

Group Composition and Intelligent Dialogue Tutors for Impacting Students' Academic Self-Efficacy

Iris Howley, David Adamson, Gregory Dyke, Elijah Mayfield, Jack Beuth, Carolyn P. Rosé
Carnegie Mellon University, Pittsburgh, PA

Abstract. In this paper, we explore using an intelligent dialogue tutor to influence student academic self-efficacy, as well as its interaction with group self-efficacy composition in a dyadic learning environment. We find providing additional tutor prompts encouraging to students to participate in discussion may have unexpected negative effects on self-efficacy, especially on students with low self-efficacy scores who have partners with low self-efficacy scores.

Keywords: Intelligent Dialogue Tutors, Collaborative Learning, Self-efficacy, Group Composition, and Discourse During Learning Interactions.

1 Introduction

We know from past research that academic self-efficacy, which is a student's perception of his academic capabilities, is beneficial in individual learning contexts (Zimmerman, 1999) as well as in collaborative learning contexts in which higher group-level self-efficacies are associated with behaviors that support learning (Howley et al, 2011). If the connection is causal, and if we can improve a student's self-efficacy, then the student may reap the associated increased learning and persistence benefits.

In this paper we test this causal connection. Specifically, we leverage conversational agents that have been used successfully as dynamic support for collaborative learning in earlier work (Kumar et al, 2007) as well as theories of discussion moves hypothesized to increase student perception of competence (Michaels, O'Connor, & Resnick, 2008) to provide opportunities for students to take a more authoritative role in a conversation in order to test the effect of that manipulation on self-efficacy. While it is possible to give a student the opportunity to participate more authoritatively, they may choose not to take it or may find themselves unable to take it. The effect of these choices in the face of these uncertainties is an open question. Furthermore, the reactions the student receives from his teammates may have also have an effect on a student's willingness to pursue a discussion opportunity, so it is also necessary to control the group's self-efficacy composition to investigate this issue systematically.

2 Prior Work

The work in this paper revolves around theories from linguistics and social psychology, most notably self-efficacy and a behavioral construct known as authoritativeness.

We focus on Bandura's (1977) theory of self-efficacy and define academic self-efficacy as a student's perceptions of her academic capabilities, interpreted from previous mastery experience, vicarious experience, verbal and social persuasions, and emotional and physiological states. Bandura (1997) also introduces collective efficacy as several individuals' combined perception of the group's capabilities to perform given tasks. Wang & Lin (2007) further investigate this group disposition in collaboration, where they report that individual student self-efficacy predicts the group's collective efficacy, and collective efficacy predicts use of high-level cognitive skills in discussion, as well as group performance.

Along with self-report measures of self-efficacy, we examine behavioral data using a framework for looking at authoritativeness of knowledge presentation. Authoritativeness provides researchers with a lens for examining students' ownership over knowledge through their behavior, rather than through self-report. For our purposes, an "authoritative" statement is a presentation of knowledge without seeking external validation for the knowledge. The Authoritativeness Framework we introduce in this paper is rooted in Martin's Negotiation Framework (Martin, 1992), from the systemic functional linguistics community. A more thorough discussion of our efforts in making this framework replicable is described in Mayfield & Rosé (2011), while our approach for automatically coding chat transcripts with the framework is further explained in Howley et al. (2011).

Our formulation of the Authoritativeness framework is comprised of two dimensions with six and three codes respectively, and is based on principles from the Negotiation framework. For this paper, we will focus on three moves in particular:

- K1, or 'primary knower'. A 'primary knower' move includes a statement of fact, an opinion, or an answer to a factual question, such as 'yes' or 'no'. It only counts as 'primary knower' if it is not presented in such a way as to elicit an evaluation from another participant in the discussion. An example: "This is the end."
- K2, or 'secondary knower'. A 'secondary knower' move includes statements where the speaker is not positioned as authoritative on the current topic, such as asking a question eliciting information, or presenting information in a context where evaluation is the expected response or formulated in such a way as to elicit feedback. An example: "Is this the end?"
- 'o' or 'other' encompassing conversational moves that do not fit within the bounds of the prior two codes described. An example: "So..."

3 Method

The data for this experiment was gathered in order to examine how student academic self-efficacy, learning, and behavior might be affected by targeted prompts from an intelligent dialogue tutor, while also manipulating the partner's self-efficacy in order to look closer at the influence of a peer's self-efficacy. 104 undergraduate students from a thermodynamics class at a private American university participated in the study by attending one of six computer lab sessions, in which time was strictly controlled. Students were given a pre-questionnaire, software training and practice (60

minutes), pretest (10 minutes), the experimental manipulation (40 minutes), and then the posttest and post-questionnaire (15 minutes).

Students were semi-randomly assigned to pairs according to a median split on their course self-efficacy scale in order to achieve homogeneous high self-efficacy pairs, homogenous low self-efficacy pairs, and heterogeneous pairs. After being assigned to pairs, each partner was randomly assigned a goal to design either an eco-friendly power plant or a power-proficient power plant. In all conditions, a tutor agent participated with the students in the chat in order to offer support. The lab session took place in a single computer lab, in which each student had her own computer and partners did not sit next to each other. The experimental manipulation took place during an online collaborative design discussion and consisted of modifying tutor behaviors only. In all other respects, the student experience in all conditions was the same.

Students used Cyclepad (Forbus et al, 1999), a computer software simulator that students use to design simulated power plant designs through a graphical interface. Specifically, students must consider trade-offs between power output and environmental friendliness in designing a Rankine cycle, which is a type of heat engine. The intelligent dialogue tutor was implemented through the Bazaar agent authoring framework (Adamson & Rosé, 2012), allowing the software agent to guide and time discussions, with additional social behaviors. Student dyads collaborated through the ConcertChat software (Stahl, 2006) which enables communication through a chat window and a whiteboard for sharing graphical information.

The experimental manipulation was a 3X3 between-subjects design. Each student pair was randomly assigned to one of nine conditions. The first independent variable manipulated tutor behavior toward high and low self-efficacy students within each pair. The three variations of the tutor behavior were: “target high” (targeting the high self-efficacy student with additional prompts for explanation), “target low” (targeting the low self-efficacy student), and “neutral” (no additional targeted prompts). An example of the tutor’s targeting behavior is “student08, I don't get it - why can't t-max be any higher?” Targeted students received two such prompts, while untargeted students received one, and students in the neutral condition received none. Task-related information such as conceptual hints and timing reminders were kept constant across all three tutoring behaviors leaving the only manipulation to be this targeting behavior. Additionally, all conditions also included context-less mini acknowledgements or encouragements such as “What do you think, student14?” In homogeneous self-efficacy teams, the student with the higher (or lower) self-efficacy score received targeted context questions as shown in **Error! Reference source not found.** That is, in a homogenous low self-efficacy pair that was assigned to “target low” the student with the lower self-efficacy score would receive these additional contextual prompts for participation. In the case that both students had identical self-efficacy scale scores, the student target would be selected at random.

The second independent variable contrasted the three team composition types (homogeneous high self-efficacy, homogeneous low self-efficacy, and heterogeneous self-efficacy), where the median split for the original assignment of “high” and “low” was determined from other similar studies, although for analysis we later reassigned the median split value to be that of this study’s cohort. As outcome measures, we

examined academic self-efficacy both before and after the experimental activity. The pre- and post-questionnaires consisted of scales for measuring collective efficacy, mastery-related beliefs (said to predict self-efficacy), and self-efficacy, constructed via the guidance in Bandura (2006). 35 isomorphic multiple choice and short answer questions were used to test analytical and conceptual knowledge on both the pre- and post-tests. And finally, a process analysis examining changes in chat behavior over time was performed.

4 Results

Data was analyzed with respect to factors including: gender, self-efficacy, dialogue tutor targeting, and self-efficacy team composition. Upon reassigning the self-efficacy median split value to match that of this study's cohort median split (i.e., after the completion of the study), our sample consisted of: 14 pairs of homogeneous high self-efficacy, 15 pairs of homogeneous low self-efficacy, and 23 pairs of heterogeneous academic self-efficacy. When analyzing this data with respect to the targeting condition, we look at individual students as "targeted", "untargeted" (if the student's partner received targeted prompts from the intelligent tutor), or "neutral" (if neither partner received targeted behaviors from the dialogue tutor).

When looking at the tutor's effect on post- academic self-efficacy, we found a significant effect of team composition type, $F(2, 97) = 4.91, p = 0.0093$. Specifically, students who were in homogeneous low self-efficacy groups ended with a self-efficacy score significantly lower than the homogenous high self-efficacy groups and heterogeneous groups, even with controlling for initial self-efficacy. The interaction term between team composition type and whether the student was targeted, untargeted, or in a neutral condition was not significant, although we did find that targeted students in homogeneous low self-efficacy pairs did significantly worse than all other combinations of tutor-conditions and team composition types (except for students in neutral-tutor homogeneous low condition who were indistinguishable from either group). This result is the opposite of what we expected. These results suggest that prompting low self-efficacy students for further participation may not be the ideal method for improving activity self-efficacy in situations where both partners have low self-efficacies compared to the rest of their classmates. More investigation is necessary to make stronger claims, but future designs should take this into account.

Investigating the relationship between students' collective efficacy and self-efficacies showed that collective efficacy is significantly positively correlated with both students' pre- ($r(104) = 0.54, p < 0.0001$) and post- self-efficacy ($r(104) = 0.59, p < 0.0001$).

In order to look at what effect the intelligent dialogue tutor has on conversational behavior, we examine authoritativeness that has been automatically coded through the process described in Mayfield & Rosé (2011). As a validation of the automatic coding, we tested our agreement and found an inter-rater reliability with the automated coding scheme of 0.65, which is close to a robust confidence in non-random agreement.

When looking at overall counts of authoritativeness codes, we find a significant main effect of team composition and targeting condition on K2 moves that is superseded by the interaction term between team composition and the targeting condition on K2 moves, $F(2, 98) = 6.01, p = 0.0034$. A post hoc analysis reveals that targeted students in homogeneous low self-efficacy groups had significantly more K2 moves than every other group. Additionally, team composition type is a significant predictor of “other” moves, $F(2,98) = 3.71, p = 0.028$ as well as authoritativeness (the number of primary knower moves over the total number of knowledge authority moves), $F(2, 98) = 3.52, p = 0.033$. Students in homogeneous low self-efficacy groups were significantly less authoritative than students in heterogeneous groups, with homogeneous high groups being indistinguishable from either. Students in homogeneous low self-efficacy groups also had significantly more “other” moves than students in homogeneous high groups.

While it may be expected that homogeneous low self-efficacy dyads would perform worse than their heterogeneous counterparts, it is interesting that the manipulation appears to have an impact on student behavior within these low self-efficacy pairs. When we look at K2 moves, we find that it is negatively correlated with pre- to post- self-efficacy residuals $r(104) = -0.255, p = 0.009$. This is consistent with our previous results from the self-efficacy analysis.

The tutor targeting conditions, team composition, authoritativeness, and pre- and post- individual self-efficacies did not have a significant effect on learning, but collective self-efficacy is marginally positively correlated with learning, $F(1, 101) = 3.11, p = 0.081$. One might think that perhaps students with lower collective efficacies were at a disadvantage because their group has access to less knowledge; however, there was no significant correlation between pretest scores and collective efficacy, nor between pretest scores and self-efficacy.

5 Conclusions

The majority of our results involved students in homogeneous low self-efficacy dyads. Targeted students in homogeneous low groups ended with self-efficacies lower than predicted compared to all other groups. Students in the low self-efficacy groups had more secondary knower authoritativeness moves and lower authoritativeness scores.

These results point to some important caveats for future work in this area. Dyad self-efficacy composition must be taken into consideration, especially since much of our results concern students in homogeneous low self-efficacy pairs. Simply providing opportunities for students to participate more in the discussion may not harm the untargeted students in the pair, but it does not seem to have the desired effect for the targeted student. And so, future work should control for self-efficacy team composition, as well as consider the dynamics within homogeneous low self-efficacy groups.

With regards to authoritativeness, we found that targeted students in homogeneous low self-efficacy groups had significantly more K2 moves, but we do not yet know if secondary knower moves are beneficial or desirable. Future work should look more

specifically at how students propose knowledge for evaluation, and if there is a beneficial or harmful side effect to doing so.

This work was supported by the Pittsburgh Science of Learning Center (#SBE 0836012) and a Graduate Training Grant from the Department of Education (#R305B040063).

6 References

1. Adamson, D. & Rosé, C. P. (2012). Coordinating Multi-dimensional Support in Collaborative Conversational Agents, in Proceedings of Intelligent Tutoring Systems.
2. Bandura, A. (1977). Self-Efficacy: Toward a unifying theory of behavioral change, *Psychological Review*, 84, pp. 191-215.
3. Bandura, A. (1997). *Self-efficacy: The exercise of control*. New York: Freeman.
4. Bandura, A. (2006). Guide for constructing self-efficacy scales. *Self-efficacy Beliefs of Adolescents*, pp. 307-334.
5. Forbus, K. D., Whalley, P. B., Evrett, J. O., Ureel, L., Brokowski, M., Baher, J., Kuehne, S. E. (1999). CyclePad: An articulate virtual laboratory for engineering thermodynamics. *Artificial Intelligence*, 114(1-2), pp. 297-347
6. Howley, I., Mayfield, E., & Rosé, C. P. (2011). Missing something? Authority in collaborative learning. *Proceedings of Computer-Supported Collaborative Learning 2010*.
7. Kumar, R., Rosé, C. P., Wang, Y. C., Joshi, M., and Robinson, A. (2007). Tutorial dialogue as adaptive collaborative learning support. *Proceedings of Artificial Intelligence in Education*.
8. Martin, J. (1992). *English Text: System and structure*. Amsterdam: Benjamins.
9. Mayfield, E. and Rosé, C. P. (2011). Recognizing authority in dialogue with an integer linear programming constrained model. *Proceedings of the Annual Meeting of the Association for Computational Linguistics (ACL)*, pp. 1018–1026.
10. Michaels, S., O'Connor, C., & Resnick, L. (2008). Deliberative discourse idealized and realized: Accountable Talk in the classroom and in civic life. *Studies in Philosophy and Education*, 27(4), pp. 283-297.
11. Stahl, G. (2006). Analyzing and designing the group cognitive experience. *International Journal of Cooperative Information Systems (IJCIS)*.
12. Veal, R. (1999). Language, knowledge, and authority in school mathematics, in Francis Christie (Ed.) *Pedagogy and the Shaping of Consciousness: Linguistics and Social Processes*, Continuum.
13. Wang, S. & Lin, S. (2007). The effects of group composition of self-efficacy and collective efficacy on computer-supported collaborative learning, *Computers in Human Behavior*, 23(5), pp. 2256-2268.
14. Zimmerman, B. J. (1999). Self-efficacy: An essential motive to learn. *Contemporary Educational Psychology*, 25, pp. 82-91.