Collaborative Lecturing by Human and Computer Tutors

Sidney D'Mello, Patrick Hays, Claire Williams, Whitney Cade, Jennifer Brown, and Andrew Olney

Institute for Intelligent Systems, University of Memphis, USA
[sdmello, dphays, mcwllams, wlcade, jlbrown7, aolney]@memphis.edu

Abstract. We implemented and evaluated a collaborative lecture module in an ITS that models the pedagogical and motivational tactics of expert human tutors. Inspired by the lecture delivery styles of the expert tutors, the collaborative lectures of the ITS were conversational and interactive, instead of a polished one-way information delivery from tutor to student. We hypothesized that the enhanced interactivity of the expert tutor lectures were linked to efforts to promote student engagement. This hypothesis was tested in an experiment that compared the collaborative lecture module (dialogue) to less interactive alternatives such as monologues and vicarious dialogues. The results indicated that students in the collaborative lecture condition reported more arousal (a key component of engagement) than the controls and that arousal was positively correlated with learning gains. We discuss the implications of our findings for ITSs that aspire to model expert human tutors.

Keywords: collaborative lecture, expert tutor, Guru, arousal, engagement

1 Introduction

There is probably nothing more boring and less effective than a lecture on a topic that the recipient has little or no intrinsic motivation to learn. Most would agree that pedagogical activities in the form of "long-winded didactic explanations" that are characteristic of lectures [1] have little to no value, at least when compared to more interactive alternatives such as scaffolding explanations and active problem solving [2, 3]. Although lectures go by many names such as transmission/information delivery [4], direct instruction [5], and didactic teaching [1], they never make the list of ideal tutoring models. Simply put, lectures are inefficient at promoting deep learning because polished deliveries of information by a teacher or a tutor makes the typical student a passive information receiver rather than an active problem solver [2].

Given this bleak sketch of the merits of lecturing in educational contexts, we were somewhat surprised to discover that lectures were abundant in our analysis of 50 naturalistic tutoring sessions between students and expert human tutors [6]. In particular, when we segmented the tutoring sessions into eight *dialogue modes* (i.e., pedagogically distinct phases in a session that last for several minutes and encompass multiple speech acts), lecturing was the second most frequent mode. Lectures comprised 22.1% of the modes and 30.2% of the turns. Lectures were only surpassed by the scaffolding mode, which comprised 27.8% of the modes and 46.4% of the turns [7].

One explanation for the somewhat counterintuitive finding of the relatively high incidence of lectures might lie in the students that were tutored. These students were seeking expert tutoring because they were having considerable difficulty in their classes. It might be the case that the expert tutors extensively lectured in order to provide the necessary common ground before collaborative problem solving can be effective or even functional. There is some evidence to support this hypothesis. First, interactive problem solving is not very effective if the students do not have the requisite knowledge base [8]. For example, it is difficult to imagine a student solving a cytokinesis problem (cell splitting) without knowing what a cell is. Second, and more importantly, problem scaffolding is most likely to follow lectures in the expert tutoring corpus [7]. Hence, it is reasonable to assume that lectures are used to establish the knowledge foundation (i.e., common ground) upon which problems can be modeled, scaffolded, and faded [3].

The fact that lectures are frequent in expert tutoring has important implications for ITSs that aspire to model expert tutors. We are currently in the process of developing a tutoring system (Guru) for high school biology based on the tactics, actions, and dialogue of expert human tutors. It is in this respect that expert tutor lectures are very relevant to our research.

The process of developing a computational model of expert tutoring for Guru, highlighted some important characteristics of expert tutor lectures. Contradictory to the popular conception of lectures primarily being a one-way information transmission stream from the tutor to the student, we were intrigued to discover that the expert tutor lectures were quite interactive [9]. Although direct instruction and explanations played central roles, the lectures were filled with opportunities for students to play a more active role by doing some of the talking.

For example, tutors attempt to keep the student engaged via comprehension gauging questions (e.g., "Do you understand?"). There is some evidence that these questions are not very useful because students cannot accurately monitor their own understanding [10-12]. However, tutors might interleave these questions into the direct instruction cycle to enhance students' engagement and also to cue students to the fact that they need to be actively comprehending the lecture. A more active form of collaboration occurs when tutors directly engage the student via hints, prompts, forced choices, and simplified problems. These activities make students active participants in the tutorial sessions despite the fact that the primary goal of lectures is to deliver information.

In summary, our analysis of lectures during expert tutoring sessions was not consistent with boring, extended, long-winded, explanations. Instead, we found that expert tutor lectures were highly collaborative, presumably because the expert tutors acknowledge that active participation, even during lectures, is key to learning and engagement [13].

We have recently implemented and evaluated a lecture module that closely mirrors the expert tutor lectures. Although we do not expect impressive learning gains from the lecture module, we predict that the collaborative lecturing strategy will boost engagement, at least when compared to non-interactive alternatives (monologues and vicarious dialogues). We tested this hypothesis in an experiment where collaborative lecturing (called dialogues) was compared to monologues and vicarious dialogues. Our prediction is that students will be more engaged in the dialogue condition than the other two less collaborative conditions.

2 Modeling and Implementing the Collaborative Lecture

2.1 Modeling the Collaborative Lecture

An extensive analysis of the lecture strategies of our sample of 50 expert tutors is discussed in [9], hence, we will focus on the major points here. In particular, there are two major clusters of dialogue moves as illustrated in Figure 1. The first cluster (information-transmission) is primarily concerned with the tutor delivering information to student (solid lines in Figure 1). The tutor may assert some information (direct instruction and explanation, die), to which the student provides backchannel feedback via an acknowledgment (ack), and the tutor asserts more information (die). Alternatively, the tutor transmits some information (die), asks a comprehension gauging question (cgq) (e.g., "Do you understand?"). The student replies with an acknowledgement (e.g., "Yes sir") or a metacomment (e.g., "No. I don't quite get it"), and more information is transmitted. These basic patterns associated with information-transmission account for 70.2% of the dialogue moves during lectures.

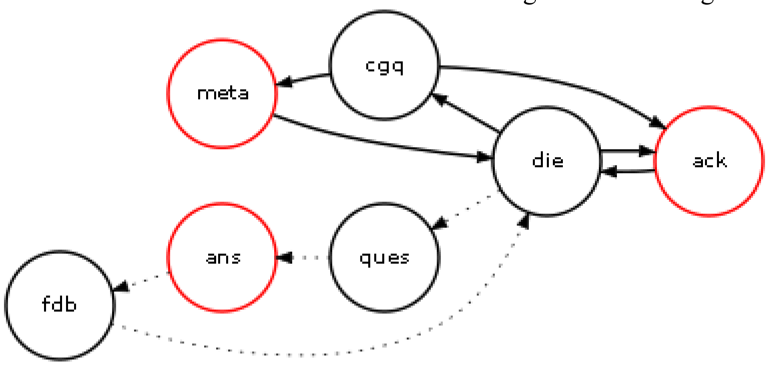

Figure 1. Information-transmission and information-elicitation clusters

The second cluster, or the *information-elicitation* cluster (dotted links in Figure 1), consists of moves associated with attempts by the tutor to elicit information from the student. These moves are variations of the Initiate Respond Evaluate (IRE) sequence [14]. The sequence begins by the tutor asking the student a question (ques) with prompts, pumps, forced choices, or simplified problems. The student responds with an answer (ans). The tutor evaluates the student's response and provides feedback (fdb) followed by more direct instruction (die). This cluster accounts for 18.6% of the moves during lectures.

In addition to these primary clusters that account for 88.6% of the dialogue moves, there is also an off-topic conversation cluster (9%), and a student-initiated question cluster (2.2%). The latter two clusters were not included in the current study because they are difficult to implement in the monologue condition (described below).

2.2 Implementing the Collaborative Lecture

We developed a lecture module in Guru for eight biology topics (cellular respiration, amino acids and RNA, etc). As previously stated, Guru's lecturing strategies were designed to closely mirror the expert tutor lectures. This was accomplished in two ways. First, the content of the lectures was obtained from transcripts of actual expert tutoring sessions. This made the lecture delivery style more conversational, informal, and presumably more engaging.

Second, the tutor closely modeled the collaborative lecturing tactics that were observed from our analysis of the human tutors (see Figure 1). In particular, Guru primarily transmitted information (68% of the time) but occasionally provided cues for acknowledgements (e.g., "Right?", "ok?"), asked comprehension gauging questions, and prompted the student for answers (e.g., "X is a type of what?"). On average, the lectures contained 32% opportunities for student involvement. The tutor:student dialogue move ratio of the tutor strongly correlated (r = .97) with the tutor:student ratio from the actual tutoring sessions. Hence, we are quite confident that Guru does indeed model the collaborative lecturing styles of the expert human tutors.

The lectures were delivered via a simple conversational interface that consisted of an animated conversational agent that delivered the content of the lectures via synthesized speech, a media panel that displayed images relevant to the lectures, and a dialogue box for students to type their responses.

We implemented two non-interactive variants of the collaborative lecture module. The collaborative module (called *dialogue*) closely mirrors the lecturing strategies of the expert tutors, as described above. Alternatively, in the *monologue* version, the tutor did all the talking and the student was a passive recipient. This module was designed to simulate a conventional non-collaborative lecture that is not expected to be very engaging.

The third version consisted of *vicarious dialogues*, where the dialogue patterns were structurally similar to the dialogue module, but with one important exception. Here, it was a virtual student, instead of the learner, that answered the tutor's comprehension gauging questions and prompts. The virtual student always provided the correct answer and the human learner simply watched the interaction. This was the only difference between the vicarious and the dialogue condition. All other aspects of the interface and interaction were equivalent.

Sample dialogues from the human tutors and Guru are presented in Table 1. In the actual lecture, the tutor introduces the topic (T1), uses a discourse marker (T2), asserts some information (T3), and then gives the student an opportunity to chime in (T4). The student provides an acknowledgment (S1), the tutor responds with a conversational OK (T5), asserts some more information (T6), and then prompts the student (T7). The student responds (S2), to which the tutor provides some feedback (T8), followed by an assertion, and so on (T9 and S3).

When Guru delivers a monologue for this sample lecture, it preserves most of the conversational style, asserts the same content, but does not give the learner an opportunity to type a response (see Table 1). In contrast, the learner in the dialogue condition has three opportunities to type in a response, which is consistent with the 1:3 student to tutor dialogue move ratio discussed above.

Although not included in Table 1, the vicarious-dialogue condition was identical to the dialogue condition. However, a virtual student, instead of the human learner, would type in (i.e., via simulated keystrokes) responses to items S1, S2, and S3. The simulated keystrokes were carefully calibrated in order to mirror the temporal dynamics of actual typing (i.e., onset delay, variable interstroke delay, and delay before hitting enter key to submit response).

On average, the expert human tutors articulated 790 words in each lecture, while the ITS articulated an average of 677, 718, and 718 words in the monologue, dialogue, and vicarious-dialogue conditions, respectively.

Table 1. Sample exerpts from lectures

N	Actual Lecture	Monologue	Dialogue
T1	Let's talk about mitosis.	Let's talk about mitosis.	Let's talk about mitosis.
T2	Ok.	Ok.	Ok.
T3	Now, let’s say here’s a skin cell, he’s just sitting around, and he needs to divide.	Now, let’s say here’s a skin cell, it's just sitting around, and it needs to divide.	Now, let’s say here’s a skin cell, it's just sitting around, and it needs to divide.
T4	Someone’s got to tell him, **right?**	Someone’s got to tell him to divide.	Someone’s got to tell him to divide, **right?**
S1	**Mm hmm.**	**<pause>**	**<student response>**
T5	Ok	Ok	Ok
T6	I mean, let’s say a skin, skin cell is sitting around.	I mean, let’s say a skin cell is sitting around.	I mean, let’s say a skin cell is sitting around.
T7	Do you think somebody needs to tell him to split, or do you think he can just say, oh, I think I’ll split?	Do you think somebody needs to tell him to split, or do you think he can just say, oh, I think I’ll split?	Do you think somebody needs to tell him to split, or do you think he can just say, oh, I think I’ll split?
S2	**Tell him it’s time?**		**<student response>**
T8	Yeah!		
T9	Because, see, now folks need to get instructions, **right?**	Someone must tell him to split because he needs to get instructions.	Someone must tell him to split because he needs to get instructions, **ok?**
S3	**Mm hmm.**		**<student response>**

3 Method

Participants were 90 college students from a mid-south university in the US who participated for extra course credit.

Participants engagement levels were tracked at multiple points in the tutorial session with the affect grid [15]. The affect grid is a validated single item affect measurement instrument consisting of a 9 × 9 (valence × arousal) grid; these are the primary dimensions that underlie all affective experiences [16]. The arousal dimension ranges from sleepiness to high-arousal, while the valence dimension ranges from unpleasant feelings to pleasant feelings. Participants indicate their affective state by marking an X at the appropriate location on the grid.

The knowledge tests (used to measure learning gains) were 24-item multiple-choice tests with three questions for each lecture. *Prompt* questions tested participants on content for which the tutor explicitly prompted the student in the dialogue and vicarious conditions. Although there were no explicit prompts in the monologue condition, we verified that the content of the prompts was explicitly covered in the monologue. *Assertion* questions tested participants on content that the tutor explicitly asserted to the student via direct instruction. Finally, there were *deep reasoning* questions that required causal reasoning, inference, etc. rather than recall of shallow facts. Participants completed alternate test versions for pretest and posttest that were counterbalanced across participants.

Participants were tested individually over a two hour session. Participants completed an informed consent followed by the pretest. They then read instructions on how to use the affect grid. On the basis of random assignment participants then completed a tutorial session with the monologue, dialogue, or vicarious tutor. There were 30 participants in each condition. The tutoring session consisted of eight lectures that were randomly ordered for each participant. Participants used the affect grid to indicate their affective state after each lecture. They completed the posttest after the tutorial session and were fully debriefed.

4 Results and Discussion

Engagement levels presumably decrease over time, hence, an analysis comparing engagement without controlling for time on task would be confounded. As could be expected, the monologue condition was shorter (M = 37.2 minutes) than the dialogue (M = 54.6) and vicarious conditions (M = 55.4). Since dialogue and vicarious were of equivalent length our first analyses compared engagement across these two conditions; monologues were considered in a follow-up analysis that equated time on task in a post-hoc fashion.

4.1 Engagement Levels

Self reported engagement trajectories were computed by averaging participants' valence and arousal scores (from the affect grid) for each lecture. These are presented in Figure 2 as a dialogue trajectory (D1, D2, …D8) and a vicarious trajectory (V1, V2,…V8). When averaged across lectures, learners reported higher levels of arousal in the dialogue condition (M = 4.10, SD = 1.80) than the vicarious condition (M = 3.00, SD = 1.30), $t(58) = 2.73$, $p = .008$, $d = .70$. There was no significant difference (p = .648) in valence levels across conditions (M = 4.48, SD = 1.66 for dialogue and M = 4.68, SD = 1.79 for vicarious).

Comparisons of arousal scores for each lecture (i.e., D1 vs. V1, D2 vs. v2, etc) indicated that participants in the dialogue condition were significantly ($p < .05$) more aroused than their vicarious counterparts for the first six lectures. The difference was marginally significant ($p = .102$) in favor of the dialogue condition for the seventh lecture. There was no significant difference ($d = .270$) for the eighth lecture, although there was a small to medium sized effect ($d = .3$) in favor of the dialogue condition.

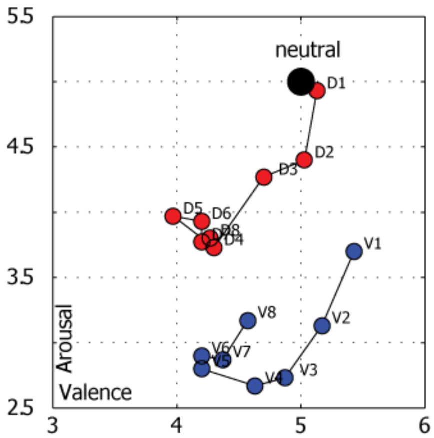

Figure 2. Engagement trajectories. (Numbering, D1, D2, etc indicates order)

We performed a follow-up analysis that controlled for time on task. Specifically, mean arousal and valence scores were computed for each participant by only including their responses for the first 37 minutes, which is the mean length of the monologues. An ANOVA indicated that there was a significant difference in arousal scores across conditions, $F(2, 87) = 5.59$, $p < .01$. As predicted, arousal scores for the dialogue condition ($M = 4.3$, $SD = 1.81$) were significantly ($p < .5$ on a one-tailed test) greater than arousal scores for the monologue ($M = 3.65$, $SD = 1.34$) condition, with an effect size of .41 sigma. There was a marginally significant difference ($p = .098$) with a medium sized ($d = .50$) effect when arousal scores for the monologue condition were compared to the vicarious condition ($M = 3.00$, $SD = 1.30$).

The ANOVA comparing valence scores for the three conditions was not significant ($p = .750$), so the interactivity afforded by the collaborative lectures impacts arousal but not valence.

4.2 Learning Gains

Proportional learning gains were computed for each of the question types (prompts, assertions, and deep reasoning questions as described in the Methods section) as (posttest – pretest) / (1-pretest). Since it is generally acknowledged that tutoring differentially benefits low versus high domain knowledge students, our analyses proceeded by dividing participant into these two groups on the basis of their pretest scores (see Table 2). There were no differences in learning gains across conditions for the high prior knowledge group, so the subsequent discussion focuses on the low domain knowledge group.

The results indicated that participants in the dialogue condition had marginally significantly ($p = .104$) higher scores for prompt questions when compared to their monologue counterparts ($d = .56$). The monologue versus vicarious comparison was not significant, however, there was a medium effect ($d = .49$) in favor of the vicarious condition. Hence, the pattern for prompt questions appears to be [Dialogue =

Vicarious] > Monologue. This pattern is intuitively plausible because prompts direct the learner's attention to specific words in the dialogue and vicarious conditions, but not the monologue condition (there were no explicit tutor prompts in this condition).

It should be noted that medium sized, marginally significant effects are meaningful for the current learning gains analyses because there was a significant loss of statistical power when the participants were split into low and high knowledge groups; these effects are likely to be significant with a larger sample.

A somewhat different pattern was observed for questions that tested participants' retention of the tutor's assertions. Here, the monologue condition was on par with the dialogue condition, but outperformed the vicarious condition (p = .051, d = .81). These results suggest that participants in the vicarious condition overlooked important assertions by the tutor, presumably because their focus was on the virtual student's responses to the tutor's prompts.

There was no difference in learning gains for deep reasoning questions. In summary, these results suggest that when it comes to low prior knowledge students, the monologue and vicarious conditions yield *inconsistent* results because they are ineffective for prompts and assertions, respectively. In contrast, low-domain knowledge students assigned to the dialogue performed *consistently* across the different question types (i.e. it was never significantly worse than other conditions).

Table 2. Mean proportional learning gains

	Low Prior Knowledge			**High Prior Knowledge**		
Question	*Monologue*	*Dialogue*	*Vicarious*	*Monologue*	*Dialogue*	*Vicarious*
Prompt	.27	.47	.43	.34	.14	.42
Assert	.43	.29	.17	.20	.32	.21
Deep	.28	.23	.22	.13	.19	.27

4.3 Correlations between Engagement and Learning Gains

Our results so far are indicative of (a) the following hierarchical ordering of arousal levels across conditions: Dialogue > Monologue > Vicarious, (b) equivalent valence levels, and (c) differential patterns in learning gains across conditions and prior knowledge. It appears that it is arousal and not valence that is most relevant to learning gains. Arousal is correlated with deep learning gains in the dialogue and vicarious conditions and overall learning gains (i.e. gains not segregated by question category) in all three conditions (see Table 3).

There was one more interesting pattern pertaining to the relationship between arousal and learning. Recall that participants in the monologue condition outperformed vicarious participants for assertion questions, while a reverse pattern was observed for prompt questions. The correlational analyses indicate that these patterns were related to self reported arousal, thereby providing further evidence that it is arousal and not valence that is relevant to learning gains.

Table 3. Correlations between engagement and learning

	Arousal				Valence			
Condition	*Prompt*	*Assert*	*Deep*	*Overall*	*Prompt*	*Assert*	*Deep*	*Overall*
Monologue	.072	.315*	.188	.363**	.002	-.280	.115	-.132
Dialogue	.022	.042	.537**	.347*	-.075	.203	.112	.119
Vicarious	.441**	.154	.347*	.492**	.073	.145	.144	.146

Notes. ** $p < .05$; * $p < .10$

5 General Discussion

As most people in the field of education will attest, the task of keeping students engaged in educational activities is extremely challenging. Establishing and maintaining student engagement is especially critical in situations with high degrees of learner control, such as in distance education, computer-based tutoring, and informal learning environments, because learners are a mouse click away from ending the session. The engagement problem is undoubtedly more severe in situations where the computer tutor does most of the talking as when lectures are delivered to remedial students.

Although we were initially surprised by the high incidence of lectures in our sample of 50 expert tutoring sessions, we hypothesized that expert tutors implement a collaborative lecturing strategy to avoid the pitfalls associated with boring, one-way, didactic instruction. This hypothesis was confirmed in our evaluation of a computer tutor that simulated the lecturing style of expert human tutors. Our results indicated that arousal, a key component of engagement, was higher in the condition that implemented collaborative lecturing when compared to less interactive alternatives. Furthermore, arousal is critical because it is positively correlated with learning gains.

The correlation between arousal and learning gains is consistent with theories that highlight the importance of affect to deep learning [17, 18]. Physiological arousal is a universal and fundamental dimension of affective experience, a component of all emotional episodes, and a signal for alertness and action [16]. Hence, it comes as no surprise that arousal was highest in the most interactive condition and that arousal was linked to learning gains.

Our implementation of the collaborative lecture strategies of expert human tutors is one important step towards the larger goal of understanding the tactics that underlie their effectiveness. However, several important questions have not yet been answered. How do the expert tutors blend lecturing and scaffolding in order to optimize learning gains? What motivational strategies do they use to enhance self-efficacy and heighten engagement? How do they detect and respond to students affective states in order to prevent students from wallowing in negative emotions and promote more fruitful trajectories of thought? It is our hope that answers to these questions will deepen our understanding of expert tutors and launch next-generation ITSs to new levels of effectiveness.

Acknowledgements. This research was supported by the by the Institute of Education Sciences, U.S. Department of Education, through Grant R305A080594. The opinions expressed are those of the authors and do not represent views of the IES.

References

1. Chi, M.: Constructing Self-Explanations and Scaffolded Explanations in Tutoring. Applied Cognitive Psychology 10, 33--49 (1996)
2. Bransford, D., Brown, A., Cocking, R. eds.) How People Learn: Brain, Mind, Experience, and School Committee on Developments in the Science of Learning. National Academy Press. (2000)
3. Rogoff, B., Gardner, W.: Adult Guidance of Cognitive Development. In: Rogoff, B., Lave, J. (eds.) Everyday cognition: Its development in social context, pp. 95--116. Harvard University Press, Cambridge, MA (1984)
4. Wheatley, G.: Constructivist Perspectives on Science and Mathematical Learning. Science Education 75, 9--21 (1991)
5. Lu, X., Di Eugenio, B., Kershaw, T., Ohlsson, S., Corrigan-Halpern, A.: Expert vs. Non-expert Tutoring: Dialogue Moves, Interaction Patterns and Multi-Utterance Turns. In Proceedings of Eighth International Conference on Intelligent Text Processing and Computational Linguistics, pp. 456--467. Springer, Berlin / Heidelberg (2007)
6. Person, N., Lehman, B., Ozbun, R.: Pedagogical and Motivational Dialogue Moves Used by Expert Tutors. In Proceedings of 17th Annual Meeting of the Society for Text and Discourse, Glasgow, Scotland. (2007)
7. Cade, W., Copeland, J., Person, N., D'Mello, S.: Dialogue Modes in Expert Tutoring. In: Woolf, B., Aimeur, E., Nkambou, R., Lajoie, S. (eds.) Proceedings of the 9th international conference on Intelligent Tutoring Systems, pp. 470--479. Springer-Verlag, Berlin, Heidelberg (2008)
8. Ausubel, D.: Educational Psychology: A Cognitive View. Holt, Rinehart and Winston, Inc., Austin, TX (1978)
9. D'Mello, S., Olney, A., Person, N.: Mining Collaborative Patterns in Tutorial Dialogues. Journal of Educational Data Mining (in review)
10. Glass, M., Kim, J., Evens, M., Michael, J., Rovick, A.: Novice vs. Expert tutors: A Comparison of Style. In: Midwest Artificial Intelligence and Cognitive Science Conference. Bloomington, IN (1999)
11. Graesser, A., Person, N., Magliano, J.: Collaborative Dialogue Patterns in Naturalistic One-To-One Tutoring. Applied Cognitive Psychology 9(6), 495--522 (1995)
12. Person, N., Graesser, A., Magliano, J., Kreuz, R.: Inferring What the Student Knows in One-to-One Tutoring - the Role of Student Questions and Answers. Learning and Individual Differences 6(2), 205--229 (1994)
13. VanLehn, K., Graesser, A.C., Jackson, G.T., Jordan, P., Olney, A., Rose, C.P.: When Are Tutorial Dialogues More Effective Than Reading? Cognitive Science 31(1), 3--62 (2007)
14. Mehan, H.: Learning lessons: Social Organization In The Classroom. Harvard University Press, Cambridge (1979)
15. Russell, J.A., Weiss, A., Mendelsohn, G.A.: Affect Grid - A Single-Item Scale of Pleasure and Arousal. Journal of Personality and Social Psychology 57(3), 493--502 (1989)
16. Russell, J.: Core Affect and the Psychological Construction of Emotion. Psychological Review 110, 145--172 (2003)
17. Meyer, D., Turner, J.: Re-conceptualizing Emotion and Motivation to Learn in Classroom Contexts. Educational Psychology Review 18(4), 377--390 (2006)
18. Csikszentmihalyi, M.: Flow: The Psychology of Optimal Experience. Harper and Row, New York (1990)