functionality for the maintenance of a domain (i.e., course) ontology, including the visual representation of the ontology, the enabled forms of interaction with it as well as and the role of students’ collaborative tags in domain ontology evolution2. Second, new scaling options were introduced for some of the questions. Since we needed finer grained feedback from participants and wanted to be more consistent with recommended practices, we used 5-point, instead of 4-point, Likert-like scale. There was one question (about information overload caused by the tool) which had 3-point scale (No, Have No Opinion, and Yes) in the first study, and 5-point scale (including strongly disagree, disagree somewhat, uncertain, agree somewhat, and strongly agree) in the second study. To be able to compare the results for this question, we needed to define a mapping schema between our 3-point and 5-point scales. In Section 4.2 and Figure 4a, we show and provide justification for the developed mapping schema between these two scales. We mapped strongly agree, and agree somewhat to Yes; Uncertain to Have no Opinion; and disagree somewhat and strongly disagree to No. The corresponding recorded data values of the scale points are shown beside each point.

In the first (2006) study, the first version of LOCO-Analyst tool was used. In addition to a face-to-face tutorial about the tool, written guidelines about the use of the tool were provided. In the second (2009) study, the improved version of LOCO-Analyst (cf. Section 2.1-2.2) was used. Instead of written guidelines and face-to-face tutorial, video demos of the tool are used. The face-to-face tutorials (in 2006) and demo videos (in 2009) covered all the kinds of feedback that tool provides (Table 1). In addition, two demo videos were aimed at presenting two new features of the LOCO-Analyst tool that were introduced in 2009: feedback about students’ comprehension of course topics based on their annotation of course content, and the tool’s support for course ontology maintenance. These videos are still available on the LOCO-Analyst’s web site4. Within the questionnaire (of the 2009 study), for each of the questions we refer the participants to corresponding videos. To exemplify these different kinds of feedback, we used the log data and the content of one learning module (Programming Languages) of the “Introduction to Computer Science for Non-Majors course”9.

3.2.4 Procedure
The participants received (via email) a document explaining the purpose of the evaluation study and outlining the steps they should take in order to contribute to the study. The participants of the 2006 study were asked to join us for a face-to-face presentation of the LOCO-Analyst functionality and features (one presentation was organized in each participating institution: SFU, BU, USASK). In the 2009 study, the demo videos were uploaded on one of the on-line repositories10 and we also uploaded stream version of the demos on the Athabasca University server. It was explicitly mentioned to the participants that all the videos referenced from the questionnaire could be either downloaded or directly streamed from the given links.

After familiarizing themselves with the LOCO-Analyst tool, in both studies, the participants were asked to download the tool and try the presented functionality. They were also encouraged to send any further clarification questions to the evaluation team. Together with the guidelines, we also supplied the participants with the evaluation questionnaire and they were asked to send the completed questionnaire back to us by email. Finally, having received the answers from all the participants, we entered answers into Excel spreadsheets for the analysis.

4. RESULTS AND DISCUSSION
In this section, we present the employed statistical techniques, the purpose of their use, the obtained results and their interpretation. First, descriptive statistics, including mean, standard deviation, and number of valid responses for each question from the questionnaire, are described. Next, hypotheses related to the research questions (RQ1-6) are tested with appropriate statistical methods. We used the JMP(https://www.jmp.com) tool to perform statistical tests on the hypotheses. The threshold of p<0.05 was chosen to designate the statistically significant level.

Before discussing the specific results, we report the internal reliability of the collected Likert scale data. For this, we used the standard Cronbach’s α coefficient which is an appropriate test when questions measure different areas within a single product, like different features within the LOCO-Analyst tool. We obtained α = 0.90 which is higher than the reliability value of 0.80, typically used as a minimal threshold for measuring reliability.

4.1 The Perceived Usefulness of the Tool (RQ 1)

The first research question (RQ1) is tested using descriptive statistics. The results obtained for the data collected in both studies (2006 and 2009) are presented in Table 2. The results relevant for the exploration of RQ1 are given in the following categories: perceived value of the supported feedback types; perceived usefulness of the tool for improving the course content/instruction; perceived value of the tool’s GUI; and general perception of the tool. For each question, Central tendency measure (mean), Standard Deviation (SD) and the number of valid answers (N) are reported. Regarding the first category – the perceived value of different feedback types – the results show that the minimal perceived value of feedback are 4.33 (SD = 0.84) and 4.27 (SD = 0.77) in the 2006 and 2009 studies, respectively. Considering the 5-level scale that was used for these questions (1-very weak, 2-weak, 3-neutral, 4-good and 5-very good), the average opinion about the feedback offered by the tool is between good and very good. Moreover, there is no significant difference between 2006 and 2009 results in the value of the provided feedback types.

---

2 The evaluation of this feature of the tool is out the scope of this paper
3 http://jelenajovanovic.net/LOCO-Analyst/videos.html
4 The course was deployed within the iHelp Courses Learning Management System at the University of Saskatchewan; iHelp Courses’ usage tracking data were used for testing LOCO-Analyst’s functionality
10 Dropbox - https://www.dropbox.com/
One can observe that results in the 2009 study revealed some slightly decreased values of the answers to the questions (Q1a-c), which are related to feedback types F1-F3. Our statistical analysis showed no significant differences though. This slight drop in the values can be attributed to the fact that our improvements of the tool did not have much focus on those three types of feedback. As explained in Sect. 2, the participants’ expressed a need for improvements of feedback types F4-F5 (i.e., also lower scores Q1d-e) in the 2006 study, and improvement of the feedback about individual students (i.e., new feedback types F6-F7 covered in questions Q1f-g). That is, given the first version from 2006, the users likely expected more modern interfaces and data representation in 2009 on those three types of feedback as well, even though they did not express that in the 2006 study.

Table 2: Descriptive statistics for both the studies: the first column of the table is the category of the questions; the second column shows the content of the questions of the questionnaires, and the third and fourth columns show the mean, standard deviation, and number of valid answers for each question. N/A means the question did not exist in the first questionnaire. (‡) – a question in the 2006 study had 4-point scale. (‡‡) – a question in the 2006 study had 3-point scale

<table>
<thead>
<tr>
<th>Category</th>
<th>Question Description in the questionnaire</th>
<th>Study 1 – 2006</th>
<th>Study 2 – 2009</th>
</tr>
</thead>
<tbody>
<tr>
<td>Perceived value of feedback types</td>
<td>Q1a: Feedback about individual lesson</td>
<td>4.65, 0.61, 17</td>
<td>4.41, 0.59, 22</td>
</tr>
<tr>
<td></td>
<td>Q1b: Feedback about a group of (related) lessons</td>
<td>4.61, 0.50, 18</td>
<td>4.36, 0.66, 22</td>
</tr>
<tr>
<td></td>
<td>Q1c: Feedback about a learning module as a whole</td>
<td>4.56, 0.51, 16</td>
<td>4.29, 0.64, 21</td>
</tr>
<tr>
<td></td>
<td>Q1d: Feedback about students’ performance on a quiz</td>
<td>4.33, 0.84, 18</td>
<td>4.73, 0.46, 22</td>
</tr>
<tr>
<td></td>
<td>Q1e: Feedback about the student’s activities in discussion forums and chat rooms</td>
<td>4.33, 0.77, 18</td>
<td>4.68, 0.57, 22</td>
</tr>
<tr>
<td></td>
<td>Q1f: Feedback about the student’s interaction with the learning content (lessons)</td>
<td>N/A</td>
<td>4.41, 0.73, 22</td>
</tr>
<tr>
<td></td>
<td>Q1g: Feedback about the student’s comprehension of the studied topics (based on his/her annotations)</td>
<td>N/A</td>
<td>4.27, 0.77, 22</td>
</tr>
<tr>
<td>Perceived usefulness of the tool</td>
<td>Q2: The tool enables me to get an insight into the students’ interactions with the learning content</td>
<td>3.71, 0.47, 17</td>
<td>4.45, 0.67, 22</td>
</tr>
<tr>
<td>for improving the course content</td>
<td>Q3: The information the tool provides helps me identify what needs to be improved in the learning content</td>
<td>3.50, 0.70, 18</td>
<td>4.27, 0.70, 22</td>
</tr>
<tr>
<td>instruction</td>
<td>Q4: The tool provides relevant information regarding the students’ interactions within the online learning environment.</td>
<td>3.72, 0.46, 18</td>
<td>4.41, 0.80, 22</td>
</tr>
<tr>
<td></td>
<td>Q5: The information provided by the tool helps me determine how to improve the students’ online interactions.</td>
<td>3.17, 1.09, 18</td>
<td>3.54, 1.01, 22</td>
</tr>
<tr>
<td>Perceived value of the tool’s GUI</td>
<td>Q6: The tool helps me identify the domain topics the students were having difficulties with.</td>
<td>N/A</td>
<td>4.57, 0.68, 21</td>
</tr>
<tr>
<td></td>
<td>Q7: LOCO-Analyst’s GUI (Graphical User Interface) is intuitive enough</td>
<td>4.22, 0.73,18</td>
<td>4.50, 0.80, 22</td>
</tr>
<tr>
<td></td>
<td>Q8: LOCO-Analyst’s GUI is overburdened with information</td>
<td>1.72, 0.58, 18</td>
<td>2.50, 1.34, 22</td>
</tr>
<tr>
<td></td>
<td>Q9: My general opinion about the GUI</td>
<td>4.00, 0.84, 18</td>
<td>4.50, 0.60, 22</td>
</tr>
<tr>
<td>General opinion of the tool</td>
<td>Q10: All in all, I found LOCO-Analyst a handy tool for feedback provisioning</td>
<td>N/A</td>
<td>4.68, 0.48, 22</td>
</tr>
<tr>
<td></td>
<td>Q11: I would like to be able to use LOCO-Analyst in my teaching practice</td>
<td>N/A</td>
<td>4.04, 0.90, 22</td>
</tr>
<tr>
<td></td>
<td>Q12: LOCO-Analyst provides me with more useful feedback than other similar tool(s) I have used/ tried</td>
<td>N/A</td>
<td>3.33, 0.86, 21</td>
</tr>
<tr>
<td></td>
<td>Q13: LOCO-Analyst is more intuitive than the other tools capable of for feedback provisioning I have used/ tried</td>
<td>N/A</td>
<td>3.25, 0.55, 22</td>
</tr>
</tbody>
</table>

For the second category of questions – the perceived usefulness of the tool for improving the course content/instruction – the minimal perceived value by the participants are 3.17 (SD = 1.09), out of 4, in 2006, and 3.54 (SD =1.01), out of 5, in 2009. These minimal values are considerably lower than those for the other questions in this category. Specifically, the difference between the participants’ average opinion regarding Q5 (“The information provided by the tool helps me determine how to improve the students’ online interactions”) and their opinion regarding other questions of this category reveals that the feedback provided by the tool is suitable for gaining insight into the weak points of students online interactions, but not for getting suggestions how to improve the identified weaknesses. This is actually what we had expected since LOCO-Analyst is currently not capable of offering suggestions for course improvement. On the other hand, some participants believed that by revealing the weak areas, the tool implicitly helps users to improve the students’ interactions during the learning process.

Considering the participants perception of the tool’s GUI (the third category of questions), we observed that the intuitiveness had slightly improved in the new version of the tool (Q7), as well as the general opinion about the tool’s GUI (Q9). However, the results also showed that the information burden posed by the tool’s GUI (Q8) was increased in the new (2009) version. This issue, as well as the use of different scales in Q8 within the two studies is discussed in detail in the next section (Section 4.2).

Regarding the last category of questions – general opinion about the tool – we have found that the study participants were satisfied with the tool, as their responses to the questions Q10 and Q11 demonstrate (both receiving mean values above 4). The other two questions (Q12 and Q13) got significantly lower values (both around 3.3, out of 5), reflecting that the great majority of participants chose the answer “Uncertain”. This is caused by the participants’ low level or lack of experience with other tools for feedback provisioning.

4.2 Improvement in the Tool’s Intuitiveness (RQ 2)

In order to explore whether the changes in the LOCO-Analyst’s GUI led to the significant increase in the tool’s
intuitiveness and ease of use as perceived by the users, we set the following research hypothesis (RH2):

RH2: There is a significant improvement in the intuitiveness and ease of use of the tool’s new GUI (cf. Sect. 2.1-2.2) over the previous version of its GUI.

To test this hypothesis, we had to compare the statistical results (i.e. the mean of the responses to GUI related questions) between the two studies (2006 and 2009). The questions (Q7-Q9) regarding the tool’s GUI and their corresponding mean and standard deviation, for both studies, are shown in Table 2.

As previously stated (cf. 3.2.3 Materials section), to be able to compare the users’ perception of the information load caused by the tool, as expressed in the two consecutive studies, we defined a mapping between the scales used in the two studies (Fig. 4a). The means of the responses related to information overload (Q8), after applying the mapping schema, are presented in Fig. 4b. We argue that the performed mapping does not affect the results of comparisons. In a pessimistic point of view, one may say that mapping Agree to Yes and Disagree to No is correct. However, the degree of agreement in Strongly Agree and Disagree Somewhat is higher than in Yes and No, respectively. Therefore, mapping these options assigns lower numeric values than real possible values for them could be. For example, if in the 2009 study we had used 3-points Likert scale instead of 5-points scale, the participants who selected Disagree somewhat, might have selected Have no opinion instead of No. Considering that we mapped Disagree somewhat to No, and No equals to 1 and Have no opinion equals to 2, the mapping provides lower mean value for the 2009 study from its actual mean value. Finally, because the lower value indicates lower overload with the information provided by the tool, the t-test results for information load question would not be correct. Although the mapping affects the values of Strongly Agree and Disagree Somewhat options, we have also the inverse situation for the Strongly Disagree and Agree Somewhat. That is, the degree of agreement in the Strongly Disagree and Agree Somewhat is lower than in No and Have no opinion, respectively. Therefore, the mapping assigns higher numeric values to the scales than real possible values for them could be. As shown in Figure 4(b), the total number of Strongly Disagree and Agree Somewhat responses (i.e. 13) is higher than the total number of Strongly Agree and Disagree Somewhat responses (i.e., 9). Therefore, based on these evidences, we assume the mapping does not affect the final results.

Results. To assure reliable results, we checked the underlying assumptions like normality of the data and equality of variance within two datasets. For all three GUI-related questions, the distribution of the data was not normal. Thus, we have run the student t-test over log-transformed data to compare the data values related to GUI. The t-test results for all three questions (i.e. Q7, Q8 and Q9) are as follow:

- There is no significant difference between intuitiveness of the first version \((M = 4.22, SD = 0.73)\) and the second version of the tool \((M= 4.50 SD= 0.80)\), \(t(37) = 0.89, p = .38\).

- There is no significant difference in the information overload produced by the tool in 2009 version \((M = 1.77, SD = 0.58)\) and 2006 version \((M = 1.72, SD = 0.72)\), \(t(37) = -0.28, p = 0.78\).

There is a significant preference for the new tool’s GUI \((M = 4.00, SD = 0.84)\) over the previous one \((M = 4.50, SD = 0.60)\), \(t(35) = 1.87, p = 0.03\).

![Fig. 4.](image)

Discussion. After the first (2006) study, we conducted a qualitative analysis of the results. The results showed that one of the main concerns of participants was the tool’s GUI. They suggested more powerful visualization techniques to convey the feedback more efficiently. The details of these findings can be found in our technical report [Ali et al, 2010]. Therefore, we have introduced visualization facilities to more effectively convey feedback produced by the tool. Our expectation was that the visualizations (and other smaller improvements in the tool’s GUI) would increase the intuitiveness of the user interface and consequently the ease of use of the tool. However, as shown above, the results did not confirm this. We believe that this might be caused by the increase in the amount of information that the new version of the tool provides to educators (through extension of existing and inclusion of a new feedback type). This interpretation is further supported by the second result (t-test on Q8) showing that there is no significant difference in the information load of the tool. It seems that the introduction of visual elements was neutralized by introducing additional information for educators. The increase in the kinds of feedback offered by the tool and the improved method of visualizing the data might be the reasons for evaluating the new interface as more effective than the previous one.

4.3 Impact of Educators’ Role on the Perceived Value of the Feedback on Learning Content (RQ 3)

In our attempt to identify potential relations between the participants’ characteristics (including their roles and experience with related tools) and their responses on the questionnaire, we first concentrated ourselves on the rela-
tion between the participants’ role and their perception of the feedback the tool provides about learning content (Table 1, F1-4). To explore this, we needed to test our research hypothesis (RH3):

RH3: The user’s role (instructor, teaching assistant, or research student/practitioner) does not significantly influence his/her perception of the LOCO-Analyst’s feedback related to the learning content (including feedback about single lesson, composite lesson, learning module, and quizzes).

In order to test the RH3 hypothesis, we first defined the following variables:

- **Role** variable contains the role of each participant; roles are categorized as instructor, teaching assistant, and research student/practitioner.
- **Single lesson feedback**, **composite lesson feedback**, **module feedback**, and **quiz feedback variables** (i.e., those obtained through Q1a-d) keep the participants’ ratings of the corresponding kind of feedback.
- The **Learning content feedback variable** is described as the sum of the variables corresponding to the above mentioned individual kinds of feedback.

Next, we defined the null hypothesis of the above research hypothesis and verified it using an appropriate test.

RH3a: The role of a person working with the LOCO-Analyst tool influences his/her evaluation of tool’s feedback on the learning content.

**Results.** As the collected data were not normally distributed (i.e., they were almost normal), a one way ANOVA test was used over log-transformed data of both 2006 and 2009 study to compare the means of the dependent variable (Learning content feedback variable) for the three roles of the participants in our study. The results of 2006 study did not reveal significant difference between research students/practitioners 19.20 (SD = 0.84), teaching assistants 18.17 (SD = 1.90), and instructors 18.50 (SD = 1.3), F (2, 14) = 0.70, p = 0.52. Likewise, in the 2009 study we did not find significant difference between different group means (research students/practitioners 17.57 (SD = 1.72), teaching assistants 17.37 (SD = 1.51), and instructors 18.66 (SD = 2.42)) F (2, 20) = 0.72, p= 0.49. Consequently, the null hypothesis was rejected and we could conclude that the role of a person working with the LOCO-Analyst tool does not influence his/her evaluation of the tool’s feedback on the learning content.

**Discussion.** As shown through testing the RH3 hypothesis, all the participants, independently of their role have a similar opinion about the feedback on learning content. On the other hand, since the minimum mean is 17.37 out of 20 for the learning content feedback variable and over 4 for its constituent variables, we can conclude that these kinds of feedback provide suitable level of information about the learning content for all potential users of LOCO-Analyst.

### 4.4 Impact of Educators’ role on the Perceived Value of the Feedback on Particular Students’ Learning Activities (RQ4)

Having explored the relations between the participants’ role and their responses about the learning content feed-
back, we have done similar investigation focusing on the role of participants and their perceived value of the usefulness of the feedback related to one specific student (Table 1, F5-7). Since we added a new kind of feedback related to individual students (Table 1, F7) and improved visual representation of the already existing ones (cf. Sections 2.1-2.2), we wanted to explore whether and how our study participants with different roles perceive these changes. We distinguished this research question from the previous one (RQ4), because in the two consecutive studies the number of feedback types for a particular student changed from two to three, whereas the number of feedback types related to learning content remained fixed. Similarly to RH3, we hypothesized the following answer for the research question (RH4):

RH4: The user’s role (instructor, teaching assistant, or research student/practitioner) does not significantly influence his/her perception of the LOCO-Analyst’s feedback related to a particular student (including feedback about student’s activity in discussion forums and chat rooms, student’s interaction with the learning content, and student’s comprehension of the studied topics).

To test the RH4 hypothesis, we operationalized the hypothesis’ concepts and defined following variables.

- **Role** variable contains the participants’ roles, which are categorized as instructor, teaching assistant, research student/practitioner.
- **Forum and chat room interaction**, **learning content interaction**, and **comprehension variables** (derived from Q1e-g). Values of these variables were assigned based on the participants’ ratings of the corresponding kinds of feedback through the questions Q1e-g.
- **Particular student activity variable** is defined as a sum of Forum and chat room interaction and learning content interaction variables for the 2006 study, and sum of Forum and chat room interaction, learning content interaction, and comprehension variables in the 2009 study.

Afterward, the null hypothesis of the above research hypothesis is defined and verified using an appropriate test (One-way ANOVA test).

**Results.** Since the collected data was not normally distributed, One-way ANOVA test was used over log-transformed data of both 2006 and 2009 study to compare the means of the dependent variable (Particular student activity) for the three groups of participants in our study. The results of the 2006 study did not reveal significant difference between research students/practitioners 8.00 (SD = 1.00), teaching assistants 9.12 (SD = 0.99), and instructors 8.80 (SD = 0.83), F (2, 17) = 2.24, p = 0.14. However, in the 2009 study we found significant difference between different group means: research students/practitioners 13.62 (SD = 1.92), and instructors 14.33 (SD = 1.03)) F (2, 21) = 3.50, p= 0.05. Therefore, the null hypothesis was not rejected for the 2006 study and it was rejected for the 2009 study. Having obtained these results, we performed One-way ANOVA on all the log-transformed variables in the hypothesis separately (i.e. for Forum and chat room interaction, learning content interaction, and comprehension variables) with respect to the role of participants for both
significant difference between teaching assistants and Tukey-Kramer test and comparing all means, we found a
mean of instructors 4.83 (SD = 0.41), F (2, 21) = 4.84, p= 0.02. After doing also the Tukey-Kramer test and comparing all means, we found a significant difference between teaching assistants and instructors in both tests for the corresponding question (Q1f) in the 2009 study. No significant difference was found between the participants based on their roles, in the 2009 study. Furthermore, we found that there was a significant difference between different roles and their perception of feedback on student’s interaction with learning content (i.e., learning content interaction variable). The means for the corresponding question (Q1f) were (research students/practitioners 4.62 (SD = 0.52), teaching assistants 3.87 (SD = 0.83), and instructors 4.83 (SD = 0.41)) F (2, 21) = 4.84, p= 0.02. After doing also the Tukey-Kramer test and comparing all means, we found a significant difference between teaching assistants and instructors in both tests for the Particular Student activity and learning content interaction variables.

Analysis. We found the results gained through the exploration of this research question very interesting. Whereas in the 2006 study no significant difference was found between the participants based on their roles, in the 2009 study (after the introduction of a new kind of feedback and improved visualization of all kinds of feedback), significant difference was observed between instructors and teaching assistants. One possible interpretation of this fact is that the level of details and sophistication, introduced in the new version of the tool, has been appreciated by the groups who understand and overlook the whole learning process in greater level of details and responsibility over those who only had some teaching assistance experience. For example, the detection of learning patterns from a visual representation of a student’s interaction with learning content, as the one shown on Fig. 1, can be more easily done by an experienced instructor than a teaching assistant; hence instructors have obviously shown more appreciation for this kind of feedback.

4.5 Impact of Educators’ Previous Experience with Similar Tools on the Perceived Usefulness of LOCO-Analyst

The next area of our exploration was the relation between the participants’ level of experience with related e-learning tools, especially Reload Content Packaging Editor, and their evaluation of LOCO-Analyst’s feedback and GUI. As previously explained (cf. Section 2), LOCO-Analyst is implemented as an extension of Reload, and its user interface is largely determined by the Reload’s interface. Accordingly, our assumption was that users already familiar with Reload would more easily familiarize with LOCO-Analyst and find it easier to work with than users who did not have any experience with Reload. We generalize this assumption into the following research hypothesis (RH4):

**RH5:** The user’s experience with related tools (e.g., Reload Content Packaging Editor) significantly influences his/her perception of the LOCO-Analyst’s GUI and the feedback it provides.

To test this hypothesis, we defined its null hypotheses (RH5a) as follows:

**RH5a:** There is no significant difference in the evaluation of the LOCO-Analyst’s GUI and the feedback it provides between participants familiar and those who were unfamiliar with Reload.

Results. Since the distribution of the data was not normal, we did the student t-test over log-transformed data for all the questions related to the feedback offered by the tool (Q1a-Q1g) and its GUI (Q7-Q9). For all the questions the t-value was small and the p-value was above the significant level (p>=.50). Accordingly, we were not able to drop the null hypothesis (RH5a).

Discussion. The results of the conducted tests did not prove our assumption (better say, our concern) that educators’ experience with the Reload tool would influence their acceptance of LOCO-Analyst and perception of its functionality. There are many tools that educators use for creating and managing their online courses and it is natural to expect that many of them would not be familiar with Reload. Therefore, we find it relevant for the LOCO-Analyst’s acceptance that educators’ experience with the tool would not be diminished due to their unfamiliarity with Reload.

4.6 Identification of Additional Variables Influencing LOCO-Analyst’s Perceived Usefulness

Aiming to identify additional variables that influence the users’ perception of the usefulness of our tool and the intuitiveness of its GUI (besides those related to the users’ role and their experience with e-learning tools), we did a correlation analysis. We analyzed the correlation between the questions from the 2009 questionnaire to find the questions that are correlated the most, and also explored the kinds of relations between those questions. The final objective was to identify the variables that we should act upon in order to improve our tool in general, and its GUI, in particular.

The correlation coefficient, which we used, was Spearman’s ρ-coefficient due to the non-normal distribution of the collected data. We listed some of correlations that we identified as strong (0.8~1) or moderate (0.6~0.8) in Table 3. The first two columns of Table 3 show the questions (Qi, i=1,13), whereas the third column presents Spearman’s ρ-coefficient and the p-value. In what follows we discuss the correlations that we have identified as significant.

Q6 - Q1a, Q1b: This correlation was something that we have expected, that is, what we aimed to achieve. Specifically, the tool’s feedback about lessons, both individual (Q1a) and composite (Q1b), was designed with the goal of providing educators with information about the course topics that are difficult for the students (Q6) and thus require the educator’s attention.

Q7 - Q1e, Q2: This correlation could be explained by the fact that the feedback about students’ activities in forums and chat rooms (Q1e) as well as about their interaction with the course content (Q2) is presented with lots of visual elements (charts and graphs, as those show on and Fig. 2). We believe that those visualizations (some of them
interactive) have contributed to the intuitiveness of the user interface. This is an important observation for the further development of not only our tool but feedback provisioning tools in general, as it indicates the importance of information visualization for facilitating the feedback comprehension by end users.

Q13 - Q12: The implication of this correlation – the way the feedback is presented (Q13) significantly influences how its usefulness is perceived by the end users (Q12) – is directly related to the previous one. It further intensifies the importance of intuitive interfaces for the perceived value and thus acceptance of a feedback provisioning tool.

Table 3. The correlation between variables where correlation coefficient is higher than .6 and p-value is less than <.05.

<table>
<thead>
<tr>
<th>First Question (Variable)</th>
<th>Second Question (Variable)</th>
<th>Correlation coefficient; p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Q6: The tool helps me identify the domain topics the students were having difficulties with</td>
<td>Q1a: Feedback about individual lesson</td>
<td>0.66; &lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Q1b: Feedback about a group of (related) lessons</td>
<td>0.71; &lt; 0.01</td>
</tr>
<tr>
<td>Q7: LOCO-Analyst’s GUI (Graphical User Interface) is intuitive enough</td>
<td>Q1e: Feedback about the student’s activities in discussion forums and chat rooms</td>
<td>0.60; &lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Q2: The tool enables me to get an insight into the students’ interactions with the learning content</td>
<td>0.61; &lt; 0.01</td>
</tr>
<tr>
<td>Q13: LOCO-Analyst is more intuitive than the other tools capable of feedback provisioning I have used/ tried</td>
<td>Q12: LOCO-Analyst provides me with more useful feedback than other similar tool(s) I have used/ tried</td>
<td>0.65; &lt; 0.01</td>
</tr>
<tr>
<td>Q9: My general opinion about the GUI</td>
<td>Q1a: Feedback about individual lesson</td>
<td>0.81; &lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Q1b: Feedback about a group of (related) lessons</td>
<td>0.66; &lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Q1c: Feedback about a learning module as a whole</td>
<td>0.69; &lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Q1g: Feedback about the student’s comprehension of the studied topics (based on his/her annotations)</td>
<td>0.72; &lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Q2: The tool enables me to get an insight into the students’ interactions with the learning content</td>
<td>0.64; &lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Q6: Helps me identify the domain topics the students were having difficulties with.</td>
<td>0.65; &lt; 0.01</td>
</tr>
<tr>
<td>Q10: All in all, I found LOCO-Analyst a handy tool for feedback provisioning</td>
<td>Q1a: Feedback about individual lesson</td>
<td>0.65; &lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Q1b: Feedback about a group of (related) lessons</td>
<td>0.68; &lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Q1d: Feedback about students’ performance on a quiz</td>
<td>0.68; &lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Q2: The tool enables me to get an insight into the students’ interactions with the learning content</td>
<td>0.69; &lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Q4: Provides relevant information regarding the students’ interactions within the online learning environment</td>
<td>0.71; &lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Q6: Help me identify the domain topics the students were having difficulties with.</td>
<td>0.79; &lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Q7: LOCO-Analyst’s GUI (Graphical User Interface) is intuitive enough</td>
<td>0.72; &lt; 0.01</td>
</tr>
<tr>
<td></td>
<td>Q9: My general opinion about the GUI</td>
<td>0.76; &lt; 0.01</td>
</tr>
<tr>
<td>Q11: I would like to be able to use LOCO-Analyst in my teaching practice</td>
<td>Q3: The information helps me identify what needs to be improved in the learning content</td>
<td>0.77; &lt; 0.01</td>
</tr>
</tbody>
</table>

Q11 – Q3: This correlation implies that users would appreciate tools that could provide them with some hints/suggestions on how to improve their courses (besides pointing out the problems). However, LOCO-Analyst was designed as a problem detection tool, not problem solving tool. Accordingly, in its present state, it is capable of indicating potential problems to the user, but is not able to suggest how to resolve those problems. We have already started exploring potential approaches (e.g., using libraries of problem-solution patterns) to extend LOCO-Analyst with the features that would facilitate the user’s task of solving the identified problems.

In order to determine which of the variables can best predict a general opinion about the perceived utility of the feedback provisioning tool for educators (Q10), the regression model was used. The following result has been discovered:

\[ Q10 = 1.02 + 0.25 Q4 + 0.37 Q6 + 0.18 Q7 \]

Q4, Q6, and Q7 are questions about the tool’s feedback regarding students’ interactions, students’ comprehension of course topics, and intuitiveness of the user interface, respectively. Reported p value (max 0.0006) for each coefficient reveals that these three variables significantly predict the participants’ general opinion about the tool. The reported statistics related to overall equation show that these variables also explain a significant proportion of variance in the participants’ general opinion about tool (\( R^2 = 0.93, F(3, 20) = 81.46 \) and \( p < 0.001 \)). This result reasserts, what we have had expected, that the key values appreciated by educators are: information about students (mutual) interactions and their comprehension of the course content, coupled with an intuitive and easy to follow presentation of that information. Similar to the perceived utility of the tool for educators, we also tested which of the variables would be the best predictors for the general impression about the graphical user interface of the tool (Q9). Again, a regression analysis is used and the following model is obtained:
5.1 Threats to internal validity

In this section, we discuss some potential threats to validity of our experimentation results.

5.2 Threats to external validity

The external validity investigates if the obtained results can be generalized. Specifically, in our case the question is whether we can generalize our results to similar tools in the e-learning domain. Two main confounding factors are population and ecology [Chin, 2001].

Our population is comprised of instructors, teaching assistants and research students/practitioners who could be considered as the target population for our feedback provisioning tool. One may say that research students and teaching assistants are not the real target users of this kind of tool and their judgment of the tool’s features is not that relevant. However, they have already been involved in online learning processes and have some experience of working with students in such settings. On the other hand, it is true that they still lack instructional skills and experience that would allow them to fully appreciate the functionality that a tool like LOCO-Analyst offers for online learning processes (as shown in Section 4.4). Of course, a replicated experiment should further investigate the validity of this analysis.

Another important issue related to the population factor is the fact that in our studies the population was predominantly composed of individuals with Computer Science or Information Systems background: 16 out of 18 participants in the 2006 study, and 21 out of 22 participants in the 2009 study. Therefore, a replicated experiment with non-computer scientists is needed to validate the general applicability of our results.

Finally, our results are applicable to a specific set of possible learning applications – university education. Our participants from the industry (in the 2009 study) expressed their concern that the same results could be applicable to the corporate training context. Still, we believe that at least some of the lessons learned could be transferred to contexts which are not universities.

6 RELATED WORK

The latest, 2010 Horizon Report [Johnson et al., 2010] has recognized the relevance of capturing, analyzing and visualizing student data as means for enabling educators to make better decision on what and how to teach. In online and to a great extent blended learning environments, user tracking data and students grades form the bases for different kinds of analysis and visualizations aimed at making it easier for educators to understand where students are successful, as well as to see where improvements can be made. Accordingly, log data analysis and visualization, for the purpose of feedback generation, have been relevant research topics in the educational research community in general, and more specifically among AIED (Artificial Intelligence in Education) and EDM (Educational Data Mining) researchers. These communities have done a substantial research work in generating student-directed feedback by leveraging user tracking data of e-learning systems and tools in general, and ITSs (Intelligent Tutoring Systems), in particular (as shown in, e.g.,

Q9 = -0.21 + 0.37Q1a + 0.37Q1e + 0.31 Q1g

Reported p values for each coefficient reveal that all three variables significantly (Q1e and Q1g, p<0.01, and Q1a, p<0.05) predict the participants’ perceived value of the tool’s user interface. That is, the user interfaces applied for presenting and interacting with the feedback about individual lessons, about the students' activities in discussion forums and chat rooms, and the students’ comprehension of the studied topics (based on their annotations) are the best predictors of the overall perceived value of the user interface of the feedback provisioning tool. The three feedback types (Q1a, Q1e, and Q1g) also explain a significant proportion of variance in the perceived value about the user interface of the tool (R2= 0.70, F(3, 21) = 22.11 and p < 0.001).

5 Threats to validity

In this section, we discuss some potential threats to validity of our experimentation results.

5.1 Threats to internal validity

With respect to internal validity of our experiment, we consider if some confounding factors would make a difference in the analyses [Chin, 2001]. In our experiment, the participants could have responded to the questions differently based on the following confounding factors: difference in the educational role, experience, and motivation. As reported and discussed in Sections 4.3 and 4.4, the participants' role (instructors, teaching assistants or research students/practitioners) did not influence their perception of certain kinds of feedback (feedback on learning content, cf. Section 4.3), whereas it did influence the perception of other kinds of feedback provided by the tool (feedback about an individual student, cf. Section 4.4). The analysis of participants experience in working with e-learning tools did not reveal significant differences between people familiar with this kind of tools and people who are not (cf. Section 4.5). We exclude the motivation as a confounding factor because the participation in the study was on a voluntary basis, and none of our participants left the experiments, while a great majority responded to the optional open-ended questions.

One factor that might be an important internal validity threat is that we have no evidence if the participants actually used our tool or based their evaluations on the face-to-face tutorial in 2006 study that is, watching the videos in 2009 study. However, the depth of the observation in open-ended questions in both studies [Ali et al, 2010] reflects a high-level of familiarity with the tool and high-likelihood of the actual use of the tool.

Another potential internal validity threat for our results is the difference in some of the question scales used in the questionnaires of the first (2006) and the second (2009) study. As explained in Section 4.2, with regard to this limitation we defined a mapping between the different scales to reduce the effect of difference and showed that the final results were most likely not affected by this mapping (cf. Section 4.2).
[Dominguez et al, 2010] [Roll et al, 2010] [Kazi et al, 2010]. However, so far, significantly less research efforts have been oriented towards educators and the feedback they require. This is somewhat unexpected since the need for educator-directed feedback generation tools is evident in the communities of e-learning practitioners\textsuperscript{14}.

In general, feedback produced by diverse feedback generation tools (both student and educator directed) can be classified into two broad categories: local and global [Melis and Ullrich, 2003]. Local feedback is designed as a direct response to a user’s action (e.g., a student’s response to a question on a test), whereas global feedback is often delayed and is designed to provide information about a whole learning process or some of its relevant parts. Our approach and a great majority of other approaches oriented towards educators (presented below) are focused on the provision of global feedback. This came as a result of educators’ feedback requirements elicited through user studies [Mazza & Dimitrova, 2003] [Zinn & Scheuer, 2006] [Jovanovic et al, 2007].

In what follows we present some of the previous research work (on log data analysis and visualization for educator-directed feedback) which we found related to and relevant for our work presented in this paper.

Kosba and his associates have developed the Teacher ADVisor (TADV) framework which uses LCMS tracking data to elicit student, group, and class models, and using these models to help teachers gain a better understanding of their distance students [Kosba et al, 2005]. It uses a set of predefined conditions to recognize situations that would require teachers’ intervention, and when such a condition is met, TADV generates an advice for the teacher, as well as a recommendation for what is to be sent to students. Whereas TADV is focused on the teachers’ day-to-day activities, LOCO-Analyst aims at helping them rethink the quality of the employed learning content and learning design.

Zinn & Scheuer [Zinn & Scheuer, 2007] have developed Student Inspector, a tool which analyzes and visualizes usage-tracking data in order to help educators learn more about their students in distance learning environments. Among other things, it provides educators with information about students with the lowest performance rate, the most frequently occurring error types, and the competences of a given student. The development of the tool was preceded by a user study aimed at identifying the information that, on one hand, is valuable for educators, and on the other, can be generated from user-tracking data [Zinn & Scheuer, 2006]. However, unlike Student Inspector which is bounded to ActiveMath\textsuperscript{15} (a Web-based, user-adaptive learning environment for mathematics) and iClass\textsuperscript{16} (an intelligent cognitive-based open learning system), our solution, thanks to its ontological foundation, is tool-independent [Jovanovic et al, 2008].

Log data analysis aimed at exploring the kinds of pedagogical support that are the most effective, either overall or for a certain group of students or in a specific situation, has been one of the key application areas of Educational Data Mining methods [Baker & Yacef, 2010]. This is exemplified in the research work of Zaine & Luo (2001), Merceron & Yacef (2005), Romero et al (2008), and Ben-Naim et al (2009), presented below.

Zaine & Luo (2001) applied advanced data mining techniques on access logs of an LMCS in order to extract patterns useful for evaluating and interpreting on-line course activities. Educators can tailor the data mining process to their needs by expressing them as constraints on the mining process (e.g., they can select a desired student or study group, the desired time period, etc). The discovered patterns are presented in the form of charts and tables. TADA-Ed (Tool for Advanced Data Analysis in Education) is another data mining platform which integrates various visualization and data mining facilities to help educators discover pedagogically relevant patterns in students’ online exercises [Merceron & Yacef, 2005]. Unlike these and similar systems that focus on a single learning activity (reading and exercises, respectively, in the aforementioned systems), LOCO-Analyst analyzes diverse kinds of learning activities typically occurring in today’s LCMSs. In addition, it is easy to use (as our user study has demonstrated), which, in general, is not the case with data mining based tools.

Romero et al [2008] have employed sequential pattern mining and visualization in order to scaffold the educators’ task of analyzing students’ navigation in a web-based learning system. The researchers’ approach was to interpret the student’s “click-stream” data in order to reconstruct their navigation through the course content. To facilitate the educators’ comprehension of the mined students’ traces, the researchers used a graph-based representation of discovered traces. Even though the usability of this tool for educators is questionable, for us, the research work it is based upon is very interesting as we are exploring approaches for detection of successful learning paths and the factors influencing them.

Similar, though more complex, approach has been suggested by Ben-Naim et al (2009) with the goal of helping educators understand student’s behavior in Adaptive Tutorials. Specifically, the objective was to help educators to improve the feedback their students receive upon ending up in ‘error-states’, as well as to identify new kinds of ‘error-states’, i.e., problem-specific error patterns. The researchers have developed two tools: one that requires educator’s involvement in the discovery process (educator identifies relevant patterns in the graph-based visualization of student’s traces through the problem space); and the other where the educator is presented with a list of the most relevant issues that might need his/her attention. Whereas the second tool is far more convenient for educators, it cannot identify all potentially relevant issues, especially those that are domain specific. This solution is valuable for problem-based Intelligent Tutoring Systems where problems to be solved by students are well-structured and precisely defined (all problems’ attributes are well defined and presented in a machine interpretable form). We aim at providing feedback in di-
verse educational contexts where students’ tasks are typically either fully unstructured or only partially structured.

Some researchers opted for using visualization techniques to help educators understand what is happening in their online courses. So, instead of relying on machine intelligence to do complex data analysis, the idea is to present the available data to the educators in an intuitive and comprehensible manner, and let human intelligence do the task where it (still) excel over machines: detection of trends, patterns, etc. An example of such an approach is CourseViz, which works with the WebCT LCMS to produce various graphical representations of student tracking data [Mazza & Dimitrova, 2007]. Another example is GISMO, which takes a similar approach for Moodle LCMS [Mazza & Milani, 2005]. Both tools aim at helping educators examine social, cognitive, and behavioral aspects of students enrolled in Web-based courses. They use students tracking data (e.g., access to resources and results on assignments and quizzes) as source data, and apply diverse visualization techniques to graphically represent them. Whereas these systems exploits visualization techniques to present row data, LOCO-Analyst goes a step further in terms that it also analyzes the data and provides educators with qualitative feedback.

Social network analysis and related visualizations have been increasingly used for gaining a better understanding of online social networks in different domains, e-learning being one of them (e.g., [de Laat et al, 2007]). For example, SNAPP (Social Networks Adapting Pedagogical Practice) is a tool that generates and visualizes a network of users’ interactions based on the data about the users’ activities in a discussion forum (e.g., who posted and replied to whom, what were the discussions about, etc.). It has been developed within the “Seeing” networks: Visualizing and evaluating student learning network” project to help educators to identify patterns of social interactions in an online learning community.

There is a plethora of web-based general purpose visualization tools which could be used for visualizing diverse kinds of learning-related artifacts, such as documents, collaborative tags, concept maps, etc. For example, Wordle, a handy visualization tool widely adopted by Web users in general, has also been recognized as beneficial for quick analyses and visualization of students’ texts and collaborative tags [Johnson et al, 2010]. By extracting and visualizing most frequently used words and phrases in students’ texts, Wordle allows both students and instructors to quickly identify the points that require additional work, as well as to get insights into the used terminology and domain conceptualization.

The use of visualization techniques as a mean for facilitating the comprehension of education-related data and what those data imply about the quality of the learning process, has already found its commercial use. Companies, such as Interactive Data Partners, have started offering different kinds of dashboards and other interactive visual tools providing relevant information not only for educators, but also school administrators.

We conclude this section with some very recent research work that we have found relevant for our future work on further improvement of the feedback offered by our LOCO-Analyst tool.

Macfadyen & Dawson [2010] have done a comprehensive analysis of usage tracking data captured by a LCMS in order to identify the students’ online activities that most accurately predict their academic achievement. They have identified the total number of posted discussion messages, sent email messages, and completed assessments as three key variables affecting students’ final grade. In addition, they have leveraged previously mentioned SNAPP tool for detection of variables influencing the development of an online learning community. The identified variables are planned to be used for the development of a reporting tool for educators. Obviously this work is highly related to our work on LOCO-Analyst and provides us with some relevant new knowledge which we plan to use to improve the feedback that our tool offers. This is primarily related to different measures of social network analysis that we intend to use to enhance the feedback about students’ (mutual) online interactions.

In workplace learning settings, traditional approaches to evaluating course design and learner performance (e.g., exams) are often not applicable and despite numerous efforts, evaluation of the true impact of [online] training is still an open issue [Murray & Efendioglu, 2007]. Aiming to address this problem, Macfadyen & Sorenson [2010] have developed Learner Interaction Monitoring System (LiMS) which captures fine-grained data about learners’ activities within online learning environments and by analyzing those data generates a descriptive profile of a learner available for inspection to the learner him/herself and managers. This profile contains a kind of ‘behavioral grade’ which LiMS computes by comparing the learner’s use of the training material with its expected use (as defined by the educator). LiMS is developed as a browser plug-in and as such available for any web-based learning platform. It can be customized to allow educators to inspect some specific aspects of learners’ behavior or course content. We have found this work highly related to and relevant for our future work, especially as it addresses the needs of workplace learning settings that we explore within the IntelLEO (Intelligent Learning Extended Organization) EU project. Unfortunately, besides the cited paper we were not able to find any more details about LiMS. We hope that more information about LiMS will be available soon as we are very interested in many of its features (e.g., computation of ‘behavioral grade’ and customization to educators’ specific needs).

7 Conclusion

To the best of our knowledge, there were no comprehensive quantitative evaluation studies of educator-directed feedback tools, which were aimed at moving beyond a

17 http://research.uow.edu.au/learningnetworks/seeing/snapp/
18 http://www.wordle.net/
19 http://www.interactivedatapartners.com/
20 http://www.intelleo.eu
basic set of descriptive statistics in order to explore more deeply different factors that are influencing the feedback provisioning systems and thus build a set of empirically validated facts about the important predictors of that process. Our initial evaluation of LOCO-Analyst [Jovanovic et al, 2007], our tool for providing educators with information-rich feedback, was no exception as we also did just some basic statistical analysis of the collected data. However, aiming to better inform our future research work, but also the work of the research community in general, we have done an in-depth quantitative evaluation of the LOCO-Analyst tool, which we have presented in detail in this paper.

The purpose of the conducted evaluation study was threefold. Firstly, to evaluate the effectiveness of the feedback offered by the tool, i.e., its ability to provide educators with useful information about the learning activities and performance of their students. Secondly, to compare the new (second) version of the tool with its first version, with special focus on the new kinds of feedback and the changes in the feedback presentation. Finally, to identify factors influencing LOCO-Analyst’s perceived usefulness. With respect to the first objective, the quantitative results showed that the participants rated most of the tool’s feedback as highly effective (the average rate was higher than 4.27, out of 5, for all the kinds of feedback). The second objective, finding out whether the improvements in the tool’s GUI made the GUI (and tool itself) easier to understand and more intuitive for the users, has also been supported with the statistical results. In pursuing the third goal of the study, we have first identified an interesting relation between the users’ role and their opinion about the feedback the tool provides. Specifically, it has been revealed that all three roles: instructors, teaching assistants, and student researchers/practitioners, have similar opinion on the effectiveness of different kinds of feedback related to the learning content. However, we have also found a significant difference between instructors and teaching assistants when it comes to the perception of the feedback about individual students’ interaction with learning content; this most likely originates from the instructors’ higher ability to identify useful patterns and draw relevant inferences from the presented feedback (as explained in Section 4.4). Yet another important insight is that the users’ experience in working with other e-learning tools did not affect their evaluation of the LOCO-Analyst’s feedback.

Through correlation analysis, we have also found that the way the feedback is presented significantly influences how its usefulness is perceived by the end users. This further implies the importance of intuitive interfaces for the acceptance of a feedback provisioning tool. Additionally, we have identified the users’ appreciation of tools that could provide them with some hints/suggestions on how to improve their courses (besides pointing out the problems). Finally, using regression analysis, we confirmed our expectation that the key values appreciated by educators are: information about students (mutual) interactions and their comprehension of the course content, coupled with an intuitive and easy to follow presentation of that information. We have also further strengthen the validity of our finding that visualization-rich user interfaces (specifically, in LOCO-Analyst those are user interfaces for presenting and interacting with the feedback about individual lessons, the students’ activities in discussion forums and chat rooms, and the students’ comprehension of the studied topics) are the best predictors of the overall perceived value of the user interface of a feedback provisioning tool.

In our future work we are going to use the study results to improve the LOCO-Analyst tool and provide more relevant and efficient feedback for educators. In addition, we plan to experiment with the feedback provisioning when multiple e-learning tools are in use, which is the case in the increasingly popular Personal Learning Environments [Attwell, 2007]. We are also going to explore and define metrics and predictors that can help an educator identify whether a certain pedagogy or planned learning objectives are achieved and to what extent. Finally, we intend to move LOCO-Analyst further from being problem diagnosis tool, and extend it with a guidance mechanism for educators that would indicate them potential improvements for the identified problems.

ACKNOWLEDGMENT
This project is in part supported by NSERC-funded LORNET Research Network and CIDA’s grants nos. 629 and 778. The research is also support through the NSERC Discovery Grants – Individual program.

REFERENCES


