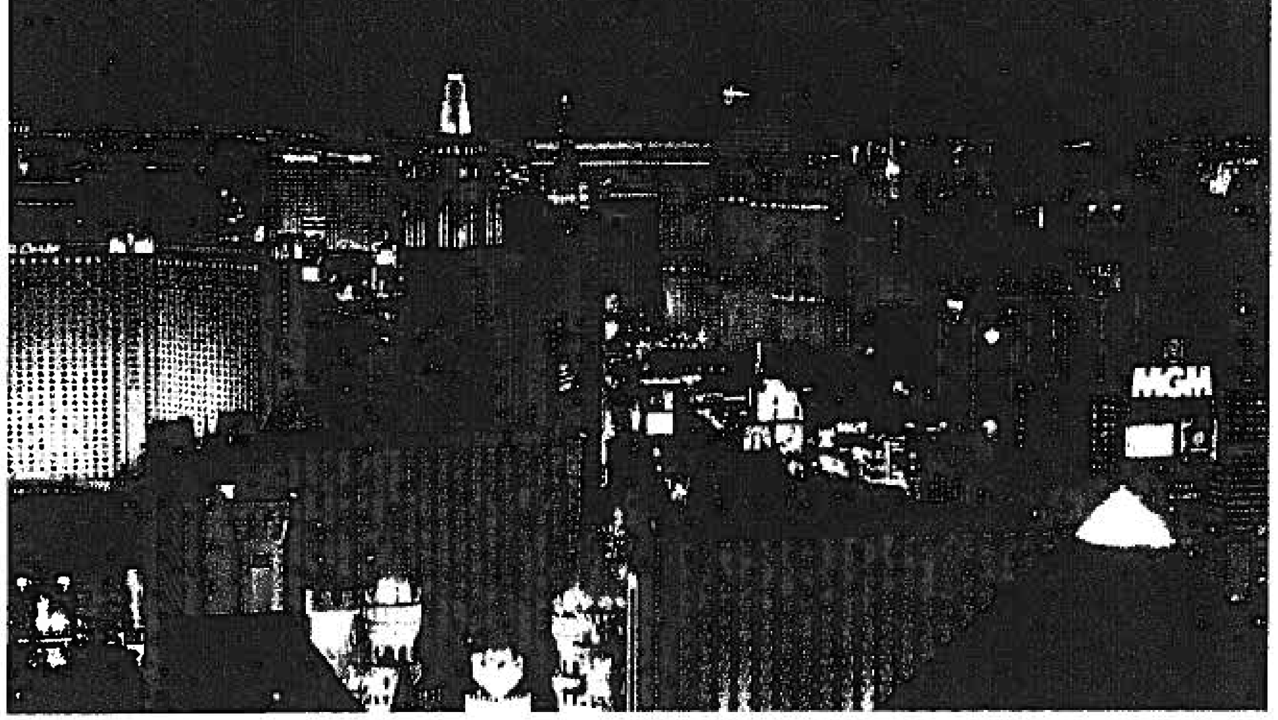
E-Learn 2008

World Conference on E-Learning in Corporate, Government, Healthcare, & Higher Education

November 17-21, 2008 • Las Vegas, Nevada, USA



AACE
Association for the Advancement of Computing in Education
Celebrating 25+ Years of Service to the IT in Education / E-Learning Community

A CONFERENCE OF AACE
www.aace.org

Edited by
Curtis J. Bonk
Mimi Miyoung Lee
Tom Reynolds

Tailoring of Feedback in Online Assessment: Lessons Learnt

Ekaterina Vasilyeva, Paul De Bra, Mykola Pechenizkiy
Department of 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: Feedback plays an important role in the learning process. Particularly, the functions of the elaborated feedback (EF) in online assessment may include assisting students in understanding their mistakes and misconceptions, motivating the student for further learning, suggesting directions for improvement, and others. Many studies suggest that different kinds of feedback can be differently effective (or harmful) depending on numerous factors. During the academic years 2006-2007 and 2007-2008 we have conducted a series of online assessments of students (as integral parts of several bachelor and master courses) and have studied the possibilities of tailoring the feedback (presented to a student as a result of his/her response to questions of an online test) taking into account the individual learning styles (LS), certitude in a response and correctness of this response. In this paper we summarize our major findings from the assessments data and lessons learnt from the organization of online assessments with tailored feedback. We also discuss the advantages and limitations of our studies, and outline the prominent directions for further research.

Introduction

Lack of interaction between students and teachers is one of the main problems in web-based learning systems (WBLS) (Mory, 2004). During the learning process the student performs a number of (inter)actions where feedback is crucial, for example in assessments or in task solving. In the traditional in-class learning the student could get the feedback from the teacher almost at any time and feedback can be given in different forms, depending on the student's personality, the learning task and the situation, whereas in distance learning the student receives feedback that is limited by the availability of certain information in the WBLS and the functionality of the system.

Feedback in WBLS performs all the functions that it performs in traditional learning. For example, during the assessment feedback provides information about the performance, it assists the student in understanding his mistakes and misconceptions, feedback motivates the student for further learning and can suggest some directions for improvement. Feedback is considered as one of the most important sources of information for assistance in students' knowledge restructuring (Mason & Bruning, 2001; Gouli *et al.*, 2004).

The increasing number of users of WBLS as well as the existence of different types of feedback and the ways of its presentation emphasize the necessity of feedback personalization. The same feedback could have different influence on different students. Tailoring and personalizing the feedback for students offers possibilities to deliver feedback that is the most effective for the student and is the most appropriate for the student's expertise and cognitive abilities in general and, in particular, to the performance, current mood and attentiveness.

Although there have been studies of different types of feedback and their influence on the learning and interaction in WBLS (Mory, 2004; Shute, 2008), feedback personalization remains an almost totally neglected area, even though its importance has been widely acknowledged in the educational research (Vasilyeva *et al.*, 2007a). The results of the existing research on the effectiveness of feedback (and its different types for different types of learners) in WBLS remain fragmented and unimplemented in the design and authoring of feedback in existing e-learning applications. There is no clarity about to which characteristics of the student feedback should be personalized and how personalization of feedback and authoring of the personalized feedback in WBLS should work in practice.

In this paper we summarize our experience and lessons learnt from a series of studies organized during the academic years 2006-2007 and 2007-2008 at Eindhoven University of Technology, the Netherlands. Particularly, we discuss the possibilities of tailoring the feedback (presented to a student as a result of his/her response to a question of an online test) taking into account the individual learning styles (LS), certitude in a response and correctness of this response. First, we give an overview of the main functions of feedback we studied, we discuss what the students' preferences were, and report on the lessons we have learnt. Then we describe the basic ideas behind our studies of tailoring elaborated feedback with respect to response certitude and correctness and the students' LS. We gradually move to the issues of elaborated feedback adaptation, presenting the theoretical background and empirical evidence as motivation for designing particular EF adaptation rules and discussing the results of their evaluation in subsequent studies. We conclude our paper with a brief summary and directions of further research.

Functions of Feedback and Student Preferences

Feedback can play different functions in WBLSs according to its learning effect: feedback can (1) inform the student about the 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 (Vasilyeva *et al.*, 2008a).

The functions of the feedback imply the complexity of information that can be presented in immediate feedback: verification and elaborated feedback (EF) (Kulhavy & Stock, 1989). 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 (Shute, 2008). Elaboration can address the topic and/or the response, discuss the particular errors, provide examples or give some gentle guidance (Mory, 2004). 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 they can thus be differently effective in terms of learning and interaction and (when used inappropriately) can even be disturbing or annoying to the student and have negative influence on the learning and interaction processes (Hatie & Timperley, 2007). These observations emphasize the necessity of careful design of feedback in WBLSs.

Knowing Response Certitude beside the Correctness

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.

We conducted three different online multiple-choice tests (partial exam and mandatory individual exercises for three different courses). Each question (answered strictly one after another) in a test was accompanied by the compulsory question about response confidence that affected the grade[1] (2 points for a HCCR, 1 point for a LCCR, -1 point for a HCWR, and, 0 for a LCWR). After giving the response (with a specified certainty) the student could either go directly to the next question or request immediate (KCR/KR/EF) and delayed (EF) feedback. Students could optionally indicate with each question whether EF was useful or not. The analysis of the results of these three tests is reported in (Vasilyeva *et al.*, 2008a, Vasilyeva *et al.*, 2008b, Vasilyeva *et al.*, 2008c). Here, we present the main lessons learnt from these experiments.

Most of the students were eager to get KCR and/or KR. There were usually only a few students who did not check their answers for most of the questions in each test. In two tests we analyzed whether the students were eager to get only KR feedback or also KCR+EF by separating these possibilities. In this scenario more students requested only the KR, without KCR+EF. In another test where KCR/KR could only be obtained from studying the EF students did request the EF to extract the KCR/KR. After students knew whether their answer was correct they tended not to request any EF (and this was independent of the response certitude). For incorrect responses, especially for the HCWRs, the frequency of "ignoring" the EF was lowest.

Consequently, hiding (embedding) KCR (or KR) in EF is beneficial for students. This allows EF to gain the distinct role of an additional instructional element, which enhances its power. Hiding KCR in EF forces the students to examine EF for acquiring KCR information. It is beneficial since students are paying more attention to study EF and could either confirm the correctness of their way of thinking, or learn new bits of information (fill the knowledge gap) or understand a particular misconception and hopefully "patch" it.

Response certitude together with response correctness allows to more clearly determine what role EF should play in a certain situation. For the uncertain responses (LCCR and LCWR) feedback mainly "fills the knowledge gap". It acts as an additional instruction and provides the student with some information he has not learnt well or forgotten. Most of the HCWR responses are given due to misconceptions (or sometimes simply due to inattentiveness) of the students. In such cases EF assumes function of "patching" the student's knowledge and helps the student to overcome the misconceptions. For HCCR responses the main role of feedback is to inform about the results of the response, i.e. to confirm their correctness. Consequently, it may be enough to provide only KR and/or KCR.

Overall, the results strongly suggested that EF may sufficiently improve the performance of students within the online tests. Knowing that the students are able to estimate the certainty of their responses fairly well and that knowledge of response certitude together with response correctness allows determining what kind of feedback is more preferable and more effective for the students, we concentrated on defining rules (summarized in Table 1; suggested actions are divided into two categories, *A* that defines whether EF should be shown directly to the learner, and *B* that defines how strongly EF should be recommended; of course if only one type of EF is available and it is already shown directly then this EF will not be recommended in even if a rule may suggest this) for tailoring EF to response correctness and response certitude. The evaluation of these rules was reported in (Vasilyeva *et al*, 2008c). The results of the experiment demonstrated feasibility of tailoring feedback to the response correctness and response certitude.

Table 1. Effect of Response Correctness and Response Certitude on EF Tailoring

Response (CC)	Theoretical background according to (Kulhavy *et al*, 1979; Kulhavy & Stock, 1989; Mory, 2004)	Empirical evidence KCR, KR, EF	Suggested action *A* - (not) show directly; *B* - recommendation
HCCR	- Students spend the shortest time for studying the EF; - feedback receives only cursory attention and no EF is needed.	- students requested KCR or KR for 65-85% of their responses; - the number of EF requests was low (8-17% of all HCCR; 10-15% of all EF requests).	**CC1** { *A*: not shown *B*: not recommended }

HCWR	- require the most time for the feedback processing;
	- EF has the greatest corrective effect.
		- students requested KCR or KR more often than for HCCR (for 75-90% of responses);
		- The highest percentage of the EF requests (43%-62% of all HCWR; 30-40% of all EF requests).
			CC2 {
			A: shown;
			B: strongly recommended
			}

LCCR	- caused by student's low level of understanding of the task and/or learning material; - EF serves as an additional instruction; - time for reviewing EF is shorter than for HCWR, but longer than for HCCR;	- students requested KCR or KR feedback more frequently than for HCCR or HCWR (for 80-95% of responses); - students reviewed EF in (15-30% of all LCCR; 10-20% of all EF requests).	**CC3** { *A*: not shown; *B*: mildly recommended }
LCWR	- same as for LCCR	- students requested KCR or KR feedback more frequently than for HCCR or HCWR (for 85-95% of responses); - students reviewed EF more frequently when giving LCWRs (16%-40% of all LCWR; 20-30% of all EF requests).	**CC4**{ *A*: shown; *B*: mildly recommended }

Effect of Learning Style on the Feedback Preferences

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).

We analyzed several dimensions of LS and stated the first hypothesis about the influence of individual LS onto feedback presentation parameters in (Vasilyeva *et al*, 2006).

The pilot study to explore the interrelation between several types of immediate feedback presentation

[1] H(L)CC(W)R – high (low) confidence correct (wrong) response.

(KCR/KR and EF) and LS of users was reported in (Vasilyeva *et al*, 2007c). In our recent series of online assessment studies we analyzed possibilities of feedback tailoring to LS more in detail.

In our experiments we used the Index of Learning Styles Quiz designed by Felder and Silverman (Felder & Silverman, 1989). This quiz allows determining the student's LS according to four dimensions: active/reflective, sensing/intuitive, visual/verbal, and, sequential/global. In our experiments we used only 3 dimensions excluding visual/verbal as most of the EF provided was verbal and did not contain many illustrations. The obtained results (Vasilyeva *et al.*, 2007c; Vasilyeva *et al.*, 2008b) as well as the description of the LS dimensions allowed us to later decrease the number of the LS dimensions which are important for EF personalization to just active/reflective and sensing/intuitive.

According to (Felder & Silverman, 1989) active learners tends to "try things out and see how they work", while reflective learners prefer to "think it through first". The results of our pilot study (Vasilyeva *et al.*, 2007c) demonstrated the tendency that users with the bias towards reflective LS perform better (score and time) in the tests where EF is provided. On the other hand the active learners (and those who have tendency to be active) were better (score and time) in the tests where KCR was presented. We have proposed a hypothesis that active learners most likely would prefer to get KCR, while EF could be more helpful for reflective learners. In a more recent experiment (Vasilyeva *et al.*, 2008b) the students with active LS quite often (in 12,5% of the cases) did not request any type feedback at all. They were often (in 80% of the correct responses) not willing to confirm their way of thinking through EF after getting KCR (unless it was HCWR) in comparison with reflective learners. Reflective learners requested KR/KCR in 95% of the cases and in many cases (40-80%; increasing with decrease of certainty and correctness) requested EF.

The sensing/intuitive dimension of LS is described as follows: "Sensors <students with sensing LS> tend to be patient with details and good at memorizing facts and doing hands-on (laboratory) work; intuitors <students with intuitive LS> may be better at grasping new concepts and are often more comfortable than sensors with abstractions and mathematical formulations." (Felder & Silverman, 1989). Following this description we initially formulated a hypothesis that sensing learners most likely would prefer to get KCR, while for the intuitive students EF is necessary. In our pilot study (Vasilyeva *et al.*, 2007c) the performance of the sensing students was in average better in the tests with KCR, while the intuitive students performed better in the tests where EF was provided. In a more recent experiment (Vasilyeva *et al.*, 2008b) sensing students were usually satisfied with KR and/or KCR for their correct answers (98% of the cases), but they did request the explanations for the incorrect responses (in 70% of the cases). Intuitive learners did not request any type of feedback in 7% of the cases. However, in case they pressed "Check Answer" button, they requested explanations for both correct (in 40%) and incorrect responses (in 75%); more frequently for the incorrect responses.

These experimental findings together with LS theory were taken into consideration for the development of further EF adaptation rules (with regard to the LS), which we summarize in Table 2 (suggested actions are divided into two categories as in Table 1; *A* defines whether EF should be shown directly to the learner, and ***B*** specifies here, which type of EF should be directly shown/recommended).

Tailoring Elaborated Feedback with Adaptation and Recommendation

The results described in the previous section naturally suggest combining both types of characteristics (response-certitude and student personality-related) for designing a more general EF personalization strategy. We extended our EF personalization rules (originally based on response correctness and response certitude) with two LS dimension – active/reflective and sensing/intuitive. The obtained rules as well as the preliminary designed rules (described in Table 1 and Table 2) were used for the design of the rules summarized in Table 3 (*, **, and *** correspond to light, mild, and strong recommendations of particular type of EF). For example, for the case when a student with active and sensing LS gives LCWR example-based EF would be shown directly to the student (according to the preliminary rules *CC4A, LS3A, LS2B,* and, *LS3B*) and theory-based EF would not be recommended (according to the preliminary rules *LS2B, LS3B*).

We have implemented the obtained EF personalization rules for our recent experiment. The results of the study (one test with 73 students) demonstrated 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 to them than for the feedback which they had to request (even though that took just one mouse click); (4) gave sufficiently more positive than negative ratings to 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.

Table 2. Effect of Individual LS on EF Tailoring

LS	Theory on LS (Felder & Silverman, 1989)	Empirical evidence (Vasilyeva *et al.*, 2007c, Vasilyeva *et al.*, 2008b)	Suggested action *A* - (not) show directly; *B* – type of EF to show/recommend
Reflec-tive	- prefer to think about the leaning materials quietly first; - need to have more time for processing new information.	- performed better (score and time) when EF (not KCR) was provided; - requested KR/KCR in 95% of the cases; - requested EF in 40-80% increasing with decrease of certainty and correctness.	**LS1** { *A:* shown}
Active	- tend to retain and understand information best by doing something active with it—discussing, applying, or explaining it to others.	- performed better when KCR (but no EF) was presented; - in 12,5% of the cases did not request any type of feedback at all; - in 80% of the cases did not request EF after getting KCR.	**LS2** { *A:* shown for the HCWR or LCWR; **B:** Example-based EF is preferable}
Sen-sing	- learn first concrete and practical information; - oriented toward facts; - understand ideas and theories better via practical applications; - focus on details and may ignore "the big picture".	- performed better when KCR (but no EF) was provided; - always requested KR and/or KCR (100% of the cases); - requested EF only for the incorrect responses (in 70% of the cases) after getting KCR.	**LS3** { *A:* shown for the HCWR or LCWR; *B:* Example-based EF is preferable}
Intui-tive	- prefer conceptual and innovative information oriented toward theories and meanings; - dislike repetition; - act on their hunches and sometimes missing a key part of learning materials.	- performed better when EF (not KCR) was provided; - did not request any type of feedback in 7% of the cases; - requested EF for both correct (in 40%) and incorrect responses (in 75%) after getting KCR.	**LS4** { *A:* shown for the HCWR or LCWR; *B:* Theory-based EF is preferable}

We analyzed which of the rules appeared to be reasonable and summarized the results in Table 4 which resembles the structure of the table above. Each cell contains EF utility coefficient. Utility of directly shown EF for each case (i.e. each cell in the table) is calculated as the difference between the number of times directly shown EF was read and the number of times it was just scanned (i.e. spent less than 7 seconds on):

$$CoefShown = EF_{read}^{shown} - EF_{scanned}^{shown} \tag{1}$$

Thus, if *CoefShown* is close to 1 then students *read* EF in 100% of cases and this provides the evidence that the suitable EF was presented. If *CoefShown* is close to 0, it means that students *scanned* EF in a half of cases. Consequently, if *CoefShown* is close to -1, it means that students *scanned* directly shown EF almost in all cases. Thus, if *CoefShown* > 0.5 than personalization rules could be considered as reasonably good (corresponds to more than 75% of success); in other cases additional examination of the rules is needed.

The utility of recommended EF is also calculated separately for each case that is equal to the sum (divided by the total number of cases in the cell) of differences of "positive" and "negative" behavior of student for read, scanned and unseen EF:

$$CoefRcom = \frac{(EF_{not\_seen}^{notrecom} - EF_{not\_seen}^{recom}) + (EF_{scanned}^{recom} - EF_{scanned}^{notrecom}) + (EF_{read}^{recom} - EF_{read}^{notrecom})}{EF_{not\_seen} + EF_{scanned} + EF_{read}}. \tag{2}$$

Table 3. Resulting EF adaptation rules

LS	HCCR *RuleCC1*		LCCR *RuleCC3*		LCWR *RuleCC4*		HCWR *RuleCC2*	
	Show:	*Recommend:*	*Show:*	*Recommend:*	*Show:*	*Recommend:*	*Show:*	*Recommend:*
No L/S	-	-	-	Theory (*)	Theory	Example (*)	Theory	Example

	CC1A	*CC1B*	*CC3A*	Example (*) *CC3B*	*CC4A*	*CC4B*	*CC2A*	(***) *CC2B*
Active/ Balanced ***RuleLS2***	- *CC1A,* *LS2A*	- *CC1B*	- *CC3A,* *LS2A*	Example(**) *CC3B,* *LS2B*	Example *CC4A,* *LS2B*	Theory(*) *CC4B*	Example *CC2A,* *LS2B*	Theory (**) *CC2B*
Reflective/ Balanced ***RuleLS1***	- *CC1A*	Theory (*) *CC1B,* *LS1A*	Theory *LS1A*	Example(*) *CC3B*	Theory *CC4A,* *LS1A*	Example (**) *CC4B,* *LS1A*	Theory *CC2A, LS1A*	Example(***) *CC2B*
Balanced/ Sensing ***RuleLS3***	- *CC1A,* *LS3A*	- *CC1B*	- *CC3A,* *LS3A*	Example(**) *CC3B,* *LS3B*	Example *CC4A* *LS3A,* *LS3B*	- *LS3B*	Example *CC2A,* *LS3A,* *LS3B*	Theory(**) *CC2B*
Balanced/ Intuitive ***RuleLS4***	- *CC1A,* *LS4A*	- *CC1B*	- CC3A, LS4A	Theory(**) *CC3B,* *LS4B*	Theory *CC4A, LS4A,* *LS4B*	- *LS4B*	Theory *CC2A, LS4A,* *LS4B*	Example(**) *CC2B*
Active/ Sensing ***Rules:*** ***LS3+LS2***	- *CC1A,* *LS3A, LS2A*	- *CC1B*	- *CC3A,* *LS3A, LS2A*	Example(**) *CC3B,* *LS2B,* *LS3B*	Example *CC4A,* *LS3A,* *LS2B,* *LS3B*	- *LS2B,* *LS3B*	Example *CC2A, LS3A,* *LS2B,* *LS3B*	Theory(**) *CC2B*
Active/ Intuitive ***Rules:*** ***LS4+LS2***	- *CC1A,* *LS4A, LS2A*	- *CC1B*	- *CC3A,* *LS4A, LS2A*	Theory (**), Example (*) *CC3B,* *LS4B,* *LS2B*	Theory *CC4A, LS4A,* *LS4B*	Example (*) LS2B	Theory *CC2A, LS4A,* *LS4B*	Example (**) *CC2B,* *LS2B*
Reflective/ Sensing ***Rules:*** ***LS3+LS1***	- *CC1A,* *LS3A*	Example (*) *LS1A, LS3B*	- *CC3A,* *LS3A*	Example (**) Theory (*) *CC3B,* *LS3B*	Example *CC4A,* *LS3A,* *LS3B*	Theory (**) *CC4B,* *LS1A*	Example *CC2A,* *LS3A,* *LS3B*	Theory (***) *CC2B,* *LS1A*
Reflective/ Intuitive ***Rules:*** ***LS3+LS1***	- *CC1A,* *LS4A*	Theory (*) *LS1A*	Theory *LS1A*	Example (*) *CC3B,* *LS4B*	Theory *CC4A, LS4A,* *LS4B,* *LS1A*	- *LS4B*	Theory *CC2A,* *LS4A,* *LS4B,* *LS1A*	Example (***) *CC2B,* *LS1A*
Balanced/ Balanced	- *CC1A*	- *CC1B*	- *CC3A*	Theory (*) Example (*) *CC3B*	Theory *CC4A*	Example (*) *CC4B*	Theory *CC2A*	Example (***) *CC2B*

The intuition is that the more often students read or at least scanned the EF when it was recommended and the less often students requested EF when it was not recommended the better. *CoefRcom* $\in$ [-1;1]; the higher *CoefRcom* is the better people followed EF recommendations.

It may be valuable to distinguish with which strengths EF was recommended. We scale the coefficient of recommendation by the strength of recommendation *RcomStr* used in certain rules. We set $RcomStr_0 = 0$ when EF was not recommended; $RcomStr_1 = 0.3$ when EF was recommended with one star; $RcomStr_2 = 0.6$ when EF was recommended with two stars; and $RcomStr_3 = 1$ when EF was recommended with three stars. So, normalized coefficients are calculated similarly to (2), but $EF^{recom}_{\{not\_seen, scanned, read\}}$ counts are affected by corresponding *RcomStr* value, that is e.g. $EF^{recom}_{read} = \sum_{RcomStr=0.3}^{1} EF^{recom}_{read}$, while $EF^{notrecom}_{read}$ counts remain unaffected.

The coefficients scaled according to the recommendation strengths are more illustrative in case two types of EF (when not directly shown) were recommended with different strengths. So we can find out whether the students preferred to request EF that was recommended with higher strength.

In Table 4 the normalized utility coefficients of recommended EF are given in the brackets where applicable (i.e., if the recommendation strength is equal to 0, the coefficient is also equal to 0). Dash denotes the complete absence of corresponding cases in the test; n/a denotes the cases when no type of EF is directly shown to the student according to the rule.

The rule for recommendation can be considered reasonable when its utility coefficients are positive. But it is important to take into account for which situations the utility is calculated – for cases where EF was shown directly and another type of EF was recommended, or, for situations, when no EF was shown and students had a choice between 2 types of EF recommended with certain strength.

In general, we can conclude that EF personalization rules for direct presentation of EF were designed reasonably well as utility of directly shown feedback is in most of the cases is close to 1. There are two rules that should be checked because of low utility coefficients of "show" rules: (1) for HCWR responses of intuitive

learners when theory-based EF was directly shown (the choice of the type of EF to be shown was based on *LS4B*), and, (2) for LCWR responses of active/sensing students when example-based EF was presented (the choice of the type of EF to be presented was based on *LS3B*). However, these problems can be explained by insignificant number of corresponding cases being analyzed.

The EF recommendation rules were also followed by the students quite well in case of HCCR and LCCR. But for the incorrect responses we can observe the negative values in places. This is due to the fact that in many situations students did not request an additional (though recommended) type of EF after examining the directly shown type of EF (likely because the subject matter was clear already after reading just one type of EF).

Table 4. Utility of the EF adaptation rules

LS	HCCR		LCCR		LCWR		HCWR	
	Show:	*Recommend:*	*Show:*	*Recommend:*	*Show:*	*Recommend:*	*Show:*	*Recommend:*
Active/ Balanced	n/a	0.06	-	-	1	-1 (-0.3)	1	-0.5 (0.3)
Reflective/ Balanced	n/a	0.82 (0.19)	1	-1 (-0.3)	1	-1 (0.6)	1	-1 (-1)
Balanced/ Sensing	n/a	0.1	n/a	1 (0.3)	0.71	0.86 (0)	1	-0.78 (-0.47)
Balanced/ Intuitive	n/a	-0.09	n/a	0 (-0.3)	0.75	-0.5 (-0.15)	-1	1 (1)
Active/ Sensing	n/a	0.13	n/a	1 (0.6)	0.43	1(0)	1	-1 (-0.6)
Active/ Intuitive	n/a	-0.06	n/a	0.75 (0.9)	1	0.57 (0)	1	0.56 (0)
Reflective/ Sensing	n/a	0.07	n/a	0	1	-0.2 (-0.12)	1	0 (0)
Reflective/ Intuitive	n/a	0.5 (0.3)	1	0	-	-	1	0 (0)
Balanced/ Balanced	n/a	0.16	n/a	0 (0)	0.87	-0.48 (-0.14)	0.85	-0.7 (-0.7)

Conclusions

Designing and authoring tailored feedback, adapting it to the students' needs and personality is an important and challenging problem in WBLS, and particularly the online learning assessment. We have approached this problem from different angles, trying to identify key research and practical issues, designing possibilities for authoring and adapting (tailored) feedback in multiple-choice tests, and conducting a series of studies in which students participated for a partial grade for a course (8 tests for 4 different courses).

The main lessons learnt from our studies are the following; (1) personalization of feedback is feasible in online assessment with multiple-choice testing; (2) hiding KR and KCR feedback in EF is helpful for the students (although they may not always like this) and results in a better learning curve; (3) response certitude assessment allows for better understanding of students' mistakes (gaps in knowledge, misconceptions, and inaccuracies) and can thus be utilized for feedback adaptation; and (4) knowledge of student's individual LS allows to personalize EF selecting the most appropriate form of presentation.

In this paper we were particularly focusing on the estimation of the utility of the designed personalization/recommendation rules for EF adaptation. Our study showed that the majority of adaptation rules have high utility. We also identified a set of rules, which require further inspection and additional experimentation. In general, further research is needed to extend the designed prototype of feedback adaptation to be implemented within WBLS and to increase the number of personalization/recommendation rules supported. We see further experimental research as one of the main directions of the future work.

Particularly, experimentation with other characteristics than response certitude and individual LS is needed in order to analyze their suitability and influence on the effectiveness of (different types of) feedback. Some of those characteristics are listed in (Vasilyeva *et al.*, 2007a, Vasilyeva *et al.*, 2007b). Taking into account combinations of different characteristics can increase the accuracy of feedback personalization.

It should be noted that we had a rather limited number of subjects for the uncontrolled experiments (from 20 to 70) where it was not always feasible to identify and eliminate all the possible factors that may have an effect on the studied phenomena. Therefore, further studies with a larger number of students and with a more controlled settings are needed for elaboration of the personalization rules.

It was relatively easy to introduce feedback adaptation mechanisms into the Moodle LMS. Integration of adaptation mechanisms with other popular LMS like Sakai and Claroline is among our future work directions.

References

Felder, R.M. and Silverman, L.K. (1988). Learning and teaching styles in engineering education, *Journal of Engineering Education*, 78(7) (pp. 674-681).

Gouli, E., Gogoulou, A., Papanikolaou, K., and Grigoriadou, M. (2006). An Adaptive Feedback Framework to Support Reflection, Guiding and Tutoring. In G.D. Magoulasand S.Y. Chen (Eds.) *Advances in Web-based Education: Personalized Learning Environments*, Hershey, PA: Idea Group Inc (pp. 178-202).

Hatie, J. & Timperley, H. (2007). The power of feedback, J. *Review of Educational Research*, 87 (1) (pp. 81–112).

Kulhavy, R. W., & Stock, W. A. (1989). Feedback in written instruction: The place of response certitude. *Educational Psychology Review*, 1(4) (pp. 279–308).

Kulhavy, R. W., Yekovich, F. R., and Dyer, J. W. (1979). Feedback and content Felder, R.M. and Silverman, L.K. 1988. Learning and teaching styles in engineering education", Journal of Engineering Education, 78(7) (pp. 674-681).

Mason, B. J. & Bruning, R. (2001). Providing feedback in computer-based instruction: What the research tells us, http://dwb.unl.edu/Edit/MB/MasonBruning.html.

Mory, E. H. (2004). Feedback research revisited, In Jonassen, D. (eds.) *Handbook of research on educational communications and technology* (pp. 745–783). Mahwah, NJ: Erlbaum Associates.

Shute, V.J. (2008). Focus on formative feedback, *Review of Educational Research*, 78(1) (pp. 153-189).

Vasilyeva, E., Pechenizkiy, M., De Bra, P. (2008a). Tailoring of feedback in web-based learning: the role of response certitude in the assessment, In: Proc. of *Int. Conf. on Intelligent Tutoring Systems ITS'08*, LNCS 5091, Springer (pp. 771-773).

Vasilyeva, E., De Bra, P., Pechenizkiy, M., Puuronen, S. (2008b). Tailoring feedback in online assessment: influence of learning styles on the feedback preferences and elaborated feedback effectiveness. In: Proc. of *IEEE ICALT'08 Conf.* (pp. 834-838).

Vasilyeva, E., Pechenizkiy, M., De Bra, P. (2008c). Adaptation of Elaborated Feedback in e-Learning. In: Proc. of 5th *Int. Conf. on Adaptive Hypermedia and Adaptive Web-Based Systems AH'08*, LNCS 5149, Springer (pp. 235-244).

Vasilyeva, E., De Bra, P., Pechenizkiy, M. (2008d). Immediate Elaborated Feedback Personalization in Online Assessment, In: Proc. of *European Conf. on Technology Enhanced Learning EC-TEL'08*, LNCS 5192, Springer (pp. 449-460).

Vasilyeva, E., Puuronen, S., Pechenizkiy, M., Räsänen, P. (2007a). 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).

Vasilyeva, E., Pechenizkiy, M., De Bra, P. (2007b). Adaptation of Feedback in e-learning System at Individual and Group Level, In: Proc. *Personalization in E-Learning Environments at Individual and Group Level* (UM'07) (pp. 49-56).

Vasilyeva, E., Pechenizkiy, M., Gavrilova, T., Puuronen, S. (2007c). Personalization of Immediate Feedback to Learning Styles, In: Proc. of 7th IEEE *Int. Conf. on Advanced Learning Technologies*, IEEE CS Press (pp. 622-624)

Vasilyeva, E., Pechenizkiy, M., and Puuronen, S. (2006). The Challenge of Personalization of Feedback to Learning Styles in Web-based Learning Systems, AWELS Workshop, Proc. of *IEEE ICALT'06 Conf.*, IEEE CS Press (pp. 1143-1144).