Data Analytics in Web-based Education in the Higher-education Classroom

Master's Thesis

Sambit Praharaj

Data Analytics in Web-based Education in the Higher-education Classroom

THESIS

submitted in partial fulfillment of the
requirements for the degree of

MASTER OF SCIENCE

in

COMPUTER SCIENCE

by

Sambit Praharaj
born in Rourkela, India

TUDelft

Web Information Systems
Department of Software Technology
Faculty EEMCS, Delft University of Technology
Delft, the Netherlands
http://wis.ewi.tudelft.nl

Data Analytics in Web-based Education in the Higher-education Classroom

Author: Sambit Praharaj
Student id: 4512960
Email: `s.prahraj@student.tudelft.nl`

Abstract

Attention span of students in a classroom is very short. To overcome this, different active learning methodologies have been used in the past. Active learning keeps the students busy and engaged throughout the lecture. It breaks the lecture into certain time intervals by intermixing breaks, demonstrations and questions after each interval. For using active learning, clickers and laptops are commonly used in higher education classroom. Most experiments in higher education classroom studying different characteristics of students like learning performance and attention, use clickers and laptop. But, most of these experiments are in a controlled setting, not scalable and compromise the privacy of students. We overcome these problems in an active learning setup in the higher education classroom where we use a web-mediated teaching tool called `ASQ`. `ASQ` is a web application that helps to give presentation in a classroom where the presenter has control over the flow of the presentation. `ASQ` also allows the presenter to interleave the presentation with questions, videos and other interactive JavaScript components. Anyone can anonymously join a presentation in `ASQ` using a web browser. `ASQ` tracks the activity of every student interaction by generating event logs each second. In the previous work using `ASQ`, it has been shown that these logs could be used to infer the attention level of students in the classroom. The goal of this thesis is to gather insights about the fine-grained study behaviour of students in a higher education classroom by analyzing these event logs.

We investigate (i) the effect of lecture elements (like the difficulty, relative positioning and spacing of questions; and duration of discussion in the slides) on study behaviour (like attention level, performance and reaction time while answering questions) of students; (ii) the relationship that might exist between attention percentage of students and their participation in the in-class questions; (iii) if students are taking external help when answering questions during the lecture and the relationship that might exist between their tendency to take external help with the difficulty of questions. We conduct our study in a classroom of around 300 students, for 15 lectures in the Web and Database Technology course at TU Delft taught by 2 instructors. We find significant effect of (i) spacing of questions on reaction time and instructor on performance; (ii) length of discussion time associated with a slide on the attention level of students which agrees with past studies; (iii) relative positioning of questions on the performance of students. However, we do not find significant effect of difficulty of questions on performance and reaction time of students while answering these questions. We also find significant effect that students with more attention percentage participate more in the in-class questions. Finally, we find that students take external help while answering questions but the tendency to take external help does not depend on the difficulty of questions.

Thesis Committee:

Chair:	Prof. dr. ir. Geert-Jan Houben, Faculty EEMCS, TUDelft
University supervisor:	Dr. ir. Alessandro Bozzon, Dr. ir. Claudia Hauff, Faculty EEMCS, TUDelft
Committee Member:	Dr. ir. Marco Zuniga, Faculty EEMCS, TUDelft

Preface

I would like to thank my parents for always acting like a powerhouse to recharge me whenever I needed any help. I would also like to thank Dr. Bozzon for making me aware about the details of their experiments in higher education classroom. This sparked my interest as I came to know that the research involved live experimentation in a higher-education classroom. Although it seemed tedious to think of such a study in the beginning, it was really exciting to undertake a experiment of such stature.

During the course of my thesis work, I met Dr. Hauff and Dr. Bozzon frequently. They supervised me and guided me throughout my thesis journey. I would like to thank both of them for their immense support, encouragement and feedback. They stood like a rock even during my tough times in the beginning and showed me the right path to follow. It helped me a lot for the timely completion of my thesis.

I would also like to thank Vassilis and other researchers from the Università della Svizzera Italiana (USI) in Lugano, Switzerland who helped me a lot in the beginning with all the technicalities. They guided me with different aspects like integration, testing, etc. Last but not the least, I would like to thank Prof. Houben for setting up my base in research and teaching me the art of successful scientific writing during the seminar in the first year of my Masters.

Lastly, I am extremely happy to be able to undertake such a large scale study as a part of my thesis. I learned a lot during this time: designing the experiments; cleaning and analysis of data; the art of presenting a good scientific work; and the ability to critique my own work successfully. I would be happy to continue my research to improve the existing teaching and learning in the classroom.

Sambit Praharaj
Delft, The Netherlands
August 14, 2017

Contents

List of Figures

Chapter 1

Introduction

Higher education (or post-secondary education) generally occurs in large classrooms. Large classrooms pose significant challenges for both the teachers and the students [47, 15]: teachers cannot give right amount of attention to every student; and it is also difficult to help every student participate simultaneously. Student population in a classroom is very diverse. The motivation level and approach (like surface, deep or strategic [1]) to learn different things is also quite different [17, 36]. Motivation level in the classroom has an impact on the attention rates of students [44, 29]. Attention rates of students vary significantly during lectures with regular instances of inattention (or daydreaming and mind wandering) [49, 10]. The length of attention span in a classroom can be as low as 30 seconds [10] or can be as high as 10-20 minutes [56, 65].

To deal with the problem of short attention span, different approaches have been adopted in the past. One of the methods suggested by Johnstone and Percival [27] was to intermix the lectures with breaks. Another method is active learning. It is a mode of learning where students are not only listening passively to the lectures but also busy in reading, writing, discussing, solving problems (or questions) or watching demonstrations during the lectures [8]. These activities involve the students in the learning process and makes them think about their own actions [8]. Active learning is known to be effective when attention span of students is short[45]. Evidence also suggests that active learning helps in achieving positive learning results [24, 41, 6] like higher retention of lecture content and better performance in the final exam.

Active learning can be aided by ICT (like clickers, laptops, smart phones, etc.). Using ICT in classroom is common[63, 59]. A popular class of devices used in classroom is clickers [2] [11, 35, 42]. Clickers are otherwise known as audience response systems (ARS) or personal response systems (PRS) which have been invented long back but it's frequent use started only in 2003 [32]. Higher education students feel more engaged [18, 55] while using clickers. They report fewer attention lapses (i.e., loss of attention during lectures) while using it [10]. The learning performance of students measured by a final exam is better while using clickers as compared to the other group not using clickers [37].

On the other hand, several studies [20, 67] in the classroom show that laptops can act as a major source of distraction. The students can use it for different purposes like gaming, social networks, etc., which promotes aimless use of technology. During aimless use almost two-third of the time is spent on doing activities not related to study (like chatting, browsing lecture unrelated content, etc.) [46]. But, laptops are not banished [38, 69] as they are an integral part of the tech-enabled modern day classroom. We can ward off this distraction by reducing the aimless use [31, 12] of laptop. For instance, students who used laptop in a structured manner (i.e., with a definite purpose) spent significantly more time in activities related to the lecture like note-taking as compared to the students who used laptop in an

---

[1] Make calculative effort to do whatever it takes to achieve higher grade.

[2] Clickers are polling devices which can be a hardware or software installed in a laptop or smart phone that can help in question and answer in a classroom.

unstructured manner [31]. Active learning can also help in reducing the aimless use of laptop [19, 70]. Besides, students perceive the directed use of laptop in classroom (i.e. for classroom activity like question and answer or note-taking) to have a positive effect on their attentiveness, engagement and learning performance [72, 52].

1.1 Motivation

In addition to the study on measuring learning performance (or learning outcome) and attention lapses of students, many other studies have been conducted in the higher education classroom to study other characteristics of students: inattention based on pattern of paper notes taken [54]; and inattention based on heart-rate, skin temperature and eye-gaze [7, 4]. But, these past studies in higher-education classroom suffer from many limitations. Most studies conducted in the past were with controlled classroom conditions [67, 53, 1]. It does not reflect the environment of a real world higher education classroom. Furthermore, controlled conditions can have a detrimental impact on the learning experience of students which can affect the outcome of the study.

Some studies on inattention in classroom used volunteers to detect attention lapses [53]. Other studies to determine inattention in classroom track the paper notes [54] of every student. They have a significant drawback of not being scalable for large studies. Besides, others [7, 4] used different physical and physiological measures like EEG (Electroencephalogram), heart rate monitor, temperature sensors, etc., to determine inattention in the classroom. It can be very intrusive and pose additive cognitive burden on the students in the classroom.

Most studies on inattention in higher-education classroom do not track the attention of students second by second [10, 49]; instead, they use self-reports or probes at certain intervals of time during the lecture. It gives us an idea about the attention states of students at those particular times only. This process can hamper or improve the concentration level of students. On one hand they are continuously intervened, so they may feel disturbed but on the other hand they are also alert because of the frequent interruptions which may give us a false representation of the attention levels. There have been many studies to determine learning outcome of students in the classroom by using different question placement strategies[66, 13]. But, these studies are mostly limited to comparing the in-between questions with the at-the-end condition [66] and do not try for other variations of placement of questions during the lecture. Besides, the privacy of the students are also not preserved in most of these experiments [49, 48, 57] which is a matter of concern today as no one wants to compromise their personal space and privacy.

Some of the challenges in detecting inattention and other characteristics of students in the higher education classroom can be summed up as:

- Can we effectively determine the characteristics of the students' in an active learning setup in the classroom in an unobtrusive fashion?

- Can we implement this setup in a real world higher education classroom?

- Can it work effectively on a large scale?

To overcome these challenges, Triglianos et al. used ASQ[3] [61] for a study on Computer Science students for 3 lectures in the higher education classroom. ASQ is a web-mediated tool which allows synchronous flow of lecture between the presenter (instructor) and the students (viewers) where anyone can join a lecture with the help of a connection URL. ASQ has a rich set of different questions (like multiple-choice, multiple-answer, fill the blanks, SQL queries and highlight) which can be interleaved in the presentation. ASQ assigns a unique connection token to every student; so, their privacy is preserved. ASQ also has a

---

[3]http://asq.inf.usi.ch/

mechanism to collect the event logs each second based on the interactions of every student with it. They used these logs to analyze and determine the attention variation in the classroom based on the the *inferred attention states* (i.e., bored, distracted, working on a question, interacting with a non-question slide, idle, etc.) obtained by analyzing the logs. But, the usability of `ASQ` to determine other characteristics of students like participation in the questions posed during the lecture, reaction time while answering questions, etc., is yet to be fully understood.

1.2 Problem Statement

Inspired by previous work from Triglianos et al. [61], we aim at furthering the understanding of how technology can be used to support active learning and help us to understand study behaviour of students in the classroom in detail. We do this by means of a large scale longitudinal study of around 300 undergraduate students of first year Computer Science in the Web and Database Technology course at TU Delft for 14 lectures (each of 90 minutes duration). We also make use of the `ASQ` platform. We interleave the lecture slides in `ASQ` with videos, questions, interactive JavaScript components and SQL extensions. Instead of inferring only attention based on the event logs, we try to analyze the event logs to infer different raw characteristics like reaction time, performance, attention level, participation in questions, etc.

We seek answer to this main research question:

RQ: What can we effectively determine regarding the *fine-grained study behaviour* of students in an *active-learning* setup in the real-world classroom while using a *privacy-preserving* web-mediated tool for teaching large student groups?

To explain this question further, we need to elaborate certain terms in detail. *Fine-grained study behaviour* refers to dis-aggregated characteristics of students like performance, reaction time while answering the questions, change of attention level, etc. *Privacy preserving* means that we do not store the details of any student; instead, `ASQ` assigns every student a random unique identifier for the lecture.

1.3 Research objectives

To answer the main research question above, we organize the work around three main sub-questions.

- **RQ 1: What evidence can we find regarding the *study behaviour* (i.e., performance, attention level and reaction time) of students when *lecture elements* (like relative positioning of questions, difficulty of questions, duration of slides and different type of question placing strategies) vary in a continued manner?**

  Objective 1: We want to study the effect of these *lecture elements* on the *study behaviour*. The study also investigates the effect of different question strategies used during the lectures. This question will be pursued in Chapter 2 and Chapter 4.

  As mentioned above, *study behaviour* of students comprises of the attention level of students, time required to submit response to a question i.e., reaction time and the performance in the questions posed during the lecture. *Lecture elements* comprises of the components used during the lecture like the difficulty of questions (i.e., easy and hard) used in the lecture, duration of discussion time associated with the slide, relative positioning of questions (i.e., proximity of questions to the content related to them) and the different types of question placement strategies.

- **RQ 2: Does the measured *attention* of students using `ASQ` have any impact on their *participation* in the in-class questions?**

  Objective 2: We want to find the effect of the attention of students on their participation in questions during that lecture. This question will be pursued in Chapter 4.

  Attention is measured as *Attention percentage* i.e., the percentage of time a student is paying attention during the lecture. *Participation* is measured as the percentage of questions to which a student submits response during the entire lecture.

- **RQ 3: Do students take help of *external resources* while answering questions? Does it relate to the difficulty of questions?**

  Objective 3: We want to find if the students take help of external resources based on our assumptions below. Then, we want to find it's dependence on the difficulty of questions. This question will be pursued in Chapter 4.

  This research question is based on the following *assumptions*:

  - Each student uses their own laptop and do not search from their peers laptop while answering: this was also evident as reported by the field observers in each lecture.
  - If a student looses focus [4] while answering a question before submitting it, then he/ she might be taking help from external resources.

1.4 Contributions

This thesis provides the following original contributions:

- **An extended and customized version of the `ASQ` platform**

  We contributed to the development of the `ASQ` platform by improving the conversion of slides (from pdf to HTML) to overcome the following issues:

  - Overlap of text and figures
  - Overlap of some slides when there are multiple animations
  - Repetition of certain text

  Proper display of slides in `ASQ` is a necessity as the lectures need to convey clear information to the students so that they understand the lectures easily. These issues were resolved using different techniques which are described in detail in chapter 3.

- **An analysis of the impact of *lecture elements* on the *study behaviour* of students**

  Different question strategies have significant effect on the reaction time of students while responding to questions. The percentage of correct responses also vary with different question strategies. Different instructors have an impact on the performance of students. We do not observe significant effect of question difficulty on the performance and reaction time of students in most lectures. But, we do observe that attention level of students fluctuates significantly more in slides whose length of discussion time is in some minutes as compared to some seconds. Proximity of questions to the content related to that question has a significant effect on the performance of students.

---

[4] As most questions were of the length: 30, 60 or 90 seconds; so we assume that they do not get involved in other study unrelated browsing or chatting when they loose focus for such a short duration in-between questions although this *cannot* be verified.

- **An analysis of the impact of measured *attention* percentage of students (using `ASQ`) on their *participation* in the in-class questions**

  Students with more attention percentage participate more in the questions posed during the lecture. We also observe significant differences of the participation in the in-class questions for students with lower and higher attention percentage during the lecture.

- **An investigation over the adoption of *external help* while answering questions during the lecture, and an analysis of it's relation to the difficulty of questions**

  Students take external help while answering questions based on our assumptions explained in the objectives. We find that more than half of the students take external help in each of the lectures. We do not find significant impact of question difficulty on the likelihood of students taking external help for most of the lectures.

The preliminary results of this work are part of a long paper accepted for publication at *UMAP* (User Modeling, Adaptation and Personalization) conference. It is titled as: *Measuring Student Behaviour Dynamics in a Large Interactive Classroom Setting* [62].

1.5 Outline

The remainder of the thesis is organized as follows: First there is a short description about the literature and the background in chapter 2; then we move on to know about ASQ in detail and the web engineering done to prepare and render the lecture slides in ASQ in chapter 3; next, we discuss the context of using ASQ in our experiments, the experimental design and the methodology that we use for analyzing the data set where we explain it based on manual annotation and the sub-research questions in chapter 4; then we discuss the results in chapter 5; next, we have a discussion of the results in chapter 6; finally, we conclude in chapter 7 where we also discuss the future scope of our work. Figure 1.1 gives an overview of the outline of the thesis.

Introduction

| Motivation | Problem Statement | Research Objectives | Contributions |

Background and Related work

Classroom study characteristics

- Active Learning
- Methodology of detection
- Spacing of questions

ASQ

Architecture — Event Aggregation — Web Engineering

Experimental Design and Methodology for Analysis

Question spacing strategy | Context of using ASQ | Manual annotation | Methodology for Analysis (RQ1, RQ2 and RQ3)

Results

RQ1: Find and report the effect of lecture elements on study behaviour | RQ2: Find and report the relation between attention percentage and participation in the in-class questions | RQ3: Find and report if students take external help while answering questions

Discussion

Discuss the results | Threats to internal validity

Conclusions

Conclusions and Future Work

Figure 1.1: Outline of the thesis

Chapter 2

Background & Related work

In this chapter, we provide an analysis on the state-of-the-art literature relevant to our research goals. First we discuss different studies that determine the instances of [in]attention (or mind-wandering) and also discuss the attention level of students in the classroom. Next, we discuss regarding active learning which helps in improving the attention level of students in the classroom and reduces task unrelated thoughts. Then, we move on to the next section where we discuss the use of technology like clickers and laptop in classroom which is used in most studies as an active learning methodology. Next, we discuss regarding the different methods (i.e., intrusive, non-intrusive and combined methods) that have been employed in higher education classroom in the past studies to determine [in]attention and other study characteristics of students. It gives us an idea about the different characteristics (like varying attention level, learning performance, etc.) of students measured in the higher education classroom. Then, we move on to the studies on spacing of questions which also affects learning performance. Finally, we conclude the chapter with an overview of the related work and it's relevance in our research context.

Figure 2.1 shows the interconnection and relationship of different components in this chapter that we are going to discuss now. There can be different classroom study characteristics of students like inattention, attention levels, learning performance, amount of lecture notes, browsing style, etc. Active learning can help in improving the attention level and learning performance of students in a classroom. Technology like clickers and only laptop is used in a classroom to implement active learning. Clickers have positive effects on students. On the other hand, use of only laptop can have positive or negative effects. Moving onto the next study characteristic i.e., learning performance, spacing of questions (i.e. *in-between* the lecture or *at-the-end* of the lecture) can have an effect on learning performance. To measure these study characteristics of students in the classroom, different detecting methods like intrusive, non-intrusive and a combination of both intrusive and non-intrusive have been used.

2.1 Inattention in the classroom

Inattention in the classroom is defined as instances of loss of attention during the lecture. It is synonymous to mind-wandering in the studies that we will discuss. In some studies authors measure [in]attention with the help of different concentration levels of students or the time spent by students in thoughts unrelated to the ongoing lecture.

Regular instances of in-attention of students is very common in the classroom. Bunce et al. [10] performed a experiment on the undergraduate chemistry classroom to find the duration for which students can be attentive in the classroom. They used clickers to record the attention of students self-reported by students. They found that attention span declines after 15 minutes in the lecture and can be as low as 30 seconds.

Risko et al. [49] found similar effects of higher attentiveness in the first half of the
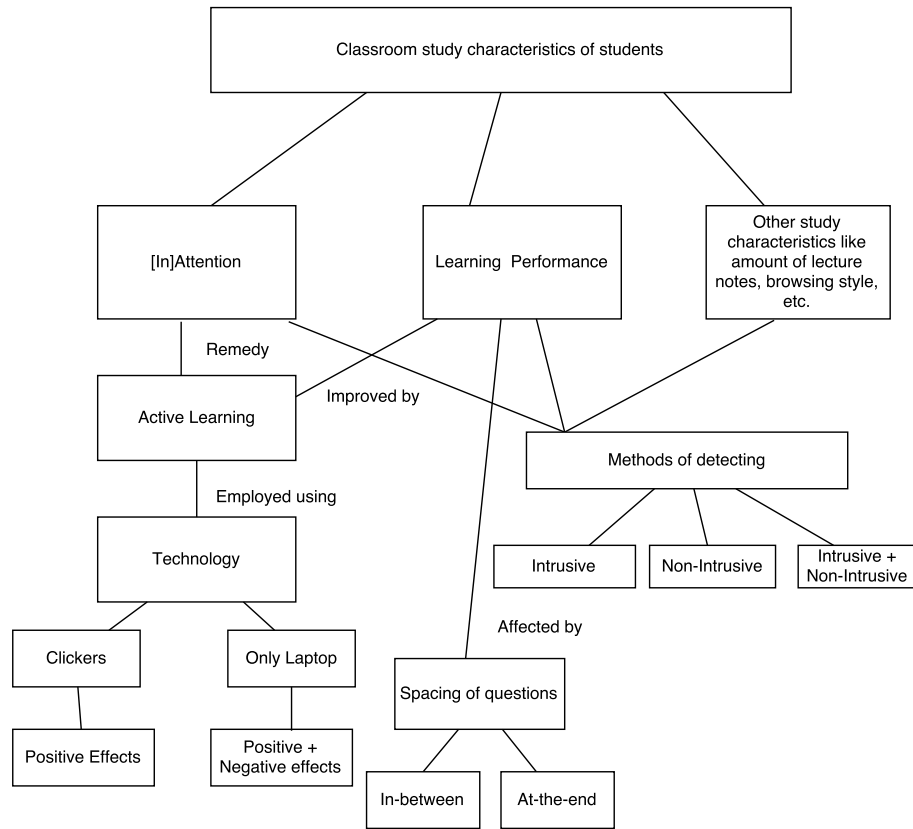
Figure 2.1: Overview of the background and related work

lecture. They performed an experiment on 60 students in an undergraduate classroom in University of British Columbia. The students were asked with the help of 4 probes (i.e., 2 in the first half of the lecture and 2 in the second half of the lecture) to answer "Yes" or "No" to the question asking them if they are mind-wandering. Students had less instances of mind-wandering in the first half of the lecture as compared to the second half of the lecture; the instances of mind-wandering also increased with time. Students who were having more instances of mind-wandering performed poorly in a test immediately after the lecture.

Studies on measuring attention levels in the classroom has also be done in the past [57]. Stuart used a questionnaire to determine the student attention during the lectures[57] at periodic intervals of 5 minutes across 12 lectures in a classroom. The questionnaire consisted of a scale from 1 to 9 ranging from no concentration to maximum concentration in the class. He found that the concentration level of students increased during the start of the lecture and was maximum at 10-15 minutes into the lecture. But, then it declined slowly as the lecture progressed to the end. Next, we discuss the impact of in-attention due to task unrelated images and thoughts which affects the learning performance of students.

Lindquist and Mclean [33] experimented on 463 undergraduate psychology students in the classroom. Students were asked to record self-reports if they are in-attentive due to task unrelated images and thoughts. The responses of students followed a simple "Yes" or "No" format. These self-reports were made after the use of a auditory probe to alert the students at periodic intervals in the lecture. It was found that students who had least interest in the course had more task unrelated images and thoughts. Even, the amount of notes taken by those students was less. Another important finding was that students who had more task unrelated thoughts and images performed poorly in the course exams too.

To summarize, attention span in the classroom can be as low as 30 seconds. Instances of inattention increases as the lecture progresses. Students with less instances of inattention perform better in the final exam as compared to students with higher instances of inattention.

2.2   Active learning

In this section, we discuss about active learning which can improve the attention level and learning performance of students. Active learning is defined in [8] as a mode of learning where students not only passively listen to the lectures but also actively participate in the lecture. They are busy in reading, writing, discussing, solving problems or watching demonstrations during the lecture. In [16], a list of active learning strategies have been described. One of the strategies is the use of in-class questions which is combined with the feedback techniques that use formative assessment (i.e., instructor and the students get the complete view of the whole class performance but the grade is not part of their final course grade) which is similar to the strategy of in-class questions and feedback (and assessment) incorporated in our active learning setup.

Gauci et. al. [21] found the benefits of active learning by using PRS [1] (or clickers) in the classroom. They performed a experiment on 175 undergraduate science students of a second year physiology course. Before starting the use of PRS, they recorded the grades of a prerequisite course (i.e., a physiology course in a previous semester) for every student. It was found that students who used PRS performed better than the students who did not use it. Also, students who had the lowest score in the prerequisite course (i.e., <60%) showed significant increase of performance in the end semester as compared to the ones having mid (60-75%) and high (>75%) score in the prerequisite course. It means that the use of PRS improved the performance of students and the impact was most on the students with lower prerequisite scores.

Yoder and Catherine [70] also got similar results of better performance while using active learning methodology. They performed an experiment in a women psychology course where they compared the performance of groups of students, one of which was exposed to only normal lecture, video representations and autonomous representations; and the other group was exposed to active learning techniques along with the normal lecture. It was found that in both "between subjects" and "within subjects" design, the group that was exposed to active learning performed better in the multiple choice questions.

Szpunar et al. [58] also observed the benefits of active learning. 80 students in an introductory statistics course (shown as a 21 minutes video lecture) were part of two experiments conducted by them to study regarding mind wandering (or loss of attention), note-taking and learning of students. The 21 minute video lecture was divided into 4 segments. In the first experiment, they compared two groups of students where one half was exposed to a memory test after each segment (i.e., the "tested condition") and the other half in the "non-tested condition" was exposed to a memory test only after the fourth segment. The students also reported the number of times they mind wander and whether it increased as the lecture progressed before the final test and after the last segment. It was found that the "tested group" had less instances of mind wandering and it was significantly less likely to increase with the progress of the lecture as compared to the "non-tested group". Even the percentage of notes taken was more for the "tested group".

For experiment 2, there were 3 groups (i.e., "tested", "restudy" and "non-tested"). Only addition from first experiment was the "restudy" group which was showed the question and solution that was used for testing the "tested" group after each of the segments. Another difference was that the mind wandering probes with a ("Yes/No") question was posed after each segment to each of the groups. The findings were similar to first experiment with the

---

[1]PRS are polling devices which can be a hardware or software installed in a laptop or smart phone that can help in question and answer in a classroom.

tested group having lowest mind wandering followed by a increase for restudy and non-tested group. For the notes taken, the findings were complete opposite of the instances of mind wandering. But for the performance, the order in decreasing fashion was: tested, non-tested and restudy. One-way ANOVA also showed a significant effect of condition (tested, restudy and non-tested) on mind wandering. So, we can conclude from their study that interpolated memory tests (i.e., tests in-between the lecture) improve attention, note taking and learning.

Use of active learning shows positive results on the learning performance of students and less instances of inattention during the lecture. It also makes the students engaged during the lecture. Now, we discuss about different ICT like clickers and laptop which are used in higher-education classroom to implement active learning.

2.3   Use of clickers and laptop in higher-education classroom

Clickers and laptop are the most common ICT that have been used in the classroom to implement active learning. Bunce et al. [10] used clickers and demonstrations in an under-graduate chemistry classroom. They used three chemistry courses (i.e., chemistry I, chemistry II and chemistry III). One group of students in chemistry I course were exposed to only demonstrations; another group in chemistry II used only clickers; and the third group in chemistry III used both clickers and demonstrations. Students were given a clicker to report attention lapses in each of the chemistry groups. In addition to it, only students in chemistry II and III brought their own clickers to participate in the classroom questions using clickers. They found that both the groups using clickers (i.e., the participants of chemistry II and chemistry III) had reported significantly less number of attention lapses over the lecture. It means that the use of clickers reduced the number of attention lapses during the lecture.

But, Martyn [37] did not find a significant impact for the use of clickers on performance of students. Martyn [37] performed an experiment on 92 students in an introductory computer information class. They students were divided equally in four groups: the first 2 groups used clickers; and the other 2 groups participated in classroom discussions instead of using clickers. There was a pretest before the start of the experiment. There was no significant difference in the performance of students who used clickers and who used class discussions. After the experiment, there was a final exam. It was found that the performance of the group using clickers did not differ significantly from the group that did not use clickers; contrary, to most studies pointing to the benefit of using clickers. But, the mean perception of students using clickers (based on a survey questionnaire to rate responses from 1 to 5) was much higher as compared to the other group. It shows that students perceive the impact of clicker use in classroom as advantageous and positive.

Contrary to Martyn's results [37] , Gauci et al. [21] found a increase in performance of students using PRS (or clickers) in higher-education classroom as compared to students not using PRS. We have already discussed it in detail in the *Active learning* section above. Mayer et al. [39] also found the advantage of using clickers in a psychology class; students using clickers to answer in-class questions performed significantly better in the final exam as compared to the other groups (i.e., the one not using clickers and the other acting as control group who did not answer any in-class questions).

Use of only laptop in the classroom also affects the learning outcome of the students and their peers. In [1], the effect of laptop use on the learning outcome of the students was studied. For this, around 800 students were part of control and experimental group: the experimental group comprised of students who were divided into two zones (i.e., laptop permitted and laptop free area); and the control group comprised of students who can sit anywhere they like. The experiment showed that non-laptop users performed significantly better in an exam conducted later as compared to the laptop users when compared to the estimated score obtained from their pre-academic indicators. It was also found that the

performance of non-laptop users was not significantly different in either the experimental (i.e., zoned) or the control (i.e., not zoned) group indicating that the use of laptop by some students had no effect on the surrounding students. So, it is understood that the use of laptop has a detrimental effect on the learning performance of students but not on the learning performance of the neighbouring students. Ravizza et al. [48] found that most of the time laptop is used for non-academic purposes in the classroom.

Internet usage while using laptop is examined by Ravizza et. al. [48] on undergraduate students in a introductory psychology course, where students are logged by a proxy server to connect to the internet. So, the pattern of internet usage of each student is tracked during the lecture but, is kept confidential. Later, this content was partitioned into two groups as academic (lecture related content) and non-academic (lecture unrelated content). Non-academic internet use was common among the students and it was inversely proportional to the performance (in an exam conducted after the lecture) of students who were busy in non-academic internet search for most of the time. This problem arises due to *unstructured* use of laptop.

Unstructured use of laptop means that we do not care for what purpose students are using their laptop in the classroom [31]. Students control their own laptop use. It leads to the undirected use of laptop [20]; students are busy doing all lecture unrelated (browsing, chatting, playing games, busy with the emails, etc.) activities with their laptop in the classroom. Unstructured use of laptop makes the students dis-satisfied [68], they tend to be involved in more off-task behaviour [20, 22] and also have less learning [20, 22] of lecture material.

To examine this, Kay et al. [31] performed an experiment on 177 university students in two different courses of philosophy or family issues: one course was followed with an *unstructured* laptop use; and another was by *structured* laptop use. At the end of each of the course, there was a anonymous survey where each student was asked to rate the percentage of time they were busy in either on-task (note-taking, academic activities) or off-task (personal email, personal messaging, games, movies) behaviour from 5 options as 0%, 1-25%, 26-50%, 51-75% and 76-100%. It was found after a paired-t-test that the students in *structured* laptop condition spent a considerable more time in "on-task" activities as compared to the "off-task" activities when compared to the *unstructured* laptop condition. So, directed use of laptop shows promise.

Multitasking while using laptop is common in a classroom[67, 28, 53], occurs frequently and has a detrimental impact on student learning (or performance) [67, 28, 53]. Wood et al. [67] examined the impact of multitasking in lectures on 145 undergraduate students. Each lecture was followed by a 15-item multiple choice quiz. Students were randomly assigned to one of the seven experimental conditions (explained in table 2.2). It was found that students who multitask had significant negative impact on their learning performance. Sana et. al. [53] found another impact of multitasking on the nearby peers when they conducted study on 44 students of a introductory psychology course. They concluded that multitasking has detrimental impact on the students and also their peers. Although multitasking is a negative impact of laptop use, active learning can reduce the aimless use of laptop[45, 19]. We have also tried to use an active learning set-up with the help of ASQ in such a way that students are constructively engaged throughout the lecture with their laptop.

Clicker use had a lot of positive effects like improved learning performance and less attention lapses (loss of attention) during the lecture. But, the use of only laptop has both positive and negative effects. Structured (or directed) use of laptop can help to negate the negative effects of using laptop.

Table 2.1: Comparative overview of the studies in higher-education classroom: **#L(Duration)** column shows the number of lectures and duration of each of them. Column **Controlled** indicates whether the experiment was conducted in a controlled classroom (✓) or a real-world classroom (✗) setting. Column **Privacy** indicates whether the study preserved the privacy (✓) or not (✗). **Logging type(s)** indicates how we determine the students' behaviour: allotted condition (allotted experimental condition is an indicator of the behaviour), self-reports (students report on their own behaviour/ concentration level/ distraction), volunteers (observers sit with the students and observe them) or online (i.e., through a proxy server or web browser as ASQ in our work). The **Evaluation** shows how students were evaluated on their learning performance. **Outcome** indicates the main conclusion of the study.

	#L(Duration)	#Students	Controlled	Privacy	#Exp. conditions	Logging type(s)	Evaluation	Class	Reward	Outcome
[49]	3 (60 minutes)	64	✓	✗	2: High mind-wandering – Low mind-wandering	Alotted condition	4 MC questions immediately after the lecture	Psychology, Economics & Classics	$5 or course credit	Mind-wandering is more in second half of the lecture and inversely proportional to performance
[48]	15 (100 minutes)	84	✗	✓	3: Academic (class-related) internet usage – Non-academic internet usage – No internet usage	Online	Final exam	Introductory Psychology	Course credit	Non-academic internet usage is in majority and inversely proportional to performance
[33]	1 (50 minutes)	463	✗	✗	2: Low task unrelated images & thoughts (TUIT) – High task unrelated images & thoughts	Self-reports	Final exam	Introductory Psychology	None	TUIT is inversely proportional to the course interest, amount of notes taken & performance
[21]	36 (50 minutes)	175	✗	✓	2: Personal response system (PRS) usage – No PRS usage	Self-reports by PRS	Midterm, final exam	Psychology: Control of Body Function	None	PRS use improved performance
[1]	13 (50 minutes)	800	✓	✓	2: Zoned (laptop free & laptop permitted) – Not Zoned	Allotted condition	Final exam	Bio 93: DNA to Organisms	None	Laptop use has a detrimental impact on performance
[37]	4 (75 minutes)	92	✓	✓	2: Clickers – No clickers (Classroom discussions)	Allotted condition + Self-reports	Pretest + Final exam	Introductory Computer Information Systems	None	Use of clickers had no significant impact on performance but student perception favoured the use of clickers
[31]	NA	177	✓	✓	2: Structured (Active) laptop use – Unstructured laptop use	Self-reports by online survey	None	Philosophy or Issues in the Family	None	Structured use of laptop promoted more percentage of time to be spent in note-taking and other academic activities
[57]	12 (60 minutes)	113	✗	✗	2: Low concentration – High concentration	Self-reports by questionnaires	None	Medical	None	Concentration level is maximum at 10-15 minutes in the lecture and then decreases slowly towards the end

Table 2.2: Continued — Comparative overview of the studies in higher-education classroom.

	#L(Duration)	#Students	Con-trolled	Pri-vacy	#Exp. conditions	Logging type(s)	Evaluation	Class	Reward	Outcome
[67]	3 (20 minutes)	145	✓	✗	7: Facebook — Texting — Normal Technology Use — Word Processing — pen and paper — MSN — email	Allotted condition	15 MC questions immediately after the lecture	Research methods, Statistics	$15 or course credit	Multitasking has significant negative impact on the performance
[53]	1 (45 minutes)	44	✓	✗	2: Multitasking — Non-multitasking	Allotted condition	20 MC questions immediately after the lecture	Introductory Psychology	Course credit	Multitasking in laptop has significant negative impact on performance and also affects performance of peers
[20]	20 (75 minutes)	137	✗	✓	2: Open laptop — Closed laptop	Weekly Self-reports	4 exams and 10 homework assignments	General Psychology	None	Laptop use has a detrimental impact on student learning and performance
[39]	NA	385	✓	✗	3: Use of no clickers + no group discussions — Use of clickers — Use of group discussions	Allotted condition	Mid-term exam + Final exam	Educational Psychology	None	Group using clickers performed significantly better than other groups
[58]	1 (21 minutes)	80	✓	✗	2: Tested — non-tested	Allotted condition + Self-reports	In-class questions & Final exam	Statistics	None	Tested group have significantly improved performance in the end segment of lecture. Also they take more notes and have less mind-wandering.
[70]	NA	45	✓	✗	2: Active learning — Non-active learning	Allotted condition	Final exam	Psychology	None	Performance in active learning was much better to the other group
Our study	14 (90 minutes)	89-319	✗	✓	Multiple: Determined based on the interactions with ASQ	ASQ event logging	123 questions (MC, MA, highlight, fill-in-the-blank) interspersed in lectures	Web & Database Technology	None	ASQ can be used as an effective tool to implement active learning in a large higher-education classroom to give fine-grained insights to the instructor

2.4 Comparative overview of the studies in higher-education classroom

Table 2.1 and 2.2 shows the comparative overview of the studies in higher-education classroom discussed till now. All the experiments are conducted in a classroom similar to our experiments. But, most of them are in a controlled setting and do not preserve the privacy of students. The duration of most of the lectures is also similar to our study. Most of these experiments use testing between the lectures, mid-term exam or a final exam to measure the learning performance of the students. When we observe the "outcome" column in the table, then we can find the positive effects of active learning in the classroom. We can also observe in the "outcome" column in all the experiments that the use of clickers and structured use of laptop improved the performance of the students. Next, we discuss different methods that have been employed in the past experiments in higher-education classroom to determine the study characteristics of students.

2.5 Methods of detecting different study characteristics of students in the higher-education classroom

Broadly, we classify the methods into *three* categories based on the nature of it's implementation: *intrusive* (which has possibility of causing disruption or annoyance), *non-intrusive* (which does not cause disruption) and *combined* (which can be both intrusive and non-intrusive). These methods are used to study different characteristics of students like *attention level*, *learning performance* in the questions, *amount of notes* taken, etc.

2.5.1 Intrusive

First, we discuss the different *physiological* measures like heart rate measurement, temperature sensors, gaze detector and EEG that help us to detect study characteristics of students in the classroom. Then we discuss another intrusive measure i.e., use of self-reports or probes at certain time intervals.

Heart rate measurement: Bligh [7] monitored the heart rate during the lecture of 40 minute duration and found that heart rate decreases as the lecture progresses. According to him, heart rate is an indicator of arousal which also shows the *attentiveness* of an individual. So, the decrease in heart rate indicated the dipping of attention levels with time. But, this could not be verified as an immediate test showed that students performed better on the content taught in the last 20 minutes of the lecture.

Skin temperature sensor and gaze detector: Bixler et al. [4] performed a experiment to detect mind wandering on 178 candidates by using the data obtained from the temperature sensor and gaze detector. They found that mind wandering is linked to a rise in skin temperature and with fewer and longer eye fixations.

EEG (Electroencephalogram): EEG helps in the detection of electrical waves inside the brain to determine the activity in the brain. It is required to connect the head with the electrodes to perform EEG. Liu et al. [34] performed a study to determine the degree of human attention by using mobile brainwave sensors. They extracted the necessary raw features from the data collected by EEG and then used a SVM (Support vector machine) classifier to determine the best set of features which can determine the attention level of the students. They concluded from the study that attention levels of students decline in the second half of the lecture and is also dependent on the type of instructor.

Self-reports or response to probes: Some studies [10, 49, 57, 58] used self-reports (or probes) at specific intervals of time to determine the attention variation in class; this process hampers the concentration of students frequently as they have to respond to these probes. Stuart et al. [57] performed a study on medical students who were asked to record their concentration levels at the beginning of the lecture and then at frequent 5 minute

intervals throughout the lecture by filling a questionnaire containing questions based on a scale. Bunce et al. [10] obtained self-reports with the help of clickers as discussed earlier to determine the duration of attentiveness of students.

2.5.2 Non-intrusive

Here, we discuss different non-intrusive measures like WiFi networks, web-browser and pattern of paper notes that have been used in the past studies to study the classroom behaviour of students.

WLAN (WiFi networks): Tracking classroom behaviour on a large scale in the classroom by analyzing the internet access data generated from WLANs (WiFi networks) and mobile phone sensors of the students was done in [71]. It helped to gather insights of attendance, delay in attending lectures and performance in a non-intrusive fashion for each student but, compromised the privacy of the students as it kept track of the student ID which was mapped to the MAC address of their mobile phones. Based on the usage of smart phone in the class, they calculated the distraction percentage for every students as the percentage of time they were busy with their phones. They found that distraction percentage was least in the first 10 minutes of the lecture but increased as the lecture progressed. They found that students who attended more lectures had higher GPA in that semester but did not find any effect of the punctuality of arrival to lectures on the performance of the students; the correlation of delayed arrival or early escape with the performance was not clear, contrary to the studies performed later by [30] by using smartphone sensors who showed that early arrival to a class strongly correlates with performance.

Web-browser: ASQ [61], a web-mediated tool allows synchronous flow of lecture between the presenter (instructor) and the students (viewers) where anyone can join the lecture with the help of the connection URL via a web-browser. The advantage of using web-browser is that anyone can connect easily without much hassle and it is not cost-prohibitive as the use of sensors to track the heart rate or brain activity. Another advantage is that it can be scaled up easily as everything is handled by the server and the database without the need of human volunteers. They [61] logged the events of all the users in the class every second and then deduced different inferred attention states like "Bored", "Distracted", "Searching for a solution", "Following", etc. These attention states helped them to conclude that attention breaks occur regularly and increase as the lecture progresses. They used questions in between the slides during the lecture. They concluded that the introduction of interactive teaching elements like questions in-between the lectures helped to improve the attention level of students.

Similar to the study of using ASQ, Niwa et al. [43] used reactive presentations in which they varied the animation of the objects used in the presentation slides in real time to control the attention of the audience. They were also able to react to unexpected questions by highlighting important objects with varying animations.

Analyzing the pattern of notes: In [54], they analyzed the pattern of notes taken by 160 students who participated as a part of a course. Videotaped lectures were shown to the students and some of the statements were highlighted by spoken or written cues. It was found that the information noted down by the students in their notes declined with the progress of lecture and also written cues were mostly found in the notes. The comparison of notes was made with the script of the recorded lecture to determine the amount of information noted down. The retention of material tested in the form of multiple choice questions was almost the same for different portions of the lecture. So, the conclusion regarding the attention level of students was mixed as the notes showed that attention level declined with time as the lecture progressed but, the retention tests do not reveal any such change in the attention levels. A study in a real world classroom setting (as mentioned by the authors) will give promising results.

2.5.3 Combined (both intrusive and non-intrusive)

Here, we discuss strategies of determining classroom study behaviour which use both intrusive and non-intrusive measures.

Interpolated memory tests and probes: 80 students in an introductory statistics course (shown as a 21 minutes video lecture) were part of two experiments conducted by Szpunar et. al. [58] to study regarding mind wandering, note-taking and learning of students as discussed above in detail. They used both interpolated memory tests (i.e., tests in-between the lecture at certain time intervals) and probes in the form of self-reports by students to find the in-attention level and learning performance of students.

2.5.4 Comparative overview of different methods of study in higher-education classroom

Table 2.3 shows the overview of the different methods (i.e., intrusive, non-intrusive and combined) that we discussed till now in the higher-education classroom. We find that most studies do not preserve the privacy and use external volunteers to monitor the note pattern and other activity of students which do not make them scalable. Most of the studies are also in controlled settings. But, using Web-browser and Wifi to track student activity is efficient and scalable as it can track a large number of students simultaneously.

Table 2.3: Comparative overview of the studies in the higher-education classroom based on the different methods i.e., *intrusive*, *non-intrusive* and *mixed*. The **Nature** column shows these 3 methods. Column **Privacy** indicates whether the study preserved the privacy (✓) or not (✗). Column **Controlled** indicates whether the experiment was conducted in a controlled classroom (✓) or a real-world classroom (✗) setting. Column **Volunteers** indicates whether the study used the help of human volunteers (✓) or not (✗) **Duration** column shows the duration of each lecture. **Testing** shows the strategy employed to evaluate the learning outcome of students during the lecture.

Nature	Study	Privacy	Controlled	Volunteers	Duration	Students	Testing
Intrusive	[4]	✗	✓	✗	NA	178	Multiple-choice pre-test and post-test
	[34]	✗	✓	✓	NA	24	NA
	[10]	✗	✓	✗	50	23-74	NA
Non-intrusive	[71]	✗	✗	✗	45 or 90	700	NA
	[48]	✓	✗	✗	100	127	Final exam
	ASQ in our exp.	✓	✗	✗	90	89-319	In-class questions
Combined	[58]	✗	✓	✗	20	80	Interpolated memory tests
	[53]	✗	✓	✓	45	44	Multiple-choice test

2.6   Spacing of questions

In this section, we look at the literature on spacing (or placement) of questions. There has been limited study on spacing of questions; most of these studies concentrated on two types: 1) in-between the lecture and 2) at-the-end of the lecture. Moreover, the spacing of questions can also affect the learning performance of students. As we are also studying the impact of spacing of questions on the learning performance of students in the classroom, so we need to understand the works that have been done in the past. Weinstein et al. [66] made a study of the placement of practice questions where they compared the performance of students in two types of question spacing strategy i.e., practice questions placed in between the content in the slides after every 1-2 slides or all the questions were placed at the end of the presentation. 175 online (via Amazon Mechanical Turk) subjects and 62 subjects in the undergraduate classroom participated in the experiment. The learning performance of students was measured in the end when everyone appeared for a final test after their lectures. It was found that students who were exposed to *in-between* condition of practice questions during lecture performed better than the *at-the-end* condition and had a significant difference in their mean score. But, there was not much difference in performance of these two groups in the final test which was given after a long time of the lecture. The reason for this reduction in difference in the final test can be attributed to the fact that when question is asked at the end of the lecture then it requires more effort to retrieve the information as it was conveyed long back in the lecture; thereby improving the performance in the final test to bring it at par with the students exposed to the *in-between* condition. Earlier study by Bjork et al. [5] also found that retrieval of information which required more effort to retrieve increased learning.

According to [13], spaced questions are generally considered more recent as they occur at multiple locations throughout the lecture. So, they have close proximity to the material taught recently as compared to their massed (i.e., all the questions are placed together) counterparts. But, the effect of recency [14] (i.e., closeness of related content to the questions) also needs to be taken into account while designing the spacing of questions for a lecture otherwise it invalidates the claim of any difference that we can observe in the *in-between* questions and *at-the-end* questions. Recent studies made by Bloom and Toppino [60] similar to the study conducted by Greene et. al. [23] brought into light a very important fact that even a minor *recency bias* can increase or negate the effect of the question spacing strategies on learning performance. In reality, it is difficult for the instructors to redistribute the content and the questions while designing the lectures.

Verkoeijen et. al. [64] examined the effects of spacing (or placement of questions) on two categories of students i.e., having higher and lower memory-performance for college students where they discovered spacing effect is more on students with higher memory-performance as compared to the lower memory-performance ones. So, spacing can have different effects on different types of students. Figure 2.2 makes it clear where the solid arrow lines show the different placement strategies of practice questions used in the past and the dotted arrow line shows the effect of the placement strategies on different types of students and also the effect of recency on the placement of questions.

To summarize, most of the studies on spacing (or placement) of questions compare the *in-between* placement of questions with the *at-the-end* placement of questions. The proximity of questions to the lecture content results in better performance of the students. Spacing of questions also have different impact on students with different memory-performance.

2.7   Chapter conclusion

Attention level of students in a classroom varies. The attention span can be as low as 30 seconds. We study this by using `ASQ` as shown in table 2.4. To improve the attention level of students, active learning has been used in the classroom. It has shown promising
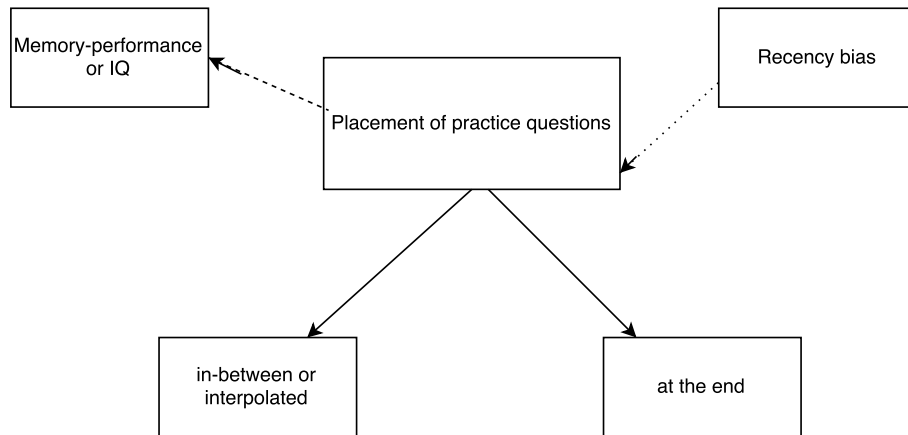
Figure 2.2: Placement of practice questions: types and effects. The different types of placement strategies used for practice questions are shown in solid arrow. The dotted arrow shows the effects i.e., effect of recency bias on the placement of questions and the effect of placement of questions on students with higher and lower memory-performance.

Table 2.4

Focus of past studies	Relevance in our context
Short attention span in classroom	Using `ASQ` to check the attention span in a classroom.
Use of active learning improves the attention level and learning performance in a classroom.	Use of questions in between slides in `ASQ` to check the impact of active learning on attention level of students in a classroom.
Examine the effect of placement (or spacing) of questions on the learning performance of students.	Examine the effect of different types of question spacing strategies on the learning performance and reaction time of students.
Regulated and directed use of laptop can improve learning in classroom.	Students using `ASQ` for most time of the lecture should participate in the in-class questions.

results to improve the learning performance and attention level of students. Clickers and laptop are the common active learning methodologies used in classroom. Clickers reduce the instances of in-attention and improve the learning performance. But, use of only laptop can have negative impact on learning performance due to it's aimless use. So, it can be used in a regulated constructive manner to reduce it's undirected use and improve learning outcome of students. In our context, `ASQ` helps for regulated use of the laptop. Tests in-between the lecture can also be used to improve the learning outcome of students. Similarly, we use active learning components like questions and other interactive components in our context. Most studies in the past examine the effect of placement of questions during the lecture on learning performance in only two categories (i.e., in-between and at-the-end). To measure these learning outcomes, [in]attention in the classroom and other study characteristics, different measures have been adopted as: intrusive, non-intrusive and a combination of both intrusive and non-intrusive. Most of these studies in higher-education

classroom are controlled, intrusive and not privacy preserving. However, we use `ASQ` in a non-intrusive manner which also preserves the privacy of students. Apart from different study characteristics (also studied in the past) like attention level and learning performance, we study other characteristics like reaction time, attention percentage, participation in the in-class questions. To the best of our knowledge, no study in the past has tracked these characteristics of students *simultaneously*.

Table 2.5 shows the comparison of our work with some of the past studies in higher-education classroom and also the study that used `ASQ`. The comparison is based on different criteria as mentioned in the table. As compared to the past work without `ASQ`, our work takes into consideration a real-world classroom setting where we integrate questions and other components within the lecture. We track different characteristics of the students each second instead of few minutes which also makes our monitoring mechanism with `ASQ` superior. Although the study using `ASQ` [61] tried to measure inferred attention states along-with the performance of students, the inferred attention states can be misleading and may give us a false depiction of attention in the classroom [2]. So, we adopt a *novel* methodology of analyzing the event logs to determine the raw characteristics of students. Moreover, compared to the past study using `ASQ`, we conduct a large scale study spanning for around 2 months where we also employ *novel* question spacing strategies for testing the students between the lectures.

Table 2.5: Comparison of our work to some past studies that did not use `ASQ` and the study that used `ASQ` based on multiple criteria

Criteria	**Our work**	**Past work without ASQ**	**Past work with ASQ**
Nature of study	Real-world class-room settings	Controlled classroom settings	Real-world class-room settings
Methodology	Questions, videos and interactive components integrated with the lecture	Questions not integrated with the lecture	Questions integrated with the lecture
Tracking frequency	Each second	Every few minutes	Each second
Privacy	Preserved	Not preserved	Preserved
Scalability	Scalable	Not scalable	Scalable
Size of study	Large	Large and Small	Small
Testing during the lectures	Use of different question placement strategies	Use of questions in-between lecture or at the end of the lecture	Use of questions which are randomly placed in the lecture

---

[2]For instance, some students with inferred attention state as bored were actually following the lecture from the beamer and having minimal interaction with `ASQ`.

Chapter 3

ASQ

`ASQ`[1] project was launched by researchers at the UniversitÃ della Svizzera Italiana (USI) in Lugano, Switzerland to support web-based education in the classroom. The `ASQ` research project is still being developed in collaboration with TU Delft. We collaborated with the researchers from Lugano for thesis study to extend the `ASQ` platform and generate insights of the students in the classroom. `ASQ` is a web-based platform which helps an instructor (otherwise known as *presenter* in `ASQ` terminology) to create and deliver interactive presentations using HTML5. It allows presenters to stream live lecture slides to an audience (otherwise known as *viewers* in `ASQ` terminology). The presentation slides can contain questions, videos, interactive elements (like JavaScript elements and SQL lite) embedded in between them which can be answered by the students. The instructor has the option to view the statistics of the responses in real time. So, the responses from the audience are collected in real-time and the presenter can give timely feedback to the students. The aim of this application is to improve the efficiency of active learning between a large audience and the instructor.

Here, we discuss about the `ASQ` architecture in detail and then we discuss about the mechanism used by `ASQ` to capture and aggregate the events. These events are generated by students every second which helps us to analyze the student interactions later. Finally, we discuss the web engineering involved in converting the slides to the desired format which can be displayed using `ASQ`.

3.1 ASQ architecture

Figure 3.1 shows an abstract overview of `ASQ`'s architecture. The front end of `ASQ` renders the slides which can be viewed by the instructors and students using their web browser. Now, Google Chrome is the only supported web browser. These slides are made using impress.js and reveal.js in HTML5. The presenter has the option to login in `ASQ` as shown in figure 3.2 to upload the slides. Rich set of questions like multiple-choice, multiple-answer, fill the blanks, SQL Queries, highlight, classify, code, rating, etc., are supported by `ASQ`. The presenter can embed these questions with the help of custom HTML elements defined in `ASQ`. The presenter and the viewers (students) have two separate views as shown in figure 3.3 and 3.4; the viewers can login with the URL shown using laptop, tablet, or smart phone. Each student is given a unique random identification token by `ASQ` which lasts till the presentation is live (or active). Once the presenter stops the presentation from `ASQ`, then the token expires. This token helps us to associate each student with the events generated by them without compromising their *privacy*. The student is disconnected on closing the tab that shows the `ASQ` presentation. The flow of presentation is controlled by the presenter and it is synchronous with all the viewers.

---

[1]http://asq.inf.usi.ch/
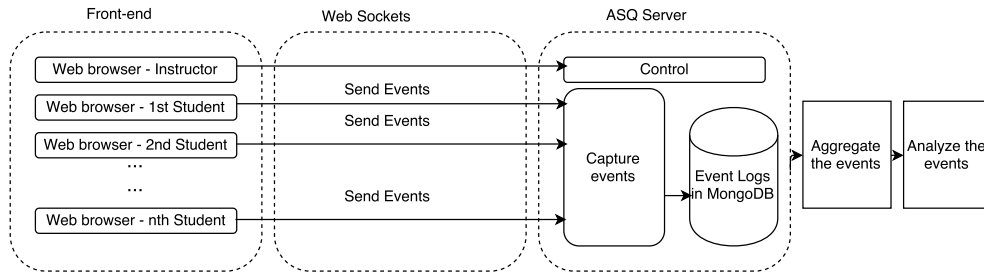
Figure 3.1: High level view of `ASQ` architecture. Modified from [62].
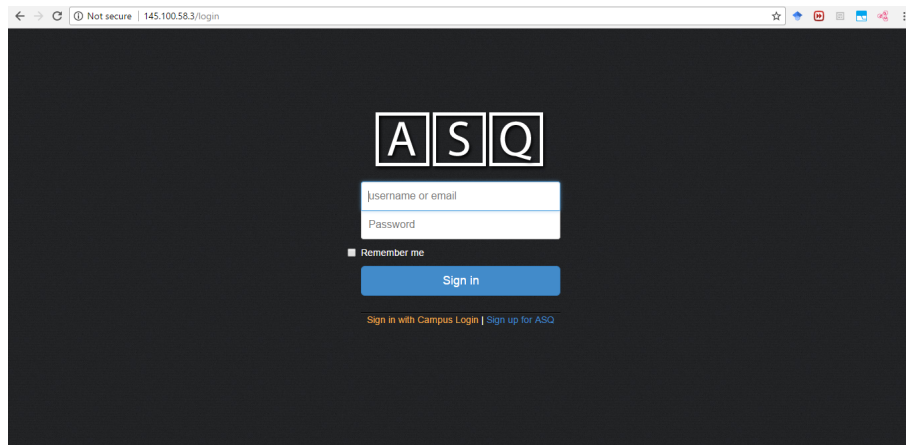
Figure 3.2: Login page of `ASQ` for the presenter.
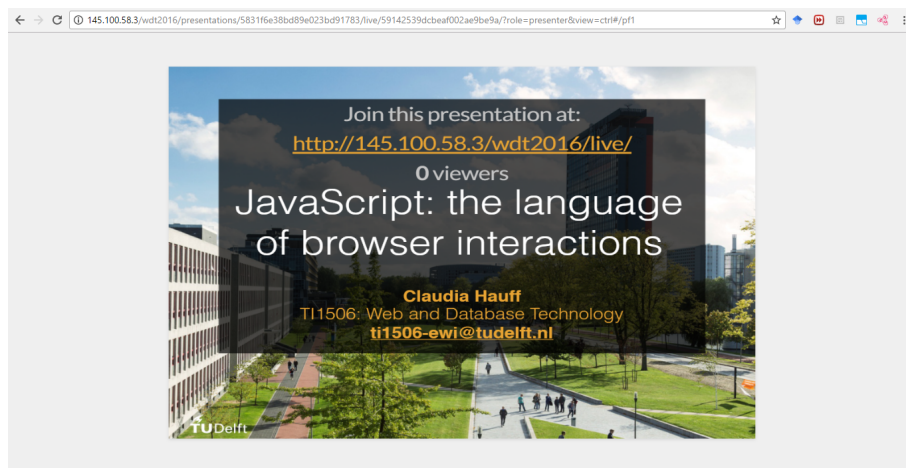
Figure 3.3: Sample window for the presenter.

The viewers (students) and the presenter (instructor) are connected by *Web sockets* to the `ASQ` server. Web sockets help to open a new interactive session for communicating between the browser and the server. The web sockets carry event-driven reply from the server immediately after receiving a request from the browser. They do not work on the queuing mechanism where the server addresses the requests from the browser based on the priority given to the queues. So, the responses received by the browser are quite fast. `ASQ` uses these web sockets. So, the statistics of responses are immediately shown after few students

Figure 3.4: Sample window for the viewer.

answer a question. It also helps the presenter to control the flow of the presentation.

The events that are generated every second by the students are captured as event logs and stored in MongoDB. It contains details about each student's activity like the time spent in answering the question, focus activity, etc. We will look into it in details in the next section and also discuss regarding the aggregation of events.

3.2 Capturing and aggregation of events

`ASQ` captures the interactions of students as they happen and immediately stores them as event logs in the database. These events help us to gather insights on the activity of students every second like whether they are active or idle with `ASQ`, answering questions correctly, etc. Here, we will explain in detail how `ASQ` captures the events and what is the meaning of these events. Next, we discuss how we aggregate these events.

3.2.1 Capturing events

Table 3.1 shows the various events captured by `ASQ`. These events are captured immediately after being generated during the presentation and stored. Examples of events include: 1)connected and disconnected event indicates if a user has `ASQ` open in any tab or has closed the tab rendering `ASQ`. 2) tabhidden and tabvisible events indicate if the tab rendering `ASQ` is visible or not. Similarly, all the captured events are explained in detail in the table 3.1.

3.2.2 Event aggregation

`ASQ` captures and aggregates the events each second. Table 3.2 shows the details of aggregated events which are formed by aggregation of the events shown in table 3.1. The aggregated events are binary in nature i.e., they can either have a true or false value; which eases the analysis of data later. In total there are 7 aggregated events i.e., *connected*, *exercise*, *visible*, *focus*, *idle*, *input* and *submitted*; *exercise*, *input* and *submitted* are valid only during a *question* (otherwise known as *exercise* in `ASQ` terminology). The *input* aggregated event is also valid during non-question interactive slides like the interactive JavaScript components.

3.3 Web engineering during slide conversion

Here, we discuss how we performed the slide conversion from pdf to HTML and embedded questions, videos and other interactive components in between the presentation slides in

Table 3.1: Web browser events captured by ASQ. Modified from [61]

Event Name	Description
tabhidden	0 pixels from the browser tab that displays the ASQ web app becomes visible on the user's screen.
tabvisible	At least 1 pixel from the browser tab that displays the ASQ web app becomes visible on the user's screen.
windowfocus	The browser window that displays the ASQ web app receives focus.
windowblur	The browser window that displays the ASQ web app loses focus (blurs in HTML terminology).
focusin	the DOM document emits an focusin event. Similar to windowfocus but the event does not bubble (propagated to ancestors)
focusout	the DOM document emits an focusout event. Similar to windowblur but the event does not bubble (propagated to ancestors)
exercisefocus	An ASQ exercise (can have one or more questions) HTML element receives focus.
exerciseblur	An ASQ exercise HTML element blurs.
input	There is student input in the browser window that displays ASQ
questioninput	ASQ question types emit this event when there is student input.
exercisesubmit	A student submits the solution to an ASQ exercise.
answersubmit	A student submits an answer for an ASQ question (an exercise can have multiple questions).
idle	Emitted by the browser window that displays the ASQ web app when none of the above events has occurred for 10 seconds.
connected	A student connects to the ASQ server.
disconnected	A student disconnects from the ASQ server.

ASQ. ASQ has a provision which allows us to upload the pdf version of the slide and then convert it to the desired HTML version required for the presentation in lectures. Figure 3.5 and 3.6 shows the upload mechanism and home screen respectively after uploading a presentation in ASQ.

But, the HTML version that is initially generated by ASQ does not work properly for slides with animations and slides containing overlapping images and text. The possible issues are listed as:

- **Multiple text overlap:** Different text elements overlapped on each other in the slide. An example image is shown in figure 3.7.

- **Multiple text and image overlap:** Different text elements overlap on image elements in the slide. An example image is shown in figure 3.8.

- **Overlay of empty box:** Sometimes empty boxes overlay on some slides containing animations without displaying the content inside the box. An example image is shown in figure 3.9.

Table 3.2: Aggregated events based on the browser events. Modified from [61]

Name	Description
`exercise`	True when the current slide has an exercise.
`connected`	True when the student browser is connected
`visible`	True when the tabvisible event occurs. It means that the tab is visible.
`focus`	True when the browser has focus on the tab or exercise related to the lecture i.e., focusin, windowfocus or exercisefocus event occurs.
`idle`	True from the time of an idle event until one of tabhidden, tabvisible, windowfocus, windowblur, focusin, focusout, exercisefocus, exerciseblur, input, questioninput, exercisesubmit and answersubmit occurs.
`input`	True when an input or questioninput event occurs. This state is valid only on slides that contain exercises.
`submitted`	True when the student has submitted at least once this exercise (as indicated by an exercisesubmit or answersubmit event). This state is valid only on slides that contain exercises.
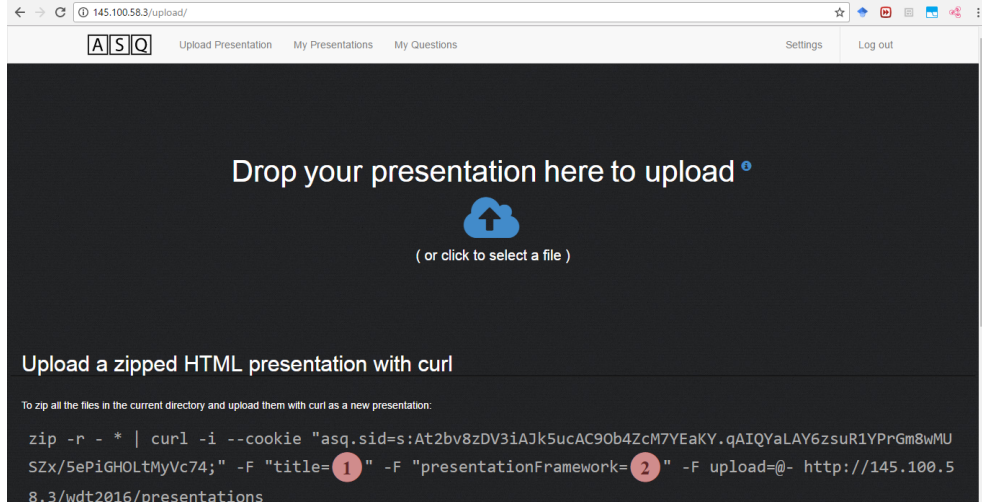
Figure 3.5: Upload mechanism in `ASQ`.

Therefore, we describe a *semi-automatic* methodology, organized in 3 steps, to fix the possible issues with automatic translation in `ASQ`. Figure 3.10 shows the abstract view of the adopted methodology.

The details of the methodology are as follows:

- **Analysis**

  First, the important thing is to scan the HTML code generated by `ASQ`[2] and determine the reason behind the inconsistencies (i.e., overlap and overlay). To achieve this, we need to determine the problem by making small changes in the HTML version and

---

[2]The HTML code generated by `ASQ` can be obtained by downloading the initially uploaded presentation.
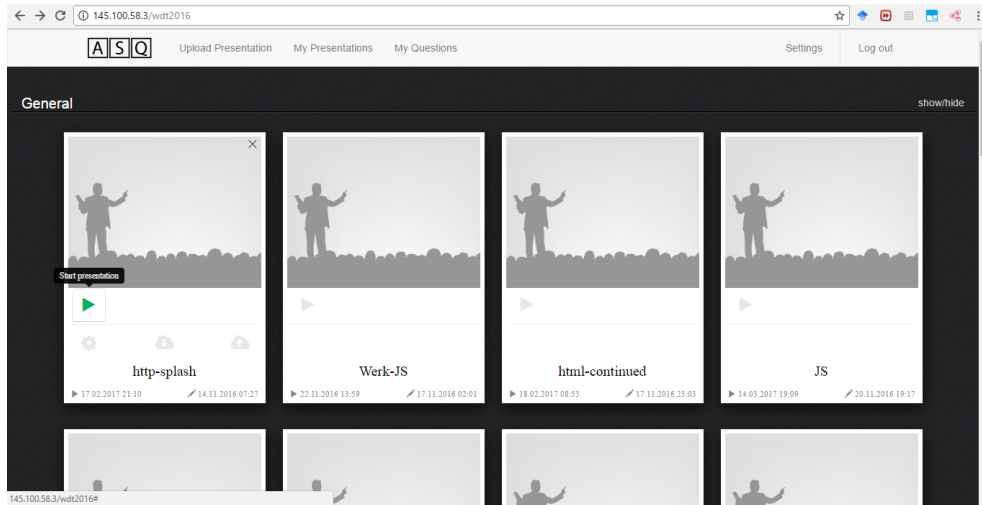
Figure 3.6: Screen after presentation upload in `ASQ`.

HTML5 overview

Figure 3.7: Multiple text overlap after conversion in `ASQ`

Figure 3.8: Multiple text and image overlap after conversion in `ASQ`

Appcache "workflow"
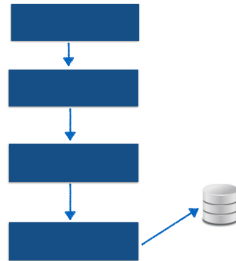
55

Figure 3.9: Overlay of empty box after conversion in `ASQ`

Analysis

| Substep problem | Repetition problem |

Design

| Resolve overlap and overlay problem | Insert videos | Insert questions and other interactive components |

Testing

| All overlaps and overlays are resolved | Questions work properly in real-time | Video works uninterrupted |
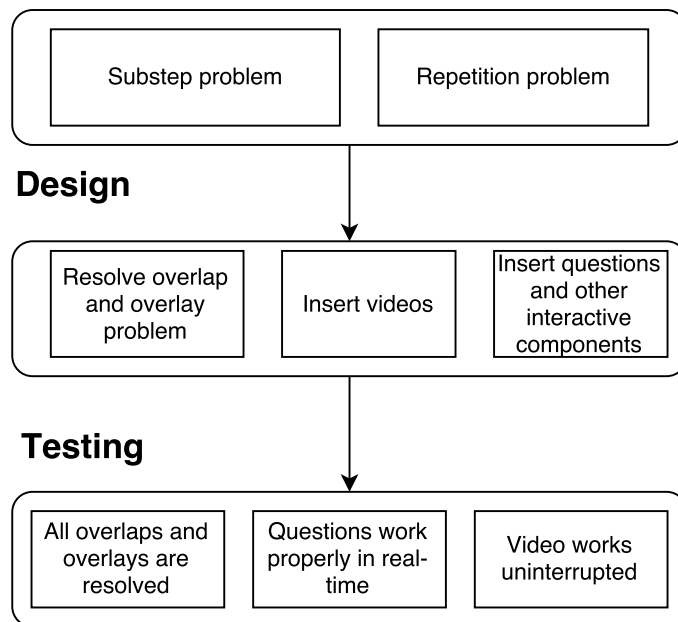
Figure 3.10: Abstract view of the methodology used to prepare the slides for the presentation in `ASQ`.

see it's result in the browser so as to determine the root cause of the problem and find the potential solution. We discovered 2 main problems in the analysis:

– **Substep problem:** Each unique slide in HTML code generated by `ASQ` has a class named as "step" and the animations of that slide are named as "substep" class in that slide. A "step" is a slide, a "substep" is the content within a slide that is revealed with an extra keypress. This content in the "substep" class overlaps with the content of the "step" class. It causes the overlap and overlay problems. Figure 3.11 shows a sample of the "step" and "substep" class generated by `ASQ`.

– **Repetition problem:** Due to some internal technical fault during slide conversion in `ASQ`, sometimes text was repeated in HTML code in case of animations
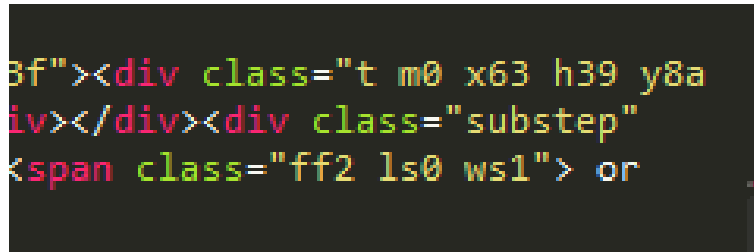
Figure 3.11: Generation of "substep" class after conversion in `ASQ`

in the slides causing overlap of text in the slides.

- **Design**

  The main aims of the design are:

  – **Resolve the overlap and overlay problems:** We need to solve the overlap and overlay problems by manually digging into the HTML code which is a time-intensive process. First, we remove all the "substep" class and create a separate slide with the "step" class as shown in figure 3.12 which is the solution for the problem shown in figure 3.11. To solve the repetition problem, we need to selectively remove the text which is repeated and also replace some text with white spaces where they overlapped on images or boxes in the slides.
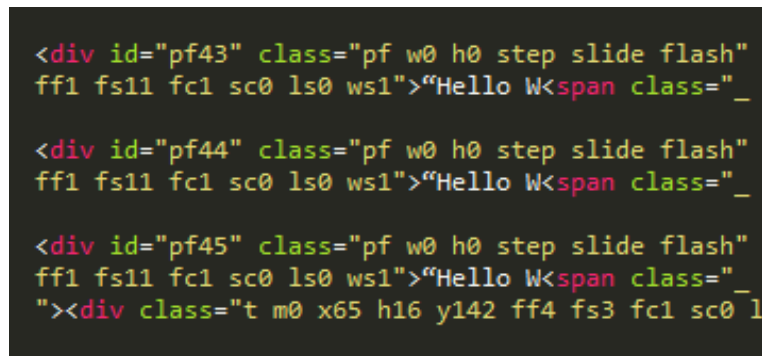
Figure 3.12: Replacement of all "substep" with "step" in the class in HTML

  – **Insert videos in between slides:** We insert the video tags in HTML based on the videos provided by the instructor.

  – **Insert questions and other interactive components in between the slides:** Details about question, video tags and other parts of slide conversion for all the lectures are provided in the GitHub repository mentioned in Appendix B.

- **Testing**

  We upload the modified presentation slide design to `ASQ` and check whether:

  – All overlaps and overlays are resolved.

  – The correct answers of each question is shown.

  – Video works uninterrupted for the presenter but not shown to the viewers. [3]

---

[3]Videos are not shown in laptop screen of each student (viewer); rather, shown on the beamer to avoid load on the server.

To summarize this section, by using these steps of analysis, design and testing, we can convert the slides from pdf to the HTML version that can be rendered properly in `ASQ`. The testing phase is iterative as we need to test each conversion and then again start the analysis if the testing does not fit our requirements.

To summarize the chapter, we discussed about a web-mediated tool called `ASQ`. We have used it for our experiments in the higher education classroom. We discussed about it's architecture, event generation and event aggregation mechanism. These aggregated events will be used for analyzing and answering the sub-research questions later. Besides, we also discuss about the methodology employed for converting the pdf version of the slides to the HTML version that can be rendered properly in `ASQ`.

Chapter 4

Experimental Design & Methodology for Analysis

We use `ASQ` for a longitudinal study of around 300 undergraduate students of first year Computer Science in the Web and Database Technology course at TU Delft. The course has two instructors; one instructor for the Web part and another for the Database part. It continued from 14th November 2016 till 20th January 2017. At the end, in total 19 lectures used `ASQ`. But, only 15 of them are classroom lectures and 4 of them are lab sessions. We only consider these 15 lectures for our analysis and do not include the labs because the labs are mostly dominated by successive question and answer periods without the style of a lecture in the classroom with fewer question periods at certain time intervals. Out of the 15, we exclude one lecture due to problems in logging the data for that lecture. All the lectures are of 90 minutes duration; there is a break of 15 minutes in between two sessions of 45 minutes for the whole lecture.

We interleave the lecture slides in `ASQ` with videos, interactive components (like Java Script interaction, cookie interaction and SQL lite) and different types of questions like multiple-choice, multiple-answer, fill the blanks, SQL queries and highlight. Figure 4.1, 4.2, 4.3 and 4.4 shows the example of different types of questions used in `ASQ`. Although videos and interactive components are part of the lectures for one of the instructors, questions are included in lectures of both the instructors.
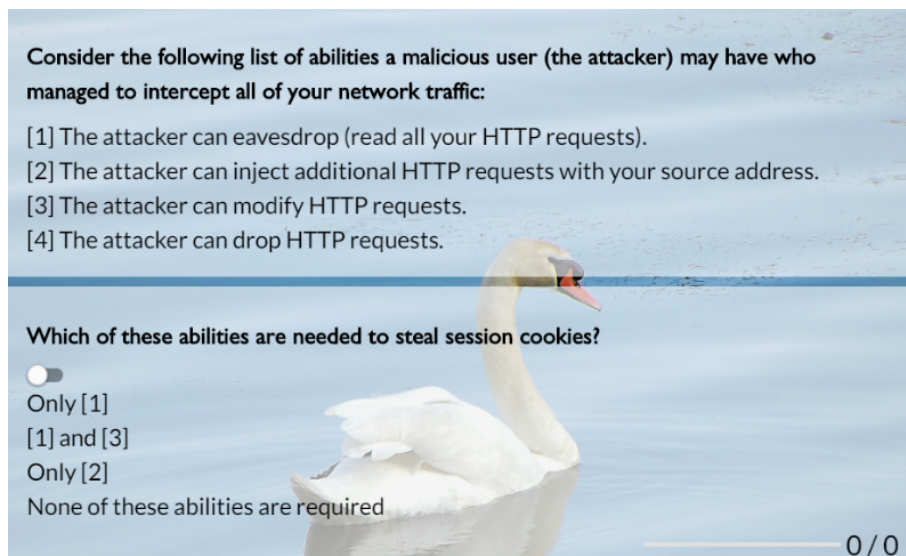
Figure 4.1: Example of multiple choice question.
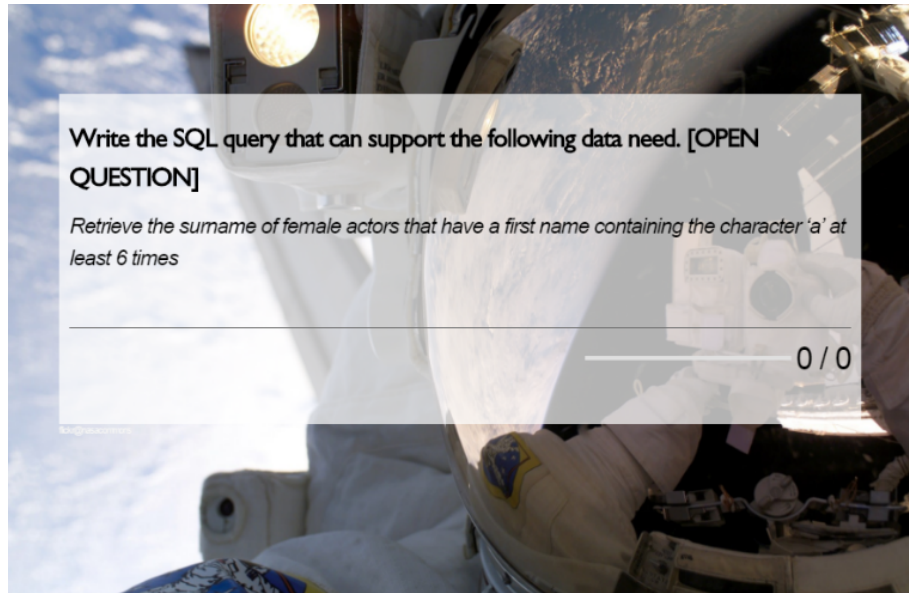
Figure 4.2: Example of fill the blanks question.

Figure 4.3: Example of SQL queries question.

The aim of our study was to find different kind of fine-grained insights about the study behaviour of students in the classroom by using `ASQ`. These fine-grained insights can be anything like the *attention level*, *performance*, *reaction time* while answering the questions, *change of attention level*, etc. To answer this main research question, we had 3 sub-research questions as: (i) What evidence can we find regarding the *study behaviour* of students when *lecture elements* vary in a continued manner?; (ii) Does the measured *attention* of students using `ASQ` have any impact on their *participation* in the in-class questions?; and (iii) Do students take help of *external resources* while answering questions? Does it relate to the *difficulty* of questions? To answer these questions, we design 2 experiments as: (i) Using different question spacing strategies across all the lectures; (ii) Enabling the option to save
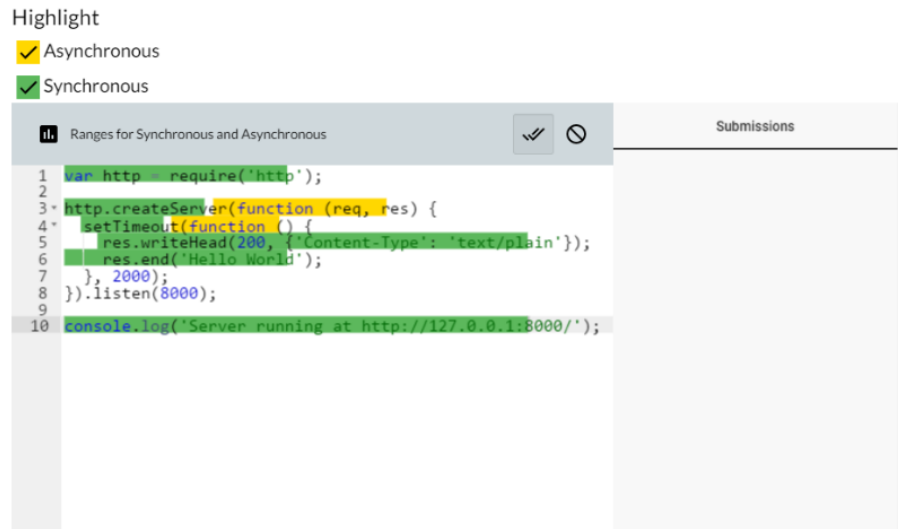
Figure 4.4: Example of highlight question requiring to highlight certain components of the code.

the cookies in `ASQ` for performing the within-subjects study for last few lectures.

First, we discuss about this experimental design, then we discuss regarding the methodology that is used for analyzing the data. In the experimental design, we discuss regarding the design of the experimental setup and then about the collection of data. Next, we discuss the methodology of analyzing the data after pre-processing the data. The pre-processing of data is described elaborately in Appendix A.

4.1 Experimental Design

In this section, we discuss the design of our experiment and then move on to discuss our data collection approach and ER model. While designing the experiment, we used different question spacing strategies across all the lectures. It helped us to vary the questions at different times of the lecture. First, we discuss regarding these strategies. Then, we discuss about enabling the option to save the cookies.

4.1.1 Use of different question spacing strategies

We used 3 different question spacing strategies i.e., *burst*, *uniform* and *increasing*. Figure 4.5 shows the three question strategies used during the lectures. Each red dot in the figure 4.5 represents the placement of each question with time. In the burst strategy there are at least 2 questions stacked back to back and these bunch of questions appear at certain intervals throughout the lecture. During our study, we used about 3 to 4 bursts in every lecture. In uniform strategy, a single question appears at nearly equal intervals of time throughout the lecture. The increasing strategy was meant to show questions which was supposed to increase with time. For example, in first 30 minutes one question was posed and then after 15 minutes 2 questions are posed, again after 15 minutes, 3 questions are posed. The instructors of the course found it difficult to implement the increasing strategy while preparing the lectures, so it was dropped later. It was only used in the first two lectures.
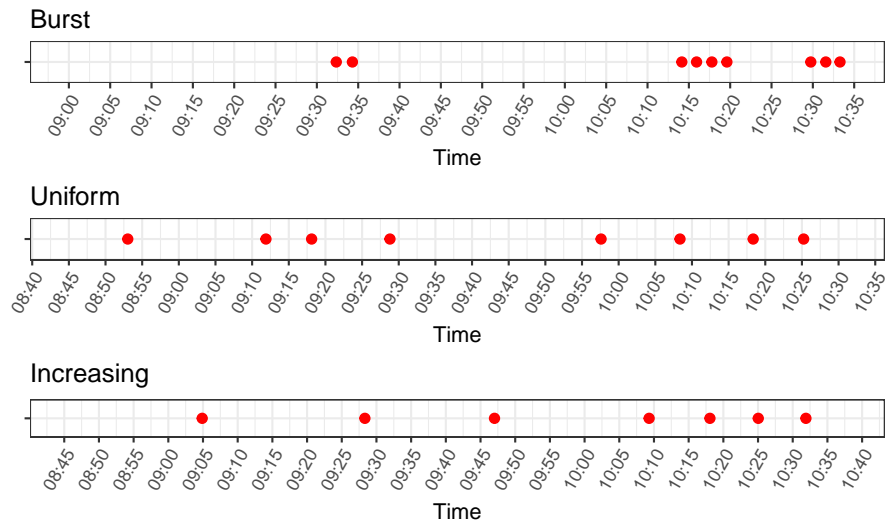
Figure 4.5: Temporal view of the question spacing strategy

4.1.2 Saving Cookies for the Within-subject study

We also conduct a *within-subject* study for last 3 lectures where we have enabled the option in ASQ to save the cookies of the users. It can help us to determine the common set of users who are present in all the 3 lectures based on the common user IDs. Although we find the same user set, we *cannot* associate them to their *real name* and *identity*. We only track the unique token (or user ID) assigned to them. So, there is no *privacy breach* by this design.

4.1.3 Data collection

This data collection helps us while processing and analyzing the data collected by using ASQ. For each of the lectures, we took different live field observations as:

- **Start, end and break time:** These times are important to determine manually as they cannot be determined from the event logs of ASQ. ASQ only keeps track of the lecture start time and lecture end time on the server which is not the true start time of the lecture as the lecture is started in ASQ long before the real start and ends long after the real end in the classroom.

- **Feedback time:** We note down the feedback time for each question to the nearest seconds to know the end time of each question. Basically, when the instructor starts showing the solution and explaining it then it marks the beginning of the feedback.

- **Number of students:** We count and note down total students present in the class before and after the break.

- **Ids:** We note the Presentation and session id which is useful for the analysis of data.

- **Miscellaneous:** We note the activity (like noise, murmuring, disturbing entry into classroom) of students and also observe other activities like how do they pay attention to the slides shown in the laptop. Mostly we observe the laptop screens of some students and try to determine in general whether they are busy in playing games or in social networking or paying attention to the lecture slides in their laptop.

4.1.4 Entity relationship model

Here, we describe the Entity Relationship model which sheds some light on the important attributes and entities that are involved in our experiment. It will help us while designing the methodology for the analysis of data.

The figure 4.6 shows all the entities, their attributes and the relationship between these entities involved across the lectures.
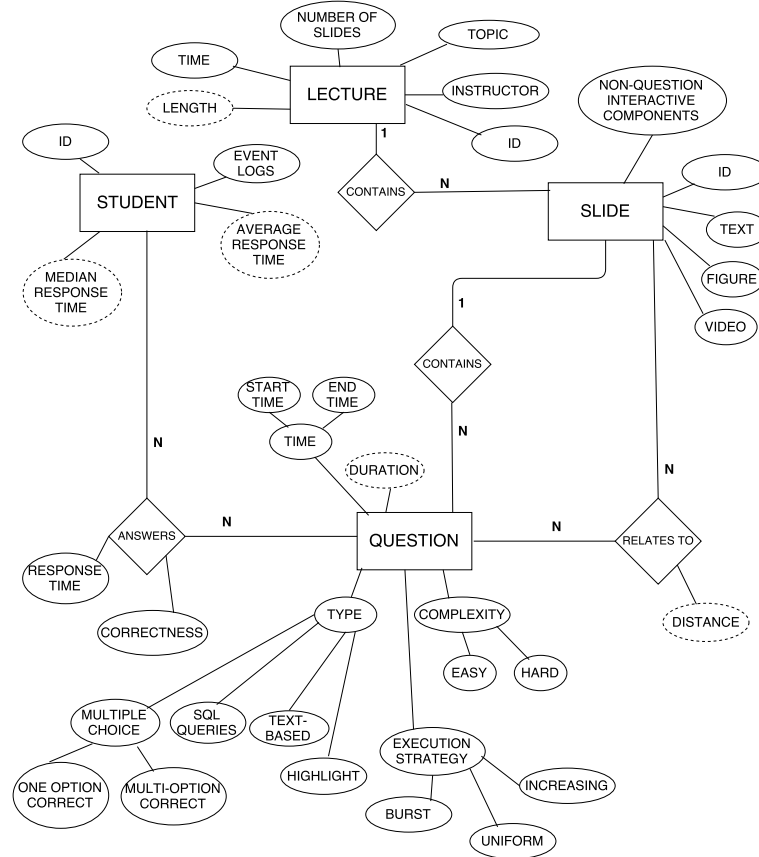
Figure 4.6: ER diagram for the in-class experiments using `ASQ`

The main entities of each presentation are *lecture*, *slide*, *student* and *question*. Every *lecture* contains slides. The attributes of each *lecture* are the "ID" (i.e., presentation ID and session ID), "instructors", "topic of the lecture", "total number of slides" and the "start" and "end" time. The "length" of the lecture is derived from the start and end time.

Similarly, each *slide* has "ID", "video", "figure", "question" and "non-question" interactive components. Furthermore, a slide contains question and one or more slides may also relate to a question which brings into picture the distance of the slide from the question; this "distance" is derived from the position of the slide and the position of the question. A *question* can have different attributes like "start" and "end" time which helps in the derivation of the "duration" of the question. It also has "complexity" (i.e. easy and hard), "types" and the "execution strategy". Each *question* is answered by many *students* who have a unique "ID" and generate the "event logs". The time taken by them to respond to a question helps us to derive the "mean" and "median reaction time". We also get the "correctness" of response from their answers to the question.

4.2   Methodology for Analysis

In this section, we define the methodology that we adopt for processing and analyzing the data based on each research question in detail. We also define the hypothesis associated with each research question.

4.2.1   Manual annotation

We discuss a *manual annotation* procedure that is used to calculate the *distance* of each question from the most recent content slide to which it is related. *Distance* is nothing but the number of slides separating the question slide from the slide with the related content. We need the distance to design the methodology for analyzing some research questions. We calculate the nearest distance of the last slide whose content matches with the content in the question based on the position of the question and that slide. If a question is not related to any slide preceding it then we mark the distance for that question as NA. It may be because sometimes the content of the question appears in the slide later or may not be present in the lecture. We design a ".csv" file of these annotations by associating the distances of each question with their respective question IDs. This annotation is necessary to answer a part of RQ1.

4.2.2   RQ 1: Effect of lecture elements on study behaviour of students

What evidence can we find regarding the *study behaviour* of students when *lecture elements* vary in a continued manner?

   We want to determine the effect of *lecture elements* (like the difficulty of questions, relative positioning of questions, strategy of questioning and discussion time associated with the slides) on *study behaviour* (like attention level, performance and reaction time) of students. Here, lecture elements are the *independent variables* and study behaviour is the *dependent variable*. To explain this research question, we have 4 sub-research questions as:

   RQ 1.1:   Is there any marked *difference* in the *percentage of correct responses*, *reaction time* and *performance* of students when different question spacing strategies (like *"Burst"* and *"Uniform"*) and different instructors are used?

   H 1.1: The percentage of correct responses in burst should be more than the percentage of correct responses in uniform strategy because questions in burst are closer to the content in the slide that relates them. The normalized reaction time in burst should be more than the normalized reaction time in uniform because students get tired due to "back to back" placement of questions in burst strategy and tend to take more time while responding to questions.

   We determine the percentage of correct responses for each question strategy (i.e., burst or uniform) associated with each of the instructors.

   We find the normalized reaction time as follows:

   We do not select all the responses of students to a question because some responses to the questions come from students very late after the feedback with the answer was shown. So, we select the responses of the students for each question that come before this feedback time. Then, we calculate the normalized reaction time. Before that, first we need to define the terms that will be used while calculating the normalized reaction time.

   For every student (viewer) v and each question q we denote the submit time of answering a question as $A(v,q)$. Start time of a question is $S(q)$. Feedback start time for the question is $F(q)$. Then, we compute the *reaction time* of a student for answering each question (denoted as $R(v,q)$) and *duration* of each question (denoted as $D(q)$) as follows:

$$R(v,q) = A(v,q) - S(q) \tag{4.1}$$

$$D(q) = F(q) - S(q) \tag{4.2}$$

Then, the *normalized reaction time* is calculated as follows:

$$NR(v,q) = \frac{R(v,q)}{D(q)} \tag{4.3}$$

This normalization is a necessary step as it helps in ensuring uniformity while comparing the reaction time with questions of varying duration. Mean normalized reaction time ($MNR(q)$) is calculated as follows:

$$MNR(q) = \frac{1}{T(v)} \sum_{t=1}^{T(v)} NR(v,q) \tag{4.4}$$

Here, $T(v)$ is the total number of students answering a question.

To find the effect of question spacing strategy on *reaction time* of students for both the instructors, we divide the students in these four groups as shown in table 4.1.

Table 4.1: **Group** shows the name of each group of the strategy and instructor combination. I1 and I2 represents the 2 instructors that taught all the lectures.

Group	Strategy + Instructor
A	Burst + I1
B	Burst + I2
C	Uniform + I1
D	Uniform + I2

These four groups cover in total 12 lectures [1] with each group covering for 3 lectures. Each group contains the normalized reaction time of the students in all the questions. We compare these groups by Kruskal Wallis test to check for any significant differences between these groups. As a post hoc test we conduct a Dunn's test with Bonferroni correction to check for significant differences between the individual groups i.e., Group A vs Group B, Group B vs Group C and so on. This is done for all the possible pairwise comparison between the groups. Similarly, we also find the effect of question spacing strategy on *performance* of students for both the instructors.

We conduct a *within-subject* study as discussed in the experimental design for the last 3 lectures where we save the cookies of the users to determine the common set of users who are present in all the 3 lectures. We perform a Wilcoxon's signed rank test to compare the performance of these common set of students across these 3 lectures.

RQ 1.2: Does the discussion time of the slides affect the *attention level* of the class before and after the slide was shown? (*Attention level* of the class can be defined as the percentage of students who are *connected*, *visible* and *focus*)

H 1.2: Slides that take minutes to discuss may show greater variability in the attention level as compared to the slides that take seconds because the attention span of students can be as low as 30 seconds [10]. With the increase in duration of discussion time in the slides, the attention level change should be more, implying the same short attention span of students.

---

[1]As 2 lectures out of the 14 lectures which used increasing strategy are not considered in this grouping.

For answering this question, we use the low level aggregated states i.e., *focus*, *connected* and *visible* as the indicator for *attention level* of the students using ASQ. First, we compare two types of slides: (i) slides with discussion time more than 1 minute i.e., $M(t)$; and (ii) slides with discussion time less than 30 seconds i.e., $S(t)$. Total number of students connected to ASQ each second is denoted as $T(s)$ and total number of students who are simultaneously *connected*, *visible* and *focus* is denoted as $F(s)$. Then, we calculate the percentage of students who are simultaneously *connected*, *visible* and *focus* (denoted as $A(s)$) at any second as:

$$A(s) = \frac{F(s)}{T(s)} \times 100\% \qquad (4.5)$$

The attention level before the slide discussion started is denoted as $A(s_n)$ which depicts the attention level at nth second. The attention level after the slide discussion ended is denoted as $A(s_{n+k})$ which is the attention level after k seconds of the slide discussion till it is over. We calculate the *difference* in *attention level* before the slide started and after the slide ended as:

$$DI(s) = A(s_{n+k}) - A(s_n) \qquad (4.6)$$

We calculate $DI(s)$ for both $M(t)$ and $S(t)$ and then compare them by a box-plot to see how $DI(s)$ varies for slides that took minutes and slides that took seconds.

To have a general view of the effect of discussion time of slides on the *difference* (or change in attention level), next we move on to find the correlation between discussion time of the slides and attention level change of all the slides for each lecture.

RQ 1.3: How does the *performance* and the *reaction time* vary with different level of *difficulty* of questions (i.e., easy and hard)?

H 1.3: The reaction time increases (i.e., more time is required to submit a response to a question) and the performance declines when answering the difficult questions as compared to the easy questions.

We do not select all the responses of students to a question because some responses to the questions come from students very late after the answer is shown as a feedback for the question. So, we select the responses of the students for each question that come before this feedback time. Then, we calculate the normalized reaction time. Before that, first we need to define the terms that will be used while calculating the normalized reaction time.

For every student (viewer) v and each question q we denote the submit time of answering a question as $A(v,q)$. Start time of a question is $S(q)$. Feedback start time for the question is $F(q)$. Then, we compute reaction time for a student for each question (denoted as $R(v,q)$) and duration of each question (denoted as $D(q)$) as follows:

$$R(v,q) = A(v,q) - S(q) \qquad (4.7)$$

$$D(q) = F(q) - S(q) \qquad (4.8)$$

Then, the normalized reaction time is calculated as follows:

$$NR(v,q) = \frac{R(v,q)}{D(q)} \qquad (4.9)$$

This normalization is a necessary step as it helps in ensuring uniformity while comparing the reaction time with questions of varying duration. Mean normalized reaction time ($MNR(q)$) for each question is calculated as follows:

$$MNR(q) = \frac{1}{T(v)} \sum_{t=1}^{T(v)} NR(v,q) \qquad (4.10)$$

Here, $T(v)$ is the total number of students answering a question.

Mean normalized reaction time ($MNR(s)$) for each student is calculated as follows:

$$MNR(s) = \frac{1}{T(q)} \sum_{t=1}^{T(q)} NR(v,q) \qquad (4.11)$$

Here, $T(q)$ is the total number of questions answered by a student.

To examine the effect of difficulty of the questions on reaction time, we compare two groups using Wilcoxon's signed rank test. One group contains the mean reaction time of each student for easy questions. Another group contains the mean reaction time of each student for hard questions. We do this across all the lectures. Table 4.2 shows a sample view of the two groups for 3 different students in a lecture.

Table 4.2: **Student_ID** reports the unique 16 digit random ID assigned to each student by ASQ. **#Count_E_MNR** reports the mean normalized reaction time of each student for all the easy questions in that lecture. **#Count_H_MNR** reports the mean normalized reaction time of each student for all the hard questions in that lecture.

Student_ID	#Count_E_MNR	#Count_H_MNR
Student 1	0.41	0.67
Student 2	0.78	0.65
Student 3	0.90	0.50

Similarly, to examine the effect of difficulty of questions on the performance of students, we compare two groups using Wilcoxon's signed rank test. One group contains the percentage of questions each student has answered correctly for easy questions. Another group contains percentage of questions each student has answered correctly for hard questions. We do this across all the lectures. Table 4.3 shows a sample view of the two groups for 3 different students in a lecture.

Table 4.3: **Student_ID** reports the unique 16 digit random ID assigned to each student by ASQ. **#Count_E_percentage** reports the percentage of questions that the student answers correctly out of all easy questions in that lecture. **#Count_H_percentage** reports the percentage of questions that the student answers correctly out of all hard questions in that lecture.

Student_ID	#Count_E_percentage	#Count_H_percentage
Student 1	40.40	60.70
Student 2	48.90	63.80
Student 3	80.10	40.20

RQ 1.4: Is the *performance* in questions linked to the *distance* of questions? (Distance or (relative positioning) refers to the position of questions i.e., the proximity of the questions to the most recent slide related to the question.)

H 1.4: Students perform better when the questions are nearer to the recent content (related to the question) instead of the questions being placed far away. This hypothesis is based on the study by Weinstein et. al. [66] where they found that when questions are nearer to the related content, then students tend to perform better.

For each question, we calculate the *distance* of most recent slide related to the question as discussed in *Manual Annotation* subsection. For checking the correctness of response, we only consider the "multiple-choice" and "multiple-answer" questions as the responses to these type of questions are automatically checked by `ASQ` from the answers stored in it. We do not consider the "SQL Query" and the "fill the blanks" question as the responses to these questions cannot be automatically checked for correctness by `ASQ`. The *distance* ranged from 1 to 29 across all the lectures. Then, we plot a box-plot showing how slide distance varied for correct and wrong responses of the students across all the lectures. To check the effect of distance on the performance of students, we use a Kruskal Wallis test. By using this test, we check for significant differences between the performance of students in questions with different distances from the related content in each lecture.

Next, we find the effect of *maximum* and *minimum distance* of questions on the *performance* of students in the questions with that particular distance. We use a Wilcoxon's signed rank test to find if the performance of students (i.e., percentage of correct responses in the questions with that particular distance value) is significantly greater in the question slides with *minimum distance* as compared to the question slides with *maximum distance* across each lecture. Table 4.4 shows the two columns reporting the percentage of correct answers given by each student for the questions with the minimum and maximum distance respectively for each lecture. We compare these two columns for each lecture using Wilcoxon's signed rank test.

Table 4.4: **Student_ID** reports the unique 16 digit random ID assigned to each student by `ASQ`. **#Count_min_percentage** reports the percentage of questions that the student answers correctly for questions with that "minimum distance" for the particular lecture. **#Count_max_percentage** reports the percentage of questions that the student answers correctly for questions with that "maximum distance" for the particular lecture.

Student_ID	#Count_min_percentage	#Count_max_percentage
Student 1	40.40	60.70
Student 2	48.90	63.80
Student 3	80.10	40.20

4.2.3   RQ 2: Effect of attention percentage of students in classroom on the participation in the in-class questions

Does the measured *attention* of students using `ASQ` have any impact on their *participation* in the in-class questions? (Attention is measured as the *attention percentage* of a student which can be defined as the percentage of time a student is *connected*, *visible* and *focus* with `ASQ`.)

H 2: More attentive student (or a student with higher attention percentage in the lecture) participates more in the in-class questions.

Our assumption is that a session s (full lecture from the time a presentation starts in the classroom and ends) of length $T(s)$ starts at second 1 and ends at second $T(s)$. For every viewer (student) v, for every second t of session s we create a binary variable $attention(v,s,t)$. It is 1 if we get the visible and focus aggregated events as true and 0 otherwise.

The *attention percentage* $AP(v,s)$ of a student across s is the number of seconds the visible and focus events are true, normalized by the session length:

$$AP(v,s) = \frac{1}{T(s)} \sum_{t=1}^{T(s)} attention(v,s,t) \qquad (4.12)$$

A student v's question *participation* is defined only over the question slides in s; it is the fraction of questions $q \in Q(s)$ (with $Q(s)$ being the set of all questions in session s) that v submitted an answer for:

$$P(v,s) = \frac{1}{|Q(s)|} \sum_{q \in Q(s)} submit(v,q) \tag{4.13}$$

Here, $submit(v,q)$ is 1 if v submitted an answer to q and 0 otherwise.

Then, we find the correlation (ρ(AP,P)) between $AP(v,s)$ and $P(v,s)$.

To make a more detailed analysis, we divide the students into 3 groups (i.e., high, medium and low) based on their attention percentage $AP(v,s)$ during the lecture. We find that the common highest attention percentage in each lecture ranges between 50% and 60%. So, we divide these students into three groups as in table 4.5 in each of the lecture. This type of grouping accounts for certain number of students in each group without making any group devoid of students.

Table 4.5: **Group** reports the name of the group. $AP(v,s)$ reports the attention percentage range for that particular group

Group	$AP(v,s)$
1 (Low)	Less than or equal to 20
2 (Medium)	Greater than 20 and Less than or equal to 40
3 (High)	Greater than 40

We compare these 3 groups' participation in the in-class questions by performing a Kruskal Wallis test. We choose Kruskal Wallis over ANOVA (Analysis of Variance) because the data sample is non-parametric (i.e., it does not satisfy the properties of Gaussian distribution). Next, Dunn's test with Bonferroni correction is used for post hoc analysis as it is suitable for samples with unequal sizes. It helps to find the relation between two groups. In this case we can find the relation between 1) Group 1 and Group 2, 2)Group 2 and Group 3, and 3) Group 1 and Group 3.

4.2.4   RQ 3: Students taking external help

Do students take help of *external resources* while answering questions? Does it relate to the *difficulty* of questions?

RQ3 is based on the following *two* assumptions:

- Each student uses their own laptop and do not search from their peers laptop while answering which was also evident as reported by the field observers in each lecture.

- If a student looses focus while answering a question before submitting it, then he/she might be taking help from external resources.

H 3: Students will take more external help in the difficult questions as compared to the easy questions as they may have difficulty in answering difficult questions.

For every student (viewer) v and each question q we calculate the percentage of time a user is emitting a *focus* event while answering a question as $P(v,q)$. This $P(v,q)$ is calculated from the point a user starts answering a question till he submits the response to the question. Then, we find $F(P(v,q))$ as:

$$F(P(v,q)) = \begin{cases} 1, & \text{if } P(v,q) = 100 \\ 0, & \text{otherwise} \end{cases} \tag{4.14}$$

So, $F(P(v,q))$ is 0 if a student looses focus (in other words takes external help) while answering that question.

Next, we calculate the percentage of students who are taking external help which is denoted as $PE(q)$. Total number of students who answered a question is denoted as $T(q)$ and the total number of students who take external help for that question is denoted as $E(q)$. For these students, we have $F(P(v,q)) = 0$. $PE(q)$ is calculated as follows:

$$PE(q) = \frac{E(q)}{T(q)} \times 100\% \tag{4.15}$$

To check the relationship between the number of students taking external help (based on our assumptions) and the complexity of the questions (i.e. easy and hard questions), we use the Wilcoxon's signed rank test. First, we find the percentage of easy questions a student takes help in a lecture. Then we find the percentage of hard questions a student takes help in that lecture. Then we compare this percentage for each student in that lecture using a Wilcoxon's signed rank test as a pairwise comparison. We prefer Wilcoxon's test over the paired t-test because our sample does not satisfy the requirements of a normal distribution. We do this pairwise comparison for each of the lecture.

Table 4.6 shows the sample of the pairwise comparison of 3 students used for the analysis in one lecture. In the real data-set there are n students for each lecture where n is the total number of students using ASQ in each lecture. We use the columns #Count_E_percentage and #Count_H_percentage in table 4.6 to make the pairwise comparison using the Wilcoxon's signed rank test for each of the lectures.

Table 4.6: **Student_ID** reports the unique 16 digit random ID assigned to each student by ASQ. **#Count_E_percentage** reports the percentage of easy questions that a student takes external help. **#Count_H_percentage** shows the percentage of hard questions that a student takes external help.

Student_ID	#Count_E_percentage	#Count_H_percentage
Student 1	30.40	50.60
Student 2	45.90	60.80
Student 3	80.10	40.20

Chapter 5

Results

We used `ASQ` for a longitudinal study spanning for 2 months in an undergraduate Computer Science classroom at TU Delft to gain a fine-grained insight into the study behaviour of students in the classroom. In this chapter, we will explain in details the results of these experiments. As we discussed earlier, `ASQ` helps in capturing the event logs. These logs contain the second by second interactions of each student using `ASQ` during the lecture. We use these logs to analyze the student behaviour in classroom and try to gain a deeper knowledge of their study activity and interactions during the lecture.

Table D.1 shows the total events collected by `ASQ` in each lecture and also the total number of slides converted so that it can be displayed in `ASQ`. The number of events is directly proportional to the number of students using `ASQ`. Events in HTTP lecture is highest because more than 300 students used `ASQ`. It slowly reduced towards the end of the course with around 90 students using `ASQ`. But, in some lectures more events are generated with less students when there are questions of more duration or more non-question interactive components are present.

Table 5.1: **#TE** reports the total number of events collected during each lecture, **#TS** reports the total number of slides converted in `ASQ`, **#TST** reports the total number of students in the class and **#TSTA** shows the total number of students in the class using `ASQ`.

Lecture	#TE	#TS	#TST	#TSTA
HTTP	104477	139	319	318
HTML	91715	168	238	225
JavaScript	85846	122	192	175
Database introduction	55939	101	204	171
node.js	34647	104	208	135
CSS	50443	176	163	135
SQL introduction	51051	66	196	147
SQL continued	36656	64	157	138
Cookies & session	31103	135	133	129
Advanced SQL	56562	54	89	78
Web security	37002	151	151	107
Conceptual design	18162	55	83	54
ER logical design	20692	62	125	85
NoSQL	13899	65	89	68

Figure 5.1 and 5.2 shows the extensive logging capabilities of `ASQ` in a short interval from the beginning and the end of the JavaScript lecture. Here each event type is shown as a coloured dot. We can see that figure 5.1 shows the events generated in the first 15 minutes of the JavaScript lecture by 3 types of student. We select students based on the

number of events (i.e., low, medium and high) generated. We observe that *Student 1* who is least active with ASQ disconnects while answering the first question. *Student 2* remains idle most of the time. *Student 3* is quite active with very few instances of being idle and also actively interacts with the question slide shown with a purple line. This trend continues till the end of the lecture as we observe the same 3 students again for last 15 minutes of the same lecture in figure 5.2. In the end, we can see that *Student 1* does not generate any events because he has already disconnected from ASQ. We can see that *Student 3* generates the most number of events as compared to the other students. Lot of overlapping orange dots are shown in the activity by *Student 3* and also few are shown in the activity of other student before the start of the last question; it indicates the presence of *input* event which is generated due to the presence of slides with non-question interactive components like the examples of JavaScript events: "mouseover", "mouseout", "onchange" and "onblur".
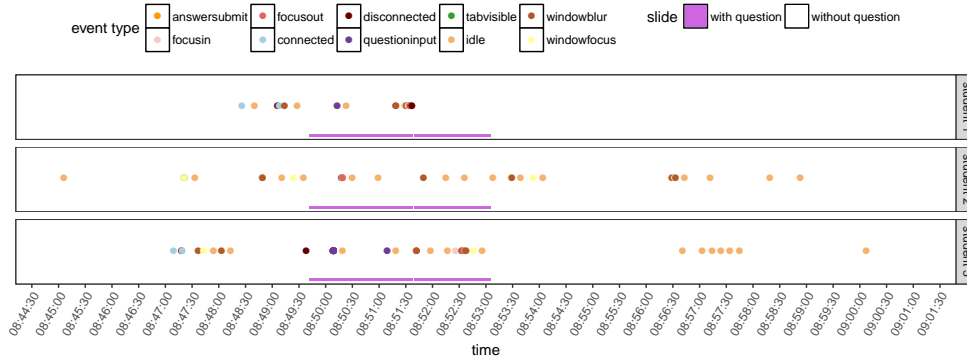
Figure 5.1: Activity of three students in the first 15 minutes of the JS lecture. Conceptually similar to [62]
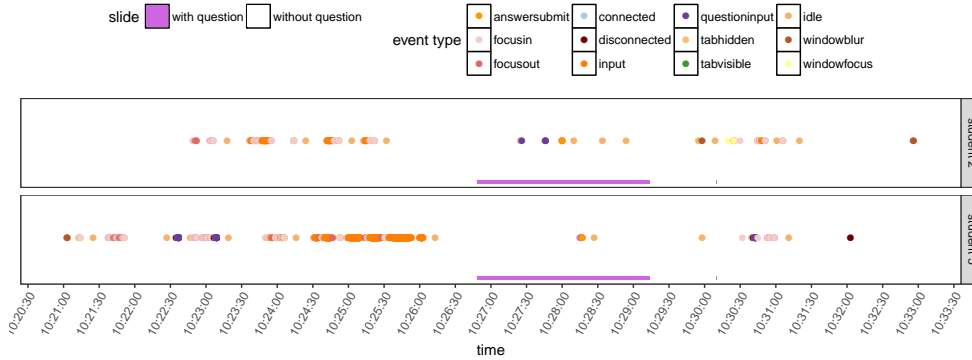
Figure 5.2: Activity of three students in the last 15 minutes of the JS lecture. Conceptually similar to [62]

Now, the fluctuation of low level aggregated events can be seen in figure 5.3 for the JavaScript lecture. During the interval where questions are posed (i.e., shown in purple lines parallel to x-axis), then there is a spike in the number of students who are in *focus*. Also during some interactive components (shown in yellow around 10:30), number of students in focus increases suddenly meaning that active learning helps a lot in the higher education classroom. Another observation that can be made is that right after answering a question, lot of students loose focus. It indicates that students take rest or become busy in other activities after answering a question. So, we can observe from the example figure 5.3 that active learning strategy like questions and other interactive components help in

improving the *focus* level of students. This similar observation was also made in all the lectures when we use active learning with `ASQ`.
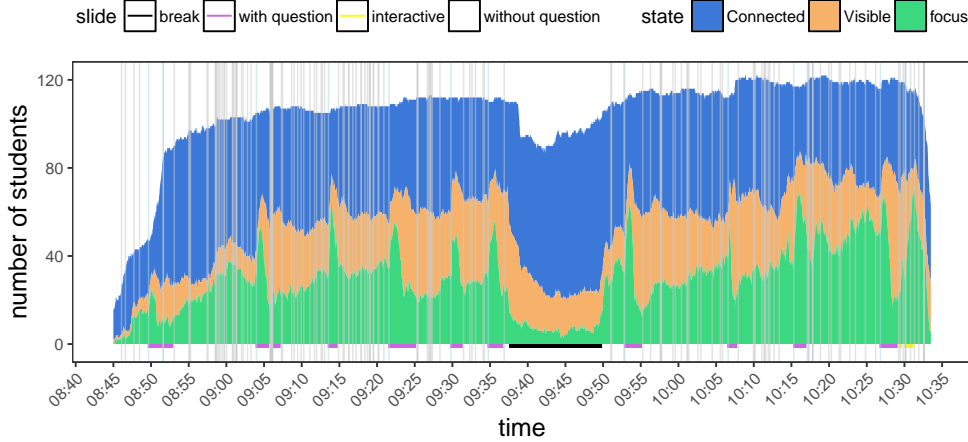
Figure 5.3: Temporal overview of aggregated events of students during JavaScript lecture. The parallel vertical gray lines show the slide transitions.

Now, we discuss regarding our problem statement in detail based on the 3 research questions defined earlier. Revisiting the main aspects of the question, first question is to determine the effect of *lecture elements* (like the difficulty of questions, relative positioning of questions, strategy of questioning) on *study behaviour* (like attention level, performance and reaction time) of students; second, we want to find if there is any relationship of *attention level* of students with their *participation* in the in-class questions; finally, we want to find if students are taking *external help* and if there is any relationship between the tendency to take external help and the difficulty of the questions.

5.1 RQ 1: Effect of lecture elements on study behaviour of students

Here, we discuss the results regarding the effect of *lecture elements* on *study behaviour* of students.

### 5.1.1 RQ 1.1: Effect of question spacing strategies and instructor on percentage of correct responses, performance and reaction time of students

We have used 3 different question spacing strategies (i.e., burst, uniform and increasing). But, as mentioned earlier, the increasing question strategy was dropped by the instructors after 2 lectures due to problem in execution. So, we compared the effect of only burst and uniform strategy. Here, we discuss the effect of lecture elements on the study behaviour of students.

We determine the percentage of correct responses for each question strategy associated with each of the instructors. In total each instructor has used burst and uniform strategy 3 times each. From table 5.2 we can see that percentage of correct responses in burst question strategy outperforms the uniform strategy for both the instructors.

In total we get 21 common users in the last 3 lectures for the *within-subjects* study where we have enabled the option so that cookies of these users are saved across these lectures. From the *within-subject* study, we find in figure 5.4 that the percentage of correct answers given by these students is more in both the lectures using burst strategy as compared to the lecture using uniform strategy. We can find after performing a Wilcoxon's

signed rank test that there is significant difference between Conceptual design (using burst question strategy) and ER logical design (using uniform question strategy) lecture but no significant difference between ER logical design (using uniform question strategy) and NoSQL (using burst question strategy) lecture.

Table 5.2: **Strategy and Instructor type** shows the type of strategy used and the type of instructor teaching in that lecture. **#%Correct** shows the percentage of correct responses across all the lectures with that particular strategy and instructor.

Strategy and Instructor type	#%Correct
Burst and I1	60.85
Uniform and I1	50.25
Burst and I2	56.75
Uniform and I2	55.58
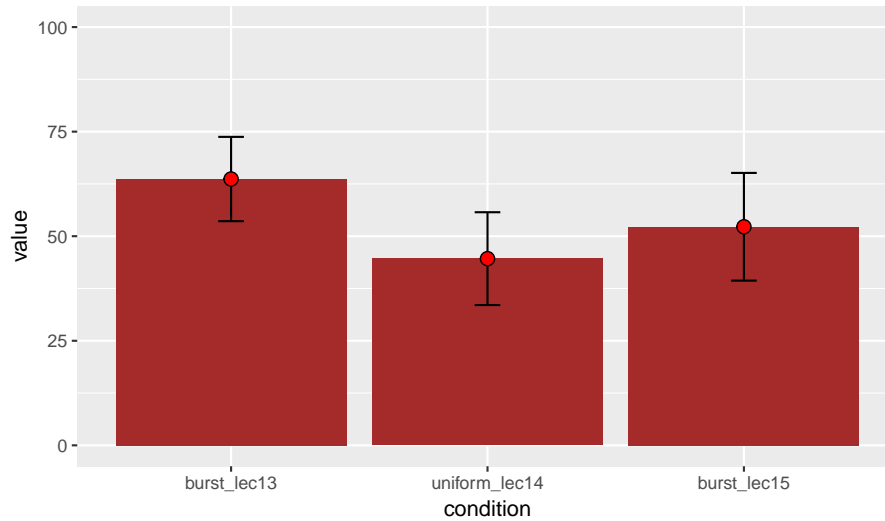
Figure 5.4: Percentage of correct responses comparison for burst and uniform strategy for the same set of users in the last 3 lectures i.e., Conceptual Design, ER Logical Design and NoSQL. The error bars for the 95% Confidence interval are also shown in each bars. The first lecture (from left) uses burst strategy, second uses uniform and the third one again uses burst.

To find the effect of *question spacing strategy* on the *performance* of students for both the instructors, we compare the reaction time of students among these four groups in table 5.4 as already defined in the *methodology for analysis*.

We use a Kruskal Wallis test to compare these groups and find if there is a significant difference (i.e., $p - value < 0.01$) between these groups. To check exactly how each group is different from the other one, we use a post hoc analysis with the help of Dunn's test with Bonferroni correction. Table 5.3 shows that we have a significant difference between the groups when the instructor of the groups are different i.e., when comparing Group B with Group A, Group D with Group A and Group D with Group C. But, if the instructor remains same and the question spacing strategy changes i.e., when comparing Group C with Group A and Group D with Group B, then we do not see significant differences between the groups.

Table 5.3: The values show the difference of means of one group to the other obtained from Dunn's test. For instance, in the first row of the table, we obtain difference in means of Group A from Group B (i.e., Group B - Group A) and likewise for the other rows report the difference in means after Dunn's test. Each statistically significant value i.e., $p-value < 0.01$ is marked with a †.

	A	B	C
B	-2.60 †		
C	-0.60	1.96	
D	-3.53 †	-1.24	-2.93 †

To find the effect of *question spacing strategy* on the *reaction time* of students while answering questions for both the instructors, we compare the reaction time of students among these four groups in table 5.4 as already defined in the methodology.

Table 5.4: **Group** shows the name of each group for the combination of question spacing strategy and instructor. I1 and I2 represents the 2 instructors that taught all the lectures.

Group	Strategy + Instructor
A	Burst + I1
B	Burst + I2
C	Uniform + I1
D	Uniform + I2

We use a Kruskal Wallis test to compare these groups and try to find if there is a significant difference (i.e., $p-value < 0.01$) between these groups. To check exactly how each group is different from the other one, we use a post hoc analysis with the help of Dunn's test with Bonferroni correction. Table 5.5 shows that we have a significant difference between the groups when the question strategy of the groups are different i.e., when comparing Group C with Group A, Group C with Group B, Group D with Group A and Group D with Group B. But, if the question strategy remains same and the instructor changes i.e., when comparing Group B with Group A and Group D with Group C, then we do not see significant differences between the groups. We can also observe from the table that the *mean normalized reaction time* for uniform question strategy is less than the mean normalized reaction time for burst question strategy and these values are statistically significant.

Table 5.5: The values show the difference of means of one group from the other obtained from Dunn's test. For instance, in the first row of the table, we obtain difference in means of Group A from Group B (i.e., Group B - Group A) and likewise, the other rows report the difference in means after Dunn's test. Each statistically significant value i.e., $p-value < 0.01$ is marked with a †.

	A	B	C
B	-2.37		
C	-7.58 †	-6.67 †	
D	-5.70 †	-4.40 †	2.84

5.1.2   RQ 1.2: Effect of discussion time of slide on change of attention level of students

Now, we examine the effect of discussion time (i.e., *duration*) of slide on the change in attention level (i.e., defined earlier as the difference in the percentage of students who are *connected*, *visible* and *focus* in `ASQ` before the start of the slide and after the slide ended) of students. We observe in figure 5.5 that the variability of *difference* in attention level across all the lectures is more for slides whose duration of discussion was in minutes as compared to slides whose duration of discussion was in seconds; indicating that attention level (or attention span) has wider variations in slides having discussion duration in minutes.
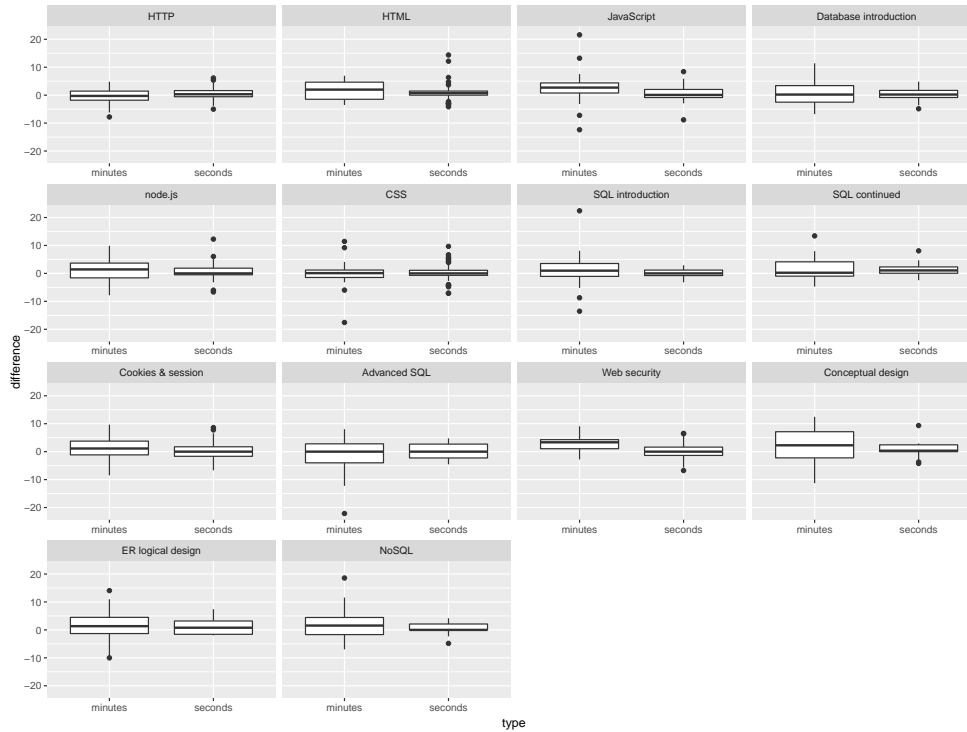
Figure 5.5: Change of focus percentage (i.e., *difference* in attention level) in slides whose duration is in minutes and seconds

To have a general view of the effect of discussion time (i.e., *duration*) of slides on the *difference* in attention level, we move on to find the correlation between duration of all the slides with the difference in attention level associated with it. We observe in table 5.6 that JavaScript, SQL introduction and Web security only showed a statistically significant correlation. All but one show weak positive correlation (although some are not statistically significant) meaning that with the increase of discussion time of slide, the *difference* in attention level also increases.

5.1.3   RQ 1.3: Effect of difficulty of questions on performance and reaction time of students

Figure 5.6 shows the performance variation with different levels of difficulty of questions where we do not find any pattern in the mean score for each type of question in any of the lectures. We do not find any effect of question difficulty on the performance of students using `ASQ`.

Table 5.6: $\rho(\boldsymbol{D},\boldsymbol{DI})$ shows the Spearman Rank Correlation of slide *duration* and the *difference* of attention level. Each statistically significant value i.e., $p-value < 0.01$ is marked with a †.

Lecture	$\rho(\boldsymbol{D},\boldsymbol{DI})$
HTTP	-0.06
HTML	0.06
JavaScript	0.21†
Database introduction	0.04
node.js	0.12
CSS	0.13
SQL introduction	0.31†
SQL continued	0.07
Cookies & session	0.11
Advanced SQL	0.02
Web security	0.24†
Conceptual design	0.19
ER logical design	0.12
NoSQL	0.22
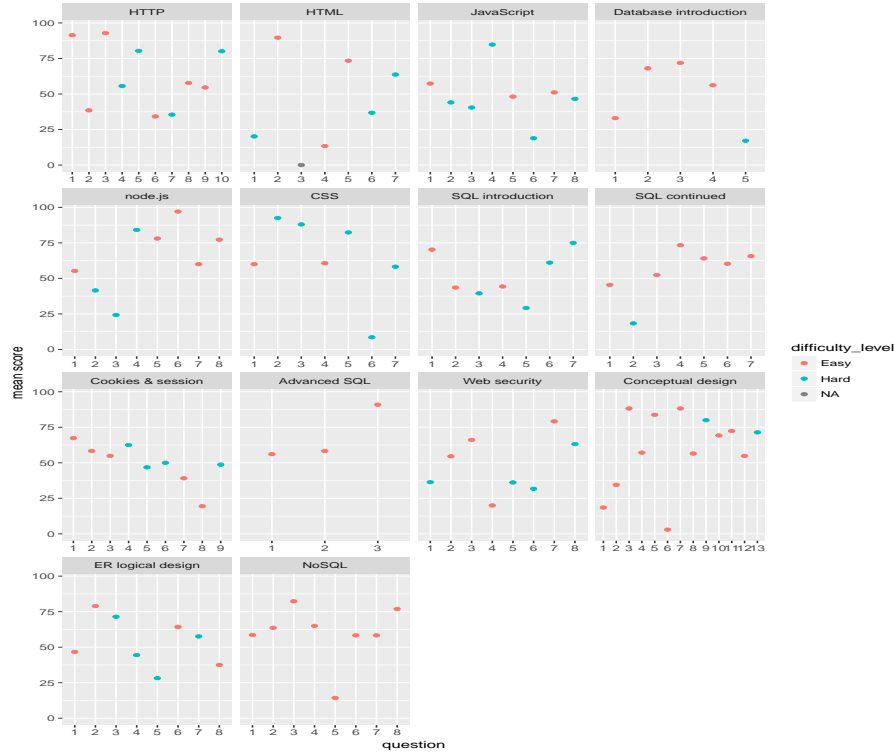
Figure 5.6: Distribution of mean score for each question with difficulty across all the lectures.

To examine the effect of *difficulty* of questions on *performance*, we compare two groups using Wilcoxon's signed rank test: one group contains the percentage of questions each student has answered correctly for *easy* questions; another group contains percentage of questions each student has answered correctly for *hard* questions. For most of the lectures, we do not observe any significant differences between the two groups as seen in table 5.7. It

means that we cannot find any relation between difficulty of questions and the performance of students.

Table 5.7: **Significance** reports the significance value after the Wilcoxon's signed rank test across each lecture. Each statistically significant value i.e. $p - value < 0.01$ is marked with a † otherwise it is marked with a ⋆. Advanced SQL and NoSQL are marked with NA because they have only easy mcq questions and no hard mcq questions. **#Easy mcq** and **#Hard mcq** shows the total number of easy and hard multiple choice (or multiple answer) questions used in each lecture.

Lecture	#Easy mcq	#Hard mcq	Significance
HTTP	6	4	⋆
HTML	3	3	†
JavaScript	3	5	⋆
Database introduction	4	1	†
node.js	5	3	†
CSS	2	5	⋆
SQL introduction	3	4	⋆
SQL continued	5	2	†
Cookies & session	5	4	⋆
Advanced SQL	3	0	NA
Web security	4	4	†
Conceptual design	11	2	⋆
ER logical design	4	4	⋆
NoSQL	8	0	NA

Next, we try to examine if there is any relationship between the normalized reaction time and the difficulty of questions. Figure 5.7 shows the variation of mean normalized reaction time with the difficulty of questions. It can be seen that the mean normalized reaction time for difficult questions is more than the mean normalized reaction for the easy questions in most of the lectures but, in many lectures the result is mixed.

To examine the effect of *difficulty* of questions on *reaction time*, we compare two groups using Wilcoxon's signed rank test: one group contains the mean normalized reaction time of each student for *easy* questions; another group contains the mean normalized reaction time of each student for *hard* questions. For most of the lectures, we do not observe any significant differences between the two groups as seen in table 5.7. It means that we cannot find any relation between difficulty of questions and the reaction time of students while answering questions.

So, we observe the distribution of normalized reaction time of every student for easy and hard questions in detail for each of the lectures with the help of a box-plot.

Figure 5.8 shows that all the lectures (except 3 i.e., JavaScript, node.js and SQL Introduction) have higher median normalized reaction time in hard questions as compared to the easy questions. It can be seen from figure 5.8 that all but one lecture have more than 75 % of the students requiring around $NR(v, q)$ (normalized reaction time) greater than 0.4 for answering the easy questions. It means that most students need 40 percent of the entire time duration to answer a question. In Advanced SQL, the normalized reaction time for hard questions is close to 1 due to the complex nature of writing long SQL queries.

5.1.4   RQ 1.4: Effect of distance (or relative positioning) of questions on the performance of students

We have explained in the *experimental design* about the procedure to calculate the *distance* of question slide; it is calculated as the number of slides a question slide is away from
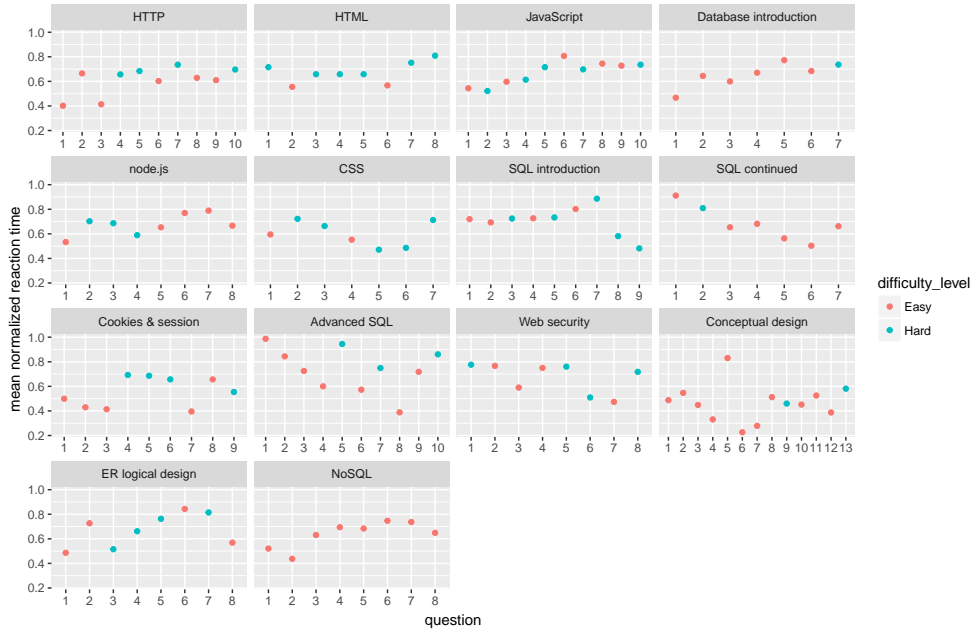
Figure 5.7: Mean normalized reaction with difficulty
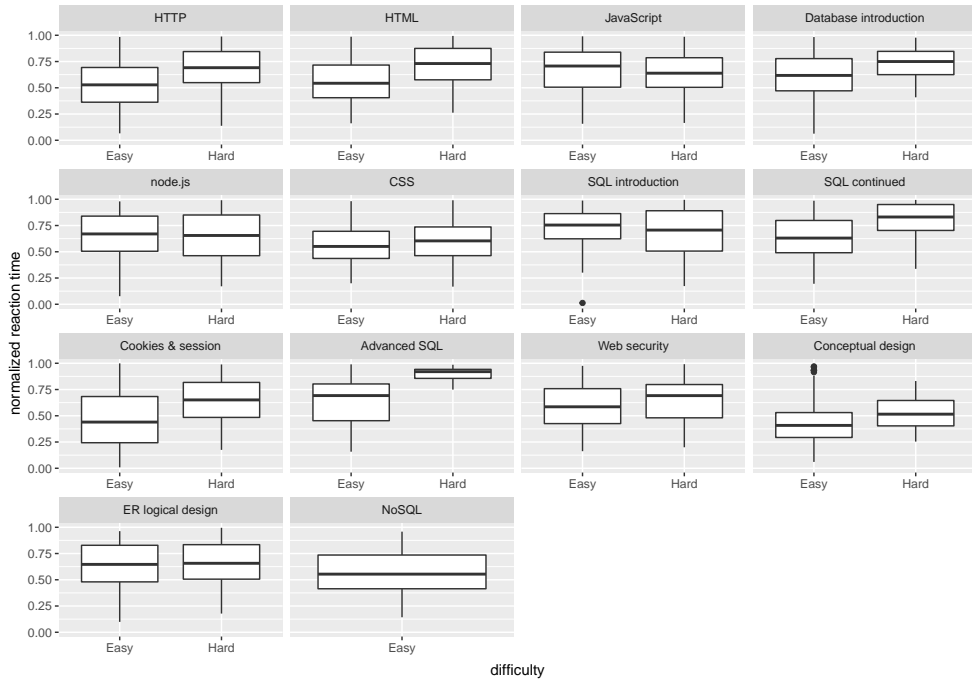
Figure 5.8: Distribution of normalized reaction with difficulty

the most recent content slide related to it. Now, the relation of slide distance with the correctness of the responses is shown in figure 5.9. As can be seen, most lectures had the questions immediately after the content (i.e., with slide distance equal to 1); this is the main reason causing it difficult to see any pattern due to the limited distribution of distance in the box-plot.

So, we perform a Kruskal Wallis test to check the effect of relative positioning of ques-

Table 5.8: **Significance** reports the significance value after the Wilcoxon's signed rank test across each lecture. Advanced SQL is marked as NA because of zero responses in most question. NoSQL is marked with NA because the lecture has only easy questions and no hard questions.

Lecture	Significance
HTTP	†
HTML †	
JavaScript	⋆
Database introduction	†
node.js	⋆
CSS	⋆
SQL introduction	⋆
SQL continued	†
Cookies & session	†
Advanced SQL	NA
Web security	†
Conceptual design	†
ER logical design	⋆
NoSQL	NA
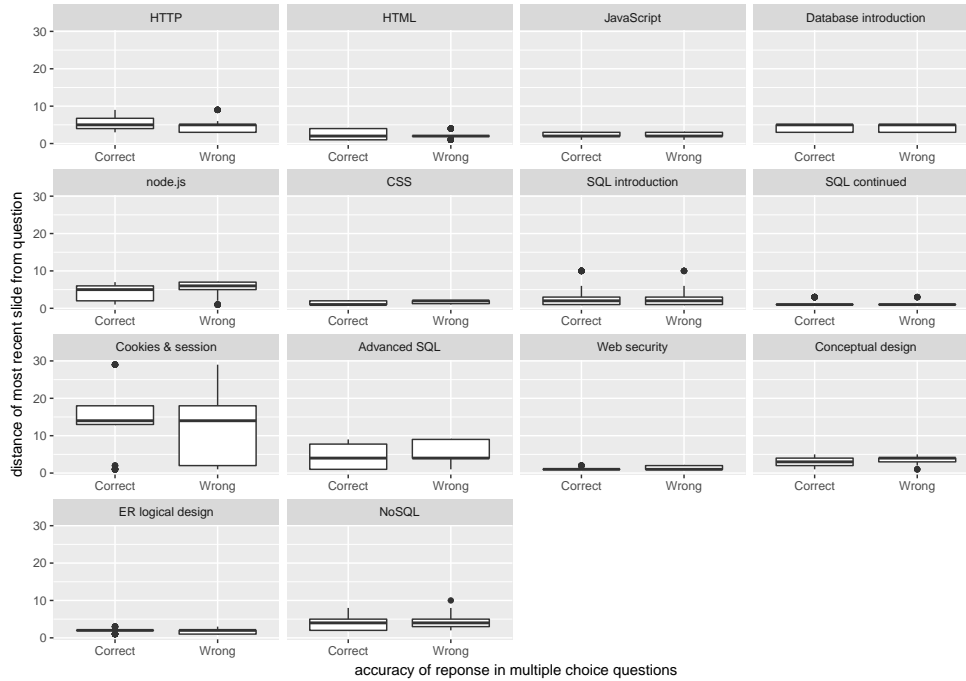
Figure 5.9: Distribution of the distance of the most recent content from the questions shown for each category of response i.e. correct and wrong

tions on the performance of students. Except 3 lectures, we find statistically significant differences in the performance of students with different distance of questions as observed in table 5.9. We do not consider Advanced SQL for the analysis because most of the questions have zero responses. So, most questions have zero score for each student.

Table 5.9: **Significance** reports the significance value after the Kruskal Wallis test to check the effect of distance of questions on the performance of students across each lecture. Each statistically significant value i.e. $p-value < 0.01$ is marked with a †, otherwise it is marked with a ⋆. Advanced SQL is marked as NA because of zero responses in most questions.

Lecture	Significance
HTTP	†
HTML	†
JavaScript	⋆
Database introduction	⋆
node.js	†
CSS	†
SQL introduction	†
SQL continued	†
Cookies & session	†
Advanced SQL	NA
Web security	†
Conceptual design	†
ER logical design	†
NoSQL	⋆

Next, we find the effect of *maximum* and *minimum distance* of questions on the *performance* in the questions with that particular distance. We use a Wilcoxon's signed rank test to find if the mean performance of students (i.e., percentage of correct responses in the questions with that particular distance value) is significantly greater in the question slides with *minimum distance* as compared to the slides with *maximum distance* of questions across each lecture. We can see in table 5.10 that the lectures where the *difference* in maximum and minimum slide distance is at least *3 slides*, then we can see noticeable difference in performance as evident by the statistically significant value obtained in Wilcoxon's signed rank test. So, for large differences in slide distance, we can see significant differences in performance of students; the performance is also better for question slides which have less distance (or comparatively closer to the content slide related to it).

5.2 RQ 2: Effect of measured attention of students in classroom on the participation in the in-class questions

The attention of students in classroom is measured as the attention percentage of students using ASQ. The attention percentage of students has a significant positive effect on their participation in the in-class question as observed in table 5.11. Towards the end of the lectures i.e., in the last 3, we can observe that the correlation strength is moderately low because most students did not participate in the questions. It may be due to the complexity of database lectures. But, most of the lectures show a moderately strong correlation. Students who are more attentive in ASQ also answer most of the questions in the class.

As mentioned earlier in the *methodology for analysis*, we calculate the attention percentage of every student during the lecture. Then we analyze the relation between attention percentage of students and their participation in the in-class questions in depth. We do this by dividing the students with different attention percentages in the lecture into different groups. Group 1 consists of students with $AP(v,s) <= 20\%$. Group 2 consists of students with $AP(v,s) > 20\%$ & $<= 40\%$. Group 3 consists of students with $AP(v,s) > 40\%$. We compare these 3 groups' participation in the in-class questions by performing a Kruskal Wallis test. Except the last few lectures, all the other lectures show significant differences

Table 5.10: **Significance** reports the significance value after using the Wilcoxon's signed rank test to compare the effect of *maximum* and *minimum distance* of questions on performance of students (i.e., percentage of correct responses in the questions with that particular distance value) across each lecture. Each statistically significant value i.e. $p-value < 0.01$ is marked with a †, otherwise it is marked with a ⋆. Advanced SQL is marked as NA because of zero responses in most questions. Columns **#max** and **#min** report the maximum and minimum distances of the questions respectively.

Lecture	#max	#min	Significance
HTTP	9	4	†
HTML	2	1	⋆
JavaScript	3	1	†
Database introduction	5	3	⋆
node.js	7	1	†
CSS	2	1	†
SQL introduction	10	1	†
SQL continued	3	1	⋆
Cookies & session	29	1	†
Advanced SQL	NA	NA	NA
Web security	2	1	⋆
Conceptual design	4	1	†
ER logical design	3	1	⋆
NoSQL	10	2	†

Table 5.11: $\rho(AP,P)$ shows the correlation of attention level (AP) in `ASQ` and participation percentage (P) in the in-class questions. Each statistically significant value i.e., $p-value < 0.01$ is marked with a †.

Lecture	$\rho(AP,P)$
HTTP	0.64†
HTML	0.51†
JavaScript	0.59†
Database introduction	0.48†
node.js	0.44†
CSS	0.47†
SQL introduction	0.51†
SQL continued	0.41†
Cookies & session	0.64†
Advanced SQL	0.66†
Web security	0.47†
Conceptual design	0.31†
ER logical design	0.29†
NoSQL	0.35†

between them as observed in table 5.12. It means that these groups' participation differ significantly from each other.

Next, Dunn's test with Bonferroni correction is used for the post hoc analysis. It helps to find the relation between between 1) Group 1 (Low) and Group 2 (Medium), 2)Group 2 (Medium) and Group 3 (High), and 3) Group 1 (Low) and Group 3 (High). We can observe in table 5.13 that all the lectures except last 3 have significant differences in participation between the groups with low and medium attention percentage. The difference between

Table 5.12: This shows the p-value obtained in the Kruskal Wallis test across the 3 groups (i.e., Group 1, 2 and 3). Each statistically significant value i.e., $p-value < 0.01$ is marked with a † in the column **#p-value** and the rest are marked with a ⋆. Column **#A** reports the total number of students who belong to Group 1. Similarly, **#B** and **#C** reports the total number of students in Group 2 and Group 3 respectively.

Lecture	#p-value	#A	#B	#C
HTTP	†	165	55	68
HTML	†	125	53	20
JavaScript	†	114	47	7
Database introduction	†	96	35	26
node.js	†	87	30	7
CSS	†	89	25	8
SQL introduction	†	102	24	14
SQL continued	⋆	101	21	5
Cookies & session	†	87	16	13
Advanced SQL	†	59	11	1
Web security	†	67	22	12
Conceptual design	⋆	34	16	3
ER logical design	⋆	56	21	4
NoSQL	⋆	36	16	9

the means of Group 2 and Group 1 is always negative which means that the mean participation of students having medium attention percentage is more than the students with low attention percentage. When we compare the groups with medium and high attention percentages, then we do not get any significant difference between them. So, participation of students paying *lower* attention differ significantly from the students paying *higher* attention. But, the similar difference is not observed when we compare two groups of students with *medium* and *higher* attention percentage in the classroom.

Table 5.13: **#BA**, **#CA** and **#CB** reports the difference between the means of Group 2 & Group 1, Group 3 & Group 1 and Group 3 & Group 2 respectively after the posthoc analysis using the Dunn's test. Each statistically significant value i.e. $p-value < 0.01$ is marked with a †.

Lecture	#BA	#CA	#CB
HTTP	-5.90†	-9.02†	-2.10
HTML	-4.67†	-2.91†	0.25
JavaScript	-5.08†	-2.19	0.07
Database introduction	-3.30†	-5.01†	-1.76
node.js	-3.71†	-2.85†	-0.80
CSS	-2.73†	-1.63	0.04
SQL introduction	-3.59†	-4.10†	-1.06
SQL continued	-2.66†	-0.38	0.93
Cookies & session	-2.99†	-4.51†	-1.41
Advanced SQL	-3.96†	-1.03	0.26
Web security	-4.39†	-2.09	1.18
Conceptual design	-1.41	0.60	1.25
ER logical design	-2.21	-0.97	0.12
NoSQL	0.04	-2.36	-2.14

5.3   RQ 3: Students taking external help

We assume that students loosing focus before answering a question indicates that they take external help. Based on this assumption, we can see in the figure 5.10 that students may be taking help externally in every question across all the lectures. The percentage of students taking help from external resources is always greater than 50 % of the students that respond to a question. It is true for each of the questions in all of the lectures. The highest percentage of students taking external help can be seen in Advanced SQL due to the complex nature of writing the long SQL queries.
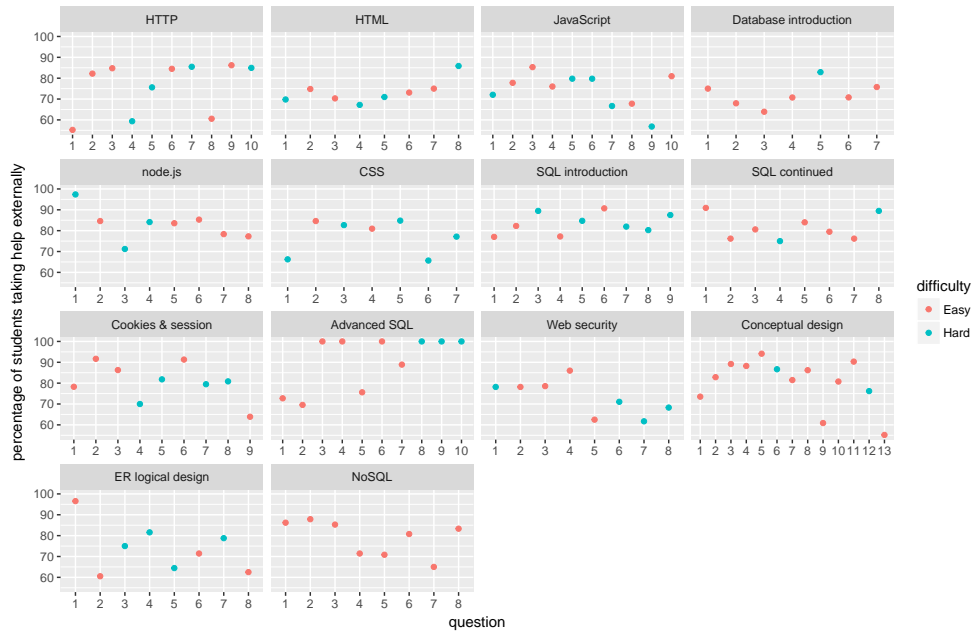
Figure 5.10: Percentage of students taking help externally in each lecture

To check the relationship between the number of students taking external help with the complexity of the questions (i.e. easy and hard questions), we use the Wilcoxon's signed rank test. We compare the percentage of questions each student takes help in both categories of questions (i.e. easy and hard). For this we create 2 groups; one group has the percentage of *easy* questions a student takes help; and another one has the percentage of students that take external help in *hard* questions. We do not find significant differences between the two groups in most of the lectures as seen in table 5.14. So, we cannot find any relation between the difficulty of questions and the likelihood of students taking external help for most of the lectures.

Table 5.14: **Significance** reports the significance value after using the Wilcoxon's signed rank test across each lecture. Each statistically significant value i.e. $p-value < 0.01$ is marked with a † otherwise a ⋆ indicates statistically non-significant values. Advanced SQL is marked as NA because of zero responses in most questions. NoSQL is marked with NA because the lecture has only easy questions and not a single hard question. **#Easy** and **#Hard** shows the total number of easy and hard questions used in each lecture.

Lecture	#Easy	#Hard	Significance
HTTP	6	4	⋆
HTML	4	4	⋆
JavaScript	5	5	†
Database introduction	6	1	⋆
node.js	5	3	†
CSS	2	5	⋆
SQL introduction	4	5	⋆
SQL continued	6	2	†
Cookies & session	5	4	†
Advanced SQL	7	3	NA
Web security	4	4	†
Conceptual design	11	2	⋆
ER logical design	4	4	†
NoSQL	8	0	NA

Chapter 6

Discussion

We used `ASQ` in an active learning setup in a large classroom to obtain fine-grained insights into the study behaviour of students. In this chapter, we discuss the results obtained from these experiments. We can observe from the example of the JavaScript lecture that the use of active learning components like questions and other interactive components (like JavaScript components) improve the attention level of the students; it is evident by the sharp rise of *focus* activity of the students using `ASQ`. Past works have also supported the fact that use of active learning components improves the attention level of students [58, 37, 10].

In the first research question (RQ1), we study the effect of *lecture elements* (like length of discussion time associated with slides, question strategy, difficulty of questions, positioning of questions) on the *study characteristics* of students (like reaction time, attention level, performance). First, we discuss the effect of *question spacing strategies* (i.e., burst and uniform) and *instructor* on the *performance* and *reaction time* of students (RQ 1.1). We have used burst and uniform strategy for 6 lectures each for both the instructors. Students answer an upper percentage of questions correctly while using burst strategy which may be due to the proximity of the content slides related to the question in the burst strategy. Weinstein et al. [66] also found similar effects of better performance due to proximity of related content to the question. But, we do not find significant effect of question spacing strategy on the performance of students; rather, we find significant effect of instructors on the performance of students. It is in accordance with past work where instructors have an impact on the attention level [9] and learning outcome [40, 51] of students in the classroom. The reaction time of students while using burst strategy is significantly more as compared to the uniform strategy which may be because students get tired when answering back to back questions in burst format.

The claim regarding *short attention span* of 30 seconds by Bunce et al. [10] was confirmed to some extent by our experiments. We find indirect evidence that attention span of students (i.e., change of *focus* activity in `ASQ`) is short and less than a minute (RQ 1.2). It is established by the fact that we can see *wider variation* of attention of students in slides where the discussion time is greater than one minute as compared to the slides where the discussion time is less than 30 seconds. Moreover, as there is an increase in the length of discussion time associated with the slides, then the fluctuation of attention level (i.e., difference in the *focus* activity of the students using `ASQ` before and after the slide display) also increases.

Contrary to our expectations, the *performance* and *reaction time* of students do not change as per the *difficulty* of the questions used (RQ 1.3) in most of the lectures. It may be due to some other factor like previous knowledge on the topic or motivation level [44, 29] which is also affecting the performance of the students while answering the questions. Another possible reason for the discrepancy in the results may be due to the way we classify each question as either *easy* or *hard*; this is in a way tricky because we have to consider

different aspects (like variation in the options, difficulty in understanding the question, difficulty and clarity in the options, etc.) of the question before classifying it.

We find that the *distance* [1] of a question slide has a significant effect on the performance of students (RQ 1.4). To add to it, students perform significantly better in question slides having minimum distance in a lecture. This is similar to the results obtained by Weinstein et al. [66]; they found that students perform better when the questions are nearer to the content related to it. It is because students find it easier to recollect a recently taught concept.

The second research question (RQ2) was concerned with finding if the *participation* of students in the class questions is dependent on the attention level of students who are using ASQ. Students with higher attention percentage in ASQ participate more in the class questions as compared to students who are less attentive. It is consistent with the findings in past literature; more use of laptop for course-related activities leads to a higher level of attention and engagement among students [72, 52]. Similarly, in our experimental scenario, as students get more involved with ASQ during the lecture (i.e., using laptop for course-related activities), they have higher attention percentage during the lecture; these students also participate more with the in-class questions.

The third research question (RQ3) is related to finding if students are taking external help while answering the questions and it's relation to the difficulty of questions. We find indirect evidence of students taking help from external resources. But, it is based on the assumption that students loosing *focus* in ASQ are taking external help which is difficult to neither accept nor deny. We also do not find any relation between the students taking external help and the difficulty of questions for 50% of the lectures.

So, we can effectively get fine-grained insights into the study behaviour of students in a classroom second by second by using a web-mediated tool like ASQ as indicated by RQ1 and RQ2. However, findings in RQ3 are limited: to evaluate RQ3 in detail we need help of other modalities like eye-tracking and different sensors to complement the ability of ASQ to obtain fine-grained insights; this will be explained in future work in the next chapter.

6.1   Threats to internal validity

- We do not control the *extraneous* variables like the previous knowledge of each student before attending the course as we do not do any pre-assessment before the start of the course. Previous knowledge regarding the subject can affect the performance of students in the in-class questions. For instance, a student with previous knowledge on the topic of the lecture may also score well while answering the difficult questions.

- We do not know the motivation level and interest of students in the course which may act as *confounding* variables. This can also affect the attention level [44, 29] and performance of students while answering questions in the classroom.

- We do not check the ability of each student to act under the pressure of time, how fast they react or how they are affected by fatigue which may be a *maturation* threat. For instance, some students may handle pressure better as compared to others; this may help them to perform better in the difficult questions during the lecture. For some students, their performance may decline with time due to fatigue towards the end of the lecture.

- Towards the end of the lecture, we get fewer data samples due to reduction in the student population which is a *mortality* threat. This happens in most courses at TU

---

[1]It is calculated as the number of slides a question slide is away from the most recent related content slide.

Delft but it has the possibility of affecting the data collection. If the duration of data collection is less then it will be difficult to compare characteristics of students across the lectures with fewer student samples.

- Mismatch between the functionality of the user interface and the user needs can cause aesthetic displeasure and effect the *learning* of students and also the *teaching* of the instructor. For instance, a static generic user interface can effect the learning and performance of the user [3]. In our case, the static user interface of ASQ can have similar effects on the students and also hamper the teaching of the instructor [50].

Chapter 7

Conclusions and Future Work

7.1 Conclusions

We used `ASQ` for a longitudinal study of around 300 students in an undergraduate classroom at TU Delft for around 10 weeks. `ASQ`'s extensive logging capabilities allowed us to collect student interactions in the classroom as event logs second by second. One advantage of using `ASQ` is that the privacy of the users is not compromised. Now, we will explain how we gain *fine-grained* insights about the study behaviour of students based on the different research questions.

First (RQ1), we investigate the effect of *lecture elements* (like the difficulty of questions, relative positioning of questions, strategy of questioning, type of instructor) on *study behaviour* (like attention level, performance and reaction time) of students. We find significant effect of: instructor on the performance and question spacing strategies on the reaction time of students. The attention span of students is less than a minute as was also observed in the past [10]. We do not find significant effect of question difficulty on the performance and reaction time of students in most lectures. But, we find a significant effect of relative position of questions on the performance of students. So, we gain a overview of how the study behaviour varies with the effect of different lecture elements.

Next (RQ2), we examine the relation of *attention percentage* of students on their *participation* in the in-class questions. We find that the more students are attentive with `ASQ`, the more they participate in the in-class questions. So, we gain deep understanding regarding the attention percentage, participation of students in the classroom questions and the relation between them.

Finally (RQ3), we determine if students take *external help* while answering questions and how is it related to the difficulty of questions. It is based on the assumption that students who loose *focus* while answering a question are taking external help. We find that more than 50 % of the time students take external help in all the questions across every lecture. But, we do not find significant effect of difficulty of questions on the possibility of students taking external help.

So, `ASQ` can be used effectively as an active learning setup to determine the *fine-grained study behaviour* of students in a large classroom as reported by the results in RQ1 and RQ2. We find from RQ1 that different instructor have significant impact on the performance of students while answering questions during the lecture. The reaction time of students vary significantly when we vary the spacing of questions. We find insights about the short attention span of students. We find a significant relationship between the distance (or relative positioning) of questions and the performance of students. RQ2 sheds light on the participation of students in the in-class questions; this participation increases with increasing attention percentage of students in the classroom. However, RQ3 and certain aspect of RQ1 (like the effect of difficulty of questions on the performance and reaction time of students) merits further investigation. We have touched some of the aspects in our future work discussion.

7.2 Future Work

As part of the future work, we can enable the option in `ASQ` to save the cookies of students right from the beginning of the experiments in the classroom. It can help us to characterize the students based on different study characteristics of students that we study in this thesis. It can help us to gain insights into how the same student behaves with change of instructor, difficulty of questions, change of topic, change of question strategy and help us for student profiling. This can give valuable insights to the instructor to design the courses accordingly. We can use different level of difficulty of questions with the help of Bloom's Taxonomy [2] to measure the understanding of the students. Similar to past experiments, we can also determine the effectiveness of using `ASQ` by creating two groups of students: control group (i.e., the group that does not use `ASQ`) and the experimental group (i.e., the group that uses `ASQ`). Then we can compare the performance of these two groups by using in-class questions. We can do this in another way where both the groups use `ASQ` but, we vary the question strategy for both the groups. Then we can find the effectiveness of the question strategy by comparing the performance of the students in both the groups.

One problem that needs to be solved is to monitor the *off-task* activity of each student in detail by using a proxy server just like [26, 48]; it can track the browsing of students when they do not use `ASQ` in classroom. It is important because some students using `ASQ` are not active in it so they are termed as *idle* by `ASQ` but, they follow the lecture from the beamer. We can know what these students are doing based on their browsing patterns. Then, we can be certain whether they are really idle in the classroom or they are following other study related material when they are termed as *idle* by `ASQ`.

Besides, we can design and test *real time* feedback for both students and instructors using `ASQ`. As a first step, we can form a regression model based on different predictors like the activity indicators (i.e., focus, visible, connected, idle, etc.), performance, time spent with `ASQ`, varying instructors, varying lecture strategy, etc., obtained from the data of past lectures to predict the attention level of students. Then, we can integrate this with `ASQ` so that the predictive model will give real time time feedback to the students and the instructors. It will help the students to understand their activity and also help the instructors to gain deep understanding of the classroom activity.

We can also use other *multimodal* data capturing mechanisms along with `ASQ` like eye tracking glasses, noise level monitor to aid the research. Although it is not scalable, it can further provide detailed insights into the activity of students from their posture, gesture and noise levels. It can also help to answer our RQ3 by giving us a certainty that students are taking external help. We can implement `ASQ` in other courses to gain a rich understanding how the use of `ASQ` varies across multifarious classrooms. For instance, use of `ASQ` may be different for a Computer Science course as compared to a course in Mechanical or Civil Engineering. It can help to generalize the impact of `ASQ` across multiple classrooms. So, these recommendations can provide impetus for ongoing research in classroom using `ASQ`.

Bibliography

[1] Nancy M Aguilar-Roca, Adrienne E Williams, and Diane K O'Dowd. The impact of laptop-free zones on student performance and attitudes in large lectures. *Computers & Education*, 59(4):1300–1308, 2012.

[2] Lorin W Anderson, David R Krathwohl, P Airasian, K Cruikshank, R Mayer, P Pintrich, J Raths, and M Wittrock. A taxonomy for learning, teaching and assessing: A revision of bloomâs taxonomy. *New York. Longman Publishing. Artz, AF, & Armour-Thomas, E.(1992). Development of a cognitive-metacognitive framework for protocol analysis of mathematical problem solving in small groups. Cognition and Instruction*, 9(2):137–175, 2001.

[3] Zuriana Abu Bakar, Phillip Long, and Ralf Muhlberger. The effects of dynamic aesthetic interfaces on students' emotion and performance. In *User Science and Engineering (i-USEr), 2011 International Conference on*, pages 7–12. IEEE, 2011.

[4] Robert Bixler and Sidney DâMello. Automatic gaze-based detection of mind wandering with metacognitive awareness. In *International Conference on User Modeling, Adaptation, and Personalization*, pages 31–43. Springer, 2015.

[5] Elizabeth Ligon Bjork and Robert A Bjork. Making things hard on yourself, but in a good way: Creating desirable difficulties to enhance learning. *Psychology and the real world: Essays illustrating fundamental contributions to society*, pages 56–64, 2011.

[6] Lorena Blasco-Arcas, Isabel Buil, Blanca Hernández-Ortega, and F Javier Sese. Using clickers in class. the role of interactivity, active collaborative learning and engagement in learning performance. *Computers & Education*, 62:102–110, 2013.

[7] DA Bligh. What's the use of lectures? 2000.

[8] Charles C Bonwell and James A Eison. *Active Learning: Creating Excitement in the Classroom. 1991 ASHE-ERIC Higher Education Reports.* ERIC, 1991.

[9] Neil A Bradbury. Attention span during lectures: 8 seconds, 10 minutes, or more?, 2016.

[10] Diane M Bunce, Elizabeth A Flens, and Kelly Y Neiles. How long can students pay attention in class? a study of student attention decline using clickers. *Journal of Chemical Education*, 87(12):1438–1443, 2010.

[11] Jane E Caldwell. Clickers in the large classroom: Current research and best-practice tips. *CBE-Life sciences education*, 6(1):9–20, 2007.

[12] Susan Payne Carter, Kyle Greenberg, and Michael S Walker. The impact of computer usage on academic performance: Evidence from a randomized trial at the united states military academy. *Economics of Education Review*, 56:118–132, 2017.

[13] Robert G Crowder. *Principles of Learning and Memory: Classic Edition*. Psychology Press, 2014.

[14] Peter F Delaney, Peter PJL Verkoeijen, and Arie Spirgel. Spacing and testing effects: A deeply critical, lengthy, and at times discursive review of the literature. *Psychology of learning and motivation*, 53:63–147, 2010.

[15] Daniel J Exeter, Shanthi Ameratunga, Matiu Ratima, Susan Morton, Martin Dickson, Dennis Hsu, and Rod Jackson. Student engagement in very large classes: The teachersâ perspective. *Studies in Higher Education*, 35(7):761–775, 2010.

[16] Jennifer L Faust and Donald R Paulson. Active learning in the college classroom. *Journal on Excellence in College Teaching*, 9(2):3–24, 1998.

[17] Richard M Felder and Rebecca Brent. Understanding student differences. *Journal of engineering education*, 94(1):57–72, 2005.

[18] Kathleen A FitzPatrick, Kevin E Finn, and Jay Campisi. Effect of personal response systems on student perception and academic performance in courses in a health sciences curriculum. *Advances in Physiology Education*, 35(3):280–289, 2011.

[19] Scott Freeman, Sarah L Eddy, Miles McDonough, Michelle K Smith, Nnadozie Okoroafor, Hannah Jordt, and Mary Pat Wenderoth. Active learning increases student performance in science, engineering, and mathematics. *Proceedings of the National Academy of Sciences*, 111(23):8410–8415, 2014.

[20] Carrie B Fried. In-class laptop use and its effects on student learning. *Computers & Education*, 50(3):906–914, 2008.

[21] Sally A Gauci, Arianne M Dantas, David A Williams, and Robert E Kemm. Promoting student-centered active learning in lectures with a personal response system. *Advances in Physiology Education*, 33(1):60–71, 2009.

[22] Geri Gay, Michael Stefanone, Michael Grace-Martin, and Helene Hembrooke. The effects of wireless computing in collaborative learning environments. *International Journal of Human-Computer Interaction*, 13(2):257–276, 2001.

[23] Robert L Greene. Immediate serial recall of mixed-modality lists. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(2):266, 1989.

[24] Jo Handelsman, Diane Ebert-May, Robert Beichner, Peter Bruns, Amy Chang, Robert DeHaan, Jim Gentile, Sarah Lauffer, James Stewart, Shirley M Tilghman, et al. Scientific teaching. *Science*, 304(5670):521–522, 2004.

[25] Sandra G Hart and Lowell E Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology*, 52:139–183, 1988.

[26] Helene Hembrooke and Geri Gay. The laptop and the lecture: The effects of multitasking in learning environments. *Journal of computing in higher education*, 15(1):46–64, 2003.

[27] Alex H Johnstone and Frederick Percival. Attention breaks in lectures. *Education in chemistry*, 13(2):49–50, 1976.

[28] Terry Judd. Task selection, task switching and multitasking during computer-based independent study. *Australasian Journal of Educational Technology*, 31(2), 2015.

[29] Daniel Kahneman. *Attention and effort*. Citeseer, 1973.

[30] Valentin Kassarnig, Andreas Bjerre-Nielsen, Enys Mones, Sune Lehmann, and David Dreyer Lassen. Class attendance, peer similarity, and academic performance in a large field study. *arXiv preprint arXiv:1702.01262*, 2017.

[31] Robin H Kay and Sharon Lauricella. Unstructured vs. structured use of laptops in higher education. *Journal of Information Technology Education*, 10(1):33–42, 2011.

[32] Robin H Kay and Ann LeSage. Examining the benefits and challenges of using audience response systems: A review of the literature. *Computers & Education*, 53(3):819–827, 2009.

[33] Sophie I Lindquist and John P McLean. Daydreaming and its correlates in an educational environment. *Learning and Individual Differences*, 21(2):158–167, 2011.

[34] Ning-Han Liu, Cheng-Yu Chiang, and Hsuan-Chin Chu. Recognizing the degree of human attention using eeg signals from mobile sensors. *Sensors*, 13(8):10273–10286, 2013.

[35] Nicole Mareno, Marie Bremner, and Christie Emerson. *The use of audience response systems in nursing education: best practice guidelines*. DigitalCommons@ Kennesaw State University, 2010.

[36] Ference Marton and Roger Säljö. On qualitative differences in learning: Iâoutcome and process. *British journal of educational psychology*, 46(1):4–11, 1976.

[37] Margie Martyn. Clickers in the classroom: An active learning approach. *Educause quarterly*, 30(2):71, 2007.

[38] Nancy G Maxwell. From facebook to folsom prison blues: How banning laptops in the classroom made me a better law school teacher. *Rich. JL & Tech.*, 14:1, 2007.

[39] Richard E Mayer, Andrew Stull, Krista DeLeeuw, Kevin Almeroth, Bruce Bimber, Dorothy Chun, Monica Bulger, Julie Campbell, Allan Knight, and Hangjin Zhang. Clickers in college classrooms: Fostering learning with questioning methods in large lecture classes. *Contemporary educational psychology*, 34(1):51–57, 2009.

[40] Joseph P Mazer and S Andrew Stowe. Can teacher immediacy reduce the impact of verbal aggressiveness? examining effects on student outcomes and perceptions of teacher credibility. *Western Journal of Communication*, 80(1):21–37, 2016.

[41] Joel Michael. Where's the evidence that active learning works? *Advances in physiology education*, 30(4):159–167, 2006.

[42] Cody Nelson, Lisa Hartling, Sandra Campbell, and Anna E Oswald. The effects of audience response systems on learning outcomes in health professions education. a beme systematic review: Beme guide no. 21. *Medical Teacher*, 34(6):e386–e405, 2012.

[43] Yusuke Niwa, Shun Shiramatsu, Tadachika Ozono, and Toramatsu Shintani. A collaborative web presentation support system using an existing presentation software. In *Advanced Applied Informatics (IIAIAAI), 2014 IIAI 3rd International Conference on*, pages 80–85. IEEE, 2014.

[44] Harold E Pashler and Stuart Sutherland. *The psychology of attention*, volume 15. MIT press Cambridge, MA, 1998.

[45] Michael Prince. Does active learning work? a review of the research. *Journal of engineering education*, 93(3):223–231, 2004.

[46] Eric D Ragan, Samuel R Jennings, John D Massey, and Peter E Doolittle. Unregulated use of laptops over time in large lecture classes. *Computers & Education*, 78:78–86, 2014.

[47] VSV Laxmi Ramana. Teaching large classes: Problems and prospects. *Sino-US English Teaching*, 10(10):743–750, 2013.

[48] Susan M Ravizza, Mitchell G Uitvlugt, and Kimberly M Fenn. Logged in and zoned out: How laptop internet use relates to classroom learning. *Psychological Science*, page 0956797616677314, 2016.

[49] Evan F Risko, Nicola Anderson, Amara Sarwal, Megan Engelhardt, and Alan Kingstone. Everyday attention: variation in mind wandering and memory in a lecture. *Applied Cognitive Psychology*, 26(2):234–242, 2012.

[50] David Roldán-Álvarez, Estefanía Martín, Manuel García-Herranz, and Pablo A Haya. Mind the gap: impact on learnability of user interface design of authoring tools for teachers. *International Journal of Human-Computer Studies*, 94:18–34, 2016.

[51] Sharron L Ronco and John Cahill. Does it matter who's in the classroom? effect of instructor type on student retention, achievement and satisfaction. *Online Submission*, 2004.

[52] Perry J Samson. Deliberate engagement of laptops in large lecture classes to improve attentiveness and engagement. *Computers in Education*, 1(2):1–19, 2010.

[53] Faria Sana, Tina Weston, and Nicholas J Cepeda. Laptop multitasking hinders classroom learning for both users and nearby peers. *Computers & Education*, 62:24–31, 2013.

[54] Mark W Scerbo, Joel S Warm, William N Dember, and Anthony F Grasha. The role of time and cuing in a college lecture. *Contemporary Educational Psychology*, 17(4):312–328, 1992.

[55] Keng Siau, Hong Sheng, and FF-H Nah. Use of a classroom response system to enhance classroom interactivity. *IEEE Transactions on Education*, 49(3):398–403, 2006.

[56] DA Sousa. How the brain learns3rd ed, 2006.

[57] John Stuart and RJD Rutherford. Medical student concentration during lectures. *The lancet*, 312(8088):514–516, 1978.

[58] Karl K Szpunar, Novall Y Khan, and Daniel L Schacter. Interpolated memory tests reduce mind wandering and improve learning of online lectures. *Proceedings of the National Academy of Sciences*, 110(16):6313–6317, 2013.

[59] Howard Tanner and Sonia Jones. Using video-stimulated reflective dialogue to learn from children about their learning with and without ict. *Technology, Pedagogy and Education*, 16(3):321–335, 2007.

[60] Thomas C Toppino and Lance C Bloom. The spacing effect, free recall, and two-process theory: A closer look. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 28(3):437, 2002.

[61] Vasileios Triglianos, Cesare Pautasso, Alessandro Bozzon, and Claudia Hauff. Inferring student attention with asq. In *European Conference on Technology Enhanced Learning*, pages 306–320. Springer, 2016.

[62] Vasileios Triglianos, Sambit Praharaj, Cesare Pautasso, Alessandro Bozzon, and Claudia Hauff. Measuring student behaviour dynamics in a large interactive classroom setting. In *Proceedings of the 25th Conference on User Modeling, Adaptation and Personalization*, UMAP '17, pages 212–220, New York, NY, USA, 2017. ACM.

[63] Richard L Venezky. Technology in the classroom: steps toward a new vision. *Education, Communication & Information*, 4(1):3–21, 2004.

[64] Peter PJL Verkoeijen and Samantha Bouwmeester. Using latent class modeling to detect bimodality in spacing effect data. *Journal of Memory and Language*, 59(4):545–555, 2008.

[65] Phillip C Wankat. *The effective, efficient professor: Teaching, scholarship, and service*. Allyn and Bacon, 2002.

[66] Yana Weinstein, Ludmila D Nunes, and Jeffrey D Karpicke. On the placement of practice questions during study. *Journal of Experimental Psychology: Applied*, 22(1):72, 2016.

[67] Eileen Wood, Lucia Zivcakova, Petrice Gentile, Karin Archer, Domenica De Pasquale, and Amanda Nosko. Examining the impact of off-task multi-tasking with technology on real-time classroom learning. *Computers & Education*, 58(1):365–374, 2012.

[68] Christian Wurst, Claudia Smarkola, and Mary Anne Gaffney. Ubiquitous laptop usage in higher education: Effects on student achievement, student satisfaction, and constructivist measures in honors and traditional classrooms. *Computers & Education*, 51(4):1766–1783, 2008.

[69] Kevin Yamamoto. Banning laptops in the classroom: Is it worth the hassles? *Journal of Legal Education*, 57(4):477–520, 2007.

[70] Janice D Yoder and Catherine M Hochevar. Encouraging active learning can improve students' performance on examinations. *Teaching of Psychology*, 32(2):91–95, 2005.

[71] Mengyu Zhou, Minghua Ma, Yangkun Zhang, Kaixin SuiA, Dan Pei, and Thomas Moscibroda. Edum: classroom education measurements via large-scale wifi networks. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 316–327. ACM, 2016.

[72] Erping Zhu, Matthew Kaplan, R Charles Dershimer, and Inger Bergom. Use of laptops in the classroom: Research and best practices. *Centre for Research on Learning and Teaching*, 30:1–6, 2011.

Appendix A

Extraction and Preprocessing

Here, we explain about the data extraction and the pre-processing of the collected data.

A.1 Data extraction from MongoDB

The raw data is extracted from MongoDB running in the ASQ server which is in BSON (Binary JavaScript Object Notation) format. Then we restore the MongoDB database in the local host so that we can view the entire data schema. Next, data in ".csv" format is generated from this local host version of the data in MongoDB. Data frames that are extracted as ".csv" from MongoDB after initial pre-processing are *slideshows*, *sessions*, *answers*, *exercisesubmissions*, *sessionevents* and *whitelistentries* as shown in table A.1, A.2, A.3, A.4, A.5 and A.6. Complete details about these ".csv" files can be found in the Github Repository mentioned in the URL in *Appendix B*.

A.1.1 Overview of data frames (or tables)

Here, we show the important columns of the data frames extracted after initial pre-processing.

Table A.1: Data frame of *slideshows*. This contains the details about the slides used in the lecture. **id** shows the session id of the lecture. **title** shows the name of the lecture. **slidesTree** shows the name and id of all the slides in JSON format. **questions** shows the id of all the questions used in the lecture in JSON format. **exercises** shows the id of all the exercises used in the lecture in JSON format. **questionsPerSlide** and **exercisesPerSlide** associates the name of the slide with the id of the questions.

id	title	slidesTree	questions	exercises	questionsPerSlide	exercisesPerSlide

Table A.2: Data frame of *sessions*. This contains the abstract details about the session i.e., the lecture. **id** shows the session id of the lecture. **title** shows the name of the lecture. **startDate** and **endDate** shows the start and end date and time of the lecture. **breakStartDate** and **breakEndDate** shows the start and end time of the breaks during the lecture.

id	startDate	endDate	title	breakStartDate	breakEndDate

A.2 Preprocessing

In this section, we discuss data cleaning, reformatting and aggregation as a pre-processing measure before starting the data analysis.

Table A.3: Data frame of *answers*. This data frame contains the details about the answers given by the students during the lecture. **question** contains the id of all the questions. **answeree** contains the id of the user who answered the question. **submitDate** contains the date and time of the submission of answer to a question. **submission** contains the response given to the question with the correct answer for the question in JSON format. **type** shows the type of question like multiple-choice, fill the blanks, SQL Query, etc. **startDate** contains the date and time in which the user started answering the question. **difficulty_level** shows whether a question is "easy" or "hard".

question	answeree	type	submitDate	submission	StartDate	difficulty_level

Table A.4: Data frame of *exercisesubmissions*. This data frame contains the details about the exercises and submissions. **exercise** contains the id of the exercise. **answeree** contains the id of the user who answered the question. **session** contains the session id of that lecture. **submitDate** contains the submission date and time.

exercise	answeree	session	submitDate

Table A.5: Data frame of *sessionevents*. This data frame contains the details about the sessions i.e., the lectures which are part of our experiments. **session** contains the session id of the lecture. **type** contains the type of event like connected, tabvisible, etc., which helps us in event aggregation later. **time** shows the date and time at which the event is fired. **user** shows the id of the user associated with that event.

session	type	time	user

Table A.6: Data frame of *whitelistentries*. This data frame contains the details about each unique user who connected with ASQ during the lecture. **session** shows the session id of the lecture. **user** shows the id of the unique user who connected with ASQ during the lecture. **joinDate** shows the date and time at which a new user joined the lecture. **role** shows the role played by the user i.e., whether the user is a presenter (or instructor) or a viewer (or student).

session	user	joinDate	role

A.2.1   Data cleaning:

The main aim of cleaning the raw data is to remove various inconsistencies in the data and fix some of the content in data fields wherever necessary. In our case we change all the *timestamps* from ISO 8601 format to POSIXct object in R so that it can be analyzed in R unambiguously. We also check the type of event for each user and if the user is a "viewer" (student) and not a "presenter" (instructor), then we change the "ctrl-disconnected" event to "folo-disconnected" event. The reason of changing it is to differentiate between the disconnected event for presenter and the viewer.

A.2.2   Data reformatting and aggregation:

First, we select the sessions that were part of the live classroom presentation from the list of all the presentation and testing sessions. Then we rename the "_id" field in all the raw data frames (like slideshows, sessions, answers, exercisesubmissions, sessionevents and whitelistentries) with "id" field. Next, start and end time of each session (lecture) is filtered

based on the true start and end time obtained from the live observations. We also add each user who triggered the event as a data field to the sessionevents. Finally, for each question in the answers data frame we find the start date of the corresponding question by looking for question-activated events in sessionevents and save the result to a new data frame *answers_with_start_date*.

Appendix B

Script

Here, we will include the reference to the Github repository which gives an overview of the important sample of HTML, css and JavaScript implemented while designing the slides to be used in `ASQ`. It includes examples of different types of `ASQ` questions, videos and the modifications made while converting the slides into proper HTML format which will be displayed using `ASQ`. We also include the Github repository for the analysis of data done using R.

Github repository containing script used for preparing the slides for all the lectures:
`https://github.com/sambit2/delft-web-technology-and-database-slides-2016-2017`

Github repository containing the data and R script used for analysis of the logs obtained from the lectures:
`https://github.com/sambit2/DataAnalytics-Web-based-Education`

Appendix C

Debugging

Some lectures of one of the instructors had videos embedded in the presentation. During one lecture containing around 7 videos there was a problem in between the presentation when the presentation freezed for few minutes and also started to flicker back and forth from previous slides till the next slides. We checked the error log in Chrome browser but could not ascertain the root cause of the problem. After the end of the lecture we tried to determine the cause of the problem based on different assumptions. Either the internet was disconnected during the lecture or there was some problem with the limit to the total number of videos or may be a effect of both.

So, we decided to recreate the same problem by playing the presentation used for the lecture again from the beginning by employing certain conditions. First, we disconnect the internet in the middle of ongoing presentation and see the effect on the flow of the presentation. The presentation freezed in the middle. Then, we reduce the number of videos from 7 to 6 and continue with the same internet disconnection experiment which again caused disturbance in the presentation. But, when we reduce the number of videos to 5 then we do not see any interruption in the flow of the presentation. This gives us an idea that the cause of the problem was a mixed effect of disconnection of the internet and the total number of videos. But, to be assured we check Stackoverflow and find the same where it is mentioned that Chrome has a limit of open sockets which is "6 sockets can open at a time". So, for future presentations we reduce the number of videos to less than 6 and also change the "preload" attribute of the video to *none* so that all the videos do not load on loading the page thereby reducing load on the server too. Instead, when we reach the slide containing the video then it starts to load and play.

Appendix D

Usability Rating of ASQ in NASA TLX Scale

NASA TLX scale [25] helps us to determine the perceived workload while executing a task based on six different indicators like mental demand, physical demand, temporal demand, performance, effort invested and level of frustration. First the weight of each of the indicator is determined and then the raw rating is obtained which helps us to calculate the aggregated adjusted weighted rating. It is on a scale of 0-100. In our case, as we can see from the 14 lectures, both the instructors had higher workload when they first used ASQ (may be due to in-familiarity with the interface) which decreased with the progression of the course.

Table D.1: Average Weighted Rating (Scale of 0-100) on NASA TLX scale for both the instructors (I1 and I2)

Lecture	Instructor	#Avg	Other notes
HTTP	I1	68.66	Cockpit view (i.e., the view which shows the next incoming slide) did not work for next slide preview
HTML	I1	28.66	Cockpit view worked with next slide preview
JavaScript	I1	36.33	
Database introduction	I2	62.33	
node.js	I1	NA	
CSS	I1	36.00	
SQL introduction	I2	32.66	
SQL continued	I2	28.33	
Cookies & session	I1	35.33	
Advanced SQL	I2	32.33	
Web security	I1	56.33	
Conceptual design	I2	23.66	
ER logical design	I2	20.66	
NoSQL	I2	14.00	

Specifically, in the first lecture (HTTP) the workload is very high because the cockpit view which shows the preview of the next slides did not work. The average workload for both the instructors through out the course was lower (around 40 and 30 respectively for I1 and I2) indicating that ASQ can be used by teachers at ease in a classroom.