Steps Towards a Challenging Teachable Agent

Annika Silvervarg[1], Camilla Kirkegaard[1], Jens Nirme[2], Magnus Haake[2], Agneta Gulz[1]

[1] Dept. Computer and Information Science, Linköping University, Linköping, Sweden
{annika.silvervarg, camilla.kirkegaard, agneta.gulz}@liu.se
[2] Cognitive Science, Lund University, Kungshuset, Lundagård, Lund, Sweden
{jens.nirme, magnus.haake}@lucs.lu.se

Abstract. This paper presents the first steps towards a new type of pedagogical agent – a Challenger Teachable Agent, CTA. The overall aim of introducing a CTA is to increase engagement and motivation and challenge students into deeper learning and metacognitive reasoning. The paper discusses desired design features of such an agent on the basis of related work and results from a study where 11-year old students interacted with a first version of a CTA in the framework of an educational software for history. The focus is on how students respond when the CTA disagrees and questions their suggestions, and how groups of students, differing in response behavior and in self-efficacy, experience the CTA.

Keywords. Teachable agent, challenge, interaction, learning, experience.

1 Introduction and Background

This paper explores students' interaction with and responses to a new type of pedagogical agent – a challenger teachable agent (CTA). In brief, a teachable agent (TA) is an embodied computer agent which is taught or trained by a student where AI techniques guide the agent's behaviour based on what the agent is taught [1]. Importantly, a TA has no knowledge to begin with, but the knowledge that it gains reflects, more or less, what it is being taught by the student.

Overall, teachable agents have proven pedagogically powerful as an implementation of the *learning by teaching* pedagogy [2, 3]. Our goal is to boost this pedagogical power even further by introducing a teachable agent with a more explicit agency or "will of its own". To our knowledge, the TAs developed so far do not show much of a "will". They do not, for example, argue with a student on whether a piece of information is adequate for a task or not, or indicate that they find a particular topic uninteresting.

Regarding pedagogical agents in general, Frasson and Aïmeur [4] proposed *troublemaker agents* as a subset of learning companions that would question and challenge a student. Such an agent suggests a solution and then asks the student if she agrees or not. If the student does not agree, the troublemaker will argue for her or his own solution – whether it is correct or not – until the student either agrees or the troublemaker runs out of arguments. The students are thereby encouraged to question their own knowledge

and be more motivated as teachers. Several studies show learning gains from trouble-maker agents, particularly for high-achieving students, e.g. [4, 5].

Even one of the seminal papers on teachable agents by Brophy et al. [1] proposed a teachable agent that "may be impetuous, not listen or collaborate well". The implementation and study presented in this article is, however, the first practical attempt in this direction. In an educational software for learning history, we introduce a CTA that, during learning activities where the CTA and student work together and take turns, questions and challenges the student in various ways.

The main motive for introducing a CTA is to stimulate deep learning (c.f. [5]). By being questioned and challenged at times by his or her TA we hope the student becomes encouraged to think once more and perhaps reorganize or rephrase the material he/she is teaching [6]. We also wish to stimulate metacognitive abilities, i.e. reflection on problem solving and learning, abilities that have a transitive value for students when faced with future challenges [7]. In addition, our previous studies have shown that students frequently ask for a TA "with more of an attitude". It can be boring to interact with an agent that is always positive, compliant and cheerful – and such agents are weak in believability [8].

In the longer run, we are implementing the following set of challenging behaviours in our CTA.

1. The CTA inducing confusion or cognitive disequilibrium by contradicting the student with the aim to provoke the student to reflect on what is true, thereby processing the material at a deeper level [9].
2. The CTA requesting clarification of a solution in a learning activity, thus creating opportunities for debating the study material before accepting it, potentially prompting a desire in the student to share meaning and be understood [10]. This kind of teachable agent behaviour is represented in *SimStudent* [11] where the TA interacts with the student in natural language while solving equations. The TA tries to solve a problem step-by-step and the user verifies the correctness of each step. The TA can ask follow-up questions during this process, which forces the student to reflect on the concepts in the current problem and to show how well he/she understands the material.
3. The CTA occasionally introducing erroneous facts during the learning activities. This will hopefully provoke the student to justify his/her answers [4]. Training to distinguish between right and wrong solutions is also a means to achieve higher confidence in a domain.
4. The CTA prompting the student to choose a task at a more challenging level.

For the study presented in this paper we focus on this first aspect, i.e. the CTAs questioning of the students' proposals – both in cases where the students' proposal is correct and when it is not. To question someone you are collaborating with means to introduce a conflict. Several authors discuss the educational potential of conflict or dissonance in other areas than that of teachable agents. For instance, one of the five modes that Weinberger and Fischer [12] lift forth in their analysis of different "social modes of co-construction" in students' collaborative activities is that of "conflict-oriented

Fig. 1. The Castle of Time with the Guardian of History to the left and the Time elf (the teachable agent) to the right.

consensus building". In this mode, a conflict is the starting point wherefrom critical reasoning at some point leads to a further step taken together, i.e. in consensus.

Therefore, as a first step in the design of a CTA, this paper presents a study of students' responses when their TA questions their proposals in a collaborative learning activity. Our two main research questions were:

1. "What do students do when the CTA questions their proposals?"
2. "How do students perceive collaboration with a CTA with respect to the agent's questioning and/or challenging behaviour?"

We predicted that the answers to both research questions would vary for students with high and low self-efficacy respectively, i.e. with a strong vs. weak belief in their own competence.

2 The Guardian of History – TA-based Learning Environment

"The Guardian of History" is an educational software for history for age 10-12 year. The narrative centres on securing a successor to the Guardian of History, who is in charge of the Castle of Time but about to retire, see Fig. 1. The student takes on a teacher role and his/her task is to teach a time elf, i.e. a teachable agent, about history so that the time elf can qualify as successor to the Guardian of History. To do this, the student must first learn for him-/herself. Thereafter the student teaches the time elf during various game-like learning activities. The learning activity used in this study focuses on basic facts where a person is linked to an event and a time period, and placed on a timeline, see Fig. 2.

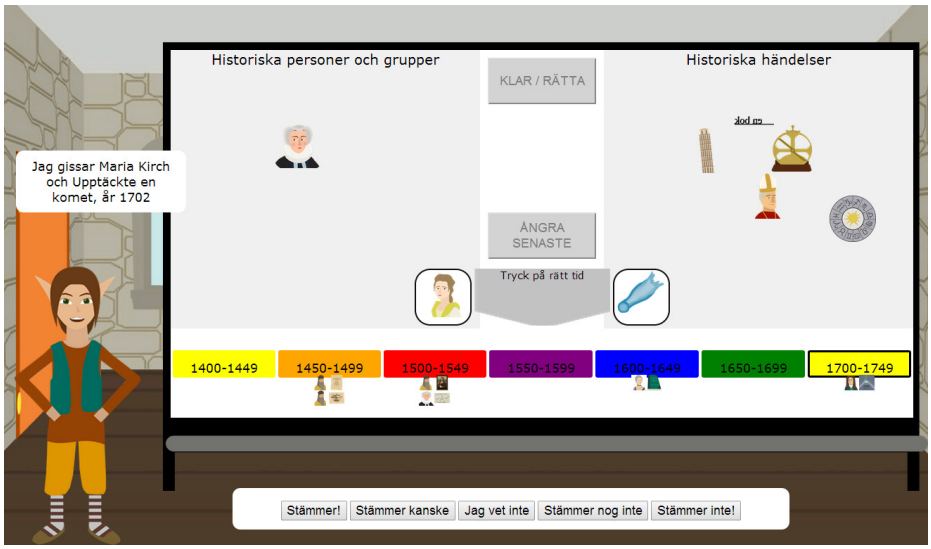

Fig. 2. The timeline learning activity. The time elf has suggested that Maria Kirch discovered a comet in 1702 and the student can choose to confirm or reject this proposal.

The time elf (TA) can be taught in two different modes:

1. By watching the student perform the learning activity on his/her own
2. Through performing the learning activity together with the student.

In the latter case, the student and TA take turns suggesting a fact and giving feedback to one another's suggestions via a multiple choice dialogue. When the TA suggests a fact the student can confirm or reject it, and similarly, when the student makes a proposal the TA can say that it believes it to be correct or incorrect. If the TA rejects the student's proposal, the student can chose to nevertheless affirm it – or the student can ask the TA to suggest an alternative instead, for example:

Student [*suggests*]: "Galileo Galilei invented the telescope between 1650 and 1700."
TA [*rejects*]: "I think the time period is wrong. Are you sure?"
Student [*affirms*]: "I am sure." [*or withdraws*]: "Do you have another suggestion?"

The TAs knowledge base consists of facts, i.e. <Person, Event, Time period>, that have a certainty between 1 and 10. The TA starts out with an empty knowledge base and learns by adding new facts with certainty 1 when it observes the student working on the timeline activity. If a fact is repeated the agent increases the certainty with 1. If the student adds facts that are incorrect or contradicting, e.g. <Galileo Galilei, Invented telescope, 1600-1649>, which is correct, and <Galileo Galilei, Invented telescope, 1650-1699>, which is incorrect, they are both added to the TA's knowledge base.

The *do together mode* is unlocked when the TA's knowledge base is sufficiently rich in facts (for the present study this happened after 5 facts). In this first version of the

system the agent uses only its existing knowledge to make suggestions or question suggestions from the student. Thus it is erroneous facts introduced by the students themselves in their teacher role that the agent uses in making incorrect suggestions or opposing correct suggestions by the students. The agent expresses rejections using the same phrasing regardless of whether the student's suggestion was correct or incorrect, e.g. "I think the time period is wrong". (In the future these behaviours will be purposely manufactured by the TA based on correct facts and using linguistic forms that express more personality and sound more challenging.)

In the *do together mode* the agent adjusts the certainty of the facts known with an amount of: +2, +1, 0, -1, -2, according to whether the student confirms them on a scale ranging over: Correct, Possibly correct, I don't know, Probably incorrect, Incorrect (see Fig. 2.) If a fact reaches a certainty of zero or below it is removed from the TA's knowledge base.

3 Study

20 female and 15 male 11-year olds from two classes in a Swedish school participated in the study. The students used the educational software and interacted with their TAs during two subsequent lessons, each with a duration of 25-35 minutes. The students' actions and choices were logged by the software. After the last session the students filled out a questionnaire regarding their experience of using the software and their view on the role of the teachable agent, the TA's willingness to cooperate, and different aspects of performing learning activities together with the TA. They were also asked about what they thought about their own ability to teach the TA (i.e. their self-efficacy). The questionnaire used a 5 point Likert scale ranging from "Strongly disagree" to "Strongly agree".

The logs of the students' behaviour showed that the students proposed a total of 161 correct propositions, of which the TA rejected 40 (25%), and 615 incorrect propositions, of which the TA rejected 137 (22%), see Fig. 3. Thus, the TA rejected approximately one fourth (23%) of all propositions suggested by the students.

3.1 Research question 1 – Students responses to CTA behaviour

Turning to our first research question: "What do students do when the CTA questions their proposals?" the log data showed that (see Fig. 3):

- *Case 1:* the students incorrectly withdrew 21 of 40 (53%) of their originally correct proposals when the proposal was rejected by the CTA.
- *Case 2:* the students correctly withdrew 117 of 137 (85%) of their originally incorrect proposals after having the proposal rejected by the CTA.

Thus, the students correctly withdrew significantly more of their proposals when they were correctly rejected by the CTA (Case 2) as evaluated by a Mann-Whitney's U test (Case 1: median = 0.5, mean rank = 19.1; Case 2: median = 1.0, mean rank = 28.8; $U = 172.5$; $Z = -2.484$; $p < .05$; $r = .35$).

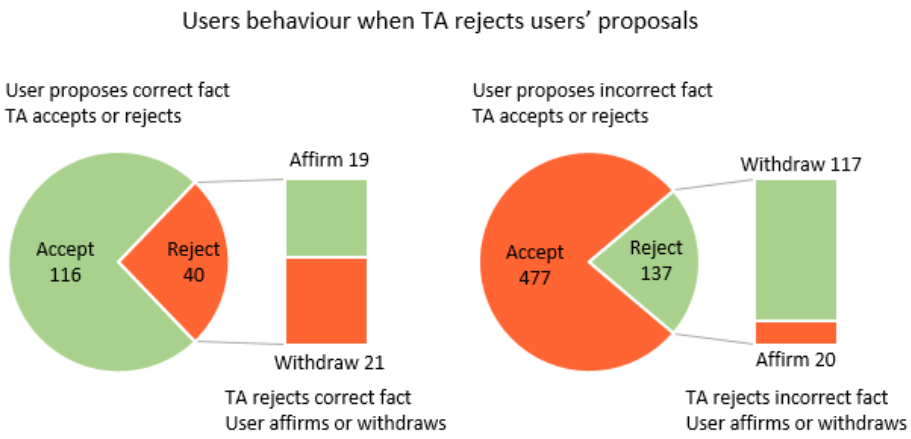

Fig. 3. The pie charts show how the TA reacts to users' correct or incorrect suggestions, and the stacked bar charts show how the user in turn reacts to the agents incorrect or correct rejection of their suggestions.

For the students that answered "Strongly disagreed" or "Disagreed" versus "Agreed" or "Strongly agreed" on the questionnaire item "The TA rejects my suggestions too often", no overall pattern could be secured due to few data points. Students who experienced that their TA rejected their suggestions "too often" withdrew 42% of their originally correct proposals compared to 74% for the students who didn't experience that the TA rejected their suggestions "too often" (see Fig. 4). Regarding the situation of an originally incorrect proposal by a student that was questioned by the TA, the students who experienced that the TA rejected their suggestions "too often" withdrew their proposals in 81% of the cases, compared to 95% for students who didn't experience that the TA rejected their suggestions "too often".

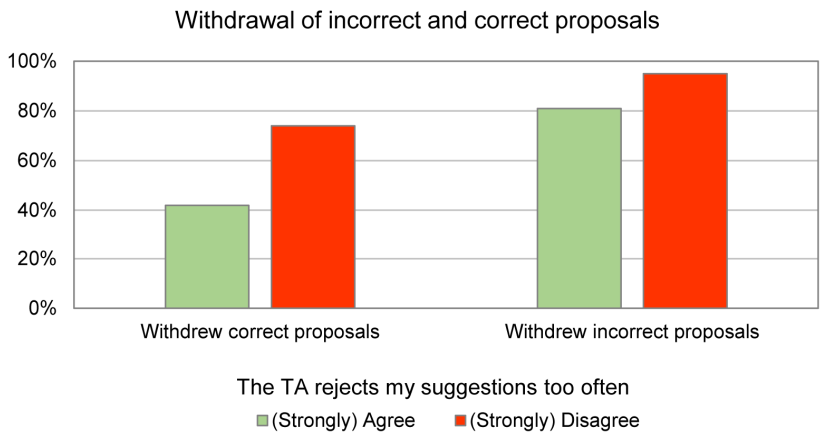

Fig. 4. The figure shows how often students withdrew correct/incorrect proposals for the two groups "(Strongly) Agree" and "(Strongly) Disagree" to "TA rejects my suggestions too often".

One striking result is how strongly inclined the students were to "go with the TA" when the TA objected to their proposal (138 withdrawals out of 177 agent rejections) giving an average of 78% for all students which is well above chance ($\chi^2(1, N = 177) = 55.4$, $p < .001$). But interestingly enough there is one diverging case where students do not "go with the TA" to the same extent. This regards students who "Agreed" or "Strongly agreed" to the item "the TA rejects my suggestions too often" and where the CTA questioned the student's correct proposal. Here the students "went with the TA" in only 42% (see above) of the situations. For the case students who "(Strongly) Agreed" and correct withdrawal there was (marginally) significant differences to all the three other cases (see Fig. 4) as suggested using logistic regression, see Table 1 below.

Table 1. Logistic regression showing the contrasts for a:i, d:c, and d:i against a:c corresponding to the groups (bars) in Fig. 4: (a:c) correct withdrawals for students who answered "(Strongly) Agree"; (a:i) incorrect withdrawals for students who answered "(Strongly) Agree"; (d:c) correct withdrawals for students who answered "(Strongly) Disagree"; (d:i) incorrect withdrawals for students who answered "(Strongly) Disagree".

Contrasts	*Z*	*p*
a:c vs. a:i	2.439	0.015 *
a:c vs. d:c	1.743	0.081 .
a:c vs. d:i	2.729	0.007 **

*. p < 0.1 * p < 0.05 ** p < 0.01 *** p < 0.001*

3.2 Research question 2 – Perception of collaboration with the CTA

Our second research question was: "How do students perceive collaboration with a CTA with respect to the agent's questioning and/or challenging behaviour?" Analysing the questionnaires we found uniform distributions regarding students' perceptions: (i) of the TAs desire to collaborate ("The TA wants to work with me") and (ii) on whether the TA rejected their suggestions to often. Regarding the question "It is more fun to work together with the TA than to work by myself" most students "Agreed" or "Strongly agreed". The distributions of the answers are shown in Fig. 5.

The actual frequency of how often the TA rejected the student's proposals did not correlate with the answers to the question "The TA rejects my suggestions too often" (Spearman's $r = -.17$, $p = .37$) for correct proposals, and (Spearman's $r = .01$, $p = .94$) for incorrect proposals.

The questionnaire item "I am good at teaching the agent" reflects a student's self-efficacy, i.e. her belief in her own ability to teach the subject in question to the agent. The scores for this item showed a negative correlation (Spearman's $r = -.59$, $p < .01$) with the item "The TA rejected my proposals too often". Students who "Agreed" or "Strongly agreed", i.e. found that their TA rejected their proposals "too often" (Self-efficacy: median = 4, mean rank = 16.7) had significantly lower self-efficacy (Mann-Whitney's U test: $U = 104.5$, $Z = 2.51$, $p < .05$) than those who

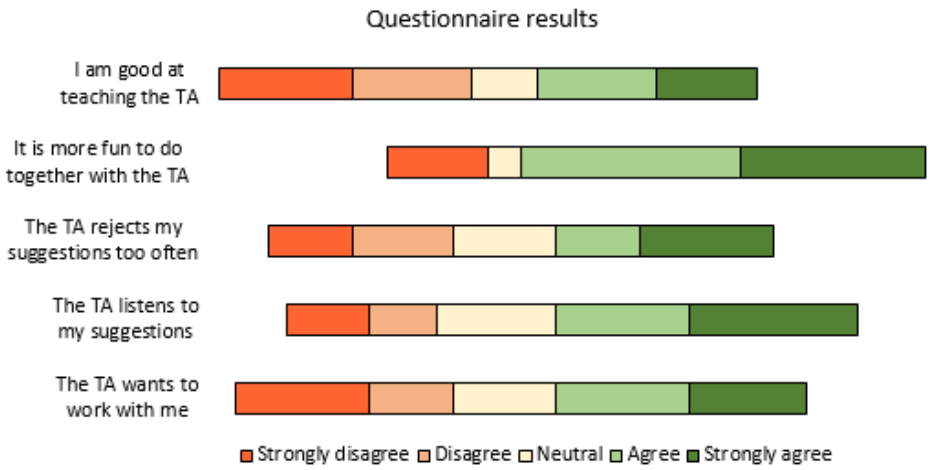

Fig. 5. The distribution of answers on a 5 point Likert scale from "Strongly disagree" to "Strongly agree" for questionnaire questions regarding the cooperation with the CTA.

"Strongly disagreed" or "Disagreed", i.e. did not find that their TA rejected their proposals "too often" (Self-efficacy: median = 2, mean rank = 9.0) [1].

4 Conclusions and Future Work

With respect to the first research question on students' responses to a CTA, the overall result was that students generally went along with their CTA's objections rather than holding on to their own initial proposal. Since they did so even when they were initially correct we did not achieve the desired result of improving the students' learning. However, most students experienced it was more fun to do the learning activity together with the CTA than to do the same activity by themselves. This is a positive result in view of our continued development of a CTA-based software, but how and when the CTA challenges the student needs to be improved. Some suggestions are discussed below as design recommendations for future development.

Regarding the students' attitudes towards their CTA, the most salient result was that students with higher self-efficacy (i.e. who saw themselves as good teachers for their CTA) did not experience that their CTA questioned them too often, whereas students with lower self-efficacy did experience this. In contrast, the *actual frequency* of challenging/questioning by the CTA did not affect the extent to which the students *experienced* that the TA was challenging/questioning them "too often". In other words, the experience of being challenged seems more strongly connected to students' personal preferences than actual occurrence of challenge by the CTA.

[1] The self-reported self-efficacy displayed a medium effect size correlation (Spearman's $r = .38$, $p < .04$) with the number of proposed correct propositions by the students, i.e. with a measurement of actual ability/performance.

In future iterations of our learning environment we intend to investigate this further by comparing effects from a CTA, a more traditional TA, and a version of the software that lacks a TA altogether. Our hypothesis is that students with higher self-efficacy will benefit more from a CTA versus a traditional TA, but that this will not be the case for students with lower self-efficacy. We will also include the student variables of goal-orientation and level of achievement and study their interaction with the type of agent (or no agent) used.

Based on the results presented here and observations during the classroom sessions we think that a problem with our current implementation of the CTA is that the students are overly confident in their agent's knowledge, which results in their accepting its incorrect suggestions. To address this we recommend the following two design considerations for future development of challenging teachable agents:

- *Clear communication of the agent's learning and knowledge base.* Encountering a computer artefact that knows less than oneself is an odd situation. The same applies to the situation where the student enrols in a teacher position. In order to handle this, we will develop a stronger narrative for the software that emphasizes that the CTA has *no initial knowledge* and that its knowledge state basically is a reflection of what the student has taught it. We will also let this be reflected in how the CTA voices its rejections to student proposals so that it does not seem too confident and knowledgeable but more as a tutee. The more the CTA is exposed to a certain fact the more certain it should become, but initially it should be quite uncertain.
- *Adjust magnitude, frequency and timing of challenging behaviour.* A student must stand on a solid ground to meet challenges in the form of rejection or errors from the CTA. Students need to have mastered the learning material to a certain degree before they are challenged, and/or the magnitude of the errors must be adjusted to the students' knowledge level so that they can be detected. We will address this by letting the student initially engage in each learning activity *without* the CTA in order to build a threshold knowledge level before starting to teach and work together with the CTA.

5 References

1. Brophy, S., Biswas, G., Katzlberger, T., Bransford, J., Schwartz, D.: Teachable Agents: Combining Insights from Learning Theory and Computer Science. In: Lajoie, S.P., Vivet, M. (eds.) Artificial Intelligence in Education, pp. 21–28. IOS Press, Amsterdam (1999)
2. Chase, C., Chin, D., Oppezzo, M., Schwartz, D.: Teachable Agents and the Protégé Effect: Increasing the Effort Towards Learning. J. Sci. Educ. Technol. 18(4), 334–352 (2009)
3. Blair, K., Schwartz, D., Biswas, G., Leelawong, K.: Pedagogical Agents for Learning by Teaching: Teachable Agents. Educational Technology: Special issue on Pedagogical Agents, 47(1), 56–61 (2007)
4. Frasson, C., Aïmeur, E.: A Comparison of Three Learning Strategies in Intelligent Tutoring Systems. J. Educ. Comput. Res. 14(4), 371–383 (1996)
5. Aïmeur, E., Dufort, H., Leibu, D., Frasson, C. (1997). Some Justifications About the Learning By Disturbing Strategy. In: du Boulay, W., Mizoguchi, R. (eds.) Artificial Intelligence in Education, pp. 1–14. IOS Press, Amsterdam (1997)

6. Schwartz, D., Chase, C., Chin, D., Oppezzo, M., Kwong, H.: Interactive Metacognition: Monitoring and Regulating a Teachable Agent. In: Hacker, D.J., Dunlosky, J., Graesser, A.C. (eds.) Handbook of Metacognition in Education, pp. 340–358. Routledge (2009)
7. Chin, D.B., Dohmen, I.M., Cheng, B.H., Oppezzo, M.A., Chase, C.C., Schwartz, D.L.: Preparing Students for Future Learning with Teachable Agents. Educ. Technol. Res. Dev. 58(6), 649–669 (2010)
8. Cassell, J., Thórisson, K.R.: The Power of a Nod and a Glance: Envelope vs. Emotional Feedback in Animated Conversational Agents. Appl. Artif. Intell. 13, 519–538 (1999)
9. D'Mello, S., Lehman, B., Pekrun, R., Graesser, A., Confusion can be Beneficial for Learning, Learning and Instruction, 29, 153–170 (2014)
10. Schwartz, D. (1999). The Productive Agency that Drives Collaborative Learning. In: Dillenbourg, P. (ed.) Collaborative learning: Cognitive and computational approaches, pp. 197–218. Permagon, Amsterdam (1999)
11. Carlson, R., Matsuda, N., Koedinger, K. R., & Rose, C. (2012). Building a Conversational SimStudent. In S. Cerri & W. Clancey (Eds.), Proceedings of International Conference on Intelligent Tutoring Systems (pp. 563-569). Heidelberg/Berlin: Springer-Verlag.
12. Weinberger, A., Fischer, F.: A Framework to Analyze Argumentative Knowledge Construction in Computer-Supported Collaborative Learning. Computers & Education, 46(1), 71–95 (2006)