Adding insight in educational lecture environments with ARS

A post-presentation analysis using an interactive visualization tool

ALEXANDRA RUNHEM
Abstract
Feedback plays an important role in evaluating and developing courses in higher education. Due to inefficient factors, the feedback cycle does not meet its full potential and can therefore be counterproductive. Current evaluation methods typically demand much effort from both students and teachers, when asking for feedback at the end of the course and analyzing the results. Lack of engagement can thus be observed in both parties. Moreover, we can see an increasing trend of ARS used in educational settings, to improve learning quality and enhance the relationship between the presenter and audience.

This study’s aim was to make the feedback process more efficient and to explore how to provide insight in lecture quality for continuous course development. To fully understand the target user, a pre-study was conducted to identify design requirements and to investigate which areas to evaluate during the courses and why. The interview sessions resulted in four main dimensions to evaluate; Meaningfulness, comprehension, knowledge and attitude. Based on these dimensions, a feedback tool was developed to gather the feedback data from students in two different cohorts. The tool was developed in a survey format, with the help of an existing ARS. The feedback was then collected after two university lectures during two courses. To explore the potential of providing useful insight to the lecturer, and to facilitate the analytical step of the process, an interactive visualization tool was prototyped to display the data. The visualization tool was evaluated, both in terms of usability and its overall concept, with a total of eight lecturers, two of which were lecturers in the courses used to gather the feedback data.

Even though the results show that it might be difficult to draw a single conclusion of the tools’ usability, the users found the concept interesting and were positive towards the idea. The perception of the tool’s intended use varied and is discussed along with future development.
Sammanfattning
Feedback spelar en viktig roll när det kommer till utvärdering och utveckling av kurser inom högre utbildning. På grund av ineffektiva faktorer uppnår feedbackcykeln inte sin fulla potential och kan därför vara kontraproduktiv. Aktuella utvärderingsmetoder kräver typiskt mycket ansträngning från både elever och lärare och brist på engagemang kan således observeras hos båda parter. Dessutom kan vi se en ökande trend av ARS som används i utbildningsinstitutioner, för att förbättra utbildningskvaliteten samt förhållandet mellan presentatören och publiken.

Syftet med studien var att göra feedbackprocessen mer effektiv och undersöka hur man kan ge insikt i föreläsningsskvaliteten för kontinuerlig kursutveckling. För att förstå slutanvändaren genomfördes en förundersökning för att identifiera deras begär och för att undersöka vilka områden som ska utvärderas under kurserna och varför. Intervjuerna resulterade i fyra huvuddimensioner att utvärdera och ge insikt inom; Betydelse, förståelse, kunskap och attityd. Baserat på dessa dimensioner utvecklades ett feedbackverktyg. Verktyget var i form av ett formulär och producerades med hjälp av ett befintlig ARS. Feedbackdata samlades sedan in efter två universitetsföreläsningar under två kurser. För att undersöka möjligheten att ge föreläsaren användbar insikt och för att underlätta det analytiska steget i processen, utvecklades ett interaktivt visualiseringsverktyg. Visualiseringsverktyget utvärderades, både vad gäller användbarhet och dess övergripande koncept, med totalt åtta föreläsare.

Även om resultaten visar att det kan vara svårt att dra en enda slutsats av verktygets användbarhet, fann användarna konceptet intressant och var överlag positiva. Uppfattningen av verktygets avsedda användning varierar och diskuteras tillsammans med framtida utveckling.
Adding insight in educational lecture environments with ARS: a post-presentation analysis using an interactive visualization tool

Alexandra Runhem
KTH Royal Institute of Technology
Stockholm, Sweden
arunhem@kth.se

ABSTRACT
Feedback plays an important role in evaluating and developing courses in higher education. Due to inefficient factors, the feedback cycle does not meet its full potential and can therefore be counterproductive. Current evaluation methods typically demand much effort from both students and teachers, when asking for feedback at the end of the course and analyzing the results. Lack of engagement can thus be observed in both parties. Moreover, we can see an increasing trend of ARS used in educational settings, to improve learning quality and enhance the relationship between the presenter and audience.

This study’s aim was to make the feedback process more efficient and to explore how to provide insight in lecture quality for continuous course development. To fully understand the target user, a pre-study was conducted to identify design requirements and to investigate which areas to evaluate during the courses and why. The interview sessions resulted in four main dimensions to evaluate; Meaningfulness, comprehension, knowledge, and attitude. Based on these dimensions, a feedback tool was developed to gather the feedback data from students in two different cohorts. The tool was developed in a survey format, with the help of an existing ARS. The feedback was then collected after two university lectures during two courses. To explore the potential of providing useful insight to the lecturer, and to facilitate the analytical step of the process, an interactive visualization tool was prototyped to display the data. The visualization tool was evaluated, both in terms of usability and its overall concept, with a total of eight lecturers, two of which were lecturers in the courses used to gather the feedback data.

Even though the results show that it might be difficult to draw a single conclusion of the tools’ usability, the users found the concept interesting and were positive towards the idea. The perception of the tool’s intended use varied and is discussed along with future development.

Author Keywords
Feedback; Audience response system; Information visualization; Usability

1. INTRODUCTION
The traditional slide presentation is often built upon a one-way transmission model, where a presenter provides information to an audience, whose main task is to listen. The one-way communication stream limits inclusiveness and thus a lot of effort is put on the presenter to create such an environment. Earlier research points out that lack of engagement, attention and interest from the audience may occur during these settings, which is why there has been an increasing trend in using audience response systems (ARS) as a presentation tool. Due to its interactive nature, the presentation is no longer just a one-way stream of information but offers the audience to participate in a greater sense. Many ARS systems focus on what happens before and during a presentation, providing useful tools to set up different features and summarizing the audience responses per slide. But, how do you know what the audience actually thought of the presentation? Is the visual feedback enough for a presenter to be able to draw a conclusion of the audience perception?

One scenario where evaluation of presentations is particularly important is in educational institutions. Here it is important to ensure that the information that is being taught has been understood and thus provide learning quality. With the help of feedback from students, the learning outcomes can be evaluated, and actions can be made for improvements. Institutions of higher education have different approaches to manage feedback and improve the quality of a course. A way of doing so includes evaluation forms that is sent out to students, after the course has finished, giving no opportunity to make changes on the go. The process can be time consuming for both students, faculties and teachers where common problems are connected to lack of engagement.

1.1 Research Question
The purpose of this study is to investigate how to provide insights regarding lecture quality in a teaching environment using an ARS, enabling the possibility to make changes during the course instead of after. The aim is furthermore to make the feedback process easier; from constructing the evaluation to analyzing the results. To accomplish this, the study aims to evaluate and develop a digital analysis tool in the form of an interactive visualization, which will display feedback from students at the end of a lecture. The research question is as follows:

How can a visualization tool be designed to enhance post-presentation analysis and increase insight of the lecture quality, when using ARS for collecting feedback amongst an audience in a teaching environment?

1.2 Principal: Mentimeter
The principal of this study was Mentimeter AB, a Swedish ARS company founded in 20121. The collaboration included the use of their product to explore the topic of feedback after a presentation

1 https://www.mentimeter.com
setting, to see the potential of integrating a final presentation slide with a summary of the audience feedback.

2. THEORY AND RELATED RESEARCH
These sections will establish related work regarding feedback and usage of ARS in higher education, as well as guidelines for developing information visualizations and evaluating its usability.

2.1 Feedback
When looking at quality assurance in educational institutions, feedback from students plays an important role in developing and improving teaching and learning outcome [9]. However, the implemented processes of gathering and analyzing student input are somewhat time consuming for the students and teachers affected. Inefficient factors include constructing evaluation surveys and analyzing the result, but maybe most importantly the resources put in to decide what to evaluate and how.

Feedback can be collected in numerous ways and include various dimensions, but if not analyzed or acted upon it suffers a counter-productive result, which can be a waste of time for those involved [9]. On that same note, if students do not understand the purpose of providing feedback nor how the feedback will be handled, lack of engagement will occur [9], [21]. Research points out that a feedback process, regardless of the application area, needs to be seen as a cycle, where the steps are transparent for them involved [21].

\[\text{Feedback is not only a courtesy to those who have taken the time to respond but it is also essential to demonstrate that the process both identifies areas of student concern and does something about them.}\]
- Harvey et al 1997

To motivate and engage an audience enough for them to want to provide feedback, it is important to look at what type of response the survey is asking for. Open-ended questions often provide more insight and useful information, but at the same time it demands more time and effort to formulate a response. Due to these aspects, open-ended questions have shown to have a higher rate of non-responses, compared to other evaluation methods in surveys [15]. Moreover, one can discuss if open-ended questions produce objective data, representing the overall crowd, or not. Putting in time and effort to write a detailed answer that provides useful information is more likely when the respondent is either really satisfied or really unsatisfied, but not when the experience was somewhere in between [15]. At the same time, a study made at the Swedish university LTH showed that teachers were more interested in free-text comments rather than quantitative data and were more likely to use visual input from students as a source for feedback during lectures. The same study also discusses the topic of emotions when receiving feedback from students. The lecturers interviewed expressed a positive attitude towards student feedback yet concerns about the potential emotional tension [17].

2.2 Audience response system
An audience response system (ARS) is a tool for enhancing the engagement and relationship between a presenter and the audience, with interactive content and features. Instead of a one-way transmission model of information, from presenter to audience, the audience gets a voice and is invited to take part in the presentation in a more active way. For instance, an ARS allows a presenter to involve the audience in activities, such as a quiz, where the mobile phone interface is used to respond to questions presented on a screen for the audience to see. The data can then immediately be displayed and analyzed on the same screen [11].

In teaching environments, ARSs have specifically been used in order to improve several dimensions such as students’ engagement, attention and learning performance, where the system has proved to contribute to active learning [11]. Moreover, ARSs provide an opportunity for the lecturer to modify the content on the go based on student input [7]. It helps to transform a static, traditional, one-way transmission model where information is passed from a speaker to a receiver, into a dynamic one that involves them both. There is a difference between receiving feedback from the audience who expresses confusion of content presented and instantly changing the teaching style, as suggested by Abrahamson [11], [11]. To actively change content and pedagogical style on the go demands a lot from a presenter, but the importance of audience input shall not be discarded on the basis of a lack of perseverance.

Other features that makes ARS valuable include preserved anonymity, which increases honesty in responses, as well as preserved data that can be processed and reflected upon afterwards, compared to a live “show of hands” session [11]. A study comparing ARSs, hand-raising and response cards in a classroom environment showed that both ARSs and response cards had higher response rate than the hand-raising group and that the students using an ARS for feedback were more likely to respond honestly [19]. Using an ARS for feedback can thus be beneficial, however little research has been done examining this field in practice.

Researchers also express flaws with current ARSs in teaching environments. For instance, the time and effort demanded from teachers to produce questions that are of good quality is considered a liability [11], [12].

2.3 Information visualization
The essence of information visualization is to provide insight from a data set. By generating visual representations and providing a layer of interactivity, users can access several dimensions of insight that might be a complex task when only looking at a raw data set. In short, the abstract information becomes more perceptible [16].

Humans are excellent when it comes to seeking patterns [6], [20]. With our capability of rapid processing of visual information, visualizations efficiently give us deeper insight and meaning of the data shown. One must keep in mind that insight comes in layers, all of them providing a piece of the puzzle that makes up the whole picture. The visual information seeking mantra suggests a design that provides the user with information in different stages; Overview first, zoom and filter, then details on demand [18]. This strategy encourages users to dig deeper and seek insight in several dimensions. The user can actively filter out irrelevant data and look at one piece at a time. All of this whilst having a good overview of the data at first sight for a general understanding.

One might say that there is no right or wrong way of visualizing data, however there are better and worse ways of doing so. With that said, it is important to create visual mappings that support the characteristics of the data provided and give the user the desired insights of the information displayed. Previous research suggests three elements for constructing visual mappings; the spatial substrate, the graphical elements and the graphical properties [6]. The first element is all about defining the space of the visualization, the
second to define how the data should be represented in the space (i.e. points, lines, areas, symbols etc.) and the third what properties the data point should have to indicate a certain relationship towards other points (i.e. color, size, orientation etc.). These elements quickly form a domain of many different visual structures that are more or less suitable depending on the use case. For instance, if visualizing comparison, suggested structures might be bar or line charts, while if visualizing compositions, gauge, pie or stacked charts might be better options [23].

It goes without saying that the design is not the only thing that matters when developing a visualization. Gathering, processing and transforming the data itself is essential to provide a solution to meet the user’s needs and desires. Transforming raw data to a visualization is a process of its own, but there are strategies to follow. The visualization pipeline is a process for converting raw data into interactive visualizations [16]. Throughout the process, the focus lies on the target group that will use the finalized product, however it can include different target groups along the way, for instance when gathering data. The user’s needs and desires are what the solution should cover throughout the process; from gathering data to transforming it to a visualization.

2.4 Evaluating usability

Usability is defined as “...the effectiveness, efficiency and satisfaction with which specified can achieve specific goals in particular environments.” according to the International Organization for Standardization (ISO) [10]. Whether or not an interface ensures usability for a user depends on different factors, such as previous experience, culture, knowledge and age. Hence, when designing an interface for usability, it is important to know who you are designing for to ensure that the level of usability matches the user’s level of previously mentioned factors.

One of the problems that comes with evaluating usability of an interface is accounting for the fact that users’ performance will most likely change over time. The more the user interacts with an interface, the easier it will be to use, as the user learns how to navigate and perform the desired tasks [10]. Hence, if an interface is evaluated with a lot of first-time users in mind, this is of importance considering the focus should lie more on learning something from the results rather than trying to learn how to use the interface itself. Focusing on usability for first-time users is therefore essential. With this taken into account, a model with the aim of evaluating usability for first-time users has been developed [10]. The goal is to see how many task repetitions are required until the user has met a certain level of expertise using the interface. The model contains three factors; Guessability, learnability and experienced user performance (EUP). The first factor, guessability, refers to the cost the user makes when solving tasks with the interface for the first time. The cost in this context can be for instance the time it takes to solve a certain task or the error rate, meaning that the lower the cost is the higher the guessability factor becomes. Learnability is the concept of evaluating the competence level of performing specific tasks when the tasks has already been solved once before. EUP refers to the cost once the user has reached a steadier level of performance.

Evaluating the usability of an interface can thus be made by measuring the effectiveness, efficiency and user satisfaction while accounting for the guessability, learnability and EUP when analyzing the results. Effectiveness can be measured by calculating the completion rate, i.e. how many tasks were completed successfully out of the total task undertaken, and efficiency by the amount of time it took to solve a task, i.e. how much effort it took to perform the task. User satisfaction can be measured through a standardized satisfaction questionnaire, for instance a system usability scale (SUS) [10], [14].

3. METHOD OVERVIEW

To answer the research question, a feedback tool was created to gather the necessary data. The data was then used to develop an interactive visualization tool to display the information for the end user. The study can be divided into two phases, where the first phase focused on the research regarding the nature of the feedback tool. The second and main phase of the study, focused on developing and evaluating the visualization tool.

The first phase included a thorough literature review as well as pre-study with the target users. Based on the results, the user needs and design requirements were defined and thus gave a solid foundation for constructing the feedback tool and enable data gathering.

Based on the data collected and the design requirements, the second phase started. The focus in phase two was visualizing the feedback. The final prototype was evaluated together with the previously interviewed subjects as well as multiple additional lecturers at the Royal Institute of Technology (KTH), located in Stockholm, Sweden. In total, eight lecturers participated in the evaluation. The following sections will further explain the development phases and results.

4. FEEDBACK TOOL

This phase included a pre-study, to define the requirements from the target group, and a data gathering session where the tool itself was developed and utilized. The sections below will further explain the outcomes.

4.1 Pre-study

Following the visualization pipeline, it is important to understand that the user is the centerpiece during the process. To fully understand the target user and its needs, this study started with conducting a pre-study, interviewing experts in the field. This included teachers and various researchers in pedagogy at KTH. The aim was to get an idea of how the teachers perceive feedback from students, how the information is processed and acted upon as well as what they desire in a potential feedback visualization. To fulfil the goal, the semi-structured interview method was chosen. The method was found appropriate due to the fact that it ensures flexibility at the same time as providing qualitative, focused data about the users’ experience and thoughts. The constructed questions were opened-end and neutral, in line with the guidelines for the method [22].

The results from the semi-structured interviews indicated that teachers see student feedback as a helpful tool for improving the quality of a lecture, in line with previous research. At multiple occasions it was pointed out that a lecture is not only supposed to be an information provider of a certain subject, but to be inspiring and
meaningful in the bigger sense. The lecture should fit in a context in which the students can understand how the information could be applicable in real life situations. The students should not only understand why they are taking the course but also how the information taught can be useful.

In a lecture setting with a large crowd, interacting with the students can be a hard task according to one of the teachers. It limits the possibility to notice potential confusion in the audience, as well as the opportunity for students to speak up. For instance, a scenario was mentioned where the lecturer asked the students if they had understood. The visual feedback received was a couple of nods, so the teacher continued with the lecture. After a while, one of the students raised a hand and expressed confusion about the subject being taught, where yet again the visual feedback from the rest of the class was agreeing to the students’ confusion. Misunderstandings in large classroom settings were not uncommon. Moreover, analyzing the general crowd’s perception and understanding during a lecture was considered a tricky task. Is it only that one student who does not follow, or is it the whole class? How can you tell?

The teachers’ current system for evaluating the courses included student representatives as well as a course evaluation survey that is sent to the students after completing the course. The questions asked in the survey were often a mix between qualitative and quantitative. When talking about what type of data provided the most value in a teacher’s point of view, qualitative responses like open-ended questions were brought up. From their experience, open-ended questions provided more insight and details about what did and did not work, in line with Miller et al. research [15]. They also experienced non-responses and raised concerns about whether the data could be seen as objective or not. Another interesting discussion that came up regarding quantitative responses was how difficult it could be to determine when to be satisfied with the result. What does a number actually mean and when is the number good enough?

Based on the interviews, a thematic analysis was used to find patterns in the responses [5] [2], [8]. Using this method, the findings were four main dimensions to evaluate and visualize in the feedback and visualization tool. The first dimension was defined as attitude, where the main focus would be to evaluate the students general attitude towards the lecture and course. The second was defined as comprehension, including evaluation of the students’ understanding of the content and how to apply it. The third dimension was defined as meaningfulness, with the focus of evaluating the relevance and value of the lecture. The fourth, and final, dimension was defined as knowledge, evaluating students’ knowledge level before and after the lecture.

4.2 Data gathering
The aspects identified in the pre-study were used to construct the feedback tool. A survey format was selected and consisted of 12 questions in total. The questions were grouped in the dimensions mentioned in the previous section, and thus each section in the survey had its own focus area. Meaningfulness, comprehension and knowledge had four respectively three and two questions, each with a five-point Likert scale. Attitude was a single multiple-choice question, with the values “Bad”, “Neutral” and “Good”. To identify the cohorts, the survey also included a question about the field of study. Here, one could choose between different educational programs or the options “Other”, if the options did not apply, and “Do not want to say”. Moreover, each question included a “Skip” button, if the students were to feel that they did not want to respond.

With the help of the existing ARS used in this study, the feedback tool was developed and accessible through a URL.

The feedback data was collected from students during two courses at KTH. Each course had two sessions, resulting in a total of four sessions and 114 responses (Figure 2). An introduction to the study and its purpose was presented to the students when half of the lecture had passed, during the break. The students were asked to fill out the feedback survey at the end of the lecture and were given a reminder as the lecture was about to end. The courses’ sessions included one introductory lecture followed by a more information dense lecture.

![Figure 2 - Number of respondents each session and course](image)

5. VISUALIZATION TOOL

5.1 Development
The development phase followed the structure of the visualization pipeline, mentioned in more detail in chapter 2.3. The sections below will further explain how this study went from raw feedback data to a complete visualization.

5.1.1 Data transformation
The used ARS platform provided the functionality to download the responses as an Excel file. When analyzing the data a few empty responses were detected, meaning that the students had accessed the survey but skipped the questions. Due to the scope of this study, it was decided to remove them from the data set and only visualize the feedback with existing data points for all questions.

Since the data included responses from two different cohorts, i.e. two different educational programs, each segment and their responses were transformed into new data sets to easier process the data in the programming phase. Each feedback dimension consisted of quantitative data sets, and so it was decided to calculate a weighted average of the responses for each dimension to display in the overview section as a general score. After this process, the data was exported as a json-object and imported into the application of the prototype.

5.1.2 Visual mapping
The four dimensions identified in the pre-study were used to make up a draft of how the visualization might look like. The draft was iterated to make up the final design of the prototype. Along with the iterations, design critique sessions took place together with ex-
perts in the field of information visualization, design and user experience (UX) to receive continuous feedback of the design choices.

The decision of what visual mappings to choose was based on the characteristics of the data as well as the desired insight that the visualization was to provide. Based on the existing recommendations for visualizations (see section 2.3), the overview focused on giving a general idea of the information provided. One efficient way of indicating a certain state is to benefit from different graphical properties, such as color indicators. Colors spanning from red to green are perceived as something going from bad to good in many cultures. Hence, the frequently used graphical property in the overview section was color attributes varying in the red-green spectrum. Moving forward and looking at the spatial substrates, meaningfulness and comprehension were mapped onto a gauge chart to display the overall score, i.e. the weighted average from the responses to each question in the dimensions. The gauge values spanned from 1 to 5, using a needle to indicate the score. The red-green color scheme was mapped to the min and max values of the gauge, meaning that the needle pointed towards a more green or red section depending on the score value (Figure 3).

\textit{Attitude} was mapped towards an area chart showing the distribution of respondents who expressed negative, neutral or positive attitude towards the lecture. The y-axis spanned from 0\% to 100\%. Depending on the distribution, the area was colored in the red-green spectrum. The more negative responses, the redder the chart was colored. The more positive responses, the greener the chart was colored. The detailed view had the same mapping, showing two areas (one for each cohort). The colors used here were not in the red-green spectrum to limit the confusion. Where the areas overlapped, an opacity change was used to enable seeing both values at the same time.

The \textit{knowledge} dimension benefited from the Likert scale used in the survey, which was used to map the data in the visualization. The weighted averages of the responses were displayed in a circle and positioned correspondingly on the rendered scale. Above the scale, an area chart was used to map the distribution of answers regarding knowledge after the lecture. Depending on the weighted average on knowledge after the lecture, the chart was colored in a more red or green value.

Number of respondents from each cohort was seen as a comparative data set and was therefore mapped to a pie chart. Each cohort received color attributes to make it easy for the user to distinguish the amounts. The colors chosen were purely a design choice and did not include red nor green to ensure as little confusion as possible.

The detail view of \textit{comprehension} and \textit{meaningfulness}, displaying all questions and Likert scale answers, was mapped to a stacked bar chart. This chart was chosen due to the fact that each question had five possible responses. Stacking the responses together creates more space and reduces the risk of visual cluttering. The color property of the bars spanned from red to green to indicate the values ranging from “Strongly disagree” to “Strongly agree”. The X-axis domain spanned from 0\% to 100\%. The bar chart displayed the total responses for each question, and to maximize the space available on the screen it was decided to map each cohort’s responses to a pie chart. The same color attributes were used to indicate the value meanings.

### 5.1.3 View transformation

The final prototype was developed as a web application, using Angular as primary framework, HTML for markup and CSS for styling. D3.js was used for creating the actual visualizations, which is a JavaScript based library that renders visual content as SVG elements \cite{D3}. The code setup followed a component-based structure, meaning each visualization on the screen is its own component.

The overall layout of the visualization focused on giving the user a solid overview and details when desired \cite{D3}. A hover interaction was applied on all components to display more information. Hovering total respondents and attitude triggered a flip animation which revealed a graph on the back side. Hovering the charts in this state revealed the data in text form. Exiting the state was done by moving the cursor out from the component, which brought back the original overview. Moving to the knowledge section, hovering the chart triggered details about the distribution of knowledge after the lecture. Exiting the state was done in the same way mentioned above. When hovering on meaningfulness and comprehension, the exact score was revealed on the gauge. What differed here, however, was that clicking on the sections redirects the user to a new page (Figure 4). Here, the user could access the questions asked, and see how the responses varied from “Strongly disagree” to “Strongly agree”. Hovering the bars revealed information about how many percent of each cohort had answered the corresponding value of the bar. To exit the view, a back button was placed in the top left corner, which led back to the overview section.

Since data had been collected during two sessions, the user could change the date using the timeline slider at the bottom of the overview section (Figure 3). Dragging the circle changed the visualized data, and new charts were rendered.

![Figure 3 - The overview section of the visualization tool displaying (a) Number of respondents (Top left), (b) Attitude distribution (Top right), (c) & (d) Overall score for Meaningfulness and Comprehension (Bottom left), (e) Mean values for knowledge before and after the lecture, with distribution of answers.](image)
5.2 Evaluation

The prototype was evaluated by interviewing eight lecturers at KTH, including the lecturers from the courses that were used to collect the feedback. When evaluating the tool with other teachers, the course name was not revealed to ensure anonymity of the lecturers.

Before the evaluation, each teacher received a consent form based on the “Usability Test Consent Form” provided by Eric Mao [13], with a few minor modifications to fit the purpose of this study. It was pointed out that the teacher’s performance was not going to be evaluated per se, but merely the purpose and usability of the prototype.

The evaluation was mainly task driven with an interview as a closing session. The task driven evaluation was based on the research by Jordan et al. about the concept of usability components, as described above [10]. To evaluate the learnability aspect, the task’s character was presented in a gradually increasing order of difficulty. The final task, however, was similar to the first one and had the same level of difficulty, to see whether or not the users’ performance had changed when solving a task on this level of difficulty for the second time. To reflect on the guessability of the interface, the time it took to solve a task was measured. The error rate was documented, as well. The time aspect was also used to reflect on in what instances, if any, the design showed flaws and confusion. In total, each participant received seven tasks to solve during the evaluation;

T1 - Find the number of total respondents.  
T2 - Find the general level of comprehension.  
T3 - Find the amount of Computer scientists who had positive attitude towards the lecture.  
T4 - Find how many answered a 4 (out of 5) on Knowledge after the lecture.  
T5 - Find how many % of Media technology students that said "Strongly agree" on the question "Going to the lecture was valuable".  
T6 - Find the mean value of "Knowledge after the lecture" on a specific date.  
T7 - Find how many Media technology students who answered the survey.

To calculate the error rate and overall effectiveness of the interface, a score was given to each participant for each task. A score of “0” meant that the user was able to solve the task without any help, hence the “0” indicates that no error was found when solving the task. A score of “1” meant that the user was able to solve the task, but with some help. Some help was defined as users asking for help once or wanted to have the task repeated. A score of “2” meant that the user was able to solve the task, but much help was given. This meant that users asked for help, or had the task repeated, multiple times. Finally, a score of “3” meant that the user was not able to solve the task or stated the wrong answer. Given these scores, the maximum possible score per task would be 24, if all eight users were given a “3”. The error rate and effectiveness were calculated accordingly:

$$ Error\ rate\ per\ task = \frac{\sum Score\ per\ task}{Score_{max}} \times 100, \ Score_{max} = 24 $$

$$ Effectiveness\ per\ task = \frac{Number\ of\ tasks\ completed\ successfully}{Total\ number\ of\ tasks\ undertaken} \times 100 $$

The efficiency aspect was analyzed by looking at the amount of time it took to solve a task. The time-based efficiency (TBE) factor per task was calculated accordingly:

$$ Time-based\ efficiency = \frac{1}{N} \sum_{i=1}^{N} \frac{m_i}{t_{ij}} $$

Whereas the task driven evaluation focused on the usability of the visualization tool, the purpose of the interview was mainly to evaluate the concept of visualizing the lecture quality for continuous development of the course. Another aspect of constructing the interviews was to receive qualitative data of general feedback and user satisfaction regarding the visualization. To be able to analyze the data, an audio recording took place during the evaluations. The sessions were transcribed and analyzed using a thematic approach to find patterns within the data. [2], [5], [8] The questions asked were the following:

Q1 - Given this visualization, do you feel that you have gotten any insight of the lecture you did not have before?
Q2 - What did you like the most?
Q3 - Is there something you feel is missing in the tool?
Q4 - Would you use this tool again? Why/Why not?
Q5 - Would you recommend a colleague to use this tool? Why/Why not?
Q6 - Do you have any additional comments or questions?

Since only two users were shown data from their own lectures, the other six were asked to imagine themselves in the scenario that they had recently held a lecture and the visualization was feedback collected from that session. Q1 mentioned above was slightly modified grammatically to fit the scenario, but the purpose remained the same.

5.3 Result

These following sections will present the outcome of the evaluation made with the visualization tool.

5.3.1 Error rate and effectiveness

Calculating the error rate and effectiveness, the results show that task T1 and T7 had no errors and a 100% effectiveness meaning each user was able to solve task without any struggle. Task T3 showed to be most tricky. Having an error rate of 54% and effectiveness of 62.5% shows that approximately half of the user group needed help to solve the task and out of the total 8, 5 people were

Figure 4 - Showing the details section of Comprehension
able to solve the task correctly. T2, T4 and T5 had the same effectiveness, however showing different error rates meaning that the same amount of people was able to solve the task correctly but needed more or less help doing so (Figure 5).

5.3.2 Efficiency
Looking at the average time spent on each task, the lowest amount of time was spent on T1 followed by T2 and T4. T1 and T2 indicate that the variance in time spent was smaller compared to the rest of the tasks, who showed a greater deviation. On average, the users spent more time on T5 compared to the rest of the tasks, at the same time demonstrating the greatest deviation (Figure 6).

Moreover, T1 had the highest TBE rate (0.55 goals/second), indicating its simplicity compared to the other tasks. T3 and T5 had the lowest TBE rate, (0.02 goals/second), however not indicating much difference from the other tasks values (Figure 7).

5.3.3 Interview sessions
The interview sessions at the end of the evaluation provided qualitative data and feedback regarding the visualization tool and its purpose. All users expressed a positive attitude towards the concept of evaluating lectures continuously, instead of having a large evaluation at the end of the course. Many users stated that there were too few tools to handle feedback and monitor the quality of the course. Thus, they were all positive to use the tool again, although the opinion of what the visualization actually communicates differed. In some cases, the data was seen as too general and abstract, lacking qualitative information. A desire to have more specifics, such as the exact time in the lecture the feedback was connected to, was expressed. Others stated that the purpose of the visualization may not be to fully understand what flaws exist, but to give an indication of in what area or cohorts the problem may occur. Other users found the tool very helpful to understand more about their own course, identifying areas in the lecture which worked and did not work.

“I have learned so much more about my course since before we started talking.” - User A

“The data is too abstract for me. I need more details. What slide in the lecture was hard to understand?” - User E

“ [...] looking at graphs does not say that much, however it can give an indication of what was good or bad which you can dig deeper into with a focus group or similar.” - User D

“I don’t think the purpose of the tool is to understand exactly what needs to change, but for quick identification of a specific cohort. “- User B

Another aspect that was brought up was the concept of teacher evaluation. Even though the feedback collected was not to evaluate a certain teacher but merely the lecture and its content, one user expressed concerns about how teachers are under constant evaluation.

“Researchers and teachers probably feel stressed about being evaluated constantly. […] When you have a lecture, it is almost a sanctuary not having to be evaluated.” - User E

General positive comments regarding the used visual mappings was about the gauge, which the users thought was most intuitive and
easy to understand. It was said that it gave a quick indication of the dimension and combined with the color scale the users hardly had to interact with the chart to understand it. The mapping that raised the most questions and concerns was the area chart used to display the attitude dimension, which many thought was hard to read. A majority also expressed concerns about what attitude actually means and pointed out that the definition is much broader than what was evaluated in the survey. One scenario that occurred in the data set was that knowledge after the lecture had a lower value compared to before. In the visualization, many users read it as a knowledge decrease which caused a lot of confusion. The users found it hard to believe that the students’ knowledge level had decreased, which was not the case. Instead it indicated that the students did not learn much during the lecture, but the visual mapping seemed to cause another perception of the manner.

Two users raised a concern regarding how to motivate students to continuously provide feedback. In their experience, the response rate for course evaluations are critically low. If the idea is to have the same feedback system after each lecture, they expressed that a challenge may be to engage the students to fill in the same form over and over again. Some users also pointed out that an experienced lecturer can gather student feedback during the lecture and may not need any tools to do so. Asking questions and looking at the crowd usually does the trick.

“A good lecturer tries to gather feedback during the lecture and adapt to the next one” - User E

During the interview sessions, the current evaluation system was brought up at multiple occasions. The lecturers talked about the course evaluation surveys’ flaws, in terms of low response rates, and its strengths, in terms of the questions asked. The impression was that they wanted the same content but visualized in the tool developed in this study.

6. DISCUSSION

The question that this study aimed to answer was “How can a visualization tool be designed to enhance post-presentation analysis and increase insight of the lecture quality, when using ARS for collecting feedback amongst an audience in a teaching environment?”. To answer this, a visualization tool was created and evaluated with two primary goals set. First of all, does the tool provide insight that the users did not have before? Let us break down the results.

6.1 Usability

At a first glance, one can see that T1 was the most efficient and effective task. However, it should be mentioned that the component was located at the top left corner and displayed the answer in pure text. With that said, the answer to the task was very easy to find and understand, not having any potential of confusion, which was also the intention, as mentioned in section 2.4. The other tasks required the users to process more information and sometimes interact with the visualization, which was according to the planned increase in difficulty of the tasks.

T3 and T5 both had higher error rate and time spent on the tasks, which gave them a lower efficiency. What differed T5 from T3 was the fact that T5 was located in the details level, meaning the users had to go deeper down the visualization to find the answer. This might explain why the time spent was higher compared to other tasks. The error rate, on the other hand, indicate that the visual mapping might have been confusing. With T3 having the highest error rate, and the lowest completion rate, the visual mapping chosen points towards not being the best choice for the purpose. The attitude visualization was mistaken as a continuous graph, when the values were in fact discrete, which was the main cause of the confusion that arose during the evaluation. To limit the confusion, a different approach to the visual mapping should have been made, choosing a more suitable mapping for discrete values such as a bar chart, or a stacked area chart. Even though the evaluation highlights a few flaws and improvement areas in the visualization tool, the fact still remains that, apart from T3, the completion rate was between 87% and 100% on the remaining tasks.

Regarding the aspects learnability, guessability and EUP, the results are too insufficient to draw any conclusions. To clearly see the learnability, the study would have to include more task repetition than what was performed. Even though T1 and T7 were similar and resulted in no error rate, the time spent on the tasks differed more than expected. If the users had truly learned the interface, the time spent might have been more similar. Moreover, the standard deviation for the time aspect is relatively large, which means that the study would have benefited from a larger test group. The data from a larger set would more likely give smaller deviations, and thus provide more insight of the learnability aspect. The same goes for the EUP, for it accounts for the time spent on each task.

The results may point towards an unstable performance level amongst the users, but the fact still remains that the deviation span is relatively large on many tasks. It would, however, not be surprising that the users have not reached a stable performance level due to the fact that they were first time users. Regarding the guessability factor, it could be assumed that the users guessed their answer on T1, since the cost was relatively low. A more likely explanation to the outcome, as mentioned, is connected to the visual representation and placement of the component. Looking at the error rate, one could argue that the tasks T2-T6 had a lower guessability due to their existing error rate. Yet, accounting for the time spent is crucial to understand if the interface showed a low guessability or not which is, as mentioned, tricky due to the deviation that exists on most tasks.

6.2 Concept

To motivate students enough to provide sufficient feedback and visualize the data in the right manner for lecturers to provide insight, has shown to be a difficult balance to maintain. From this study, one can conclude that there is room for improvement to make the feedback process more efficient, but still ensuring useful insights to the lecturer for continuous course development. By developing the feedback tool, the hope was to speed up the process itself and ask for as little effort as possible from the students and lecturers. While the results were somewhat positive and indicates great potential, it appeared to be a bit contradictory. Many interviewed lecturers wanted the same content as their current evaluation survey, visualized in the tool, but still receive a high response rate.

Since part of the results of the literature review suggested that the response rate was highly dependent on the length of the evaluation, this is a complex factor. When speculating, the current evaluation form can appear as quite complex and demands a lot of effort from a student. Furthermore, the evaluation is sent out at the end of the course, resulting in lack of motivation to provide feedback due to the fact that it only helps the students who will take the course the following year. With this in mind, the low response rate might not be so surprising. The questions asked in the developed feedback tool were therefore seen as too general, lacking the details provided in the current evaluation form. One could argue that the lecturers
are used to receive feedback in a certain manner. Gathering feedback continuously with ARS and using information visualization to enable rapid processing of information are new ways of handling the feedback process. Thus, it will take time until the balance between effort and details has become sustainable for both parties and until the concept has reached its full potential. As stated in the results, one shall not disregard that the attitude towards the concept was positive, showing that the study has started a new way of thinking when it comes to the feedback cycle in educational institutions.

What should be taken into account when analyzing the results is the fact that only two users (A and B) were looking at data from their own lectures. The other six were looking at the data from Course B [Figure 2]. This could potentially have affected the users’ feelings about the information visualized, since they did not have any emotional connections to what was displayed. These users had to imagine themselves in the scenario, which possibly influenced their opinion of whether or not the visualization provided insight that they did not have before. At the same time, user A felt that the insight about the course had increased yet user B found that the insight was more on a general level, lacking details.

In line with Roxå and Mårtenssons study [17], some users pointed out that visual feedback and asking questions to the audience during a lecture might be sufficient to understand potential problem areas. It is therefore important to discuss whether or not to put this responsibility on the lecturers themselves. As visual feedback can be misunderstood, and questions asked might not receive answers from the whole crowd, it might be tricky to draw a conclusion about the general perception. Putting the responsibility on a lecturer that does not have the experience nor the ability to adapt content on the go makes it even more difficult, as previous researchers discuss on the topic of ARS in teaching environments [1], [11].

6.3 Method discussion
As already mentioned, the study would have benefited from having a larger test group when constructing the evaluation. The results would then have been easier to determine and to see a more definite picture of the usability. Another aspect regarding the evaluation is that the user satisfaction was not measured with a SUS as research suggests, but merely as an interview session with the SUS in mind. Measuring the user satisfaction with a standard could potentially provide more insight and other conclusions compared to the method that was chosen in this study. Adding the fact that not all users interacted with data from their own lecture, one can discuss if the tool was able to provide insight about the lecture quality or not. However, the usability aspect of the interface should not have been affected by this since everyone was first time users from the target group.

The scoring system to determine the error rate gave the users a “3” if they stated the wrong answer. Yet, in some cases, the user had misunderstood the task and thought they were answering it correctly. For example, a user stated the answer for the value “Agree” instead of “Strongly Agree” on T5, which resulted in a high error rate even though the correct view and question was found and read properly. A system for handling these situations would have provided more correct data regarding the effectiveness aspect of the interface.

Some visual mappings showed confusion when evaluating the interface. With that said, the method should have included even more design iterations to limit these sorts of errors, and thus have become a more mature interface to evaluate. The same goes for the feedback tool. Even though the questions were constructed using input from various lecturers as well as the existing evaluation form at KTH, some questions were seen as too narrow according to the lecturers who took part in the evaluation. Moreover, the lecturers expressed a desire for more qualitative data during the evaluation. Hence, the study would have benefited from doing iterations of the feedback tool as well.

6.4 Future work
So, what does this study mean for the future? Given the results, the visualization has the potential of providing useful information regarding the lecture quality. To fully take advantage of the potential, this study proposes two topics for exploration. First of all, to truly see the effect of using the tool, the study should span over a whole course. This would increase the ability to analyze and see the development along the course. Yet, as a user mentioned, this might increase the teachers’ feeling of constantly being evaluated even more. Having this in mind, it is also important to account for the students’ engagement of providing feedback. As research has shown, and as the study indicates with the amount of respondents in session two, it can be a tricky task to encourage students to provide the data [9], [21], especially if the setting would be after each lecture. This study did not focus on this particular aspect, and a suggestion is to include everyone involved in the feedback cycle when developing and evaluating the concept further. To make sure students feel motivated to provide feedback, future work should also include making the data public as a transparent process has shown to be more effective when asking for student feedback [9], [21]. Further improvements of the tool itself can include visualizing the potential drop off rate from the survey. This would indicate which questions in the feedback tool that students find too hard, too demanding or too irrelevant to respond to. Including a more open-ended question in the feedback tool can possibly enhance the experience and provide more insight. As a suggestion, this can be done by asking the students to state some words about the lecture and map the data to a word cloud in the visualization tool. Since the ARS used in this study offers the ability to create a whole presentation, it would be a great opportunity to take more advantage of the system during the future development. The visualization could then be implemented as the final presentation slide, showing a summary of the feedback gathered during and after the lecture.

To take this concept even further, another topic to explore is to gather real time feedback during the presentation, to ensure more precise data. This could be done by developing a minimalistic mobile interface to be used from an audience perspective. The interface could simply be a two-dimensional spectrum, with two axes spanning from variables chosen by the presenter. For instance, one axis could be “Understand” to “Confused” and the other “Interesting” to “Boring”. The audience would then be able to click on one of the areas and be mapped to the presenter slides, giving the presenter more precise data about the presentation. The feedback tool created in this study can then be used as a complement to provide a general overview and be used less frequently. Potential aspects to have in mind is to account for students’ attention during the presentation, as the interface might result in distraction amongst the audience.

7. CONCLUSION
The concept of evaluating lectures continuously for course development was found positive and have great potential. The visualization tool developed indicates that the information presented could in most cases be found and processed, with more or less help and
time doing so. Yet, the meaning of the information presented varied amongst the users. Many users wished for more details to fully understand the feedback, while some found the tool to be useful for quick identification of cohorts and general perception. Future work regarding the concept is discussed and suggested to reach its full potential when implementing the tool as a final slide in presentation settings.

ACKNOWLEDGEMENTS
First of all, I would like to thank the lecturers for letting me gather feedback data during their courses, as well as everyone who took part in the pre-study and evaluation. Secondly, a huge thank you to Maja Jakobsson and Niklas Ingvar at Mentimeter for being amazing supervisors, helping me along the way with their expertise within ARS and UX. Last, but definitely not least, a huge thanks to Björn Thuresson at KTH for providing guidance, support and hope during this period.

REFERENCES