# Children's Interactions with Inspectable and Negotiated Learner Models

Alice Kerly and Susan Bull

Electronic, Electrical and Computer Engineering, University of Birmingham,
Edgbaston, Birmingham, B15 2TT, UK
{alk584,s.bull}@bham.ac.uk

**Abstract.** The Learner Model of an Intelligent Tutoring System (ITS) may be made visible (opened) to its users. An Open Learner Model (OLM) may also become a learning resource in its own right, independently of an ITS. OLMs offer potential for learner reflection and support to metacognitive skills such as self-assessment, in addition to improving learner model accuracy. This paper describes an evaluation of an inspectable and a negotiated OLM (one that can be jointly maintained through student-system discussion) in terms of facilitating self-assessment accuracy and modification of model contents. Both inspectable and negotiated models offered significant support to users in increasing the accuracy of self-assessments, and reducing the number and magnitude of discrepancies between system and user beliefs about the user's knowledge. Negotiation of the model demonstrated further significant improvements.

## 1 Introduction

Intelligent Tutoring Systems (ITS) routinely employ a learner model in order to provide tutoring and interaction tailored to the needs of the individual student. Conventionally this model has only been for the use of the system, and hidden from the learner. Open Learner Modelling argues that making the contents of the model visible for inspection by the student may bring opportunities for developing skills in reflection, metacognition and deep learning, e.g. [1], [2], [3], [4], [5]. Open Learner Models (OLM) may also allow the student and system to engage in a process of negotiation about the contents of the model, potentially enhancing learner reflection and model accuracy. Such negotiated learner models (e.g. [1], [2]) involve a collaborative construction and maintenance of the learner model. By requiring learners to discuss their beliefs about their knowledge with the system, argue against the system's assessment where they disagree or provide evidence for their own beliefs, it is suggested that learner reflection may be increased [1], [2]. This negotiation may also improve the accuracy of the learner model, leading in turn to improved adaptation by the ITS. OLMs may also be used as learning resources independent of an ITS, to prompt learners to reflect on their knowledge (or lack of it), to facilitate planning future learning, and to encourage users to take more responsibility for their learning [6]. Other researchers have argued that it is necessary for educational systems to model the student's meta-knowledge in addition to their domain knowledge [7]. It is this approach of modelling the student's own beliefs about their knowledge that is discussed in this paper.

B. Woolf et al. (Eds.): ITS 2008, LNCS 5091, pp. 132–141, 2008.

Educational theorists have long emphasised the importance of learner reflection [8], [9], [10]. This is now being supported in the school classroom by Assessment *for* Learning, a UK education strategy that highlights the importance of supporting the development of metacognitive skills. Promoting pupil self-assessment is regarded as an essential component of this [11]. However, it is recognised that while the most effective learners are self-regulating [12] the effectiveness of this self-regulation is reliant on accurate self-assessment of what is known [13]. It has been shown, perhaps unsurprisingly, that not all (adult) students are good at evaluating their knowledge [14], and it was suggested that allowing the student to visualize the learner model may improve self-evaluation [15]. We propose to investigate this potential for learner model visualization in improving self-evaluation in younger (primary school) learners.

This paper describes an evaluation using two versions of CALMsystem – an Open Learner Model with an integrated ***C**onversational **A**gent* for ***L**earner **M**odelling* – independent of an ITS. The inspectable version of the system offers a learner the opportunity to inspect their learner model, to view the beliefs they and the system hold about their knowledge, and to make changes to their own beliefs about their knowledge as appropriate. The negotiated version adds a conversational agent to allow learners to discuss the learner model using a natural language interface and to negotiate changes. We consider these inspectable and negotiated versions of CALMsystem in terms of facilitating self-assessment accuracy and modification of model contents.

## 2 CALMsystem

CALMsystem opens the learner model to students, allowing them to see the representations of their current knowledge level as assessed by the system, and their self-assessment for each of the topics in the subject domain. The negotiated version also offers learners an opportunity to discuss and develop their learner model. Both inspectable and negotiated versions have potential to promote metacognitive skills and improve the model's accuracy.

The CALMsystem environment is browser based, operating independently of an ITS, and allows easy access to users from a variety of platforms. It allows users to view pages that show only their own confidence in their knowledge, only the system's assessments of their knowledge, or compare these in parallel. It also allows them to answer further questions on a topic of their choice, or one selected by the system, thereby allowing both user and system to initiate further interaction to update the learner model in the usual manner. Fig. 1 shows the browser interface (common to both versions of the system) and the conversational agent used to provide negotiation.

The system tracks the student's confidence and the system's assessment of the student's knowledge in each topic using two numerical scores. These two belief sets (learner's and system's) which form the learner model are stored independently, as is necessary for comparison and negotiation of the different beliefs (as in [1]). The user's confidence in each topic is maintained by the system as a continuous value between 0 and 1. For the purpose of display to the user, this value is converted into "low", "moderate", "good" or "high" levels, based on the ranges 0 - 0.25, 0.25 - 0.5, 0.5 - 0.75 and 0.75 - 1 respectively. These four levels offer an age-appropriate model for the 10-11 year old users in this study, who are familiar with self-evaluation scales of this granularity.

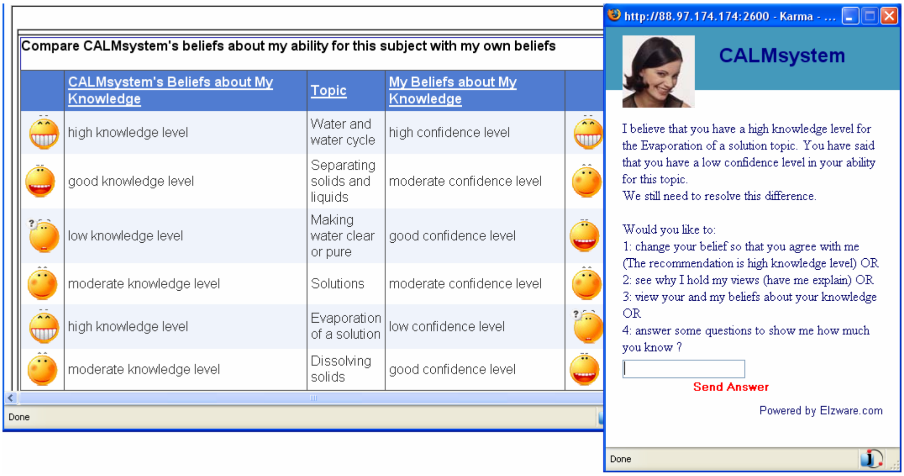


**Fig. 1.** System and learner assessments on six topics, and the conversational agent[1]

When a student first uses CALMsystem, they are required to assess their confidence in each topic by selecting the appropriate level ("low", "moderate", "good" or "high" confidence) and the initial numerical value is set as appropriate. Each time a student answers one of the multiple choice questions in CALMsystem (using the Answer Questions menu link) they are required to state the level that best matches their confidence in the topic. The system does not immediately change the numerical confidence value to match the user's new assessment, but uses an exponential filter[2] that weights most recent user assessments more strongly (so older results have a progressively lesser effect), allowing users to keep their model current.

The system's assessment of the student's knowledge is also maintained as a continuous value between 0 and 1, and uses an identical exponential filter, ensuring that the assessment represents the current knowledge level. This score for each topic is also recalculated every time the user answers a question (once past a threshold of 'sufficient evidence'). The score is increased each time a student answers a question correctly, and is reduced when a wrong answer is given. A student consistently answering questions correctly will attain a score approaching 1, and if most questions are answered incorrectly, the score will approach 0. For display, this knowledge value is also converted to four levels ("low", "moderate", "good" or "high") using the same

[1] Text reads "I believe that you have a high knowledge level for the Evaporation of a Solution topic. You have said that you have a low confidence level in your ability for this topic. We still need to resolve this difference. Would you like to: 1: change your belief so that you agree with me (The recommendation is high knowledge level) OR 2: see why I hold my views (have me explain) OR 3: view your and my beliefs about your knowledge OR 4: answer some questions to show me how much you know?"

[2] $y_t = (1-\alpha)\cdot y_{t-1} + \alpha\cdot x_t$ where $y_t$ is the output of the filter (new score) at time moment $t$; $y_{t-1}$ is the output of the filter after previous question (user's old score; $t$-1); $x_t$ is the input of the filter (1 or 0 indicating correct or incorrect answer); $0 \le \alpha \le 1.0$ is the weighting parameter. The output $y_t$ is the weighted sum of previous output and current input values. The smaller the parameter α, the longer the 'memory' of the filter and the greater the degree of smoothing.

numerical ranges as in user confidence. Both system and user beliefs are also illustrated with smiley faces (see Fig. 1) to allow easy comparison by the target users (aged 10-11) in this investigation.

### 2.1 Negotiation of the Learner Model

In the negotiated version of CALMsystem, inspection of the model is as described above, with negotiation of the learner model contents provided by a chatbot. It allows learners to use natural language to (i) query the model contents, (ii) ask for explanation or justification of the system's beliefs, (iii) offer justification of their own beliefs, (iv) change their beliefs as they refine their self-assessments, (v) modify their belief to match that of the system where they have been convinced by the model evidence, (vi) try to compromise with the system, or (vii) receive further test questions. These strategies were developed in [1] and explored in natural language in [16]. Discussion may be initiated by either the chatbot or the user.

The aim of this process of negotiation or discussion is to allow the student and system to refine their beliefs about the user's knowledge, and that this will lead to a reduction in areas of discrepancy between the learner and system beliefs. We suggest that a model refined in this way may be considered more accurate for use in driving an ITS's adaptation and may prompt further user reflection and self-assessment.

The chatbot does not attempt to be a tutor. The largest part of the chatbot's knowledge base relates to discussion of the learner model. Users may give inputs such as "why do we think differently?", "what am I good at?", "what's your belief?", "I have difficulty with [topic]", "change my belief", "what should I do next?", "why do you think I have a low level for [topic]?" or "let's talk about [topic]". They may request that the chatbot tell them the beliefs that are held, or just give the name of a topic to focus on. The learner can also ask the chatbot to give them test questions on a stated or unspecified topic. To handle many of these inputs the chatbot must connect to the database holding the learner model data, and construct its responses appropriately (see [16] for architectural details). For delivering tests it can select unanswered questions, ask the question, check the user's answer for correctness, and tell the user about the system's belief modifications made as a result. In this way it maintains focus on self-assessment and reflection on the contents of the OLM, an important benefit of independent OLMs [6]. However, the chatbot could be integrated into an ITS and provide tutoring as an extension of the system.

A short extract from a user interaction is shown below, showing the user modifying their belief after seeing the evidence from further test questions.

*Chatbot:* I believe that you have a low knowledge level for the Evaporation topic. You have said that you have a moderate confidence level in your ability for this topic. Would you like to: … [see footnote 1 for options offered]

*User:* 2

*Chatbot:* You've answered 4 questions and got 3 wrong for this topic. Therefore I believe you have a low knowledge level.

*User:* ask me quistions [sic] [OLMBot asks 2 questions; user answers correctly]

*Chatbot:* From your responses, I now believe you have a good knowledge level. Your own belief is that you have a moderate confidence level for the Evaporation topic. We will have to discuss this difference some more, or you might consider changing your belief for this topic.

*User:* i want to change my belif [sic]
*Chatbot:* What would you like to change your confidence level for Evaporation to? (Remember, this is *your* belief about your ability – not CALMsystem's assessment). You may change your beliefs to: 1: low confidence, 2: moderate confidence, 3: good confidence, 4: high confidence.
*User:* good
*Chatbot:* OK, I've updated your beliefs to show that you have good confidence in your abilities for the Evaporation topic.

The chatbot also has strategies and content to support 'smalltalk' (inputs not directly related to discussion of the learner model, but which have been found to be valuable in building rapport with users [17]) and to deal with unexpected user inputs. The intention is always to lead the user back into discussion of their learning as quickly as possible.

## 3 Experimental Evaluation

The aim of this study is to compare the effects of an inspectable and a negotiated learner model on self-assessment accuracy. As a proxy for self-assessment accuracy we compared the discrepancy between the system's assessment of the user's knowledge, and the user's assessment of their own capability. Inaccuracies in the system's modelling due to the user's accidental errors in answering questions are minimised by the use of the four broad knowledge levels, and the smoothing function of the exponential filter. It was hypothesised that using the inspectable version of CALMsystem would reduce this discrepancy, and that the discrepancy would be reduced further for participants who negotiated the learner model with the chatbot.

### 3.1 Measures of Self-assessment Accuracy

Three measures of the discrepancy between the student's confidence and system's assessments (and hence self assessment accuracy) were calculated for each user:

- **Numerical Measure of Discrepancy:** This measure sums the difference between the maintained numerical values for user confidence and system-assessed knowledge across all topics.
- **Number of Topics:** Where there is disagreement: This measure represents the number of topics that are not in agreement for a particular student. Topics are considered to be in agreement when the confidence and knowledge beliefs relate to the same level ("low", "moderate", "good" or "high").
- **Level Discrepancy:** This measure is a refinement of the Number of Topics measure outlined above, but takes into account the fact that a "low" to "high" discrepancy is more significant than, say, a "low" to "moderate" discrepancy. Adjacent levels (e.g. "moderate" and "good") are allocated a discrepancy distance of 1, those two levels apart (e.g. "low" and "good") a distance of 2 and those three levels apart (i.e. "low" and "high") are allocated 3. These distances are summed across all topics to give a measure of level discrepancy for each user (a theoretical maximum of 18). This discrepancy measure is considered to be of particular relevance, as it mirrors the typical view of a learner as to how far their own assessment differs from that of the system.

### 3.2 Participants, Materials and Methods

The study involved 25 UK Primary school children aged 10-11. CALMsystem was populated with questions on six science topics from their current study unit.

A between-subjects design was used, with the participants divided into two matched mixed-ability groups based on the results of a diagnostic test on the topics. One group was allocated to an *inspectable* learner model (LM) condition, and the other to a *negotiated* LM condition. All participants were shown how to use the system, its purpose and how it might be useful to them. Participants used the system for two sessions, three weeks apart, totalling 120 minutes. All users interacted with the system to make initial self-assessments, answer multiple choice questions, view their confidence ratings and the system's assessments, and modify their confidence records where they desired. Those in the negotiated LM condition also interacted with the chatbot to discuss their model.

As both users' confidence ratings and the system's assessments are recalculated after every question that is answered, the current values are always known and displayed by CALMsystem. The data used in this analysis was extracted from the learner model logs. The initial ('before-use') values are the beliefs held at the point where the system first had sufficient data about the user's knowledge of a topic to model the user. The final state of the learner model after both sessions gives the 'after-use' state.

### 3.3 Results

#### 3.3.1 Improvement in Self-assessment Accuracy (Numerical Measure)

Before using CALMsystem, the mean self-assessment error for all 25 participants across all six topics was 1.74 (median 1.56, range 0.69-4.31). After final use of the system this mean error was reduced to 0.82 (median 0.66, range 0.29-2.43) for all users in the inspectable or negotiated conditions. The improvement by inspectable LM users (mean reduction 0.45, median 0.55, range -0.99-1.64) was significant ($t=1.83$, $p<0.05$). Negotiated LM users made highly significant ($t=4.72$, $p<0.0005$) improvements, (mean reduction in error 1.35, median 0.93, range 0.16-3.99). Notably, this improvement was significantly greater ($t=2.38$, $p<0.025$) than that for inspectable LM users (see Fig. 2).

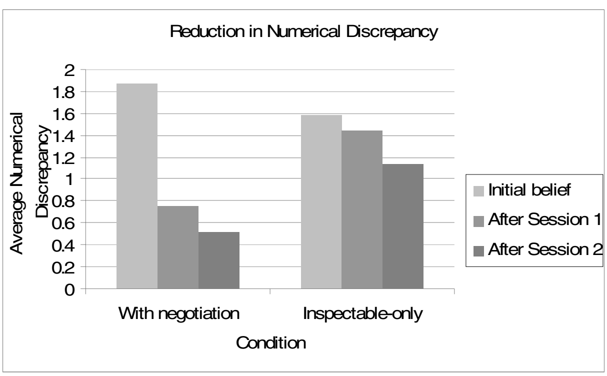


**Fig. 2.** Improvement in Self-Assessment (Reduction in Numerical Discrepancy)

### 3.3.2 Reduction in Number of Topics

The number of topics in which there was disagreement between the user and system as to the user's ability was counted. Before using the system, the mean number of topics with discrepancy was 3.88 (median 4, range 1-6). After final use of CALMsystem this average was reduced to 1.52 (median 1, range 0-6), an average reduction of 2.36. Inspection of the LM reduced the number of discrepancies significantly (mean reduction 1.5, median 2, range -3-5, ($t$=1.95, $p$<0.05)). The reducution was significantly greater ($t$=2.08, $p$<0.025) for participants in the negotiated LM condition (mean reduction 3.15, median 3, range 1-6, ($t$=8.01, $p$<0.0005)) than for those in the inspectable LM condition (see Fig. 3).

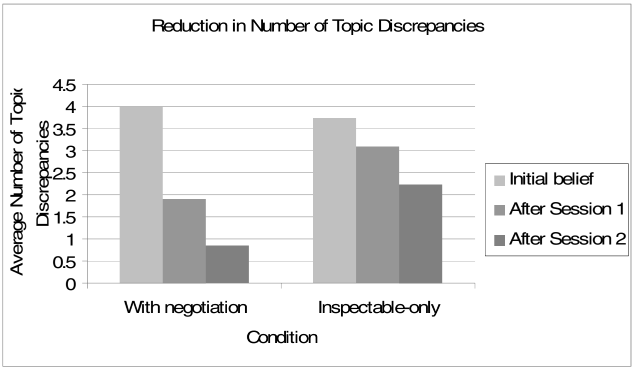


**Fig. 3.** Improvement in Self-Assessment (Reduction in Number of Topic Discrepancies)

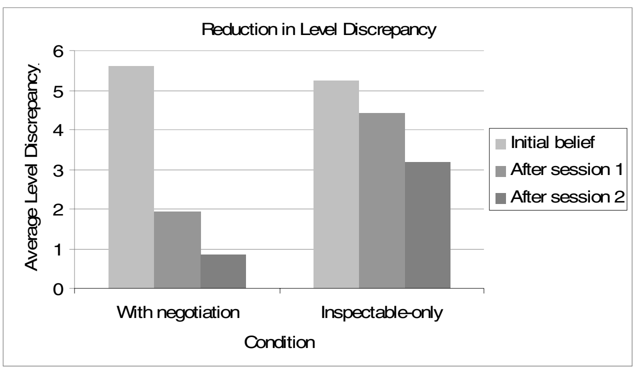


**Fig. 4.** Improvement in Self-Assessment (Reduction in Level Discrepancy)

### 3.3.3 Reduction in Level Discrepancy

The Level Discrepancy measure gives a value representing the disparity between levels ("low", "moderate", "good", "high") held by the student and system. Before using the system the mean level discrepancy was 5.44 (median 6, range 1-11). After final use of CALMsystem this average was reduced to 1.96 (median 1, range 0-9), an average reduction of 3.48. Users in the inspectable LM condition reduced the Level Discrepancy significantly (mean reduction 2.08, median 3, range -4-7, ($t=1.84$, $p<0.05$)). Again it was found that the reduction in the Level Discrepancy was significantly greater ($t=2.31$, $p<0.025$) for participants in the negotiated LM condition (mean reduction 4.77, median 4, range 1-10, ($t=7.12$, $p<0.0005$)) (see Fig. 4).

### 3.3.4 Questions Answered

Users of the negotiated LM answered an average of 35.15 questions (median 35, range 22-61). Users in the inspectable condition answered an average of 51.08 questions (median 49, range 34-79), a highly significant difference ($t=3.19$, $p<0.005$).

# 4 Discussion

The results show that after using the CALMsystem open learner model all participants (in both conditions) significantly reduced the mean error in their self assessments. Users who engaged in negotiation with the chatbot demonstrated a significantly greater improvement in their self-assessment accuracy. These results suggest that inspection of the learner model can help prompt students to re-assess their knowledge, and that the chatbot negotiation element offers further benefit. Use of the system also reduced the number of discrepancies in learner/system beliefs. There was a substantial reduction in discrepancies for all participants; again this was significantly greater for negotiation users than for inspectable LM users. This reduction in the number of topics where user and system disagree results in a model where both parties hold more similar beliefs, allowing users to help direct potential ITS adaptations which they may consider of more value. The improvement in self-assessment accuracy should allow users to better target future learning and develop greater learner autonomy.

Interestingly, the discrepancy measures reduce rapidly across the trial for negotiated LM users, but markedly less so for inspectable users. This suggests that exposure to the OLM alone was lesser of an incentive for children to substantially change their self-assessments of confidence in a topic. The more proactive chatbot element, which persuades the users to challenge their belief where there are discrepancies appears to be more effective in making them consider their ability and make changes to their self-assessments. As shown in Figures 2, 3 and 4, the interaction continued to reduce discrepancies after a second session, suggesting that there was some lasting effect over the period between sessions (three weeks). Further study would be required to ascertain whether the extended use of a negotiated learner model would improve general self-assessment and metacognitive skills, and whether the improvements in self-assessment would be maintained over time.

Users in the inspectable LM condition answered far more questions in the interaction; this was the main activity available to them. This will have given them greater opportunity to view the representations of the beliefs held. However, despite this opportunity to consider the different beliefs more often, these users' beliefs did not change as significantly as those of the users with chatbot negotiation. Answering questions, re-stating confidence, and seeing the resultant model alone appears beneficial, but a lesser prompt to reflect on the learner model than offered by negotiation.

Users of both the inspectable and negotiated systems demonstrated significant improvements in self-assessment accuracy and in reducing the number and magnitude of discrepancies. The further improvements demonstrated by the negotiated LM suggest that where negotiation can be included this would provide additional benefits. The chatbot may persuade or help users to engage with their learning by exposing them to a proactive tool that they are willing to work with. This may be an effect of the novelty, naturalness or accessibility of a chatbot, or may be due to the content it offers.

Further work is necessary to explore whether the improvements in self-assessment transfer back to normal classroom scenarios (i.e. without computer), and whether belief changes persist beyond use of the system. It will also be interesting to explore if it is the chatbot's dialogue content that is effective, or whether the presence of the chatbot is a motivational factor which keeps young users engaged with the process.

## 5 Summary

We have presented an evaluation of two versions of an Open Learner Model. One version offers inspection of the learner model, while the other is supported by a chatbot to provide discussion and negotiation of the learner model contents. This negotiation allows the user and system to collaboratively construct and maintain the learner model, providing further opportunities for the learner to reflect on their knowledge and to refine their self-assessments than was seen in users of the inspectable-only model. Improvements were seen in both conditions. The study showed that users who engaged in negotiation reduced inaccuracies in their self-assessments significantly more than those users who used the system without negotiation support.

## Acknowledgements

The first author is funded by the UK Engineering and Physical Sciences Research Council. We acknowledge Elzware Ltd's support and provision of software licences.

## References

1. Bull, S., Pain, H.: Did I Say What I Think I Said, and Do You Agree With Me?: Inspecting and Questioning the Student Model. In: Greer, J. (ed.) World Conference on Artificial Intelligence in Education. AACE, Charlottesville, pp. 501–508 (1995)
2. Dimitrova, V.: STyLE-OLM: Interactive Open Learner Modelling. International Journal of Artificial Intelligence in Education 13, 35–78 (2003)
3. Kay, J.: Learner Know Thyself: Student Models to give Learner Control and Responsibility. In: Halim, Z., Ottomann, T., Razak, Z. (eds.) ICCE 1997 International Conference on Computers in Education, Kuching, Malaysia, pp. 17–24 (1997)
4. Morales, R.: Exploring Participative Learner Modelling and Its Effects on Learner Behaviour. PhD Thesis. University of Edinburgh, Edinburgh (2000)
5. Zapata-Rivera, J.D., Greer, J.: Externalising Learner Modelling Representations. In: Workshop on External Representations of AIED: Multiple Forms and Multiple Roles. International Conference on Artificial Intelligence in Education, San Antonio, Texas (2001)

6. Bull, S., Mabbott, A., Gardner, P., Jackson, T., Lancaster, M., Quigley, S., Childs, P.A.: Supporting Interaction Preferences and Recognition of Misconceptions with Independent Open Learner Models. In: Adaptive Hypermedia 2008. Springer, Berlin (2008)
7. Aleven, V., Koedinger, K.: Limitations of Student Control: Do Students Know When They Need Help?: Intelligent Tutoring Systems, pp. 292–303. Springer, Berlin (2000)
8. Dewey, J.: How We Think: A Restatement of the Relation of Reflective Thinking to the Educative Process. D. C. Heath and Company, Boston (1933)
9. Schön, D.: The Reflective Practitioner: How Professionals Think in Action. Arena. Ashgate Publishing Limited, Aldershot (1991)
10. Kolb, D.: The Process of Experiential Learning. In: Kolb, D. (ed.) The Experiential Learning: Experience as the Source of Learning and Development, Prentice-Hall, NJ (1984)
11. Black, P., Wiliam, D.: Inside the Black Box: Raising Standards Through Classroom Assessment. King's College London, School of Education, London (1998)
12. Butler, D.L., Winne, P.H.: Feedback and Self-Regulated Learning: A Theoretical Synthesis. Review of Educational Research 65, 245–281 (1995)
13. Schoenfeld, A.H.: What's All the Fuss About Metacognition? In: Schoenfeld, A.H. (ed.) Cognitive Science and Mathematics Education, pp. 189–215. Lawrence Erlbaum, Hillsdale (1987)
14. Mitrovic, A.: Self-assessment: how good are students at it? In: AIED 2001 Workshop on Assessment Methods in Web-Based Learning Environments & Adaptive Hypermedia, San Antonio, pp. 2–8 (2001)
15. Mitrovic, A., Martin, B.: Evaluating the Effect of Open Student Models on Self-Assessment. International Journal of Artificial Intelligence in Education 17, 121–144 (2007)
16. Kerly, A., Ellis, R., Bull, S.: CALMsystem: A Conversational Agent for Learner Modelling. Knowledge Based Systems 21, 238–246 (2008)
17. Kerly, A., Hall, P., Bull, S.: Bringing Chatbots into Education: Towards Natural Language Negotiation of Open Learner Models. Knowledge Based Systems 20, 177–185 (2006)