Academically Productive Talk: One Size Does Not Fit All

David Adamson & Carolyn P. Rosé

Language Technologies Institute, Carnegie Mellon University, Pittsburgh PA, 15213
{dadamson,cprose}@cs.cmu.edu

Abstract. We present a study in which we experimentally manipulate the form of support offered to groups of three students during collaborative learning. Specifically, we contrast two forms of Academically Productive Talk (APT) facilitation, known as Revoicing and Agree-Disagree. The first form has been demonstrated effective with the target age group (i.e., 9th grade) on an earlier more difficult unit. The second form has been demonstrated effective with older kids. Results suggest that with this age group, facilitation with Revoicing may be more effective than Agree-Disagree. Implications for future work are discussed.

Keywords: dynamic support for collaborative learning, academically productive talk, discussion for learning.

1 Introduction

Collaborative learning activities, when delivered effectively, can provide significant cognitive, metacognitive, and social benefits to students [18][32][35]. Studies in the field of computer-supported collaborative learning have demonstrated the pedagogical value of social interaction [37][38]. Prior work on adaptive support for collaborative learning has adapted hint-based support originally developed for individual learning to support peer tutoring [13], and other work has grown out of earlier efforts to develop tutorial dialogue agents originally designed for individual learning [16][30][40][41]. This form of dynamic agent-based support for collaborative learning was historically tailored to specific learning populations and content domains [22], which limits its generality. More generalizable forms of support would increase the potential for impact, but as we discuss in this paper, raise new questions about principles for adaptation that would enable us as system developers to provide solutions that can be effective for diverse student populations.

Our recent efforts are in the direction of intelligent conversational agents acting as discussion facilitators, offering support behaviors that are not tied to a particular content-area or context [1][10][14]. The design of such support is in line with the literature on facilitation of collaborative learning groups [17]. In particular, it draws upon a body of work that has shown that certain forms of classroom discussion facilitation, termed Accountable Talk, or Academically Productive Talk (APT), are beneficial for learning with understanding [3][8][9][28][29][33][34][39].

In this paper we present results from a study in which we contrast two forms of APT based support. The first form, Revoicing support, has been found in prior work to achieve positive learning effects with the target student population of 9th graders [14] on an earlier and more difficult lesson. The other form of support, Agree-Disagree support, has been found to be effective with older, more advanced learners [1] in a different content domain. In this study, we show that with a 9th grade student population, Revoicing support is slightly more effective that Agree-Disagree support. These results contribute towards an empirical foundation for adapting APT based support to differences in content domain difficulty and differences in the developmental stage of target learners.

In the remainder of the paper we first review the state of the art in agent based support for collaborative learning. Next we describe two forms of APT-based support. Then we describe an evaluation study where we compare the effectiveness of these two forms of support for 9th grade biology students working on a genetics unit that is relatively easy for them. We conclude with discussion of results and future directions.

2 Prior Work

Academically Productive Talk has grown out of frameworks that emphasize the importance of social interaction in the development of mental processes. Michaels, O'Connor and Resnick [26] describe a number facilitating moves that teachers can employ to promote student-centered classroom discussion. A selection of these moves are presented in Table 1. In studies where teachers used similar facilitation strategies, students showed dramatic improvement on standardized math scores, transfer to reading test scores, and retention of transfer for up to 3 years [8][9].

Table 1. Selected Accountable Talk Moves

APT Move	Example
Revoicing a student's statement	"So, let me see if I've got your thinking right. You're saying XXX?"
Asking students to apply their own reasoning to someone else's reasoning	"Do you *agree or disagree*, and why?"

Collaboration scripts are a common way to describe and structure support for collaborative learning [20] within the field of computer-supported collaborative learning. A collaboration script may describe any of a wide range of features of collaboration scenarios, including the tasks, timing, roles, and the methods and desired patterns of interaction between the participants. A script can describe the collaborative activity at the macro or micro level [12]. Macro-scripts describe the sequence and structure each phase of a group's activities, specifying coarse-grained features such as assigned tasks and roles, and the overall shape of the activity. Micro-scripts, on the other hand, are models of dialogue and argumentation embedded in the activity, and are intended to be adopted and progressively internalized by the participants [19]. Micro-scripts can be realized by sharing prompts or hints with the user, guiding or providing models for their contributions [36]. While traditional collaboration scripts such as these can provide some degree of support for conversational and reasoning practices, they fall short of delivering the active, engaged facilitation described by the APT literature.

In particular, such scripts are static, and do not respond to changes in (or awareness of) student need or ability during the activity. Such non-adaptive approaches risk detrimental over-scripting [11]. More preferable would be the delivery or adjustment of supports in response to the automatic analysis of student activity [2][31]. The collaborative conversational agents described by Kumar and Rosé [24] were among the first to implement such dynamic scripting in a CSCL setting, with demonstrable gains over otherwise equivalent static support. Likewise, recent work by Baghaei et al [6] and Diziol et al [13] show that adaptive supports can have meaningful effects on student learning and interaction.

3 Dynamic Support for Academically Productive Talk

Two dynamic conversational supports based upon APT facilitation, namely Revoicing and Agree-Disagree, were implemented and evaluated in this study. The open-source Bazaar architecture [2] was used to author and orchestrate the conversational agent and the support behaviors described below.

3.1 Revoicing Support

One of the forms of support evaluated in this paper is a Bazaar component that performs an Academically Productive Talk move referred to as Revoicing. The agent compares student statements against a list of conceptually correct statements developed with teachers. In the study described in this article, 35 such statements were developed and validated against pilot data. For each student turn, we calculate a measure of "bag of synonyms" cosine similarity against each expert statement, based on the method described by Fernando and Stevenson [15]. If this similarity value exceeds a conservatively high threshold, we consider the student's turn to be a possible paraphrase of the matched statement, and thus "revoicable" (this threshold was determined through tests against pilot data, such that at least 80% of the revoicings suggested for candidate student were on-target). The Revoicing component may respond by offering the matched statement as a paraphrase of the student's turn, for example "So what I hear you saying is XXX. Is that right?" No statement may trigger a revoice move more than once.

3.2 Agree-Disagree Support

The other support we evaluate is a Bazaar component which performs the APT Agree-Disagree move. Candidate student statements are identified using the same method as described for the Revoicing support, but with a lower threshold that allows looser matches. After detecting such a candidate, the agent waits for the other students in the group to respond to it. If another student responds with an evaluation of their peer's contribution (for example, "I agree" or "I think you're wrong", as recognized by a small list of hand-crafted regular expressions), but doesn't support the evaluation with an explanation, the agent will encourage this second student to provide one. If a student instead follows up with another APT candidate statement, the agent does nothing, leaving the floor open for productive student discussion to continue unimpeded, reducing the risk of over-scripting their collaboration. If the other students do not respond with either an evaluation or a contentful follow-up, the agent prompts them to comment on the candidate statement – for example, “What do you think about Billy’s idea? Do you agree or disagree?”

4 Method

Following the literature on APT used as a classroom facilitation technique, in this study we test the hypothesis that appropriate APT support in a computer-supported collaborative learning setting will both intensify the exchange of reasoning between students during the collaborative activity, and increase learning during the activity.

4.1 Instructional Content and Study Procedure

Participants: This study was conducted in seven 9th grade biology classes of an urban school district. The classes were distributed across two teachers (with respectively 3 and 4 classes) for a total of 143 students total, with 76 consenting. Students were randomly assigned to groups of 3. Groups were randomly assigned to conditions. Only data from consenting students was used in the analysis presented here.

Experimental Manipulation: This study was run as a 3 condition between subjects design in which the APT agents provided some behaviors in common across conditions, but other behaviors were manipulated experimentally. Across all conditions, the agent provided the same macro level support by guiding the students through the activity using the same phases introduced in such a way as to control for time on task. It was the micro-scripting behaviors that were manipulated experimentally in order to create the three conditions of the design. The first experimental conditions was Revoicing, using the behavior described above. The second was the Agree-Disagree condition, where the Agree-Disagree behavior discussed above was used. In the control condition, neither of these behaviors was used.

Learning Content: The study was carried out during a module introducing the concepts of genetics, heredity, and single-trait inheritance. In the activity, student groups were presented with a set of three problems and asked to reason about the physical and genetic traits of the likely parents of a set of siblings. Specifically, in each problem, students were shown a litter of eight kittens that varied in fur color (either orange or white), and were instructed to identify the genotypes and phenotypes of the parents, and to explain their reasoning to their teammates. This sort of “backwards” reasoning had not been explicitly addressed in the course to date – students only had prior experience with “forward” reasoning from given parental traits. The mystery parents were presented as the inputs to an unpopulated Punnett square, as shown in Figure 2. As an incentive, students were told that the best team, determined by a combination of discussion quality and post-test scores, would be awarded a modest prize. Each of the three tasks was progressively harder than the last in that fewer clues about the parent's identities were included. The collaborative task content, the macro-scripts that supported it, and the list of statements powering the APT support were all developed iteratively with feedback from teachers and content experts.

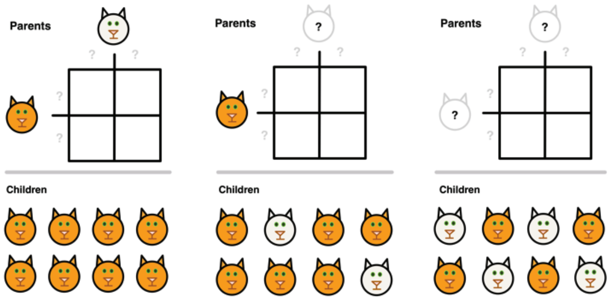

Fig. 1. Task sequence for the collaborative activity.

Study Procedure: The study was conducted over three phases, which occurred as single class periods over two school days. The first phase ("day 1") involved the teachers taking a pre-test at the end of a regular class session.

The second phase ("day 2") was centered around a 20 minute collaborative computer-mediated activity during which the experimental manipulation took place. The students performed the activity in groups of three, scaffolded by a conversational agent. Students within classes were randomly assigned to groups, then groups to conditions. The activity was introduced by a cartoon handout depicting the use of APT, and a ten-minute presentation describing the task and reviewing the basics of genetics and heredity. At the end of this second phase, the students took a post-activity test.

The computer activity was intended to equip the students with enough empirical data and attempts at reasoning to prepare them for the third phase ("day 3"), a full class APT discussion with their teacher, during which they would reconcile their different understandings and explanations. At the end of this discussion, they took a post-discussion test.

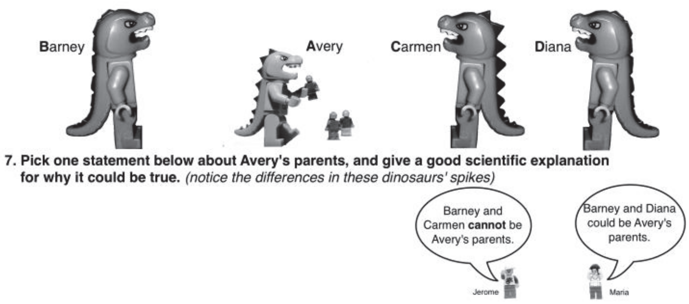

Fig. 2. Concept cartoon question from the post-activity test.

4.2 Measurement

Domain knowledge was measured at three time points using a paper based test. Each of the three tests (Pre-Test, Post-Activity Test, Post-Discussion Test) followed a similar format: a set of multiple choice problem-solving questions addressing forward and backward reasoning about single inheritance, and what we refer to as a *concept cartoon*, in which a set of potential parents for a single child was displayed, along with two hypotheses for who the child's parents might be. Students were instructed to select one hypothesis and clearly explain the conditions that would allow it to be true – either hypothesis could be correct, with different underlying assumptions. Student responses were graded with a rubric assessing the quality and depth of their explanation, including explicit displays of reasoning.

Each test covered the same knowledge but used different scenarios. The knowledge to be covered by each test was established in coordination with the teachers, with teacher trainers who identified common misconceptions, and with test results from a study run with the same content the previous year. After an initial round of consensus grading by two graders on a subset of the tests to establish a scoring guide, the remaining tests were divided and scored by one grader each.

Table 2. Total test scores (standard dev) for Pretest, Post-Activity Test, and Post-Discussion Test in the 3 conditions.

	Control	Revoice	Agree-Disagree
Pretest	5.5 (3.1)	5.5 (3.2)	3.9 (3.0)
Post-Activity Test	6 (3.4)	6.3 (3.1)	4 (3.1)
Post-Discussion Test	5.7 (3.1)	6.1 (2.9)	4.8 (3.3)

4.3 Results

First we tested whether students learned during the online activity. Test scores were divided into explanation questions and problem solving questions. Thus, for each test, each student has two scores. In order to evaluate learning, we used an ANOVA with Test Score as the dependent variable, Explanation vs Skill, Pretest vs Post-Activity Test, Condition, and Teacher as independent variables. We added Teacher as a variable because we noticed that students from one teacher learned significantly more than students from the other teacher. In this analysis, all of the independent variables were significant except Pre-test vs Post-test, which was marginal, $F(1, 270) = 3.6$, $p < .06$. There were no significant interactions between independent variables. Thus we find qualified evidence that students learned during the online activity, across conditions. However, on inspecting the average scores in Table 1, we see barely any evidence of learning in the Agree-Disagree condition. The most learning we see is about .25 standard deviations in the Revoicing condition, and about half that in the Control condition.

We also tested whether students learned during the Post-activity discussion. In this case, when comparing between the Post-Activity test and the Post-Discussion test there was no significant difference. In fact, the trend was that students scored more poorly on the Post-Discussion test than the Post-Activity test, except in the Agree-Disagree condition, where the students came into the discussion with less knowledge than students in the other two conditions, and seemed to be able to use the Post-activity Discussion to catch up, which is consistent with findings from earlier studies (Dyke et al., in press).

We compared learning across conditions between Pre-test and Post-Activity test, and between Pre-test and Post-Discussion test. In both cases, we used an ANCOVA with the posttest measure (i.e., Post-Activity test in the first comparison and Post-Discussion test in the second) as the dependent variable and the Pre-test as the covariate. We retained the Teacher variable in addition to the condition variable. In neither case do we find a significant effect of condition. However between the Pre-test and Post-activity test the trend is for adjusted posttest scores to be higher than the control condition in the Revoicing condition (by .13 standard deviations) and lower than the control condition in the Agree-Disagree condition (by .4 standard deviations), with very similar trends when comparing between Pre-test and Post-Discussion test.

We acknowledge that stronger claims could be made by conducting our analysis using multilevel modeling. However, such complex modeling techniques require larger data sets in order to avoid falling prey to type II errors during hypothesis testing. Due to the small size of our data, we employed simpler methods for our analysis.

5 Discussion & Conclusions

Overall, the results are weak. However, the results suggest a differential effect of the two experimental conditions. The trend in favor of the Revoicing condition is consistent with earlier studies with the same age group, but on a more difficult unit in the course [14]. The trend to learn less than the control condition in the Agree-Disagree condition is in contrast to earlier results with more advanced learners [1] where students in the Agree-Disagree condition learned significantly more than in the control condition. These suggestive results will need to be followed up with additional experimentation before we can have more confidence in the findings. However, they do suggest that the effect of these APT facilitation strategies on learning depend on the difficulty of the unit and the developmental stage of the learners, and that more results are needed to inform effective strategies for supporting groups of learners.

6 Acknowledgements

This work was supported in part by NSF grant SBE 0836012 to the Pittsburgh Science of Learning Center.

7 References

[1] Adamson, D., Ashe, C., Jang, H., Yaron, D., & Rosé, C. (2013). Intensification of Group Knowledge Exchange with Academically Productive Talk Agents. *Proceedings of the 10th International Conference on Computer Supported Collaborative Learning*, Madison Wisconsin, July 2013.

[2] Adamson, D., & Rosé, C. (2012). Coordinating multi-dimensional support in collaborative conversational agents. In *Proceedings of Intelligent Tutoring Systems* (pp. 346-351). Springer Berlin/Heidelberg.

[3] Adey, P., & Shayer, M. (1993). An Exploration of Long-Term Far-Transfer Effects Following an Extended Intervention Program in the High School Science Curriculum. *Cognition and Instruction, 11*(1), pp 1-29.

[4] Ai, H., Kumar, R., Nguyen, D., Nagasunder, A., Rosé, C. P. (2010). Exploring the Effectiveness of Social Capabilities and Goal Alignment in Computer Supported Collaborative Learning, in *Proceedings of Intelligent Tutoring Systems.*

[5] Azmitia, M. & Montgomery, R. (1993). Friendship, transactive dialogues, and the development of scientific reasoning, *Social Development* 2(3), pp 202-221.

[6] Baghaei, N., Mitrovic, A., & Irwin, W. (2007). Supporting collaborative learning and problem solving in a constraint based CSCL environment for UML class diagrams, *International Journal of Computer Supported collaborative Learning* 2(3), pp 159-190.

[7] Berkowitz, M., & Gibbs, J. (1983). Measuring the developmental features of moral discussion. *Merrill-Palmer Quarterly*, 29, pp 399-410.

[8] Bill, V. L., Leer, M. N., Reams, L. E., & Resnick, L. B. (1992). From cupcakes to equations: The structure of discourse in a primary mathematics classroom. *Verbum*, 1, 2, 63-85.

[9] Chapin, S., & O'Connor, C. (2004). Project challenge: Identifying and developing talent in mathematics within low-income urban schools. *Boston University School of Education Research Report* (Vol. 1, pp. 1-6).

[10] Clarke, S., Chen, G., Stainton, K., Katz, S., Greeno, J., Resnick, L., Howley, H., Adamson, D., Rosé, C. P. (2013). *The Impact of CSCL Beyond the Online Environment, Proceedings of Computer Supported Collaborative Learning*

[11] Dillenbourg, P. (2002). Over-scripting CSCL : The risks of blending collaborative learning with instructional design . *Three worlds of CSCL: Can we support CSCL*, pp. 61-91.

[12] Dillenbourg, P., Hong, F. (2008). The mechanics of CSCL macro scripts. *International Journal of Computer-Supported Collaborative Learning* 3(1), pp 5-23.

[13] Diziol, D., Walker, E., Rummel, N., & Koedinger, K. R. (2010). Using intelligent tutor technology to implement adaptive support for student collaboration. Educational Psychology Review, 22(1), 89-102.

[14] Dyke, G., Adamson, D., Howley, I., Rosé, C.P. (Under Review). Enhancing Scientific Reasoning and Explanation Skills with Conversational Agents, submitted to *IEEE Transactions on Learning Technologies.*

[15] Fernando, S., and Stevenson, M. (2008). A semantic similarity approach to paraphrase detection, *Computational Linguistics UK (CLUK 2008) 11th Annual Research Colloquium.*

[16] Graesser, A., VanLehn, K., the TRG, & the NLT (2002). Why2 Report: Evaluation of Why/Atlas, Why/AutoTutor, and Accomplished Human Tutors on Learning Gains for Qualitative Physics Problems and Explanations. *LRDC Tech Report, University of Pittsburgh.*

[17] Hmelo-Silver, C. E. & Barrows, H. S. (2006). Goals and Strategies of a Problem-based Learning Facilitator. *The Interdisciplinary Journal of Problem Based Learning*, 1(1), pp 21-39.

[18] Kirschner, F., Paas, F., & Kirschner, P. A. (2009). A cognitive load approach to collaborative learning: United brains for complex tasks. *Educational Psychology Review*, 21, 31–42.

[19] Kobbe, L., Weinberger, A., Dillenbourg, P., Harrer, A., Hämäläinen, R., Häkkinen, P., Fischer, F. (2007). Specifying computer-supported collaboration scripts. *The International Journal of Computer-Supported Collaborative Learning* 2(2-3), pp 211-224.

[20] Kollar, I., Fischer, F., Hesse, F.W. (2006). Collaborative scripts - a conceptual analysis. *Educational Psychology Review* 18(2), pp 159-185.

[21] Kumar, R., Gweon, G., Joshi, M., Cui, Y., Rosé, C. P. (2007a). Supporting Students Working Together on Math with Social Dialogue. *Proceedings of the SLaTE Workshop on Speech and Language Technology in Education*

[22] Kumar, R., Rosé, C. P., Wang, Y. C., Joshi, M., Robinson, A. (2007b). Tutorial Dialogue as Adaptive Collaborative Learning Support, *Proceedings of Artificial Intelligence in Education.*

[23] Kumar, R., Ai, H., Beuth, J., Rosé, C. P. (2010). Socially-capable Conversational Tutors can be Effective in Collaborative Learning Situations, in *Proceedings of Intelligent Tutoring Systems*.

[24] Kumar, R., Rosé, C.P. (2011). Architecture for Building Conversational Agents that Support Collaborative Learning. *IEEE Transactions on Learning Technologies* 4(1).

[25] Lison, P. (2011). Multi-Policy Dialogue Management. In *Proceedings of the SIGDIAL 2011 Conference*, Association for Computational Linguistics, pp. 294-300

[26] Michaels, S., O'Connor, C., & Resnick, L.B. (2007). Deliberative discourse idealized and realized: Accountable talk in the classroom and in civic life. *Studies in Philosophy and Education*.

[27] Rabe-Hesketh, S., Skrondal, A., & Pickles, A. (2004). *GLLAMM Manual*. University of California, Berkely. U. C. Berkeley Division of Biostatistics Working Paper Series, Paper 160.

[28] Resnick, L. B., Asterhan, C. A., & Clarke, S. N. (in press). Socializing Intelligence through Academic Talk and Dialogue. *Washington, DC: American Educational Reserach Association*.

[29] Resnick, L. B., Salmon, M., Zeitz, C. M., Wathen, S. H., & Holowchak, M. (1993). Reasoning in conversation. *Cognition and Instruction*, 11(3-4), 347-364.

[30] Rosé, C. P., Jordan, P., Ringenberg, M., Siler, S., VanLehn, K., Weinstein, A. (2001). Interactive Conceptual Tutoring in Atlas-Andes, *Proceedings of AI in Education*.

[31] Rosé, C. P., Wang, Y.C., Cui, Y., Arguello, J., Stegmann, K., Weinberger, A., Fischer, F. (2008). Analyzing collaborative learning processes automatically: Exploiting the advances of computational linguistics in computer-supported collaborative learning. *The International Journal of Computer-Supported Collaborative Learning* 3(3), 237-271.

[32] Scardamalia, M., & Bereiter, C. (1993). Technologies for knowledge-building discourse. *Communications of the ACM*, 36(5), 37–41.

[33] Topping, K. J., & Trickey, S. (2007a). Collaborative philosophical enquiry for school children: Cognitive effects at 10-12 years. *British Journal of Educational Psychology*, 77(2), 271-288.

[34] Topping, K. J., & Trickey, S. (2007b). Collaborative philosophical inquiry for schoolchildren: Cognitive gains at 2-year follow-up. *British Journal of Educational Psychology*, 77(4), 787-796.

[35] Webb, N. M., & Palinscar, A. S. (1996). Group processes in the classroom. In D. C. Berliner & R. C. Calfee (Eds.), *Handbook of educational psychology* (pp. 841–873). New York: Prentice Hall.

[36] Wecker, C., Fischer, F. (2007). Fading scripts in computer-supported collaborative learning: The role of distributed monitoring. *Proceedings of the 8th iternational conference on Computer Supported Collaborative Learning*, pp. 764-772.

[37] Weinberger, A., Stegmann, K., & Fischer, F. (2007). Scripting argumentative knowledge construction: Effects on individual and collaborative learning. In C. Chinn, G. Erkens, & S. Puntambekar (Eds.), *Mice, minds, and society: CSCL 2007* (pp. 37-39). New Brunswick, NJ: International Society of the Learning Sciences.

[38] Weinberger, A., Fischer, F. (2006). A framework to analyze argumentative knowledge construction in computer-supported collaborative learning. *Computers & Education* 46(1), 71–95.

[39] Wegerif, R., Mercer, N., & Dawes, L. (1999). From social interaction to individual reasoning: an empirical investigation of a possible socio-cultural model of cognitive development. *Learning and Instruction*, 9(6), 493-516.

[40] Wiemer-Hastings, P., Graesser, A., Harter, D. and the Tutoring Research Group, (1998). The Foundations and Architecture of AutoTutor. In B. Goettl, H. Halff, C. Redfield & V. Shute (Eds.) *Intelligent Tutoring Systems: 4th International Conference* (ITS '98) (pp334-343). Berlin: Springer-Verlag.

[41] Zinn, C., Moore, J. D., & Core, M. G. (2002). A 3-Tier Planning Architecture for Managing Tutorial Dialogue. In S. A. Cerri, G. Gouardères & F. Paraguaçu (Eds.) *Proceedings of the Sixth International Conference on Intelligent Tutoring Systems, ITS 2002* (pp. 574-584). Berlin: Springer Verlag.