

A novel feedback system for pedagogy refinement in large lecture classrooms

Pankaj CHAVAN^{a*}, Shreya GUPTA^b & Ritayan MITRA^a

^aIDP in Educational Technology, Indian Institute of Technology – Bombay, India

^bDepartment of Computer Engineering & Technology, GNDU - Amritsar, India

*pankajchavan@iitb.ac.in

Abstract: Despite the best efforts of educators to introduce active learning techniques, classrooms remain largely didactic in nature. We have developed a unique audience response system to improve pedagogy in such large-lecture didactic classrooms. A pilot study with ten participants was conducted to assess the utility of such voluntary student feedback in improving pedagogy. The web-based feedback tool lets students voluntarily indicate their perception of the lecture on four key parameters – difficult, easy, boring and engaging – in real-time. We discuss the potential of such feedback in refining pedagogy for didactic classrooms. We also discuss the usability of such a feedback system as revealed by a user perception study.

Keywords: Audience response system, cognitive-affective states, large lecture classrooms

1. Introduction

In an exhaustive compilation of instructional strategies in the United States, Stains et al., (2018) concluded that although active learning seems to have penetrated some classrooms, teaching practices in undergraduate classrooms in the United States remain disproportionately biased toward passive and lecture-based strategies. Although we are working toward introducing active learning in classrooms, a majority of instructors use lectures as the primary mode of teaching in which hour long presentations from Power Point slides are the norm. Manduca (2013) suggested that instructors reuse almost 85% of the lecture material. Therefore, if instructors have the knowledge of whether a particular set of slides, topic, figures and associated explanations work in a lecture (or not), then that can improve instruction and thereby learning in such classrooms.

Let us imagine a common scenario in a large lecture classroom. A young faculty who has received a PhD degree from an Ivy League institution in USA gets inducted to teach a large introductory classroom at a R3 public university. Coming from a background of high quality research and teaching, she assumed the responsibility of instilling in her students a sense of wonder, logic and understanding that are crucial for an introductory course. In one such class, she had to introduce the concept of oxidation and reduction. The instructor thought she had explained the concept well. However, after the exam she found that most students did not understand the concept. Additionally, she recalled retrospectively, of the few students (<1% of the class size) who had come to see her during office hours before the exam, most had trouble with this particular concept. Despite her best intentions, she was first, not able to identify the lecture segment that was ubiquitously difficult for the class to understand and second, even after receiving feedback from the few students, she was unsure of the changes that needed to be made. In this scenario, a majority of students seem to have had difficulty in understanding the instructor which prevented learning of the difficult concept. Regardless of the *actual* difficulty of the concept the *perceived* difficulty of the students was high. Pintrich (2000) defines perceived difficulty as a judgment of the difficulty level of a task and if students perceive a learning task to be difficult that perception may decrease interest and increase boredom, due to the negative affect that arises from excessive cognitive load and interruptions in information processing (Efklides, 2006). Therefore, perceived difficulty of a topic can create negative affect which in turn could hinder learning.

Graesser & D'Mello (2012) proposed a cognitive disequilibrium framework (CDF) which explains how the principal emotions arising during learning of difficult materials may be related. According to CDF, a state of engagement or flow can be impeded by a learning impasse to generate a state of confusion. The impasse could lead to frustration and eventual boredom and disengagement from future learning. As educators, we need to ensure that the state of boredom is never reached because it has been found to have negative impact on learning (Tze, Daniels, & Klassen, 2016; Vogel-Walcutt, Fiorella, Carper, & Schatz, 2012). The CDF has been mostly applied to students' learning of difficult material in computerized learning environments like intelligent tutoring systems (ITS) or with pedagogical agents (e.g. AutoTutor) (Graesser & D'Mello, 2012). However, learning in traditional classroom is not expected to differ in this regard as it should involve similar cognitive processes and affective states.

One of the problems related to the scenario discussed above is the lack of instructor's attention to the knowledge and beliefs that learners bring to the class (Bransford et al., 1999). Questioning and raising doubts related to the concepts that are difficult to understand can help students avoid obstacles during learning (Dillon, 1990). But, students often do not ask questions because of several factors such as student gender (Hall & Sandler, 1982), instructor gender (Pearson & West, 1991), class size (Crawford & MacLeod, 1990), student age range (Howard, Short, & Clark, 1996), teacher support (Karabenick & Sharma, 1994), and shyness/need for anonymity (Nadler & Porat, 1978). Moreover, for a novice instructor, it is not easy to identify ineffective lecture segments and it is even more difficult to spot what aspects of those segments (e.g. slide, figure, explanation) are central to the problem and what could be a solution for it.

A potential solution to this problem is through the use of audience response systems (ARS) or clickers which allows students to interact, vote on a topic or give feedback when asked (Caldwell, 2007). Clickers enable instructors to instantaneously collect student responses to a question posed by them. The answers are displayed to the students by the instructor and then the students and instructor can discuss them further to iron out any confusion (Caldwell, 2007). As traditional use of clickers have gained popularity among colleges and has shown positive effects on academic performance, one of its primary drawbacks is that they require students to give feedback only when asked by the instructor. Therefore, the efficacy of clickers in classroom remains tightly coupled to the instructor's perception of difficulty and may not always align with the students' perception of the same.

Rivera-Pelayo, Munk, Zacharias, & Braun (2013) extended the ARS technology by developing an app (LIM – Live Interest Meter) to obtain continuous voluntary feedback from the audience during mass lectures and conferences to improve the presentation content and skills of the speakers. The critical technological difference between LIM app and an ordinary clicker is the use of time stamps with every response that allows the app to create a time series of the requested variable thereby making it possible to collect continuous unsolicited responses and not simply discrete response to question prompts. The LIM app interface allows users to vote on any learning related variable (e.g., confusion level) throughout the duration of the lecture as and when the users felt necessary. The aggregated time series would then reflect the level of confusion (or any other parameter) of the entire class at any given time. The focus of LIM app was on reflective learning for speakers from the feedback provided by the audience. Interviews with the users of the app revealed enthusiasm and acceptance. There is a lot of potential in using such *technology* to improve teaching in the *context* of large didactic classrooms but systematic research is sparse on this particular technology-context combination. For example, Rivera-Pelayo, Munk, Zacharias, & Braun (2013) did not dwell much upon the data *vis-a-vis* its utility for improving any lecture although that premise was implicit in the article. In the following sections, we describe a novel feedback system and enumerate on the data obtained from a pilot study with a small group of students. We outline how such data when collected for large classrooms can help in refining pedagogy. We also report on the user perception of our proposed feedback system.

2. A novel feedback system

The LIM app collects continuous feedback from the audience on aspects such as comprehension (difficult or easy to understand) or performance (too fast or too slow). One of the biggest differences

between LIM app and our proposed system is the explicit accounting of cognitive as well as affective states of the learner. Learning is not simply a cognitive or an affective process. Rather, it is an interaction of the two. A feedback system that gathers data on both aspects could potentially allow us to identify how these parameters vary during a lecture and their role in the learning process. The proposed feedback system collects feedback on two cognitive states, namely, easy and difficult, and two affective states, namely, engaging and boring. The decision to capture feedback on dichotomous variables instead of continuous scales, again departing slightly from LIM App, was to reduce potential distraction as a binary decision would require less cognitive processing by the student than a decision on a continuous scale.

We have developed a web-based application (Fig. 1.a) that can be accessed by the participants on any mobile device. It utilizes the touchscreen feature of the device to make the application unobtrusive during the lecture. The interface uses front-end technologies of HTML5, CSS3, Bootstrap, JavaScript, JQuery. The backend processing and database use PHP and MySQL. The database stores email ID, password, binary responses (0/1) for each variable (easy, difficult, engaging, and boring), and time stamp for each response. When the student clicks the button, the color of the button changes to notify the participant that her response has been recorded, the button is disabled for 5 seconds and the response along with the timestamp is stored in the database.

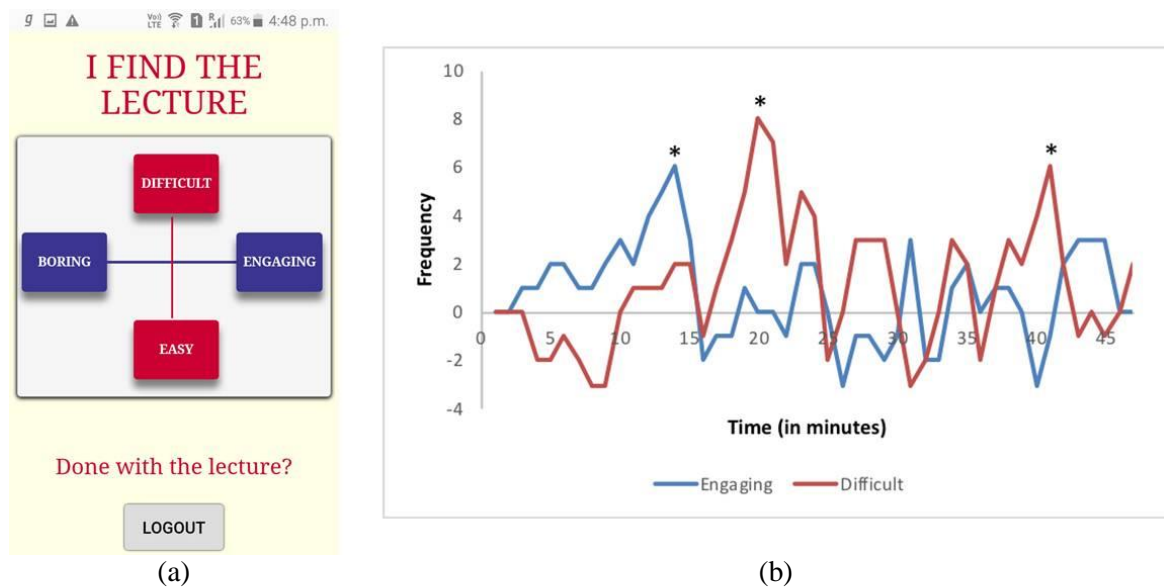


Figure 1. a) Feedback system interface on an android cell phone. Students can give voluntary feedback about their cognitive state by clicking on either the easy or difficult button and about the affective state by clicking on either the engaging or boring button. b) Data collected with the feedback system showing frequency of difficulty and engagement during a lecture session. Positive values indicate the lecture was difficult (red) or engaging (blue). Negative values indicate the lecture was easy (red) or boring (blue). Peaks marked with asterisks indicate segments where at least 60% of the students found the lecture to be of either state.

3. Method

Ten students enrolled in the course Advanced Heat Transfer in the mechanical engineering department of a large public university in India were part of the experimental group. The participants were enrolled in the Mtech or PhD program and were all males with ages between 22 and 29. A 45 minute video lecture of the faculty teaching the Advanced Heat Transfer course on the topic of boiling and condensation was used for this lab-based study. Although the lecture topic was related to the course the participants were already enrolled in, it was not a part of the syllabus for the course and none of them had any prior exposure to content in the video lecture. The students were compensated with Rs. 300 (~5 US\$) for their time. Informed consent was obtained from the student

and the study was cleared by the ethics board of the institute (IEC Approval Letter_Proposal No. IITB-IEC/2018/004). The students were asked to watch the lecture projected on a screen size of 6' by 7' and were asked to provide feedback as and when they felt necessary on a Surface tablet.

4. Results and Discussion

On an average the participants clicked 20 times during the lecture ($M=20.1$, $SD=13.82$). The maximum number of clicks was 54 and the minimum was 8. The median time of inactivity (i.e. no clicking recorded) was 111 seconds (1.85 minutes). The longest time of inactivity was 1245 seconds (20.75 minutes) and the shortest time of inactivity was 1 second across all four states (buttons).

In the design of the feedback system, we have taken care of the possibility of spurious clicking by restricting clicking *of the same button* for 5 seconds after clicks. We further applied conservative estimate of the data by considering only 1 click per minute for each of the four variables. Therefore, if someone has clicked several times in 1 minute, it was still considered as only 1 click by that person. The reason why this step was necessary instead of simply making the period of inactivity 1 minute was to differentiate between spurious or thoughtless clicks and thoughtful yet very frequent clicking activity. Frequency calculations were performed for every minute of the lecture (Fig 1b). Difficulty was calculated as the difference between the number of difficult and easy clicks. Engaging was calculated as the difference between the number of engaging and boring clicks. For example, in the 4th minute of the lecture the difference between the total number of students who found the segment engaging and those who did not was one. Similarly, the difference between those who found it easy and those who did not was two. Finally, a moving window frame of 2 minutes with 1 minute overlap was applied.

Points of interests were identified when the frequency values crossed a threshold of 60% or when six out of ten students found the lecture to be either engaging or difficult. The cutoff mark is arbitrary and would differ from study to study. However, it is important to realize that the peak value does not necessarily represent all students that found the material either engaging or difficult. That value is more reflected by the area under the curve which includes those peaks. For instance, the difficulty peak at 20 mins indicates 80 percent of the students clicked the material to be difficult at 20 minutes into the lecture. Again, this would only be a lower bound as students might have reflected upon the difficult material before and after that time. For example, those who might have identified the difficult section before and after the peak time (20 min) or those who found the content uniformly difficult throughout the entire time that particular topic was being discussed. Hence, a cutoff of sixty percent indicates that at least sixty percent of the students found the lecture either difficult or engaging. The upper limit is not revealed in this figure. Additionally, the width of the area under the curve could be also indicative of the lecture content vis-à-vis difficulty or engagement (See Fig. 2 and related discussion).

Such a feedback system could be useful for eliciting anonymous feedback from students to identify optimal/suboptimal parts of the lecture. Retrospective interviews with students and asking them to reflect upon their feedback might shed light on the source of these cognitive-affective states.

With large enough sample size, such as that possible in large undergraduate lectures, this type of feedback could also be useful to detect the relationship between cognition and affect i.e. whether the students find the lecture to be dull and therefore difficult or vice-versa (from phase lag of response curves). Such fine-grained analysis is also expected to provide high-resolution feedback to the instructor for retrospective self-reflection (Fig. 2), as the morphology of a difficulty peak arising from a very difficult figure that was mentioned in passing is likely to be different than a difficult concept that was discussed over a period of 15 minutes and across several slides. Such fine-grained analysis will assist instructor to use such data to modify future iterations of the lecture or to clarify existing doubts in the classroom. Hence, this system could be further useful as an alternative way for teaching evaluation because with every iteration of a lecture the difficulty peaks can be expected to flatten and the engagement peaks should become more prominent.

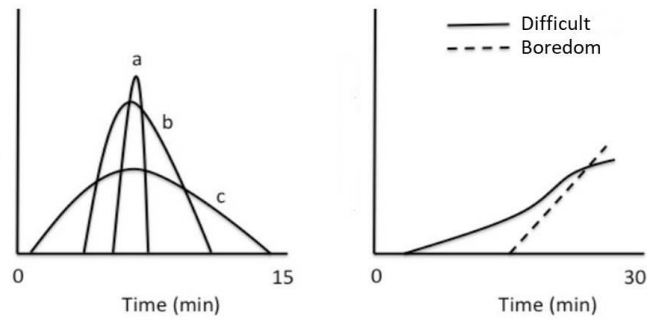


Figure 2. Hypothetical curves. (Left) Granularity of difficulty could indicate the source of a learning impasse: (a) difficulty over a figure, (b) difficulty over a lecture slide and (c) difficulty over a topic. (Right) Phase lag of boredom suggestive of prior difficulty as a probable cause

At the end of the session, a survey on the usability of the feedback system was conducted (Fig. 3). Overall the participants were very positive about the feedback system. All participants expressed their opinion regarding several usability aspects of the feedback system, like the feedback interface ‘is easy to use’ (80%: Strongly agree, 20%: Agree) or the feedback interface ‘became easier as I got more used to it (later parts of the lecture)’ (90%: Yes). The participants’ opinion about the aspect of distraction was very encouraging. There were two survey questions to measure the construct of distraction and one of the questions was reversely coded. Sixty percent of the participants said that the feedback system actually improved their attention toward the lecture (40%: Strongly agree, 20%: Agree) and when asked whether it was distracting (reverse coding) 80% of the participants disagreed (30%: Strongly disagree, 50%: disagree). The reverse coding to estimate the construct of distraction lends greater validity to the estimate.

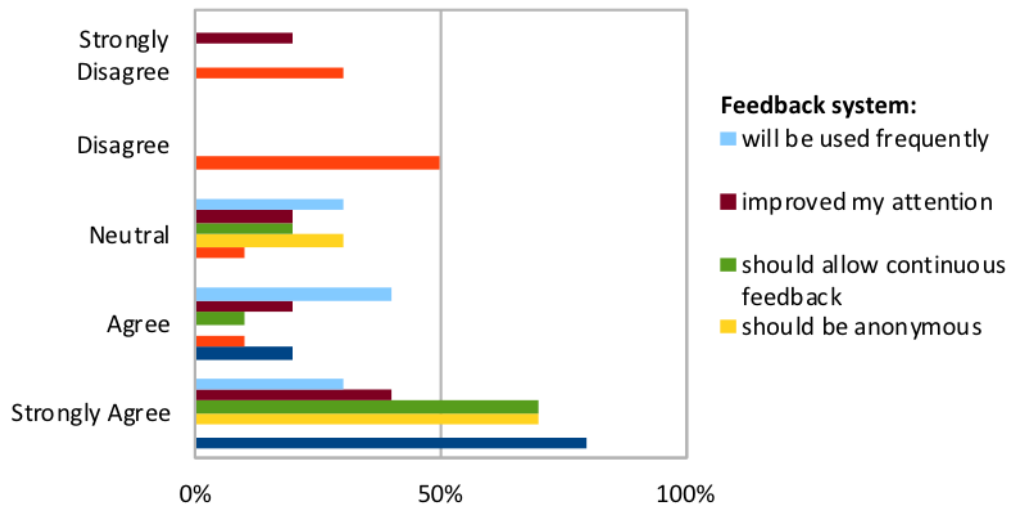


Figure 3. An overall distribution of participant’s perceptions about the usability of the feedback system

In the survey, we also asked whether giving voluntary feedback is better than giving it periodically as and when prompted by the teacher. A majority of participants agreed that voluntary feedback on cognitive and affective states without prompts by the teacher is better (70%: Strongly agree, 10%: Agree, 20%: Neutral). Additionally, the participants also felt such feedback must be anonymous in order to be reliable and honest (70%: Strongly agree, 30%: Neutral).

5. Conclusion and Limitations

The feedback system we developed enables students to report their cognitive-affective states during a lecture. This feedback provides a unique affordance for self-reflection by instructors to identify

effective and ineffective segments of the lecture and make necessary corrections. Alternatively, if the faculty is pressed for time, we can think of a model of guided self-reflection for instructors wherein a group of students/TA/near-peers evaluate synched lecture video-feedback data, identify suboptimal segments and suggest possible ways to improve those sections only. A usability survey with the participants revealed an overall positive impression of the feedback system. A critical attribute of such a system in a teaching-learning scenario is the potentially distracting nature of giving the feedback itself. An overwhelming eighty percent of the participants reported that it was not distracting for them to use and sixty percent of students reported that this intervention improved their attention toward the lecture content.

One of the important limitations of this study is the small sample size. We expect a larger classroom (sample size) would produce more robust, and possibly more nuanced (such as those in Fig 2), observations. Another limitation of the study is the need for independent validation of the observed structure in the curves through interviews with students.

6. Acknowledgements

This work was supported by an IRCC, IIT Bombay grant (17IRCCSG013) to Ritayan Mitra.

References:

- Bransford, J. D., Brown, A., & Coking, R. (1999). How people learn: Mind, brain, experience, and school. Washington, DC: National Research Council.
- Caldwell, J. E. (2007). Clickers in the large classroom: Current research and best-practice tips. *CBE-Life Sciences Education*, 6(1), 9-20.
- Crawford, M., & MacLeod, M. (1990). Gender in the college classroom: An assessment of the “chilly climate” for women. *Sex Roles*, 23(3-4), 101-122.
- Dillon, J. (1990). *The Practice of Questioning (International Series on Communication Skills)*. London: Routledge.
- Efklides, A. (2006). Metacognition and affect: What can metacognitive experiences tell us about the learning process? *Educational Research Review*, 1(1), 3-14.
- Graesser, A. C., & D’Mello, S. (2012). Emotions during the learning of difficult material. In *Psychology of learning and motivation* (Vol. 57, pp. 183-225). Elsevier.
- Hall, R. M., & Sandler, B. R. (1982). The Classroom Climate: A Chilly One for Women?
- Howard, J. R., Short, L. B., & Clark, S. M. (1996). Students’ participation in the mixed-age college classroom. *Teaching Sociology*, 8-24.
- Karabenick, S. A., & Sharma, R. (1994). Perceived teacher support of student questioning in the college classroom: Its relation to student characteristics and role in the classroom questioning process. *Journal of Educational Psychology*, 86(1), 90.
- Manduca, C. (2013, September). Teaching Geoscience: A national look at what is happening in undergraduate classrooms. *MEAS seminar series. Lecture at North Carolina State University, Raleigh, NC*.
- Nadler, A., & Porat, I. (1978). When names do not help: Effects of anonymity and locus of need attribution on help-seeking behavior. *Personality and Social Psychology Bulletin*, 4(4), 624-626.
- Pearson, J. C., & West, R. (1991). An initial investigation of the effects of gender on student questions in the classroom: Developing a descriptive base. *Communication Education*, 40(1), 22-32.
- Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning. In *Handbook of self-regulation* (pp. 451-502). Elsevier.
- Rivera-Pelayo, V., Munk, J., Zacharias, V., & Braun, S. (2013). Live interest meter: learning from quantified feedback in mass lectures. In *Proceedings of the Third International Conference on Learning Analytics and Knowledge* (pp. 23-27). ACM.
- Stains, M., Harshman, J., Barker, M. K., Chasteen, S. V., Cole, R., DeChenne-Peters, S. E., Laski, F. A. (2018). Anatomy of STEM teaching in North American universities. *Science*, 359(6383), 1468-1470.
- Tze, V. M., Daniels, L. M., & Klassen, R. M. (2016). Evaluating the relationship between boredom and academic outcomes: a meta-analysis. *Educational Psychology Review*, 28(1), 119-144.
- Vogel-Walcutt, J. J., Fiorella, L., Carper, T., & Schatz, S. (2012). The definition, assessment, and mitigation of state boredom within educational settings: A comprehensive review. *Educational Psychology Review*, 24(1), 89-111.