Adaptation of Elaborated Feedback in e-Learning

Ekaterina Vasilyeva[1,2], Mykola Pechenizkiy[2], and Paul De Bra[2]

[1] Department of Computer Science and Information Systems, University of Jyväskylä, P.O. Box 35, 40351 Jyväskylä, Finland
[2] Department of Mathematics and Computer Science, Eindhoven University of Technology, P.O. Box 513, 5600MB Eindhoven, the Netherlands
{e.vasilyeva,m.pechenizkiy}@tue.nl, debra@win.tue.nl

Abstract. Design of feedback is a critical issue of online assessment development within Web-based Learning Systems (WBLSs). In our work we demonstrate the possibilities of tailoring the feedback to the students' learning style (LS), certitude in response and its correctness. We observe in the experimental studies that these factors have a significant influence on the feedback preferences of students and the effectiveness of elaborated feedback (EF), i.e. students' performance improvement during the test. These observations helped us to develop a simple EF recommendation approach. Our experimental study shows that (1) many students are eager to follow the recommendations on necessity to read certain EF in the majority of cases; (2) the students more often find the recommended EF to be useful, and (3) the recommended EF helped to answer related questions better.

Keywords: feedback authoring, feedback personalization, online assessment.

1 Introduction

Online assessment is an important component of modern education. Nowadays it is used not only in e-learning for (self-)evaluation, but also tends to replace or complement traditional methods of formal evaluation of the student's performance in blended learning.

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. The existing great variety of the feedback functions and types that the WBLS can actually support make the authoring and design of the feedback in e-learning rather complicated [13]. An important issue is that different types of feedback can have a different effect (positive or negative) on the learning and interaction processes [2]. Badly designed feedback (and/or the lack of feedback) could distract the student from learning, it could provoke the students to stop using the e-learning system or even to drop the course (even in blended learning). Well-designed and adapted or tailored feedback can help the learning process, as we show in this paper.

Feedback personalization becomes a challenging perspective for the development of feedback in the assessment components of WBLSs as it is aimed to provide a student with the feedback that is most suitable and useful for his/her personality, the performed

W. Nejdl et al. (Eds.): AH 2008, LNCS 5149, pp. 235–244, 2008.

task and environment. The development of the personalized feedback requires having the answers to at least the following questions: (1) what can be personalized in the feedback; and (2) to which user or performance characteristics feedback should be personalized. Some answers to these fundamental issues can be found in [13].

In our earlier pilot experiment [14, 15] and more recently a series of real online assessment studies in [11, 12] we have been able to confirm that factors like the students' LS, certitude in response and its correctness, have a significant influence on (1) the feedback preferences of students and (2) the effectiveness of elaborated feedback (EF), i.e. improving students' performance during the test. These encouraging results motivated us to develop a feedback adaptation/recommendation approach for tailoring immediate EF for students' needs.

In this paper we present the result of our two most recent experimental field studies where we tested our approach in real settings during the online assessment of students through multiple-choice quizzes within the (slightly altered) *Moodle* WBLS. In each of the multiple-choice quizzes, students had to select their confidence (certitude) level and were able to receive different (adaptively selected and recommended) kinds of immediate EF for the answered questions.

The analysis of our assessment results and students' behavior demonstrate that (1) many students are eager to follow the recommendations on necessity or usefulness to read certain EF in the majority of cases, (2) after following the recommendations some students were willing to state explicitly whether particular EF indeed was useful to understand the subject matter better or not (and in most of the cases it was found helpful), and (3) recommended EF helped to answer related questions better.

The remainder of this 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 response certitude and correctness, and to students' LS. In Section 4 we first consider a very simple feedback personalization mechanism that displays the most suitable feedback (and/or provides ranked recommendations) to the students according to their knowledge of subject matter, correctness and certitude of response; then we describe the organization of and the results of our experiments. We briefly conclude with a summary and discussion of further research in Section 5.

2 Feedback in Online Assessment in Web Based Learning Systems

Feedback is usually a significant part of the assessment as students need to be informed about their (current and/or overall) performance. Feedback could play different functions in WBLS according to its learning effect: feedback can (1) inform the student about the correctness of his responses, (2) it can "fill the gaps" in the student's knowledge by presenting information the student appears not to know, and (3) it can "patch the student's knowledge" by trying to overcome misconceptions the student may have.

In traditional distance learning (external, but not computer-based learning) feedback has been examined from a number of different perspectives [3]. The studies have shown that students especially wanted detailed feedback and comments. The feedback was expected to provide positive comments on strengths, not vague generalizations. It is recommended that criticism in feedback be constructive and that students should have a chance to respond to comments [3].

In WBLS the main functions of the testing component are to evaluate the students, to give the students information about their performance, to motivate the students, and to focus the students' attention on further interaction with the system. Feedback differs from evaluation, where the main goal is simply to grade and record the result of the testing for the purpose of assessing the student.

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. EF can address the topic and/or the response, discuss the particular errors, provide examples or give gentle guidance [10]. With EF the system presents not only the correct answer, but also 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 [2]. An important issue in designing feedback is that it can draw attention away from the tasks, thereby increasing the time required to execute them. According to Oulasvirta and Saariluoma [9] interrupting messages such as feedback in human-computer interaction influence the extent and type of errors in remembering.

The effectiveness of different types of feedback in WBLS has been experimentally studied by Mandernach [5], who evaluated the educational impact of presenting various levels of computer-based, online feedback (no-feedback, knowledge-of-response, knowledge-of-correct-response, topic-contingent, and response-contingent). The results of this study have shown that the type of computer-based feedback did not have any influence on students' learning, but at the same time the students reported distinct preferences for knowledge-of-response and response-contingent computer-based feedback. This allowed concluding that the students prefer feedback that is direct and clearly addresses the correctness of their response.

Another problem of feedback is the time of its presentation. In [6] Mathan discussed the trade-off between the benefits of immediate and delayed feedback: whereas immediate feedback is more effective during the test, delayed feedback supports better transfer and retention. The advantages and disadvantages of immediate and delayed feedback can change with different learning goals and settings.

All these observations emphasize the necessity of careful design of feedback in WBLS. Our recent studies [11, 12, 13, 14] were aimed at demonstrating that the problems of feedback mentioned above could be partially solved by adaptation of feedback to the tasks and to the characteristics of an individual student. Feedback adaptation in WBLS can provide a student with feedback that is most appropriate for his or her personal characteristics, actual mood, behavior, and attentiveness.

3 Tailoring Feedback to LS, Response Certitude and Correctness

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 design of feedback rather complicated, especially in WBLSs.

Personalization of feedback to the student's personality, performance, and involved contexts (currently performed task(s), environment, etc.) may be a solution for the design of effective feedback in WBLSs. It is essential to know what can be personalized in the feedback and to which characteristics should feedback be personalized. Here, we will focus on the student's LS and response characteristics.

Response certitude (also called response *confidence* or response *certainty*) specifies the student's certainty in the answer and helps in understanding the learning behavior. The traditional scheme of multiple-choice tests evaluation, where the responses are being treated as absolutely correct or absolutely wrong, ignores the obvious situations when the correct response can be the result of a random or an intuitive guess and luck, and an incorrect answer can be given due to a careless mistake or due to some misconceptions the student may have.

Such mistakes are especially crucial in the online assessment, where the evaluation of students' real knowledge and determining students' misconceptions become an even more difficult task for the teacher than in traditional in-class settings. Our results demonstrate that not allowing for discrimination of these situations may diminish the effects of personalized assessment.

The use of feedback in certitude-based assessment in traditional education has been researched for over 30 years; see for example [4, 7, 8] for the detailed reviews. The researchers examined the student's level of confidence in each of the answers and analyzed (1) the differences in performance of students with/without receiving immediate/delayed feedback; (2) how much time the student spent on processing corrective feedback information; (3) efficiency of feedback in confidence based assessment. In spite of the intensive research, the methods and guidelines for designing and implementing feedback in confidence-based assessment remain scarce so far. It is especially important for the design of feedback in WBLSs, where "teachers" could not be as flexible as in the traditional learning.

Our studies [11, 12] demonstrated that knowledge of response certitude together with response correctness allows to determine what kind of feedback is more preferable and more effective for the students, and EF may sufficiently improve the performance of students within the online tests.

Individual LS are one of the important characteristics of the student that characterize the ways in which the student perceives information, acquires knowledge, and communicates with the teacher and with other students. 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 addressed the issue aimed at providing feedback tailored to the LS of the student except our own recent study [11].

4 Immediate Elaborated Feedback (EF) Adaptation

4.1 Generic Feedback Adaptation Framework

Figure 1 presents a generic view of feedback adaptation in a WBLS. The *Student* is identified by the system and associated with his/her profile from the repository. During the interaction with the system the student receives a *Task* (or question) from the *Tasks Repository* and provides an *Answer*. The answer is compared with an expected 'correct' answer to this *Task* by the *Evaluation Module*. The result of the evaluation as well as the user model (information about the user from the *Student Profiles Repository* and *Performance Statistic Repository*) is the input to the *Feedback Adaptation Unit*. Feedback adaptation unit includes a knowledge base containing the adaptation rules that associate user (task, environment) characteristics with certain feedback parameters from the *Feedback Repository*. In the feedback adaptation unit the most convenient form and time of feedback presentation is inferred according to the (long-term and/or short term) characteristics of the student (task, environment). The user model (*Student Profile* and *Performance Statistic Repositories*) is updated with the information obtained by the *Evaluation Module*.

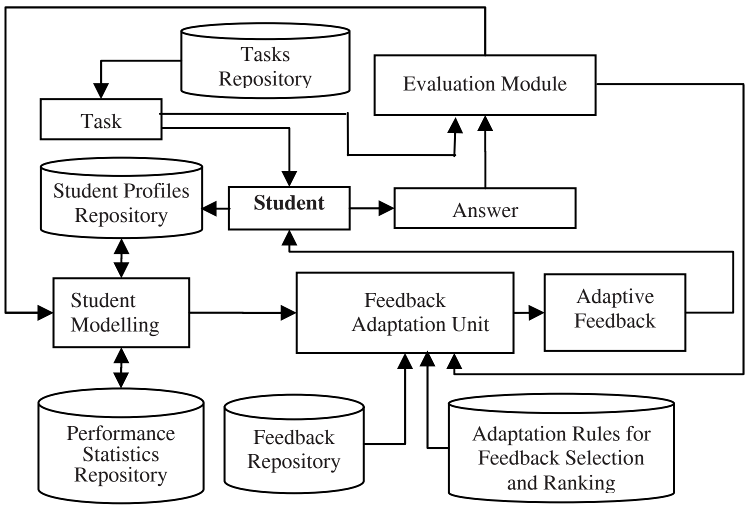

Fig. 1. A generic view on feedback adaptation in a WBLS

4.2 Authoring Adaptive EF and Overall Online Assessment Design

For our study we used a simple user model (UM) that includes information about knowledge of concepts (self-evaluated/estimated by the students[1]), and certitude and correctness of the current response (which constitute two dimensions of possible cases; high-confidence correct responses (HCCRs), high-confidence wrong response (HCWR), low-confidence correct responses (LCCRs), low-confidence wrong response (LCWR)). Other individual characteristics can be added easily of course,

[1] When the tasks/tests are part of a complete e-learning environment where the system can monitor the student's reading and learning activities the system may derive knowledge estimates itself, as is done for instance in typical AHA! applications. (See http://aha.win.tue.nl/ for more details, papers, software download, etc.) In our experiments the tests were done in a mostly traditional learning setting (with lectures and practice sessions), and we simply asked the students to estimate their own knowledge level.

however we tried to focus our study on a particular set of characteristics that allows us to verify our findings from previous experiments as well as to verify the feasibility of the EF adaptation approach and to make some new observations.

The study of EF adaptation was conducted within two tests with the students of a Databases (DB) course (30 students) and an Information Retrieval (IR) course (19 students) at the Eindhoven University of Technology during the fall semester of 2007. Before the tests the students were asked to answer a subset of the most representative (5 for each dimension) questions [16] of Felder-Silverman's index of LS quiz [1].

The tests themselves 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 or translating between English and SQL in the DB test and like reproducing decision tree learning or association rule mining in the machine learning (ML) part of the IR course). It was estimated that the students would need between 2 and 6 minutes for each question depending of its difficulty[2]. Each question was accompanied by the compulsory response confidence question: "Please evaluate your certainty about your answer (it affects your score)".

The general approach for designing the assessment procedures is depicted in Figure 2 below. In DB test, the students were able to get the KR feedback first and than to choose between theory-based and example-based EF or proceed directly to the next question. On the page, where EF was presented, the question and answers were presented with the correct and selected alternative(s) highlighted (KCR feedback). We also asked the students to express their satisfaction about the presented EF. They could optionally answer to the questions whether EF was useful or not.

With the DB experiment we were able to discover EF preference and effectiveness patterns which were the base for the construction of adaptation rules. Thus, it was evident from the analysis of the assessment data that the students requested example-based feedback more often while giving LCWRs, that the main function feedback plays after LCCR responses is "filling the knowledge gap" in the student's knowledge, and that for HCWRs EF should perform the "patching" function helping to overcome the misconceptions a student has. These and other findings resulted in the implementation of 48 adaptation rules for 3 types of EF with 2 additional rules for handling exceptional cases.

With the IR test we conducted the actual EF adaptation study aimed at confirming the feasibility of our approach. The main differences in the IR test is that the most suitable EF is adaptively selected (leaving possibilities of further study of other available EF types) and that KR was not provided separately, but had to be inferred from the EF instead. That is, students had to read the explanations of the EF to understand whether their answer was correct or not. The results of the DB test suggested that it is logical to place KR into EF to increase the overall effect of EF on the learning process during the assessment. This also made our study with the IR test more interesting since we got more EF requests (and EF was now requested for different reasons: extracting KR and learning from EF)[3].

---

[2] Tests were reasonably difficult given the amount of time to pass the test. About 40-70% of questions were answered correctly on average for different tests.

[3] Some further information, a reader may find essential regarding the implementation and organization of the experimental study, can be found in a compact Appendix on http://wwwis.win.tue.nl/~debra/ah08/

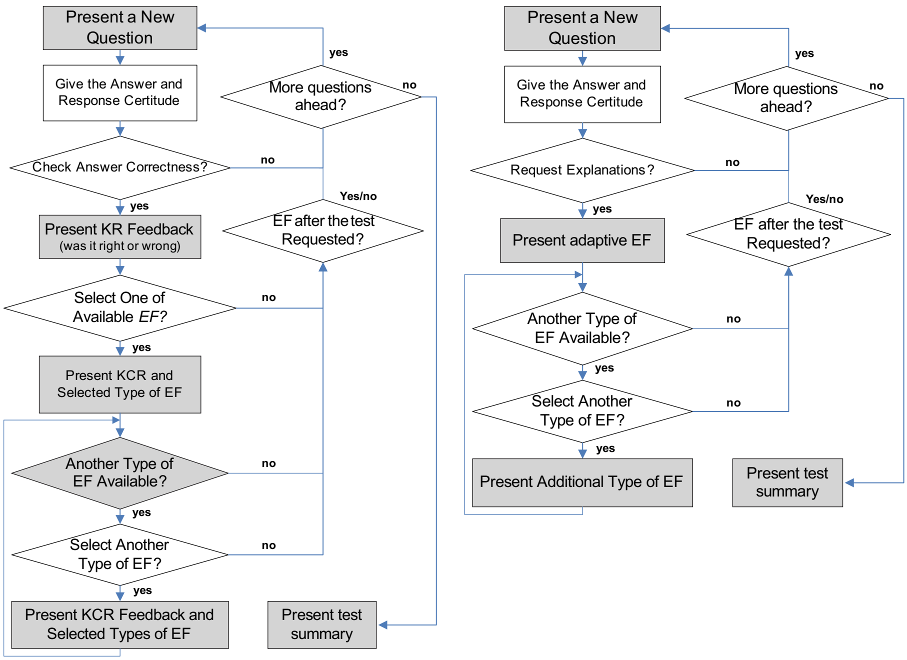

Fig. 2. Assessment process in DB (left) and IR (right) tests

4.3 Obtained Results

We evaluated the effectiveness of adaptive selection and recommendation by means of (1) the number of requests for the EF (only in the cases where the EF was not already directly shown as a result of the adaptation rules), (2) the time students spent for studying the adaptively selected or recommended EF, (3) usefulness of the EF according the students' feedback rating they have provided.[4]

Analysis of EF requests. We analyze only students' responses for which immediate EF was requested, that is 72.6% of all the responses. According to the mechanism of personalization (based on concept certitude, response certitude and correctness) the EF was provided either directly, or a student could request the explanations, selecting it from the available recommendations, which were example-based or theory-based EF. Students received both available types of EF directly in 25% of the cases, one type of EF in 54,1% of the cases, and, no EF in 20,9%. The average time the students spent on the directly received EF was 26 sec when only one type of EF was shown and 34 sec for two types of EF.

When the students received directly one of the available types of EF they could request the second available type (also with a highlighted level of importance) if they wanted. In our experiments the students did this only in 9.4% of the cases and in all

[4] The results of earlier experiments already demonstrated that EF sufficiently *improves* the students` performance during the test. Here we analyze the students' *perception* of the EF usefulness.

these cases that second type of EF was also recommended (which is evidence that our personalization rules were designed correctly). In 70% of such situations they requested example-based explanations after getting the theory-based immediate EF. In such situations the students spent more than 10 seconds for the second EF in 90% of the cases (as well as more than 10 seconds for the first EF they received).

Only in 27% of the cases when one type of EF was adaptively selected, the students spent less than 10 seconds reading the EF. Among those 27 %, the most frequently occurring situation was when the students answered correctly (81%). This means that the students quickly reviewed the explanations and analyzed whether they answered correctly or not. They did not want to spend much time for really reading the EF. For the correct responses the average time of reading EF was 15–25 sec, for incorrect responses it was 30–40 sec.

When no EF was adaptively presented automatically, students could request EF, either following our recommendation for which type of EF might be the most useful for them, or not. The students followed our recommendation in 54% of the cases, but this is actually 75% if we do not take into account the situation when we did not recommend to examine any type of available EF, but they were willing to do this anyway (in order to extract KR from EF). The first type of EF the students selected was theory-based EF in 89% of the cases overall and in 72% of cases by following our recommendations (otherwise, selecting example-based explanations instead). When we recommended studying example-based EF first, the students followed our recommendations in 100% of the cases. Only in 12% cases the students requested the second available EF after reading the first one.

In situations when the EF was selected by the students (and was not automatically shown already), they spent less time for examining it (16 sec on average), equally the same time for theory-based EF and example-based EF. This is also one of the confirmations that personalization worked correctly and those students indeed did not need EF in their situation.

Usefulness of EF. Students were willing to give both positive (73%) and negative (27%) responses regarding the perceived usefulness of the EF; in 68% of the cases for the theory-based EF and 32% for example-based EF. Among the responses about theory-based EF 20% were “not useful”, and regarding example-based EF - 35%. Only in a very few cases, when one type of the EF was directly shown to the students, they found it not to be useful (and requested the second type of the explanations instead). Interestingly, most of the students who found feedback not useful were the students who gave HCWR (this once again confirms that during the test it is more difficult for the students to analyze and to amend their misconceptions than it is to fill a knowledge gap). There was one extreme case, when a student spent more than 2 minutes for studying directly received EF (HCWR) and marked both of types of feedback he got as not useful.

Summary. The results of the study demonstrate the feasibility and effectiveness of EF adaptation. In particular, the students (1) followed our recommendations of the type of EF they could select in most of the cases; (2) only occasionally selected another type of EF when the first was selected automatically; (3) spent more time for the feedback when it was directly shown for them than for the feedback which they had to choose; (4) gave sufficiently more positive than negative responses about the EF that was shown directly or recommended to them. Besides, the analysis of assessment data confirms the generality of EF patterns and corresponding adaptation rules at least within two completely independent experiments.

5 Conclusions and Further Work

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

In this paper we presented a part of our study focused on the EF adaptation by means of adaptive selection and/or recommendation of the appropriate type of EF to students. Adaptation rules that take into account students' response certitude and response correctness, and level of their knowledge of the subject were designed according to the EF effectiveness and students' preference patterns observed during the preceding studies. The results of the assessment data analysis and well as feedback from the students provide enough evidence that our EF adaptation approach is feasible.

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", and "awareness" functions, and organization of further studies with different scenarios of feedback recommendations and personalization. In particular, the results obtained in our studies strongly advocate the benefits and necessity of taking into account LS for providing different types of feedback during the online assessment, and reveal the additional possibilities of feedback personalization [11]. There is no space in this paper for presenting our initial findings regarding the importance of taking the LS into account in feedback adaptation. These findings have be validated through experiments we performed with TU/e students in the spring of 2008, and will be presented in a forthcoming paper.

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

References

1. Felder, R.M., Silverman, L.K.: Learning and teaching styles in engineering education. J. of Engineering Education 78(7), 674–681 (1988)
2. Hatie, J., Timperley, H.: The power of feedback. J. Review of Educational Research 87(1), 81–112 (2007)
3. Hyland, F.: Providing effective support: investigating feedback to distance language learners. Open Learning 16(3), 233–247 (2001)
4. Kulhavy, R.W., Stock, W.A.: Feedback in written instruction: The place of response certitude. Educational Psychology Review 1(4), 279–308 (1989)

5. Mandernach, B.J.: Relative Effectiveness of Computer-based and Human Feedback for Enhancing Student Learning. J. of Educators Online 2(1) (2005)
6. Mathan, S.: Recasting the Feedback Debate: Benefits of Tutoring Error Detection and Correction Skills, PhD Thesis, Carnegie Melon University, Pittsburgh (2003)
7. Mory, E.H.: Feedback research revisited. In: Jonassen, D. (ed.) Handbook of research on educational communications and technology, pp. 745–783. Erlbaum Associates, Mahwah (2004)
8. Mory, E.H.: The use of response certitude in adaptive feedback: Effects on student performance, feedback study time, and efficiency. J. of Educational Computing Research and Development 40(3), 5–20 (1994)
9. Oulasvirta, A., Saariluoma, P.: Long-term working memory and interrupting messages in human-computer interaction. Behaviour and Information Technology 23(1), 53–64 (2004)
10. Shute, V.J.: Focus on formative feedback, Research Report (2007)(Retrieved January 15, 2008) , http://www.ets.org/Media/Research/pdf/RR-07-11.pdf
11. Vasilyeva, E., De Bra, P., Pechenizkiy, M., Puuronen, S.: Tailoring feedback in online assessment: influence of learning styles on the feedback preferences and elaborated feedback effectiveness. In: Proc. of Int. Conf. on Advance Learning Technologies, ICALT 2008. IEEE CS Press, Los Alamitos (to appear, 2008)
12. Vasilyeva, E., Pechenizkiy, M., De Bra, P.: Tailoring of feedback in web-based learning: the role of response certitude in the assessment. In: Proc. of 9th International Conference on Intelligent Tutoring Systems, ITS 2008 (to appear, 2008), http://wwwis.win.tue.nl/~debra/its08/
13. 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), 337–357 (2007)
14. 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.), pp. 49-56 (2007)
15. Vasilyeva, E., Pechenizkiy, M., Gavrilova, T., Puuronen, S.: Personalization of Immediate Feedback to Learning Styles. In: Proc. of 7th IEEE Int. Conf. on Advanced Learning Technologies, pp. 622–624 (2007)
16. Viola, S.R., Graf, S., Kinshuk, Leo, T.: Analysis of Felder-Silverman Index of Learning Styles by a Data-driven Statistical Approach. In: Proc. of the 8th IEEE Int. Symposium on Multimedia, pp. 959–964 (2006)