

Kerly, A., Ellis, R. & Bull, S. (2007). CALMsystem: A Conversational Agent for Learner Modelling, in R. Ellis, T. Allen & M. Petridis (eds), Applications and Innovations in Intelligent Systems XV – Proceedings of AI-2007, 27th SGA International Conference on Innovative Techniques and Applications of Artificial Intelligence, Springer Verlag 89-102.

CALMsystem: A Conversational Agent for Learner Modelling

Alice Kerly¹, Richard Ellis² and Susan Bull¹

¹Electronic, Electrical and Computer Engineering,
University of Birmingham, Edgbaston, Birmingham, B15 2TT, UK

²Stratum Management Ltd, 11 Southbrook Place, Micheldever,
Hampshire, SO21 3DE, UK

¹{alk584, s.bull}@bham.ac.uk; ²richard.ellis@stratum-management.co.uk

Abstract

This paper describes a system which incorporates natural language technologies, database manipulation and educational theories in order to offer learners a Negotiated Learner Model, for integration into an Intelligent Tutoring System. The system presents the learner with their learner model, offering them the opportunity to compare their own beliefs regarding their capabilities with those inferred by the system. A conversational agent, or “chatbot” has been developed to allow the learner to negotiate over the representations held about them using natural language. The system aims to support the metacognitive goals of self-assessment and reflection, which are increasingly seen as key to learning and are being incorporated into UK educational policy. The paper describes the design of the system, and reports a user trial, in which the chatbot was found to support users in increasing the accuracy of their self-assessments, and in reducing the number of discrepancies between system and user beliefs in the learner model. Some lessons learned in the development have been highlighted and future research and experimentation directions are outlined.

1. Background

Intelligent Tutoring Systems (ITS) provide their users with an adaptive learning environment, with personalized tutoring and testing customised to meet the needs of the individual student. This adaptation is based on the contents of the learner model, a representation of the student’s knowledge, gaps in understanding and misconceptions. Traditional ITSs have not made the contents of the learner model visible to the learner. However, it has been argued that an *Open* Learner Model (i.e. one that can be inspected by the student) can offer opportunities for learner reflection, metacognition and deep learning, which may enhance learning (e.g. [1], [2], [3], [4] and [5]), as well as improving the accuracy of the learner model. Educational theorists have emphasised the importance of learner reflection ([6], [7] and [8]). Some researchers have developed Open Learner Models (OLM) that

Kerly, A., Ellis, R. & Bull, S. (2007). CALMsystem: A Conversational Agent for Learner Modelling, in R. Ellis, T. Allen & M. Petridis (eds), Applications and Innovations in Intelligent Systems XV – Proceedings of AI-2007, 27th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, Springer Verlag 89-102.

encourage negotiation as an approach to improve learner reflection and model accuracy ([1] and [2]). In such systems the learner model is collaboratively constructed and maintained by both the system and the learner. The learner is required to discuss their beliefs about their knowledge with the system, to argue against the system's assessment if they disagree, and provide supporting evidence or argument for their own beliefs when they differ from the system. [1] and [2] found that this method of negotiating the learner model supported increased learner reflection, and produced a more accurate learner model on which to base system adaptivity. This negotiation based approach has been adopted in the current research.

Metacognition (defined as cognition about cognition: thoughts about thoughts, knowledge about knowledge or reflections about actions [9]) may be considered to capture two essential features: self-appraisal and self-management of cognition [10]. It is also recognised that the most effective learners are self-regulating [11] and that effective self-regulation is reliant on accurate self-assessment of what is known or not known [12]. In the context of Intelligent Tutoring Systems, it has been argued that it is necessary for educational systems to model the student's meta-knowledge in addition to their domain knowledge [13]. It has also been found that not all students are good at evaluating their knowledge, and suggested that allowing the student to visualize the learner model may help their self-evaluation [14].

After a period of neglect of formative assessment in classroom practice [15], modern UK educational policy is starting to recognise the importance of metacognition, and now promotes 'Assessment *for* Learning' (AfL). Pupil self-assessment is regarded as an essential component of this [15]. The aims of AfL (promoting reflection, using assessment to modify teaching, conducting pupil self-assessment and providing formative feedback) closely mirror the ethos of OLM, as reported in [16]. The practice of opening the system-held learner model for viewing by the learner has been implemented in an increasing number of cases with school age pupils (e.g. [5], [17] and [18]).

Previous learner modelling negotiation methods (menu selection and construction of conceptual graphs) may be difficult for some learners (especially younger users) or require learning a new communications method. It is envisaged that a conversational agent (or chatbot) will provide a more intuitive and convenient method for negotiation. Natural language dialogue has been employed in ITS for tutoring dialogues, avatars, and pedagogical agents. Animated pedagogical agents have been argued to engage the student without distracting or distancing them from the learning experience [19]. Natural language dialogue has not previously been used in the negotiation of an Open Learner Model. It is hypothesised that the use of a conversational agent in negotiating the learner model will be similarly engaging and non-distracting.

This paper describes CALMsystem – an Open Learner Model environment with an integrated *Conversational Agent* for *Learner Modelling*. The inclusion of a chatbot provides learners with a flexible and intuitive method with which to query the system's beliefs about their knowledge, explain or modify their own beliefs, answer questions about the topic being studied, and for the chatbot to initiate discussions.

Kerly, A., Ellis, R. & Bull, S. (2007). CALMsystem: A Conversational Agent for Learner Modelling, in R. Ellis, T. Allen & M. Petridis (eds), Applications and Innovations in Intelligent Systems XV – Proceedings of AI-2007, 27th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, Springer Verlag 89-102.

2. CALMsystem Design

2.1 Overview

CALMsystem provides an Open Learner Model to students, representing their current knowledge level and self-assessed confidence on constituent topics of the target subject. The system is intended to encourage students to discuss and reflect on their knowledge so they can be helped to develop autonomy over their learning and improve the metacognitive skills that lead to enhanced self-assessment.

The system enables students to view their own assessment of their ability in different topics, and compare these with the beliefs held by the system, based on the questions they have answered. It also allows them to answer further questions on any topic, and to update their own assessment. Figure 1 shows the user interface.

CALMsystem's Beliefs about My Knowledge	Topic	My Beliefs about My Knowledge
high knowledge level	Electricity	moderate confidence level
low knowledge level	Solids and Liquids	high confidence level
good knowledge level	Temperature	high confidence level
high knowledge level	Forces	moderate confidence level
low knowledge level	Friction	high confidence level
low knowledge level	Materials	good confidence level

Would you like to:
 1: change your belief so that you agree with me (The recommendation is low 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 ?

 Answer
 Powered by Elizware.com

Figure 1 The CALMsystem user interface showing system and user beliefs about the user's knowledge on six topics, and the system chatbot.

The provision of a chatbot enables students to use natural language to query the contents of the learner model, ask the system to explain its beliefs, justify their own beliefs if the system disagrees, change their belief if they have re-assessed their abilities, accept the system's belief (and change theirs to match), try to compromise with the system, or ask for further test questions (strategies employed by [1]).

Through the process of negotiation the aim is that both the number of topics on which the system and student disagree about the student's abilities and the magnitude of the discrepancies will be reduced, and that the student's self-assessment will become more accurate.

Kerly, A., Ellis, R. & Bull, S. (2007). CALMsystem: A Conversational Agent for Learner Modelling, in R. Ellis, T. Allen & M. Petridis (eds), Applications and Innovations in Intelligent Systems XV – Proceedings of AI-2007, 27th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, Springer Verlag 89-102.

2.2 Design

In order to provide a negotiation based on the different beliefs held by user and system, it is necessary to store these belief sets independently [1]. When a user logs in to the system for the first time, they are required to provide a self-assessment rating (one of four confidence levels) for each of the topics in the system. Every time they subsequently answer a test question the user also has to re-state their confidence in their ability to tackle questions on the given topic. This allows users to indicate changes in their confidence and keeps the model current.

The system's belief value for each topic is similarly calculated every time the user answers a question, based on whether the user answered correctly. Once a user has answered a threshold number or proportion of questions on a topic, the system calculates that it has enough evidence to begin displaying its belief.

The system maintains its model of the user's current knowledge (system inferred) and confidence beliefs (learner self-assessed) by applying a weighted recursion algorithm to each new answer or belief provided by a student. This results in a model where most recent data is weighted more strongly, and older results have a progressively lesser effect on resultant user confidence or system belief values.

CALMsystem's browser environment provides the user with the opportunity to view the learner model held about them. They may opt to view different pages that show only their own confidence beliefs or only the system's assessment, or view a page which compares the two beliefs side by side. Their own confidence belief and system assessment are shown against each topic name as either low, moderate, good or high knowledge or confidence level. For the young children (aged 8-9) who used the system in its first trial, smiley faces were also used to represent each knowledge or confidence level. This matched use of language and images allows users to quickly see where they and the system are in agreement or discrepancy. The system also allows the user to request further questions on any topic (or a topic chosen by the system) to allow their own and the computer's assessments to be updated.

2.3 Chatbot

While the web based system is fully functional as an OLM system, it is considerably enhanced by the addition of a chatbot. The chatbot has two principal functions:

- to respond to user initiated discussion on their knowledge, the reasoning for the system beliefs, and on any discrepancy in views,
- to initiate discussion on discrepancies in the system and user viewpoints.

The chatbot can also support 'smalltalk' and other interactions not directly related to the learner model.

For user initiated conversation, the chatbot 'understands' a wide range of inputs. The largest portion of the knowledge base, as may be expected, relates to discussion of learning and 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 Electricity", "change my belief", "what should I do next?", "why do you think I

Kerly, A., Ellis, R. & Bull, S. (2007). CALMsystem: A Conversational Agent for Learner Modelling, in R. Ellis, T. Allen & M. Petridis (eds), Applications and Innovations in Intelligent Systems XV – Proceedings of AI-2007, 27th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, Springer Verlag 89-102.

have a low level for Forces?”, and “let’s talk about Friction”. Users may simply request that the chatbot tells them the beliefs either they themselves or the system hold for any of the topics, or just give the name of a topic that they want to focus on. They can also request that the chatbot ask them test questions on a specific subject. If they do not state a subject the chatbot will suggest that they choose one. The chatbot will connect to the database to retrieve unanswered questions (or least recently answered if there are none), ask the user the question, check the user’s answer for correctness, and feed back any changes to the system belief as a result of the answer.

The chatbot can also recognise inputs that relate to the user ‘sounding out’ the system (e.g. ‘what are you for?’, ‘are you a computer?’, ‘what can I do?’) or to smalltalk (e.g. ‘am I cool?’, ‘do you like my name?’, ‘do you like football?’, ‘can you tell jokes?’). These interactions, though not directly related to the learner model, have been found to be valuable [20] in encouraging users to engage with the system. In order to control the conversation, after a small number of smalltalk interchanges it will remind the user that they are there to discuss their learning, and suggest a topic.

When the chatbot initiates conversation (which it will do when there is a discrepancy between learner and system beliefs, and if the user is not already in discussion or answering questions), it selects the topic for which there is the greatest disparity in beliefs, tells the user that this difference exists and suggests that they should attempt to resolve it.

The chatbot offers up to five options for negotiation (as in [1]):

- the user can accept the system’s recommendation,
- the system can justify and explain its belief,
- the user may view (be reminded of) each of the beliefs,
- the user may attempt to reach a compromise with the system,
- the user may justify their own belief by answering further test questions.

As each option is tried, it is removed from the list of available negotiation methods for subsequent use if the discrepancy was not resolved. If at any time the user decides that there is a different issue that they would prefer to discuss then they are free to direct the chatbot in this way.

The chatbot makes no attempt to teach, and does not give out correct answers, keeping the focus on self-assessment. If a user requests answers from the chatbot it will respond that it is not a teacher, cannot give the correct answers, and suggest using another revision method or asking the user’s teacher.

In order to deal with novel (unexpected) inputs, the chatbot has a variety of ‘safety nets’. Some of these are tailored to take into account the context in which the conversation was before the unrecognised input was reached, or respond based upon the user’s current activity status. Figure 2 shows a sample conversation transcript, with the uninterpretable input in italics. The iterative development process concentrated on identifying these unrecognised inputs and modifying the chatbot

Kerly, A., Ellis, R. & Bull, S. (2007). CALMsystem: A Conversational Agent for Learner Modelling, in R. Ellis, T. Allen & M. Petridis (eds), Applications and Innovations in Intelligent Systems XV – Proceedings of AI-2007, 27th SGA International Conference on Innovative Techniques and Applications of Artificial Intelligence, Springer Verlag 89-102.

code to provide an improved response. The italicised input in Figure 2 was recognised and resolved as part of this incremental development process.

User:
now friction

OLMBot:
I believe that you have a high knowledge level for the Friction topic. You have said that you have a moderate 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: see if we can agree halfway OR
5: answer some questions to show me how much you know?

User:
2

OLMBot:
You answered 3 question(s) and got 0 wrong for the Friction topic. Therefore I believe you have a high knowledge level.

User:
yes but i dint know i was guessing

OLMBot:
We still haven't resolved our different views on the Friction topic. You need to choose from one of the following options. Would you like to:
1: change your belief so that you agree with me (The recommendation is high knowledge level) OR
2: view your and my beliefs about your knowledge OR
3: see if we can agree halfway OR
4: answer some questions to show me how much you know ?

User: answer questins

OLMBot: OK, I'll ask you some test questions for you to demonstrate your knowledge. Here's the first:
...(ctd)

Figure 2 Transcript of a typical conversation fragment

There are also specialised safety nets to deal with nonsense entries, or offensive language. In this case the user is first warned not to use inappropriate language, but if they persist the chatbot will stress more forcefully that obscene content will not be tolerated, and insist that the user apologises before the conversation can proceed. In a similar fashion, counters are employed so that too many off-topic entries are trapped, as are any situations where the user causes a conversation to loop.

3. CALMsystem Implementation

The learner modelling environment, including its server-side processing facilities and browser based user interface, were implemented using ASP.NET. The system uses an SQL database for storing all model data, including subject domain topics, questions and answers, and system logs. The approach provides a clear visual

Kerly, A., Ellis, R. & Bull, S. (2007). CALMsystem: A Conversational Agent for Learner Modelling, in R. Ellis, T. Allen & M. Petridis (eds), Applications and Innovations in Intelligent Systems XV – Proceedings of AI-2007, 27th SGA International Conference on Innovative Techniques and Applications of Artificial Intelligence, Springer Verlag 89-102.

interface for users, and offers access over the Internet, which is necessary for easy deployment to schools. This architecture is shown in Figure 3, and the user interface is illustrated in Figure 1.

In order to develop the chatbot, we selected the commercial Lingubot™ [21] technology. The Lingubot technology has established a significant corpus of external scripts and applications, providing functionality relevant to conversations, such as abuse counters, theme metrics and directed conversational abilities, as well as integration with many web technologies. It has the capability to generate and manipulate variables and information regarding the conversation, and also for retrieving and posting information to other web enabled applications such as search engines and databases. The commercial nature of the technology also ensures that it is well tested. Further discussion of this technology is presented in [20].

The chatbot runtime engine was installed on the same server as the .NET system and SQL database. The chatbot is displayed in the learner model environment in a dedicated window, as shown in Figure 1. An ODBC connection allows the chatbot to extract database material in order to offer questions, test user answers for correctness, and to reflect information from the learner model to the user. It is similarly able to write to the database to maintain an accurate model for concurrent access by the learner modelling user interface. The principal interface between the chatbot and the rest of the OLM system is via the SQL database.

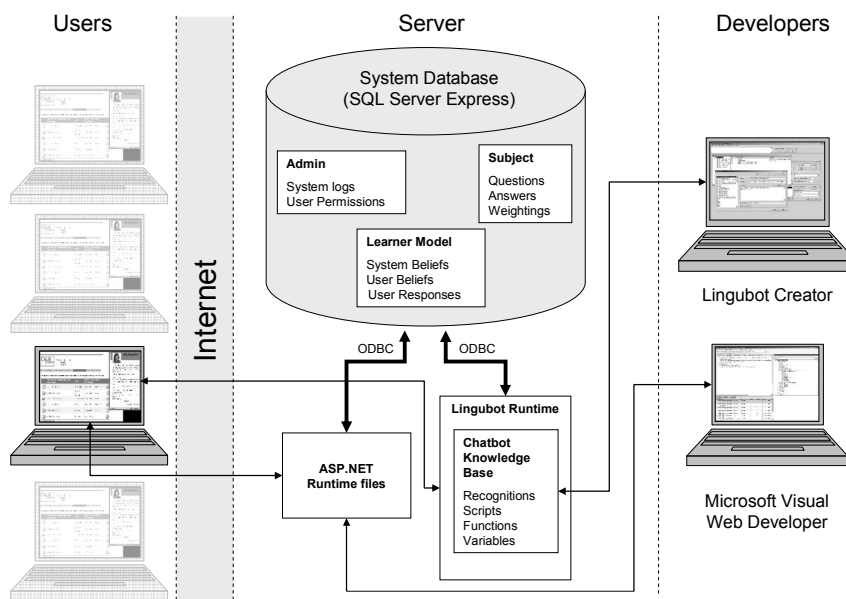


Figure 3 The CALMsystem Architecture

Kerly, A., Ellis, R. & Bull, S. (2007). CALMsystem: A Conversational Agent for Learner Modelling, in R. Ellis, T. Allen & M. Petridis (eds), Applications and Innovations in Intelligent Systems XV – Proceedings of AI-2007, 27th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, Springer Verlag 89-102.

Early design of the chatbot was informed by the results of a Wizard of Oz study [20], where users believed they were communicating with a chatbot, but which was in fact the human experimenter. This provided guidance for structuring the underlying chatbot scripts.

The CALMsystem chatbot contains over 1500 recognitions, or target patterns of user entry, which are matched to appropriate responses. Many of these recognitions are dependent on other variables including current user activity, current conversational status or current learner model status. These combinations result in almost 3000 basic answers, which again are dynamically constructed depending on, and to include, necessary variables and conditions.

4 Experimental Evaluation

4.1 Aims

A number of experiments and trials are planned for CALMsystem, in order to:

- investigate whether the basic OLM based system benefits learning,
- investigate whether the addition of a chatbot leads to improved learning outcomes.

In addition, these experiments and trials will provide valuable opportunities for feedback on the design and usability of the system in a real target environment

It is recognised that CALMsystem represents only a small part of an overall Intelligent Tutoring System, and that in initial trials the target audience would only be able to use the system for a small proportion of their learning. For these reasons, it was not possible to use direct measures of learning gain as a metric. However, CALMsystem, when used with or without the chatbot, is intended to improve the ability of the students to assess their own level of knowledge. This can be measured in terms of discrepancies between the user and the system assessments of capability for different topics. It was hypothesised that using the system would reduce both the number of and the magnitude of the discrepancies, and that this reduction would be greater when the student used the chatbot as part of the system.

4.2 Participants, methods and materials

The participants were a class of 30 UK Primary school children in Year 4 (Key Stage 2, aged 8-9). The CALMsystem database was populated with 89 multiple choice questions on Key Stage 2 Science topics (Electricity, Solids & Liquids, Temperature, Forces, Friction and Materials). All the children were given a pre-test on science topics, the results of which were used to divide the class into 2 mixed-ability groups.

The experimental design was a between-subjects study, with one group of participants allocated to a group to use CALMsystem *with* the chatbot, and the other to use it *without* the chatbot. All the children were shown a presentation of how to use CALMsystem, its purpose and how it might be useful to them.

Kerly, A., Ellis, R. & Bull, S. (2007). CALMsystem: A Conversational Agent for Learner Modelling, in R. Ellis, T. Allen & M. Petridis (eds), Applications and Innovations in Intelligent Systems XV – Proceedings of AI-2007, 27th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, Springer Verlag 89-102.

The participants each took part in sessions totalling 60 minutes use of CALMsystem (with or without the chatbot, as dictated by experimental condition), over a period of 2 weeks. All children were instructed that using CALMsystem would give them an opportunity to answer test questions on the science topics that they were studying, see what knowledge level the system believed they had reached, make their own self-assessments about their abilities for each topic, and change their beliefs or try to change the system's belief if they disagreed. The system logged all interactions.

Children using the chatbot were also shown how to use it, and given some simple suggestions as to what they might discuss, but were also told that they could ask it anything they liked regarding their learning.

When participants first used CALMsystem, they were required to make a self-assessment for each topic in the system, placing themselves in one of four confidence bands labelled low, moderate, good or high confidence. For numerical analysis, these equate to 0.25, 0.5, 0.75, and 1. These self-assessment scores were compared with the system's initial beliefs (based on questions answered by the user) to ascertain the level of discrepancy between user and system beliefs before using the system.

As participants were required to update their self-assessment for each topic whenever they answered a question, their current self-assessment value was always known by the system. The final value is therefore taken as their final self-assessment.

4.3 Outcomes

4.3.1. Self-assessment accuracy

The mean self-assessment error for all 30 participants before using the system was 2.93 (median 3, range 1.75 to 5). After using the system, this mean error was reduced to 1.77 (median 1.78, range 0 to 3.9), showing an improvement in self-assessment accuracy for all users, with or without the chatbot (see Figure 4). In addition, the improvement for users using the chatbot (see Figure 5, mean reduction in error 1.42, median 1.34, range -0.27 to 2.42) was significantly greater ($t=1.8409972$, $p<0.05$) than that for users in the without chatbot condition (see Figure 6, mean reduction 0.863, median 0.78, range -0.76 to 2.59).

Figure 4 shows the reduction in self-assessment error (discrepancy between user and system beliefs) for each individual user, after using CALMsystem. (The two negative values indicate an increase in self-assessment error by those individuals). For example, for one given topic, a user belief of 0.75 and a system assessment of 0 would give an error of .75. Given the six topics, the total maximum error (discrepancy) would be 6 (indicating completely opposing beliefs on all topics). A user who had a mean initial error (for all 6 topics) of 3.25, but who reduced this to 1.5 (again across all 6 topics) has a recorded reduction in error of 1.75 in Figure 4.

Kerly, A., Ellis, R. & Bull, S. (2007). CALMsystem: A Conversational Agent for Learner Modelling, in R. Ellis, T. Allen & M. Petridis (eds), Applications and Innovations in Intelligent Systems XV – Proceedings of AI-2007, 27th SGA International Conference on Innovative Techniques and Applications of Artificial Intelligence, Springer Verlag 89-102.

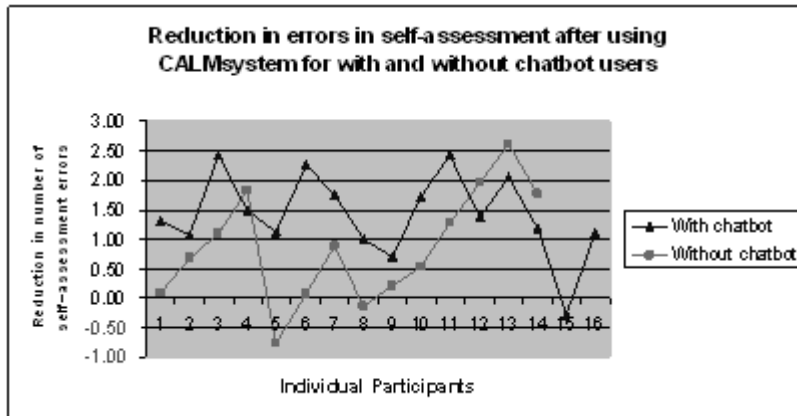


Figure 4 Reduction in number of self-assessment errors for all users

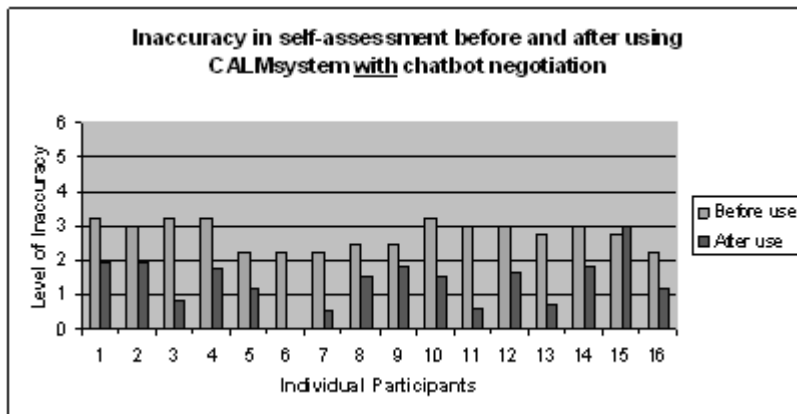


Figure 5 Reduction in self-assessment inaccuracy for with chatbot users – significantly greater than for users without the chatbot

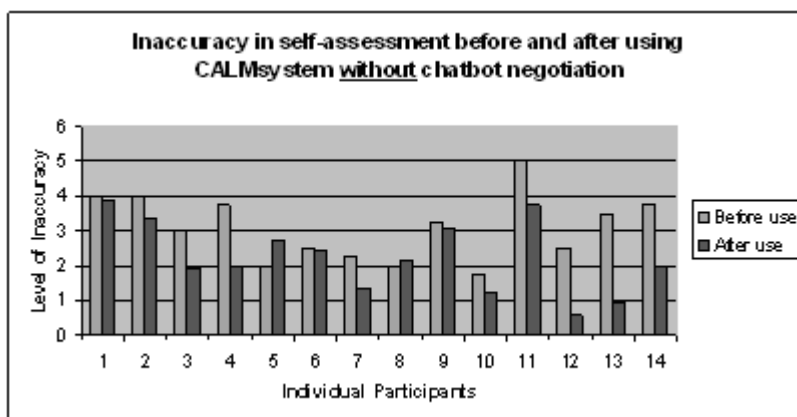


Figure 6 Reduction in self-assessment inaccuracy for without chatbot users

Kerly, A., Ellis, R. & Bull, S. (2007). CALMsystem: A Conversational Agent for Learner Modelling, in R. Ellis, T. Allen & M. Petridis (eds), Applications and Innovations in Intelligent Systems XV – Proceedings of AI-2007, 27th SGA International Conference on Innovative Techniques and Applications of Artificial Intelligence, Springer Verlag 89-102.

4.3.2. Reduction in discrepancy between user and system beliefs

A simple count was also made of the number of topics on which the user and system disagreed as to the user's ability. Before use of the system, the mean number of topics with discrepancy was 5.466 (median 6, range 4 to 6). After using CALMsystem this average was reduced to 3.066 (median 3, range 0 to 6), an average reduction of 2.4. Again it was found that the reduction in the number of discrepancies was significantly greater ($t=1.875$, $p<0.05$) for children in the with chatbot condition (mean reduction 2.875, median 3, range 0 to 5) than for those without (mean reduction 1.875, median 2, range -1 to 4) (see Figure 7).

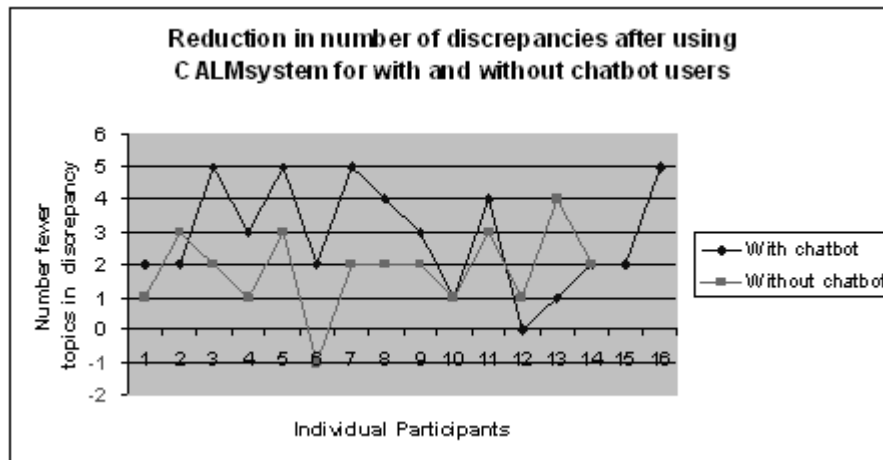


Figure 7 Reduction in number of discrepancies after using CALMsystem

4.4 Additional Observations

The trial was extremely valuable in fine-tuning both the language used by the chatbot, and its ability to interpret the inputs made by the users. The Lingubot package provides powerful tools for exploring failed recognitions and sequences of inputs that lead to the “safety net” recognitions, which the chatbot uses to fail gracefully, guiding discussion back to the topic at hand.

The chatbot includes a set of responses designed to discourage the children from swearing or using inappropriate language. The authors (and class teacher) were surprised by some of the language used by the children when unobserved, and the set of recognitions for dealing with this were extended accordingly.

5 Lessons for Other Developers

This development has highlighted a number of issues that may be useful to other developers and researchers. It drew together a number of strands of technology, including natural language, database management, web development and intelligent tutoring systems.

Kerly, A., Ellis, R. & Bull, S. (2007). CALMsystem: A Conversational Agent for Learner Modelling, in R. Ellis, T. Allen & M. Petridis (eds), Applications and Innovations in Intelligent Systems XV – Proceedings of AI-2007, 27th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, Springer Verlag 89-102.

5.1 System Development

The system was developed over a period of nine months with specialist support from Elzware Limited in the development and customisation of the chatbot element.

- **Coding Approach** - Throughout this development deliberate effort was made to take a professional approach to code development, including configuration control, code structure, documentation and commenting, across all coding environments (ASP.NET, C#, KScript, SQL). The system architecture was designed to allow it to be rapidly adapted to different subjects and environments. Subject dependent content (such as question sets, weighting rules, and the learner model) was placed in tables in the database, allowing modification and development separately from the core logic of the system. There is no doubt that this effort has been invaluable in debugging, reusing and extending the software as the project has developed.
- **Use of a Wizard of Oz study as basis for initial development** - Wizard of Oz studies (where a user believes that they are conversing with a computer, but are actually conversing with a human) offer a unique development opportunity for conversational systems. The Wizard of Oz study ([20]) carried out to support this project proved essential in providing insights into the users' likely responses, allowing appropriate conversational logic to be developed in the prototype systems. This logic was further developed following feedback from trials and use of the system.
- **Iterative Development** - experience (in this study, in commercial development and in other research (e.g. [22])) has shown that it is the nature of conversational systems that they require a number of iterations. Each user trial, such as that reported in this paper, identifies new shortfalls in both the conversational logic and the language. Examples would include identifying spelling errors that had not been envisaged originally (for example "qweetuons", "qwestens" and "qwstions" for 'questions') and adapting the language to suit a younger audience. The Lingubot development approach uses a 'Safety Net' to catch user inputs that have not been handled by the established logic. Studying the conversational flow that led to this point is a valuable element in identifying modifications required to the chatbot scripting.

5.2 User Engagement

The early and continuous engagement with the target user community has been an essential element in the development of the system.

- **Understanding User Needs and Environment** - Throughout this research programme, contact and communications were maintained with education specialists, including a survey of education professionals regarding the potential for Open Learner Modelling and its relationship with Assessment for Learning [16]. The understanding that this has developed has been an essential element in building a system that meets the users' needs.
- **Gaining User Support** - It would be impossible to conduct a study of this kind without the active co-operation of schools and other education establishments. Teachers are busy people, and could not be blamed for not supporting trials such as the one described in this paper. Understanding of the needs, environment and ethos of teachers allowed an approach to the trial that

Kerly, A., Ellis, R. & Bull, S. (2007). CALMsystem: A Conversational Agent for Learner Modelling, in R. Ellis, T. Allen & M. Petridis (eds), Applications and Innovations in Intelligent Systems XV – Proceedings of AI-2007, 27th SGAI International Conference on Innovative Techniques and Applications of Artificial Intelligence, Springer Verlag 89-102.

required minimum effort for the individuals concerned, caused minimum disruption in the school timetables and offered teachers tangible benefits (such as additional insights into the performance of pupils).

6 Further research

To continue the development of this research, further trials have been scheduled. These will involve users of various age groups (from Primary School to University), and engage them for varying timeframes, from short exposures such as the reported study, to longer-term use over an academic year. These trials will aim to

- confirm the value of the system in promoting user reflection and improving self-assessment,
- provide further feedback to support the development of the system, in terms of its functionality, logic and the language used in the chatbot scripts.

As discussed in section 5.3, it is important that the use of systems such as this is valuable to both students and their teachers. We intend to explore the views of teaching professionals towards the system and to use their feedback to ensure that using the system is beneficial to learning and to developing valuable learning skills.

7 Conclusion

CALMsystem offers support for the provision of negotiation and discussion facilities in an Open Learner Model environment. Trials of the system have shown how a chatbot can be used to encourage the development of self-assessment skills. Reflection on learning and self-assessment are critical metacognitive skills that are recognised as being beneficial to learning, and strategies to promote the development of such skills are endorsed by current UK educational policy.

A trial, involving a class of 30 UK Primary school children in Year 4 (Key Stage 2, aged 8-9), has confirmed that the system is effective in improving self assessment, and that the provision of a chatbot element further improves this ability. The development of the system and the supporting user trials have raised a number of lessons that may be of value to other developers. These include general issues relating to system development, specific lessons regarding the development of chatbots and the value of early and continuous engagement with user communities.

Acknowledgements

The first author is funded by the UK Engineering and Physical Sciences Research Council. The authors are indebted to Elzware Ltd for their support and for the provision of free user licences.

References

1. Bull, S. and Pain, H. Did I say what I think I said, and do you agree with me?: Inspecting and Questioning the Student Model. in *World Conference on Artificial Intelligence in Education*. 1995. Charlottesville, VA: AACE.

Kerly, A., Ellis, R. & Bull, S. (2007). CALMsystem: A Conversational Agent for Learner Modelling, in R. Ellis, T. Allen & M. Petridis (eds), Applications and Innovations in Intelligent Systems XV – Proceedings of AI-2007, 27th SGA International Conference on Innovative Techniques and Applications of Artificial Intelligence, Springer Verlag 89-102.

2. Dimitrova, V., STyLE-OLM: Interactive Open Learner Modelling. *International Journal of Artificial Intelligence in Education*, 2003. 13: p. 35-78.
3. Kay, J. Learner Know Thyself: Student Models to give Learner Control and Responsibility. in *ICCE97 International Conference on Computers in Education*. 1997. Kuching, Malaysia.
4. Morales, R., Exploring participative learner modelling and its effects on learner behaviour. 2000, University of Edinburgh: Edinburgh.
5. Zapata-Rivera, J.D. and Greer, J., Externalising Learner Modelling Representations, in *Workshop on External Representations of AIED: Multiple Forms and Multiple Roles. Artificial Intelligence in Education*. 2001: San Antonio, Texas.
6. Dewey, J., *How We Think: A Restatement of the Relation of Reflective Thinking to the Educative Process*. 1933, Boston: D. C. Heath and Company.
7. Schön, D., *The Reflective Practitioner*. 1983, London: Maurice Temple Smith Ltd.
8. Kolb, D., The Process of Experiential Learning, in *The Experiential Learning: Experience as the Source of Learning and Development*, D. Kolb, Editor. 1984, Prentice-Hall: NJ.
9. Papaleontiou-Louca, E., The Concept and Instruction of Metacognition. *Teacher Development*, 2003. 7(1): p. 9-30.
10. Paris, S.G. and Winograd, P., How Metacognition Can Promote Academic Learning and Instruction, in *Dimensions of Thinking and Cognitive Instruction*, B.F. Jones and L. Idol, Editors. 1990, Lawrence Erlbaum: Hillsdale. p. 15-51.
11. Butler, D.L. and Winne, P.H., Feedback and Self-Regulated Learning: A Theoretical Synthesis. *Review of Educational Research*, 1995. 65(3): p. 245-281.
12. Schoenfeld, A.H., What's All the Fuss About Metacognition?, in *Cognitive Science and Mathematics Education*, A.H. Schoenfeld, Editor. 1987, Lawrence Erlbaum: Hillsdale. p. 189-215, cited by [9].
13. Alevin, V. and Koedinger, K. Limitations of Student Control: Do Students Know When They Need Help? in *Intelligent Tutoring Systems*. 2000: Springer Verlag.
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*. 2001. San Antonio.
15. Black, P. and Wiliam, D., *Inside the black box: raising standards through classroom assessment*. 1998, London King's College London, School of Education.
16. Kerly, A. and Bull, S. Open Learner Models: Opinions of School Education Professionals. in *Artificial Intelligence in Education*. 2007. IOS Press, Amsterdam. p. 587-589.
17. Bull, S., et al. Reactions to Inspectable Learner Models: Seven Year Olds to University Students. in *Workshop on Learner Modelling for Reflection. International Conference on Artificial Intelligence in Education*. 2005. Amsterdam.
18. Bull, S. and McKay, M. An Open Learner Model for Children and Teachers: Inspecting Knowledge Level of Individuals and Peers. in *Intelligent Tutoring Systems: 7th International Conference*. 2004. Berlin-Heidelberg: Springer Verlag.
19. Giraffa, L. and Viccari, R. The Use of Agents Techniques on Intelligent Tutoring Systems. in *International Conference of the Chilean Computer Science Society*. 1998. Antofagasta, Chile: IEEE.
20. Kerly, A., Hall, P. & Bull, S., Bringing Chatbots into Education: Towards Natural Language Negotiation of Open Learner Models. *Knowledge Based Systems* 2006. 20(2): p. 177-185.
21. Creative Virtual. *Creative Virtual UK web site*. 2004-2006 [cited 1/4/2007]; Available from: www.creativevirtual.com.
22. Sparck-Jones, K., Natural language processing: a historical review, in *Current Issues in Computational Linguistics: in Honour of Don Walker*, A. Zampolli, et al., Editors. 1994, Kluwer: Amsterdam. p. 3-16