Eighth IEEE International Conference on Advanced Learning Technologies

Tailoring Feedback in Online Assessment: Influence of Learning Styles on the Feedback Preferences and Elaborated Feedback Effectiveness

Ekaterina Vasilyeva
Dept. of CS and ISs, Univ. of Jyväskylä, P.O. Box 35, 40351 Jyväskylä, Finland
ekvasily@cc.jyu.fi

Paul De Bra
Dept. of CS, Eindhoven Univ. of Technology, P.O. Box 513, 5600 MB Eindhoven, Netherlands
debra@win.tue.nl

Mykola Pechenizkiy
Dept. of CS, Eindhoven Univ. of Technology, P.O. Box 513, 5600 MB Eindhoven, Netherlands
m.pechenizkiy@tue.nl

Seppo Puuronen
Dept. of CS and ISs, Univ. of Jyväskylä, P.O. Box 35, 40351 Jyväskylä, Finland
sepi@cs.jyu.fi

Abstract

Design of feedback is a critical issue of online assessment development within Web-based Learning Systems (WBLSs). This paper examines the potential possibilities of tailoring the feedback that is presented to a student as a result of his/her preferences and responses to questions of an online test with respect to the individual learning styles (LS). The paper briefly reviews the main types of feedback that can be presented during online assessment and discusses the challenges in authoring and tailoring of feedback in WBLSs. We report the results of some recent experiments organized as online assessment of students through multiple-choice quizzes in which students were able to request different kinds of feedback for the answered questions. The experimental results have confirmed that LS have a significant influence on (1) the feedback preferences (with regard to response certitude and correctness) of students and (2) the effectiveness of elaborated feedback (EF), i.e. improving students' performance during the test.

1. Introduction

Incorporating LS in WBLSs has been one of the topical problems of WBLS design during recent years. There are currently several WBLSs that support adaptation to the individual LS (AHA!, CS383, IDEAL, MAS-PLANG, INSPIRE). However, according to our knowledge, there is no system or reported research (in the e-learning context) that addresses the issue aimed at providing feedback tailored to the LS of the student.

This paper continues our previous work [7, 8], where we stated the problem and presented the results of our pilot experiments. Our early results confirmed the necessity of further experimental research on the interrelations between the personal LS and the adaptable feedback parameters [8].

In this paper we present and discuss the results of recent experiments organized as online assessment of the students through multiple-choice quizzes. The rest of the paper is structured as follows. We briefly review functions and types of feedback that can be provided by WBLSs in Section 2. Section 3 discusses the issues of authoring and tailoring of feedback in WBLSs focusing on the problem of tailoring feedback to LS. In Sections 4 and 5 we describe the organization and the results of our experiments. We briefly conclude with a summary and the directions for further research.

2. Feedback in Online Assessment

Feedback is usually a significant part of the assessment as students need to be informed about the results of their (current and/or overall) performance. Feedback could play different functions in WBLSs according to its learning effect: feedback can (1) inform the student about correctness of his/her responses, (2) it can "fill the gaps" in the student knowledge by presenting the information unknown to the student, and (3) "patch the student's knowledge" – i.e. trying to correct/overcome misconceptions the student may have.

The functions of the feedback imply the complexity of information that can be presented in immediate feedback: *verification* and *elaborated feedback* (EF) [4]. Verification can be given in the form of *knowledge of response* (indication of whether the answer was received and accepted by the system), *knowledge of results* (KR) (information about correctness or incorrectness of the response), or *knowledge-of-correct response* (KCR) (presentation of the correct answers) feedback [5]. *Elaboration* can address the topic and/or the response, discuss the particular errors, provide examples or give gentle guidance [6]. With EF the system presents not only the correct answer, but also

978-0-7695-3167-0/08 $25.00 © 2008 IEEE
DOI 10.1109/ICALT.2008.66

additional information – corresponding learning materials, explanations, parts of problem-solutions etc.

Different types of feedback carry out different functions and thus they can be differently effective in terms of learning and interaction and can even be disturbing or annoying to the student and have negative influence on the learning and interaction processes [3]. These observations emphasize the necessity of careful design of feedback in WBLSs.

3. Tailoring of Feedback in WBLSs

Design of feedback assumes that the following questions can/must be answered: (1) when should the feedback be presented; (2) what functions should it fulfil; (3) what kind of information should it include; (4) for which students and in which situations would it be most effective. The variety of possible answers to these questions makes authoring and designing feedback rather complicated, especially in WBLSs. Existing WBLSs only support scarce possibilities for authoring and presenting feedback.

Personalization of feedback may be a solution for the design of effective feedback in WBLSs as it is aimed at providing a student with the most suitable feedback for his/her personality, the performed task, and the environment. The issues (1) what can be personalized in the feedback and (2) to which characteristics should feedback be personalized are essential in the development of personalized feedback. Some answers to these questions can be found in [9, 10].

Individual LS are one of the important characteristics of the student that should be taken into account during personalization. They characterize the ways in which the student perceives information, acquires knowledge, and communicates with the teacher and with other students. Recent research confirmed the improvement of learning applications personalized to LS [1].

In the next sections we present the results of a series of experiments in which we studied the possibilities of tailoring the feedback to the LS of the students.

4. Method

We have studied different aspects of feedback tailoring during a series of experiments in the form of seven online multiple-choice tests in the Moodle learning system organized as a complimentary yet integral part of courses (with traditional in-class lectures and instructions) at Eindhoven University of Technology during the fall semester of 2007.

In this paper we report the results of part of the study that focused on the analysis of LS within one test with the students of a database course (38 students). The subjects participated individually in the tests for partial credit for the course (i.e. as a partial exam).

Both the authoring and assessment modules of Moodle were changed for these experiments. We increased the variety of feedback types, the order and the way in which feedback was presented.

Before the test the students were asked to answer 15 questions of Felder-Silverman's [2] index of learning styles (ILS) quiz to identify their LS according to three dimensions: sensing/intuitive, sequential/global, and active/reflective. We reduced the original number of questions of the ILS (normally 11 for each dimension) to just the most representative ones (5 for each of the chosen dimensions) according to Viola et al. [11]. Completing the LS questionnaire was not compulsory for the students (85% of students have actually completed LS quiz).

The (database course) test consisted of 15 multiple-choice questions. The questions were aimed at assessing the knowledge of the concepts and the development of the necessary skills (like computing a canonical cover). During the quiz design the teachers tried to come up with believable incorrect answers and were taking into account typical mistakes students make and misconceptions they may have. It was estimated that the students would need between 2 and 6 minutes for each question depending on its difficulty[1]. Each question was accompanied by the compulsory response confidence question: "Please evaluate your certainty about your answer (it affects your score)".

The test was taken simultaneously by all students in the same room. The students used their own laptops and were allowed to use all possible sources of information, except each other. The lecturer and 2 assistants were present in the room to observe and to assist the students when needed.

Short instructions about the number of the questions in the test, the grading scheme, the functionality of the system (including the possibility of requesting feedback and providing own comments) were given to the students before the test. During the instructions the students were motivated to read EF by promising that the EF for some questions could possibly (but not necessarily) contain hints for answering some of the following related questions of the quiz. The students started each test at the same time and had to finish within 1.5 hours. The students had to answer the multiple-choice questions in a fixed order.

In Figure 1 we present the dimensions of feedback

[1] The test was reasonably difficult given the amount of time to pass the test. Between 40% and 70% of the questions were answered correctly on average.

testing study. The students were able to choose between no feedback, immediate and delayed feedback. After answering a question (and specifying the certainty of the answer) the student could either go directly to the next question by pressing "Go to the next question" button or request feedback by pressing "Check the answer" button. As feedback students could either get KCR (highlighted correct answers and explanations) and KR (grade). On the page, where KCR feedback was presented, there was a button "Get Explanations", through which the students could get the EF. There was also a possibility to tick a box, requesting delayed EF (getting the explanations after answering all of the questions), and proceed directly to the next question. Students were offered also the option to request delayed EF from the page with immediate EF. On the page, where EF was presented the question and answers were presented with the correct answer(s) highlighted (KCR feedback).

We also asked the students to express their satisfaction about the presented feedback. They could optionally answer to the question whether feedback was useful or not.[2]

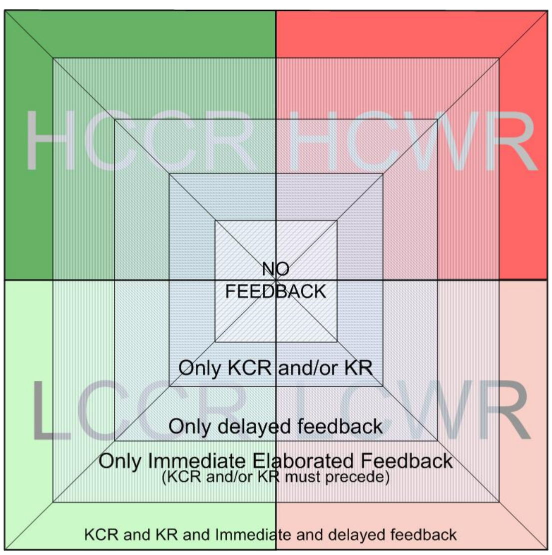

Figure 1. Dimensions of feedback tailoring study

Figure 1 also illustrates the response correctness and certitude dimensions of the study, which form four squares: high-confidence correct response (HCCR), high-confidence wrong response (HCWR), low-confidence correct response (LCCR), low-confidence wrong response (LCWR). Diagonals dividing each regions into two parts represent opposite cases (three pairs) of LS dimensions (e.g. active vs. reflective).

5. Results

The results are grouped into three sections which correspondingly highlight the major patterns in feedback preferences (Section 5.1), the effectiveness of feedback (Section 5.2) and the timing aspects (Section 5.3) with respect to students' LS, and correctness and certitude of their responses.

5.1 Differences in feedback needs with respect to LS

In Figure 2 the percentages[3] of cases where no feedback was requested at all (left column of plots) and cases where only KCR/KR feedback was requested (right column of plots) are shown. It can be seen that students with sensitive LS always requested KCR/KR feedback, whilst intuitive learners sometimes did not request any feedback at all, and this was more often with the answers in which they were not certain.

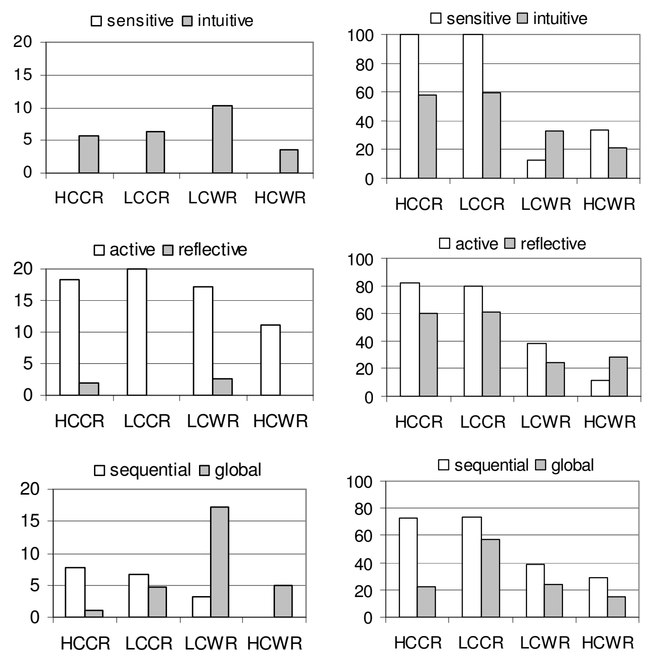

Figure 2. No feedback (left) and only KCR/KR feedback (right) requested by students with certain LS

Interestingly, students with sensitive LS never asked for EF when they knew they gave correct responses, whilst intuitive learners were requesting it in more than 40% of the cases (as it can be seen in Figure 3). Patterns for students with global LS are similar to the

[2] Due to the space limit we will omit the analysis of the corresponding results in the further text; yet we would like to mention that students with intuitive (vs. sensitive) and sequential (vs. global) LS gave many more responses about the usefulness of the feedback. These responses were almost equally distributed between active and reflective students.

[3] Each number is scaled with respect to the total number of such cases (i.e. disregarding feedback as a factor) to address uneven distribution of students across LS.

the intuitive learners in many aspects. However, surprisingly, such learners requested immediate or delayed EF with HCCR more often than others. Students with active LS quite often did not request any feedback at all. They were often not willing to confirm their way of thinking through EF (except in the case of HCWR) in comparison to reflective learners. Reflective and sequential learners share very similar behavior; they rarely avoid (or forget) requesting KR/KCR and in many cases (increasing with decrease of certainty and correctness) requested immediate or elaborated feedback.

Figure 3 shows that sensitive learners are usually satisfied with KR and/or KCR feedback for their correct answers. But they do request the explanations for the incorrect responses, especially for those from which they can clearly obtain the reason why they were incorrect (that was exactly the case in this test). Sequential and global learners are almost identical in requesting immediate EF and are very similar to intuitive learners.

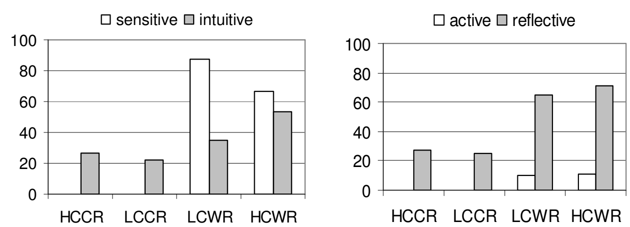

Figure 3. Immediate EF requested only (percentages)

We do not devote separate plots for the remaining possible scenarios with feedback request to save enough space for the discussion of EF effectiveness, yet we highlight the main dependencies: after getting KCR or KR feedback the delayed EF was requested only by intuitive learners mainly after the incorrect responses; sensing students never requested only delayed EF without reading immediate EF; sequential learners requested delayed EF more often than global students, especially after giving wrong answers; active learners requested delayed feedback very frequently after giving the incorrect responses (in 30-50% of the cases), whilst reflective learners requested delayed EF occasionally and mainly for the correct answers.

5.2 Effectiveness of feedback with respect to LS

In general, it is difficult to judge the effectiveness of the feedback in non-classical pretest-posttest experimental settings. However, it is still possible to measure it within each individual quiz when we know a priori which EF can help to answer which (follow-up) question. With some questions in the test we designed immediate EF in such a way that it provided some hints for answering later related questions.

Figure 4 shows what the relative difference in the performance (grades G) of students is, i.e. the ratio of how many times a "hinted" question $k+c$ was answered better than the question k that contained "hinting" feedback by the m students who read that feedback. This can be expressed as:

$$\sum_{i=1}^{m} G_{i,k+c} - G_{i,k} \Big/ m \quad \text{vs.} \quad \sum_{j=1}^{n} G_{j,k+c} - G_{j,k} \Big/ n .$$

So, each number must be interpreted as for how many cases (percent) students answered the "hinted" questions $k+c$ better or worse and on how many percent their grades G increased or decreased for the analyzed pairs of questions (k and $k+c$).

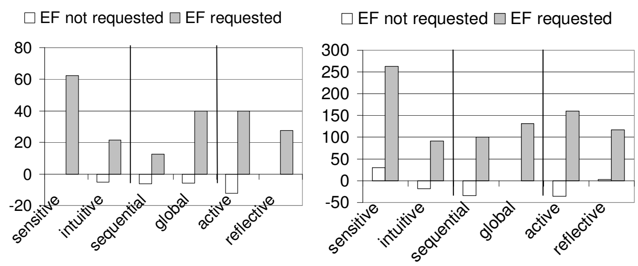

Figure 4. Correctness (left) and grade (right) improvements (percent) for students who got EF vs those who did not

The results show that the corresponding grades (that depend on both correctness and certitude) of the students were significantly higher in those cases where the EF with the hints was examined. This positive effect is due to the facts that on the one hand EF helped some students to answer correctly the later related questions, and on the other hand EF helped to choose low confidence for their answers in case they were not able to understand their "knowledge gap" or "misconception" (i.e. if EF could not help to fix the knowledge problem it was still useful to choose the "appropriate" certainty for the answers) or high certainty when EF indeed helped to fix the problem or confirmed the correctness of students thinking.

However, it can be clearly seen also from the figure that the effectiveness of EF was different for the students with different LS. So, EF was much more effective for sensitive (vs. intuitive) students, and also more effective for global (vs. sequential), and active (vs. reflective) learners.

5.3 Timing aspects

Figure 5 is aimed at demonstrating how much time students with different LS spent studying the feedback and whether students who requested immediate EF were able to answer related questions faster. It can be seen that students with sensitive and active LS tended to spend more time reading EF (and for them this resulted in the highest gain during answering the related questions, i.e. about 100 sec). There was almost no difference in the corresponding processing time for students with sensitive and reflective LS.

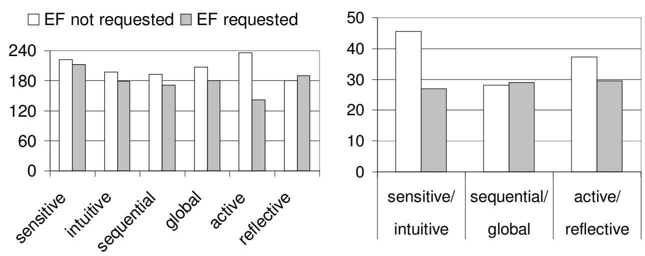

Figure 5. Time (seconds) used for a "hinted" question (left) and - for examining "hinting" feedback (right)

6. Conclusions and Further Work

Designing and authoring feedback and tailoring it to students are important problems of the online learning assessment. We have studied this problem through a series of experiments in the form of online tests organized as part of TU/e courses with traditional in-class lectures and instructions.

In this paper we presented a part of our study focused on the identification of the influence of the LS (with regard to response certitude and response correctness) to the types of feedback the students preferred to request and to the effectiveness of the immediate EF on the performance of the students during the tests.

The results of the assessment data analysis strongly suggest that LS is an important factor that highly influences (1) the feedback preferences (with regard to response certitude and correctness) of students and (2) the effectiveness of EF, i.e. improving students' performance within the test by means of helping (i) to patch the misconceptions or fill the knowledge gap a student may have, and (ii) to estimate the certainty of their responses better.

Concluding from what was stated above, the results obtained in our study provide strong evidence of the benefits and necessity of taking into account LS for providing different types of feedback during the online assessment, and reveal the additional opportunities for feedback personalization.

Our current and ongoing work includes preparation of an extended report that includes a more detailed description of the experimental settings and design, and corresponding results including the effectiveness of EF with regard to "patching" vs. "filling the knowledge gap", "awareness" functions, and organization of further studies with different scenarios of feedback recommendations and personalization (i.e. active promotion and adaptation vs. simply providing a choice of possibilities). The forthcoming experiments will be performed with TU/e students in the spring of 2008.

7. Acknowledgements

This research is partly supported by COMAS Graduate School of the University of Jyväskylä, Finland. We are thankful to the TU/e students who participated in the courses and provided their valuable comments regarding the organization of the online tests and, particularly, the feedback usefulness.

8. References

[1] N. Bajraktarevic, W. Hall, & P. Fullick, "Incorporating learning styles in hypermedia environment: Empirical evaluation", In Proc. of *Workshop on Adaptive Hypermedia and Adaptive WBSs*, Nottingham, 2003, pp. 41–52.
[2] R.M. Felder, and L.K. Silverman, "Learning and teaching styles in engineering education", *Journal of Engineering Education, 78(7),* 1988, pp. 674-681.
[3] J. Hatie, H. Timperley, "The power of feedback", *J. Review of Educational Research, 87* (1), 2007, pp. 81–112.
[4] Kulhavy, R. W., & Stock, W. A. "Feedback in written instruction: The place of response certitude." *Educational Psychology Review, 1*(4), 1989, pp. 279–308.
[5] Mory, E. H. "Feedback research review". In Jonassen, D. (eds.) *Handbook of research on educational communications and technology*, Erlbaum Associates, 2004, pp. 745–783.
[6] Shute, V.J.: *Focus on formative feedback*, Research Report, Retrieved January 15, 2008 from http://www.ets.org/Media/Research/pdf/RR-07-11.pdf, 2007.
[7] E. Vasilyeva, M. Pechenizkiy, S. Puuronen, "The Challenge of Personalization of Feedback to Learning Styles in Web-based Learning Systems", *Proc. of AWELS Workshop at IEEE ICALT 2006 Conf*, 2006, pp. 1143-1144.
[8] E. Vasilyeva, M. Pechenizkiy, T. Gavrilova, S. Puuronen, "Personalization of Immediate Feedback to Learning Styles", In: Proc. of 7th IEEE *Int. Conf. on Advanced Learning Technologies*, 2007, pp. 622-624.
[9] Vasilyeva E., Puuronen S., Pechenizkiy M., Räsänen P. "Feedback adaptation in web-based learning systems", Special Issue of *Int. J. of Continuing Engineering Education and Life-Long Learning 17*(4-5), pp. 337–357, 2007.
[10] Vasilyeva E., Pechenizkiy M., De Bra P. "Adaptation of Feedback in e-learning System at Individual and Group Level", In: *Proc. of PING 2007 Int. Workshop on Personalization in E-Learning Environments at Individual and Group Level* (UM 2007 Conf.), 2007, pp. 49-56.
[11] S.R. Viola, S. Graf, Kinshuk, T. Leo "Analysis of Felder-Silverman Index of Learning Styles by a Data-driven Statistical Approach", *Proc. of the 8th IEEE Int. Symposium on Multimedia*, 2006, pp. 959-964.