Feasibility and Acceptance of chatbots embedded in healthcare curricula:

–

This document details the evaluation of each resource in terms of the feasibility and acceptance from the end-users. There was evidence of identifying the feasibility of such resources into formal training and studies exist on the acceptance of such resources, with promising results. However, all these studies defined the need for further research in the area until the use of chatbots in healthcare education became common. Furthermore, the creation process of CEPEH resources was significantly different and had improvements to current methods, due to the co-creation process, and use of low cost but effective technology.

Table of Contents

# Introduction

Personalised Healthcare Education is needed to meet growing demand and quality maintenance. There is a growing evidence around chatbots, namely machine conversation systems- these programs have the potential to change the way students learn and search for information.

Chatbots can quiz existing knowledge, enable higher student engagement with a learning task, or support higher-order cognitive activities. In large-scale learning scenarios with a hight student-to-lecturer ratio, chatbots can help tackle the issue of individualized student support and facilitate personalised learning. However, limited examples of chatbots in European Healthcare Curricula have been utilised to combine both the continuum of cognitive processes presented in Bloom’s taxonomy, with the idea that some repetitive tasks can be done with a chatbot- to provide greater access or to scale faculty time.

Thus, CEPEH strategic partnership has co-created open access chatbots utilising artificial intelligence, promoting innovative practices in digital era, by supporting current curricula and fostering open education.

CEPEH Erasmus+ strategic partnership aimed to co-design and implement new pedagogical approaches and, in particular, chatbots for European medical and nursing schools. CEPEH used use participatory design to engage stakeholders (students, healthcare workforce staff, lecturers, clinicians, etc.) in order to co-design effective chatbots and release them as open access resources. Through CEPEH, effective use of digital technologies and open education were be incorporated into healthcare curricula. This enabled students to increase their health and medical related skills through flexible learning.

CEPEH expected that students adopted this new digital pedagogy and improve their skills and competences through flexible personalised learning, while the teaching staff enhanced their e-learning tool co-creation competences and make use of co-design best practices and recommendations for use. It is also expected increased cooperation between the partners. Thus, in the long term, CEPEH expects to influence the development of medical and nursing curricula with this digital innovation, foster the quality of the future healthcare workforce and further improve international competitiveness of the partners’ healthcare curricula. This document details the evaluation of the resources created by the CEPEH team.

The evaluation specifically explored the feasibility and acceptance from the end-users. These end-users are learners in European healthcare higher education institutions.

There was firstly evidence for the need to identify the feasibility of chatbots and similar resources into formal education and training, with a further need to improve access to these types of learning resources. Of course, studies exist on the acceptance of chatbots, virtual patients, and many other healthcare applications, with promising results. However, through various limitations, we believed there was further research to be completed to accelerate the design, development, implementation, and evaluation processes. These have financial, stakeholder, time, and efficacy benefits. The creation process of CEPEH resources was significantly different to most in the literature, and this report highlights the approach of the CEPEH team towards enhancing personalised healthcare education can be achieved.

## Background

The working practices of CEPEH are aimed at maximizing efficacy of these chatbots as learning resources, and provided a sense of shared development and ownership from all stakeholders. The process normally begins with workshops in which the project is scoped and team building occurs. The CEPEH workshops involve the widest possible team of stakeholders including tutors, students, healthcare workers, learning technologists, health service users and carers- depending on the materials being created.

For readers who are interested in using these high quality digital resources please access them for free at CEPEH.EU

The next section will now present the evaluation of all CEPEH chatbot resources.

# 1 Method

## 1.1 Participants

This dataset had 14 males and 28 females, for a total of 168. It was a repeated measure design therefore`

There were 21 females from Greece, 3 from Cyprus, 2 from Sweden.

There were 7 males from Greece, 3 from Cyprus, 2 from Sweden.

## 1.2 Procedure

## 1.3 Design

## 1.4 Materials

### 1.4.1 System Usability Scale

The System Usability Scale (SUS) was used [10] and is a widely used and adopted usability questionnaire. It is popular due to its unbiased and agnostic properties, a non proprietary, and quick scale of 10 questions.

1. I think that I would like to use this system frequently.
2. I found the system unnecessarily complex.
3. I thought the system was easy to use.
4. I think that I would need the support of a technical person to be able to use this system.
5. I found the various functions in this system were well integrated.
6. I thought there was too much inconsistency in this system.
7. I would imagine that most people would learn to use this system very quickly.
8. I found the system very cumbersome to use.
9. I felt very confident using the system.
10. I needed to learn a lot of things before I could get going with this system.

The SUS was developed with a scoring system, in which the following should be performed: For each of the odd numbered questions, subtract 1 from the score. For each of the even numbered questions, subtract their value from 5. Add up these numbers to find the total score, then multiply this by 2.5. The result is a score out of 100 and can be compared against a determined average score of 68. Further, 80.3 or higher is excellent, and 51 or under suggests significant usability problems.

### 1.4.2 Computer Self-Efficacy Scale Tool

The 10 question CSEST was based on the 32-item questionnaire by Murphy, Coover, and Owen (1989). Participants were provided with the facilitator stating ’Imagine you have found a new technology product that you have previously not used. You believe this product will make your life better. It doesn’t matter specifically what this technology product does, only that it is intended to make your life easier and that you have never used it before. I could use the new technology…

1. If there was no one around to tell me what to do as I go
2. If I had never used a product like it before
3. If I had only the product manuals for reference
4. If I had seen someone else using it before trying it myself
5. If I could call someone for help if I got stuck
6. If someone else had helped me get started
7. If I had a lot of time to complete the job for which the product was provided
8. If I had just the built-in help facility for assistance
9. If someone showed me how to do it first
10. If I had used similar products before this one to do the same job

### 1.4.3 Unified Theory of Acceptance and Use of Technology

### 1.4.4 Technology Acceptance Model (TAM)

The Technology Acceptance Model (TAM) [1] was specifically developed with the primary aim of identifying the determinants involved in computer acceptance in general; secondly, to examine a variety of information technology usage behaviours; and thirdly, to provide a parsimonious theoretical explanatory model. TAM suggests that attitude would be a direct predictor of the intention to use technology, which in turn would predict the actual usage of the technology. The only modification to the nine sub-scales of the questionnaire consists of applying the items to the context of chatbots. All the items, except those measuring attitudes, utilize a seven-point Likert scale ranging from “strongly agree” to “strongly disagree” with a middle neutral point [2].

The nine sub-scales of the questionnaire:

• Ease of use of chatbots • Perceived usefulness of chatbots • Intention of use. • Attitude toward usage of chatbots. • Perception of personal efficacy to use a chatbot resource. • Perception of external control toward chatbots. • Anxiety toward chatbot use. • Intrinsic motivation to use chatbot resources. • Perceived costs of chatbots.

### 1.4.5 Qualitative Measure- Focus Group Discussions

Focus groups are a pervasive means of market research and provides credible acceptance evaluators regarding the penetration that a product or service will have on a target demographic. Focus groups are a form of qualitative research consisting of interviews or structured discussions, in which a group of people are asked about their perceptions, opinions, beliefs, and attitudes towards a product, service, concept, advertisement, idea, or packaging. Questions are asked in an interactive group setting where participants are free to talk with other group members. During this process, the researcher either takes notes or records the vital points he or she is getting from the group. Researchers select members of the focus group carefully for effective and authoritative responses. Relevant stakeholders, then, can use the information collected through focus groups to receive insights on a specific product, issue, or topic focus [7].

A series of short focus group sessions identified the feasibility of CEPEH resources for formal curricular integration. These sessions, spanning no more than 1-1.5 hours and consisting of no more than 5-7 persons each explored all axes of curricular integration such as accessibility in the classroom, use case scenarios, technology requirements for curricular integration etc. These axes were formalized by the research team, in each evaluation site, to consider the curricular details of each institution.

## 1.5 Analysis

# 2 Results

Have users in previous years shared the HELM Open RLO catalogue?

Well, we have so much data we haven’t looked through it all yet.

33,571 learners told us how they found out- each answer is different.

We are through about 10% of this data and will update weekly.

## 2.1 Pre Usage Results

# A tibble: 41 × 2  
 profession hours   
 <chr> <chr>   
 1 Student on a Healthcare course 1-4 hours  
 2 Student on a Healthcare course Never   
 3 Student on a Healthcare course Never   
 4 Student on a Healthcare course Never   
 5 College student 1-4 hours  
 6 Student on a Healthcare course 1-4 hours  
 7 Student on a Healthcare course Never   
 8 Student on a Healthcare course Never   
 9 Student on a Healthcare course Never   
10 Lecturer Never   
# … with 31 more rows

## 2.2 System Usability Scale (SUS) Scores

*Note= The amount of ‘agreement’ is defined as the addition of ‘Agree’ and ‘Strongly agree’ responses.*

The SUS score for all data was XXX. This is within, and above the median of, 68 – which is in the range of ‘average’ usability. This is good as the resources were early demonstrations and had reduced beta alpha testing due to time constraints- future updates can improve this metric.

After reversing the scores of the negatively worded questions (odd numbered questions), participants strongly agreed the system was not complex (XX% agreements), and they did not need assistance before use (XX% agreements). All remaining questions has the most frequently observed response as ‘agree’- the lowest amount of agreement (agree and strongly agree) was XX% for question X, which was explored further in the individual Partners’ analyses.

if you don’t like boring tables, here is the same data in a graph!

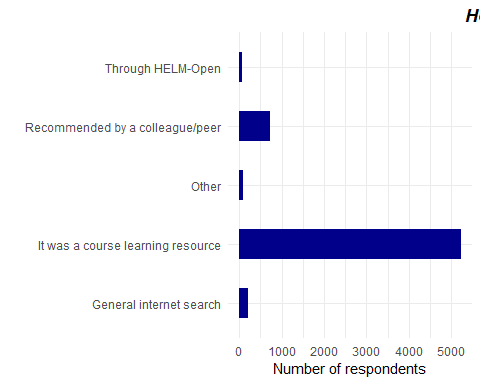
## 2.3 Technology Acceptance Model

The TAM had 3 sections (Ease of Use, Perceived Usefulness, and Intention of Use). Ease of Use results showed significant increases in Users’ usage with each Chatbot. Perceived Usefulness: There were not significant findings for the Perceived usefulness. The justification for this may be due to being early versions of applications with limited functionality and functions which can be difficult for user to experience the intended further range of features and learning exercises. Intention of Use: For users’ intentions to use within their course, the result of the Mann-Whitney U test was not significant, U = , z = , p = . in their intentions before use (m=xx, mode=xx) compared to after (m=xx, mode=x), however there was improvement therefore the chatbots may have more benefit than expected by students.

### 2.3.1 Course Learning, Recommendations, and more

The data showed that learners *strongly recommend* the RLO(s) they used to others, but how does this translate?

For the 10% of data we have, the figure below ’*How did you find out about the RLO you used’* shows 700 respondents were recommended from a friend, peer, tutor, or other.



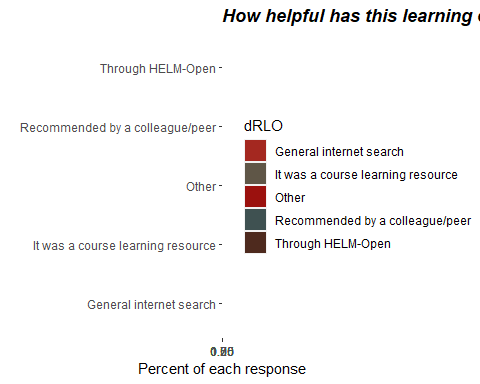
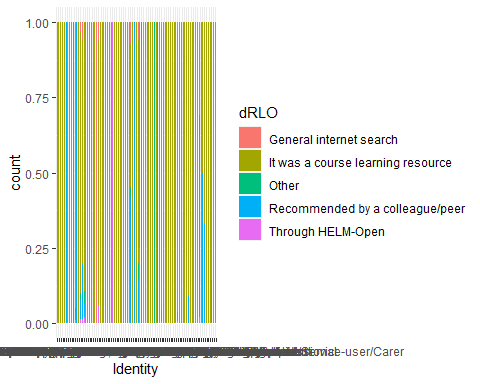
This figure also shows how more than 6000 respondents first used the RLOs as instructed by their tutors on their course.

and if you’re browsing the internet for information on a healthcare topic and come across <https://www.nottingham.ac.uk/helmopen/> :- You’re 1 of about 350 people finding about our resources from internet search. Hopefully that grows, but it seems social networking is the key to sharing these tools.

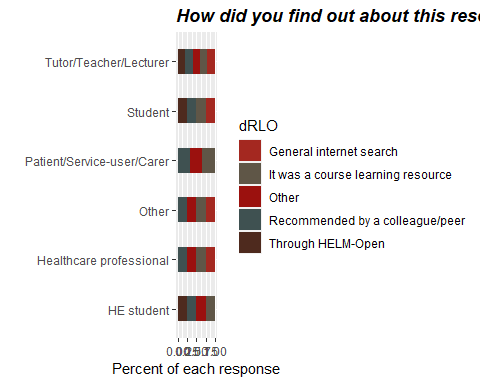
A random sample of other sources are: Twitter, Aim higher days, Barnardos ignite learning, and, well, ‘*a random Google photo*;’- our online presence seems to be in many places!

UP TO HERE1

# A tibble: 24 × 3  
# Groups: Identity, dRLO [24]  
 Identity dRLO n  
 <chr> <chr> <int>  
 1 HE student It was a course learning resource 80  
 2 HE student Other 2  
 3 HE student Recommended by a colleague/peer 4  
 4 HE student Through HELM-Open 1  
 5 Healthcare professional General internet search 7  
 6 Healthcare professional It was a course learning resource 131  
 7 Healthcare professional Other 3  
 8 Healthcare professional Recommended by a colleague/peer 10  
 9 Other General internet search 8  
10 Other It was a course learning resource 50  
# … with 14 more rows



# A tibble: 24 × 3  
# Groups: Identity, dRLO [24]  
 Identity dRLO n  
 <chr> <chr> <int>  
 1 HE student It was a course learning resource 80  
 2 HE student Other 2  
 3 HE student Recommended by a colleague/peer 4  
 4 HE student Through HELM-Open 1  
 5 Healthcare professional General internet search 7  
 6 Healthcare professional It was a course learning resource 131  
 7 Healthcare professional Other 3  
 8 Healthcare professional Recommended by a colleague/peer 10  
 9 Other General internet search 8  
10 Other It was a course learning resource 50  
# … with 14 more rows



### 2.3.2 Italics and bold

* *Italics* are done like \*this\* or \_this\_
* **Bold** is done like \*\*this\*\* or \_\_this\_\_
* ***Bold and italics*** is done like \*\*\*this\*\*\*, \_\_\_this\_\_\_, or (the most transparent solution, in my opinion) \*\*\_this\_\*\*

### 2.3.3 Hyperlinks

* [This is a hyperlink](https://www.google.com) created by writing the text you want turned into a clickable link in [square brackets followed by a](https://hyperlink-in-parentheses)

### 2.3.4 Footnotes

* Are created[[1]](#footnote-52) by writing either ^[my footnote text] for supplying the footnote content inline, or something like [^a-random-footnote-label] and supplying the text elsewhere in the format shown below [[2]](#footnote-53):

[^a-random-footnote-label]: This is a random test.

### 2.3.5 Comments

To write comments within your text that won’t actually be included in the output, use the same syntax as for writing comments in HTML. That is, .

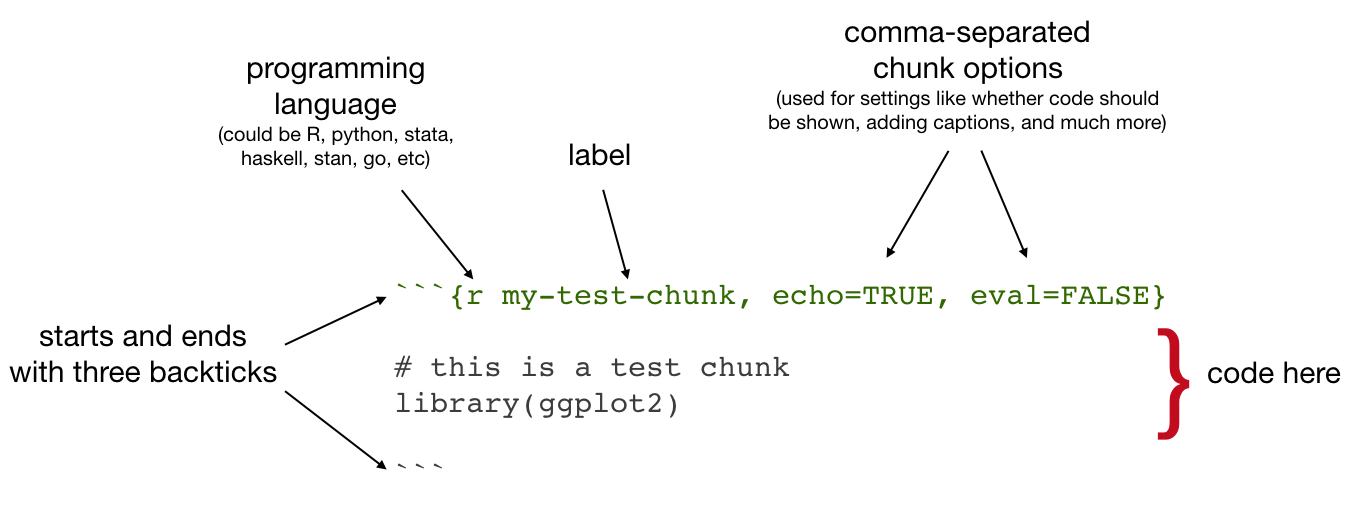


Figure 2.1: Code chunk syntax

Code chunks are also used for including images, with include\_graphics from the knitr package, as in Figure 2.2



Figure 2.2: Oxford logo

Useful chunk options for figures include:

* out.width (use with a percentage) for setting the image size
* if you’ve got an image that gets waaay to big in your output, it will be constrained to the page width by setting out.width = "100%"

#### Figure rotation

You can use the chunk option out.extra to rotate images.

The syntax is different for LaTeX and HTML, so for ease we might start by assigning the right string to a variable that depends on the format you’re outputting to:

Then you can reference that variable as the value of out.extra to rotate images, as in Figure 2.3.



Figure 2.3: Oxford logo, rotated

### 2.3.6 Including plots

Similarly, code chunks are used for including dynamically generated plots. You use ordinary code in R or other languages - Figure 2.4 shows a plot of the cars dataset of stopping distances for cars at various speeds (this dataset is built in to **R**).

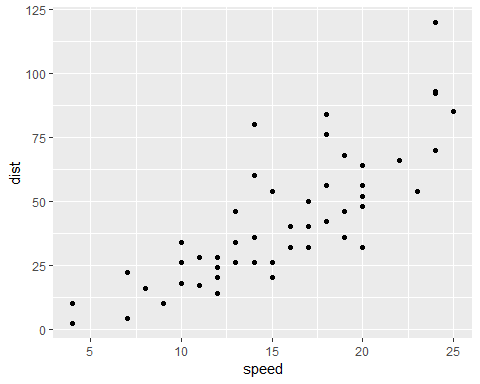


Figure 2.4: A ggplot of car stuff

Under the hood, plots are included in your document in the same way as images - when you build the book or knit a chapter, the plot is automatically generated from your code, saved as an image, then included into the output document.

### 2.3.7 Including tables

Tables are usually included with the kable function from the knitr package.

Table ?? shows the first rows of that cars data - read in your own data, then use this approach to automatically generate tables.

* Gotcha: when using [kable](https://www.rdocumentation.org/packages/knitr/versions/1.21/topics/kable), captions are set inside the kable function
* The kable package is often used with the [kableExtra](https://cran.r-project.org/web/packages/kableExtra/vignettes/awesome_table_in_html.html) package

### 2.3.8 Control positioning

One thing that may be annoying is the way *R Markdown* handles “floats” like tables and figures. In your PDF output, LaTeX will try to find the best place to put your object based on the text around it and until you’re really, truly done writing you should just leave it where it lies.

In general, you should allow LaTeX to do this, but if you really *really* need a figure to be positioned where you put in the document, then you can make LaTeX attempt to do this with the chunk option fig.pos="H", as in Figure 2.5:



Figure 2.5: An Oxford logo that LaTeX will try to place at this position in the text

As anyone who has tried to manually play around with the placement of figures in a Word document knows, this can have lots of side effects with extra spacing on other pages, etc. Therefore, it is not generally a good idea to do this - only do it when you really need to ensure that an image follows directly under text where you refer to it (in this document, I needed to do this for Figure 5.1 in section 5.1.4). For more details, read the relevant section of the [R Markdown Cookbook](https://bookdown.org/yihui/rmarkdown-cookbook/figure-placement.html).

## 2.4 Executable inline code

‘Inline code’ simply means inclusion of code inside text. The syntax for doing this is `r R\_CODE` For example, `r 4 + 4` will output 8 in your text.

You will usually use this in parts of your thesis where you report results - read in data or results in a code chunk, store things you want to report in a variable, then insert the value of that variable in your text. For example, we might assign the number of rows in the cars dataset to a variable:

We might then write:  
“In the cars dataset, we have `r num\_car\_observations` observations.”

Which would output:  
“In the cars dataset, we have 50 observations.”

## 2.5 Executable code in other languages than R

If you want to use other languages than R, such as Python, Julia C++, or SQL, see [the relevant section of the *R Markdown Cookbook*](https://bookdown.org/yihui/rmarkdown-cookbook/other-languages.html)

# 3 Training Event Results

## 3.1 CEPEH Training Event C1

The CEPEH training event C1 held at the premises of University of Nottingham aiming to prepare participants for the practical elements of co-creation and implementation of chatbots as an educational resource. It combined both theoretical and hands-on training. 15 participants were from RISE, AUTH, UoN.

Project managers of partners signposted the person involved, and relevant announcements were made though social media channels to the wider public. External to the project speakers were from University of Leeds, and Computer Science Department of University of Nottingham. It included academics, medical doctors, and researchers with focus both on clinical research and digital innovations in healthcare education and IT specialist/learning technologists 11.18 years of experiences (SD=7.2). A balance between male and female participants achieved.

# 4 Overall Training Events Evalaution

Participants were asked to highlight what they liked for each day and how each day can be improved. Findings are described below per day of the training event

Day 1  
The participants comment that they liked the design method for educational resources presented using a co-creation approach, they liked the interactions with other groups, and they liked the overview of existing chatbot resources of the partners. On the areas that can be improved, more media material were requested.

Day 2 Participants enjoyed the presentation from the invited speaker from another faculty of the University of Nottingham, the CEPEH recources presented and the storyboarding process. Participants highlighted that the participation of more clinicians in the event would be an added value in regards with the storyboarding process.

Day3 Participants liked the hands-on activities of the day also enjoyed the creativity of the groups on the online chatbot development tool. As an area of improvement, participants wanted more time on hands on sections.

## 4.1 CEPEH Training Event 2

**Pre-Training Event survey May 9th-13th 2022 Thessaloniki, Greece**

Twenty-six participants attended the Training Event, along with approximately 10 staff members. There were 21 undergraduate students and 5 postgraduate students, who completed the survey for a total of 26 responses. There were 86% of participants who stated they had not been to a similar event like the training event CEPEH facilitated. There were 90% of students who found the event schedule very organised, and 70% agreed most of the planned sessions were relevant to that interest with the remaining 30% not having enough experience to understand the context to determine if they are interested in the training event. There were 95% of students agreeing or strongly agreeing the training event location is great, the remaining person did not leave additional comments.

Table 1 suggested attendees had minimal intention to share their own ideas due to lack of previous experience of attending such events, or due to lack of knowledge on the area. However, most were interested in listening to other groups and hearing contextual cases in healthcare.

There were 77% of participants stated they were novices in experience with chatbots in healthcare and were attending to learn more. The remaining 23% (7 students) stated they were competent and had limited experience with chatbots in healthcare.

One day had several events regarding cybersecurity in healthcare. When asked before these events, 83% stated they were neutral or disagreed that they felt confident about their cybersecurity knowledge in healthcare. In addition, 80% stated they when neutral or disagreed that they felt they had strong cybersecurity safety in healthcare. Table 2 shows the main pre and post results suggesting a positive experience for more than 75% of attendees on all measures.

There were 90% (23) of students who heard about the event through a lecturer or a professor, the CEPEH newsletter (2), and 1 person was informed through the anatomy tutoring system at Karolinska Institute. Additionally, 60% suggested the training event to somebody else before the course started.

There were six individuals who stated neutral or disagree when asked if having issues on registration or finding the information for the event. This may have been due to being dependent on emails to receive the information, instead of a dedicated website where the information is available anytime.

As this was face-to-face, participants were asked about sufficient Covid-19 precautions in place at the facility, 94% agreed with sufficient precautions, two individuals stated no but did not give further information in the additional input box provided. In summary, most participants were undergraduate students with novice experience, happy with the training event location, felt the sessions were relevant to them, and most shared the event with their colleagues. The values of co-creation, chatbots in healthcare, and taking patient history were bestowed to students in an engaging and well-received manner. Notably, the highest ratings were for staff friendliness which is key to engagement and consistent interaction throughout the intense and long 5-day duration. The sessions were recorded there for the online recordings may be viewed with higher numbers over the subsequent weeks.

The usual way to include citations in an *R Markdown* document is to put references in a plain text file with the extension **.bib**, in **BibTex** format.[[3]](#footnote-84) Then reference the path to this file in **index.Rmd**’s YAML header with bibliography: example.bib.

Most reference managers can create a .bib file with you references automatically. However, the **by far** best reference manager to use with *R Markdown* is [Zotero](https://www.zotero.org) with the [Better BibTex plug-in](https://retorque.re/zotero-better-bibtex/), because the citr plugin for RStudio (see below) can read references directly from your Zotero library!

Here is an example of an entry in a **.bib** file:

@article{Shea2014,  
 author = {Shea, Nicholas and Boldt, Annika},  
 journal = {Trends in Cognitive Sciences},  
 pages = {186--193},  
 title = {{Supra-personal cognitive control}},  
 volume = {18},  
 year = {2014},  
 doi = {10.1016/j.tics.2014.01.006},  
}

In this entry highlighted section, ‘Shea2014’ is the **citation identifier**. To default way to cite an entry in your text is with this syntax: [@citation-identifier].

So I might cite some things ([Lottridge et al., 2012](#ref-Lottridge2012); [Mill, 1965 [1843]](#ref-Mill1965); [Shea et al., 2014](#ref-Shea2014)).

### 4.1.1 Appearance of citations and references section (pandoc)

By default, oxforddown lets [Pandoc](https://pandoc.org) handle how citations are inserted in your text and the references section. You can change the appearance of citations and references by specifying a CSL (Citation Style Language) file in the csl metadata field of **index.Rmd**. By default, oxforddown by the Americal Psychological Association (7th Edition), which is an author-year format.

With this style, a number of variations on the citation syntax are useful to know:

* Put author names outside the parenthesis
  + This: @Shea2014 says blah.
  + Becomes: Shea et al. ([2014](#ref-Shea2014)) says blah.
* Include only the citation-year (in parenthesis)
  + This: Shea et al. says blah [-@Shea2014]
  + Becomes: Shea et al. says blah ([2014](#ref-Shea2014))
* Add text and page or chapter references to the citation
  + This: [see @Shea2014, pp. 33-35; also @Wu2016, ch. 1]
  + Becomes: Blah blah (see [Shea et al., 2014, pp. 33–35](#ref-Shea2014); also [Wu, 2016](#ref-Wu2016), ch. 1).

If you want a numerical citation style instead, try csl: bibliography/transactions-on-computer-human-interaction.csl or just have a browse through the [Zotero Style Repository](https://www.zotero.org/styles) and look for one you like. For convenience, you can set the line spacing and the space between the bibliographic entries in the reference section directly from the YAML header in **index.Rmd**.

If you prefer to use biblatex or natbib to handle references, see [this chapter](#customising-citations).

### 4.1.2 Insert references easily with RStudio’s Visual Editor

For an easy way to insert citations, use RStudio’s [Visual Editor](https://rstudio.github.io/visual-markdown-editing/citations.html). Make sure you have the latest version of RStudio – the visual editor was originally really buggy, especially in relation to references, but as per v2022.02.0, it’s great!

## 4.2 Cross-referencing

We can make cross-references to **sections** within our document, as well as to **figures** (images and plots) and **tables**.

The general cross-referencing syntax is **\@ref(label)**

### 4.2.1 Section references

Headers are automatically assigned a reference label, which is the text in lower caps separated by dashes. For example, # My header is automatically given the label my-header. So # My header can be referenced with \@ref(my-section)

Remember what we wrote in section ???

We can also use **hyperlink syntax** and add # before the label, though this is only guaranteed to work properly in HTML output:

* So if we write Remember what we wrote up in [the previous section](#citations)?
* It becomes Remember what we wrote up in [the previous section](#citations)?

#### 4.2.1.1 Creating custom labels

It is a very good idea to create **custom labels** for our sections. This is because the automatically assigned labels will change when we change the titles of the sections - to avoid this, we can create the labels ourselves and leave them untouched if we change the section titles.

We create custom labels by adding {#label} after a header, e.g. # My section {#my-label}. See [our chapter title](#cites-and-refs) for an example. That was section 3.

### 4.2.2 Figure (image and plot) references

* To refer to figures (i.e. images and plots) use the syntax \@ref(fig:label)
* **GOTCHA**: Figures and tables must have captions if you wish to cross-reference them.

Let’s add an image: 

We refer to this image with \@ref(fig:captain). So Figure 4.1 is [this image](#fig:captain).

And in Figure 2.4 we saw a [cars plot](#fig:cars-plot).

### 4.2.3 Table references

* To refer to tables use the syntax \@ref(tab:label)

Let’s include a table:

We refer to this table with \@ref(tab:cars-table2). So Table ?? is [this table](#tab:cars-table2).

And in Table ?? we saw more or less [the same cars table](#tab:cars-table).

### 4.2.4 Including page numbers

Finally, in the PDF output we might also want to include the page number of a reference, so that it’s easy to find in physical printed output. LaTeX has a command for this, which looks like this: \pageref{fig/tab:label} (note: curly braces, not parentheses)

When we output to PDF, we can use raw LaTeX directly in our .Rmd files. So if we wanted to include the page of the cars plot we could write:

* This: Figure <a href="#fig:cars-plot">2.4</a> on page \pageref(fig:cars-plot)
* Becomes: Figure 2.4 on page

#### 4.2.4.1 Include page numbers only in PDF output

A problem here is that LaTeX commands don’t display in HTML output, so in the gitbook output we’d see simply “Figure 2.4 on page”.

One way to get around this is to use inline R code to insert the text, and use an ifelse statement to check the output format and then insert the appropriate text.

* So this: `r ifelse(knitr::is\_latex\_output(), "Figure \<a href="#fig:cars-plot">2.4</a> on page \\pageref{fig:cars-plot}", "")`
* Inserts this (check this on both PDF and gitbook):

Note that we need to escape the backslash with another backslash here to get the correct output.

## 4.3 Collaborative writing

Best practices for collaboration and change tracking when using R Markdown are still an open question. In the blog post [**One year to dissertate**](https://livefreeordichotomize.com/2018/09/14/one-year-to-dissertate/) by Lucy D’Agostino, which I highly recommend, the author notes that she knits .Rmd files to a word document, then uses the googledrive R package to send this to Google Drive for comments / revisions from co-authors, then incorporates Google Drive suggestions *by hand* into the .Rmd source files. This is a bit clunky, and there are ongoing discussions among the *R Markdown* developers about what the best way is to handle collaborative writing (see [issue #1463](https://github.com/rstudio/rmarkdown/issues/1463) on GitHub, where [CriticMarkup](http://criticmarkup.com) is among the suggestions).

For now, this is an open question in the community of R Markdown users. I often knit to a format that can easily be imported to Google Docs for comments, then go over suggested revisions and manually incorporate them back in to the .Rmd source files. For articles, I sometimes upload a near-final draft to [Overleaf](https://www.overleaf.com/), then collaboratively make final edits to the LaTeX file there. I suspect some great solution will be developed in the not-to-distant future, probably by the RStudio team.

## 4.4 Additional resources

* *R Markdown: The Definitive Guide* - <https://bookdown.org/yihui/rmarkdown/>
* *R for Data Science* - <https://r4ds.had.co.nz>

# 5 Discussion

Here is a (very large) table with all of the currently active RLOS.

Those results can be interpreted that the learning objectives of the training event was chosen appropriately for the diverse audience including clinicians, academics, researchers, and learning technologists/IT specialist resulting to a successful training event that enable participants to take the acquired knowledge back to their organisations in order to co-design and implement. As it was expected and can be depicted from self-confidence statements that some participants being very confident before the event, not all the objectives expected to be reached by everyone, since the training was targeting both technical and non-technical participants. However, on both average and individual matched responses participants self-statements showed that they improved their knowledge and understanding in using co-creation approaches to develop digital education resources and in designing and developing chatbots as educational resources.

## 5.1 Reach, Impact, and Qualatative analysis

Dealing with tables in LaTeX can be painful. This section explains the main tricks you need to make the pain go away.

(Note: if you are looking at the eBook version, you will not see much difference in this section, as it is only relevant for PDF output!)

### 5.1.1 Making your table pretty

When you use kable to create tables, you will almost certainly want to set the option booktabs = TRUE. This makes your table look a million times better:

Compare this to the default style, which looks terrible:

### 5.1.2 If your table is too wide

You might find that your table expands into the margins of the page, like the tables above. Fix this with the kable\_styling function from the [kableExtra](https://haozhu233.github.io/kableExtra/) package:

This scales down the table to fit the page width.

### 5.1.3 If your table is too long

If your table is too long to fit on a single page, set longtable = TRUE in the kable function to split the table across multiple pages.

When you do this, you’ll probably want to make the header repeat on new pages. Do this with the kable\_styling function from kableExtra:

Unfortunately, we cannot use the scale\_down option with a longtable. So if a longtable is too wide, you can either manually adjust the font size, or show the table in landscape layout. To adjust the font size, use kableExtra’s font\_size option:

To put the table in landscape mode, use kableExtra’s landscape function:

### 5.1.4 Max power: manually adjust the raw LaTeX output

For total flexibility, you can adjust the raw LaTeX output from kable/kableExtra that generates the table. Let us consider how we would do this for the example of adjusting the font size if our table is too wide: Latex has a bunch of standard commands that set an approximate font size, as shown below in Figure 5.1.



Figure 5.1: Font sizes in LaTeX

You could use these to manually adjust the font size in your longtable in two steps:

1. Wrap the longtable environment in, e.g., a scriptsize environment, by doing a string replacement in the output from kable/kableExtra
2. Add the attributes that make R Markdown understand that the table is a table (it seems R drops these when we do the string replacement)

# 6 Text Mining, Natural Language Processing, and Sentiment Analysis

## 6.1 1 Reading in texts

### 6.1.1 1.1 txt files

Here’s how you can read in one .txt file that is saved in the same location as this script (i.e. in the same folder on your computer):

If you want to read all files from a sub-folder, type the name of the folder followed by / and \* to ask R to read in all files in that folder:

### 6.1.2 1.3 Preparing data

* convert name to ID numbers with more descriptive labels

## 6.2 2 Tidy text

* One word per row, facilitates analysis
* Token: “a meaningful unit of text, most often a word, that we are interested in using for further analysis”

### 6.2.1 2.1 the unnest\_tokens function

* Easy to convert from full text to token per row with unnest\_tokens() Syntax: unnest\_tokens(df, newcol, oldcol)
* unnest\_tokens() automatically removes punctuation and converts to lowercase (unless you set to\_lower = FALSE)
* by default, tokens are set to words, but you can also use token = “characters”, “ngrams”, “sentences”, “lines”, “regex”, “paragraphs”, and even “tweets” (which will retain usernames, hashtags, and URLs)

## readtext object consisting of 2858 documents and 0 docvars.  
## # Description: df [2,858 × 3]  
## doc\_id word text   
## <fct> <chr> <chr>   
## 1 1 p1 "\"\"..."  
## 2 1 for "\"\"..."  
## 3 1 me "\"\"..."  
## 4 1 personally "\"\"..."  
## 5 1 it "\"\"..."  
## 6 1 was "\"\"..."  
## # … with 2,852 more rows

### 6.2.2 2.2 Removing non-alphanumeric characters

* str\_extract is used to get rid of non-alphanumeric characters (because we don’t want to count *word* separately from word)

### 6.2.3 2.3 Stop words

* Stop words: very common, “meaningless” function words like “the”, “of” and “to” – not usually important in an analysis (i.e. to find out that the most common word in two books you are comparing is “the”)
* tidytext has a built-in df called stop\_words for English
* remove these from your dataset with anti\_join

We can take a look:

## # A tibble: 1,149 × 2  
## word lexicon  
## <chr> <chr>   
## 1 a SMART   
## 2 a's SMART   
## 3 able SMART   
## 4 about SMART   
## 5 above SMART   
## 6 according SMART   
## 7 accordingly SMART   
## 8 across SMART   
## 9 actually SMART   
## 10 after SMART   
## # … with 1,139 more rows

## readtext object consisting of 821 documents and 0 docvars.  
## # Description: df [821 × 3]  
## doc\_id word text   
## <fct> <chr> <chr>   
## 1 1 personally "\"\"..."  
## 2 1 nice "\"\"..."  
## 3 1 week "\"\"..."  
## 4 1 ive "\"\"..."  
## 5 1 feeling "\"\"..."  
## 6 1 chatbots "\"\"..."  
## # … with 815 more rows

Define other stop words:

Break: Prepare your data with the steps above. 1) Unnest tokens, 2) Remove alpha-numeric characters, 3) Remove stopwords

## 6.3 3 Analysing frequencies

### 6.3.1 3.1 Find most frequent words

* Easily find frequent words using count()
* Data must be in tidy format (one token per line)
* sort = TRUE to show the most frequent words first

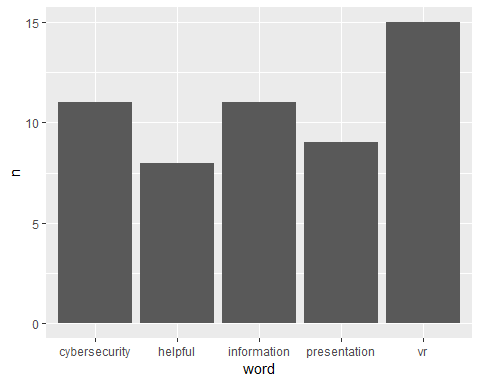
tidy\_books %>% count(word, sort = TRUE)

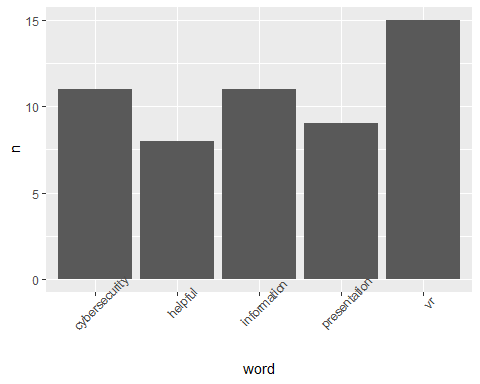
## # A tibble: 387 × 3  
## # Groups: doc\_id [1]  
## doc\_id word n  
## <fct> <chr> <int>  
## 1 1 vr 15  
## 2 1 cybersecurity 11  
## 3 1 information 11  
## 4 1 presentation 9  
## 5 1 helpful 8  
## 6 1 idea 7  
## 7 1 ideas 7  
## 8 1 lot 7  
## 9 1 workshop 7  
## 10 1 beginning 6  
## # … with 377 more rows

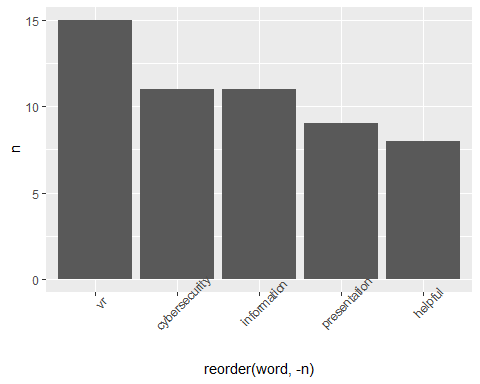
## # A tibble: 388 × 3  
## # Groups: doc\_id [1]  
## doc\_id word n  
## <fct> <chr> <int>  
## 1 1 vr 15  
## 2 1 cybersecurity 11  
## 3 1 information 11  
## 4 1 presentation 9  
## 5 1 helpful 8  
## 6 1 understand 8  
## 7 1 idea 7  
## 8 1 ideas 7  
## 9 1 lot 7  
## 10 1 workshop 7  
## # … with 378 more rows

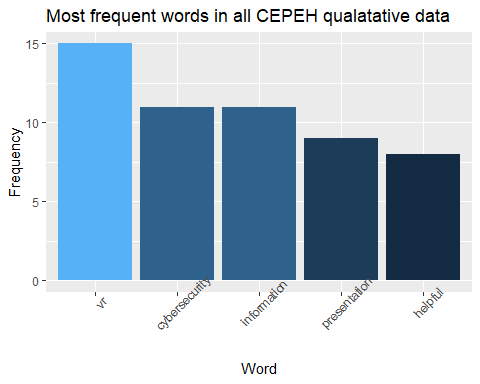
#### 6.3.1.1 Plotting word frequencies - bar graphs

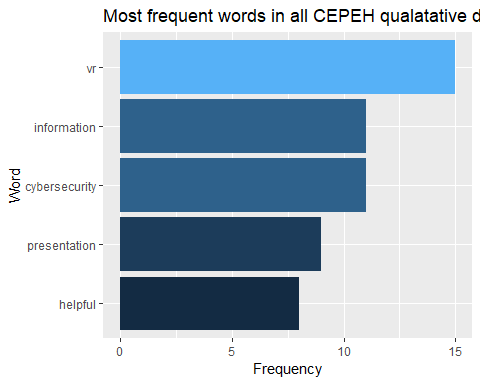
Bar graph of top words in CEPEHQ.

Basic graph: 

Readable labels: 

Descending order: 

Axis names and colors: 

Or: flip coordinate system to make more space for words 

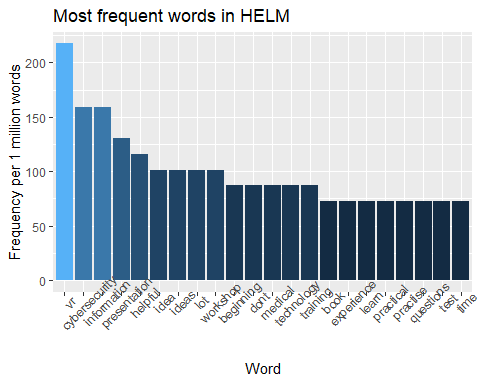
### 6.3.2 3.2 Normalised frequency

* when comparing the frequencies of words from different texts, they are commonly normalised
* convention in corpus linguistics: report the frequency per 1 million words
* for shorter texts: per 10,000 or per 100,000 words
* calculation: raw frequency \* 1,000,000 / total numbers in text

## # A tibble: 1 × 2  
## # Groups: doc\_id [1]  
## doc\_id `sum(n)`  
## <fct> <int>  
## 1 1 696

## # A tibble: 386 × 2  
## word pmw  
## <chr> <dbl>  
## 1 vr 217.   
## 2 cybersecurity 159.   
## 3 information 159.   
## 4 presentation 130.   
## 5 helpful 116.   
## 6 idea 101.   
## 7 ideas 101.   
## 8 lot 101.   
## 9 workshop 101.   
## 10 beginning 87.0  
## # … with 376 more rows

#### 6.3.2.1 Plotting normalised frequency

Now we can plot, for example, the 20 most frequent words (by pmw). 

### 6.3.3 3.3 Word clouds

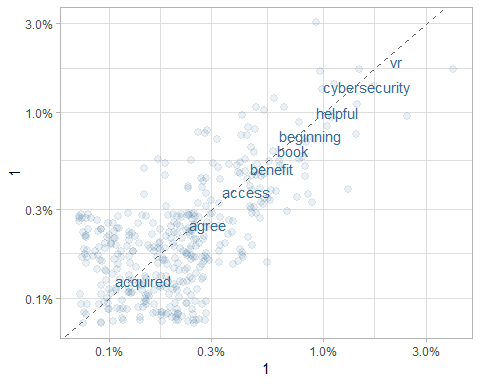
Let’s visualise the most frequent words in a word cloud. Here, the size indicates the frequency, with words that occur more often being displayed in a larger font size, but this can also be used to visualise e.g. normalised frequency (pmw) or length or anything else you pass to the freq = part of the command. 

## 6.4 4 Comparing the vocabulary of texts

Next, we’ll create two graphs to compare the vocabulary of our texts. First, we focus on Alice’s Adventures and Anderson’s CEPEHQ. The newly created comp\_2 data frame contains only the words and their frequencies in the two texts in two separate columns.

### 6.4.1 Comparing two texts

## # A tibble: 6 × 3  
## word pmw `1`  
## <chr> <dbl> <dbl>  
## 1 access 43.5 0.00435  
## 2 acquired 14.5 0.00145  
## 3 add 14.5 0.00145  
## 4 administrator 14.5 0.00145  
## 5 advance 14.5 0.00145  
## 6 advanced 14.5 0.00145

Now, we can plot the words. Their placement depends on the word frequencies. Additionally, colour coding shows how different the frequencies are - darker items are more similar in terms of their frequencies, lighter-coloured ones more frequent in one text compared to the other. We’ll discuss the interpretation in more detail once we’ve created the threeway comparison. 

# 7 Sentiment analysis

# install.packages(pdftools)  
# split PDF into pages stored in figures/sample-content/pdf\_embed\_example/split/  
# pdftools::pdf\_split("figures/sample-content/pdf\_embed\_example/Lyngs2020\_FB.pdf",  
# output = "figures/sample-content/pdf\_embed\_example/split/")  
  
# grab the pages  
pages <- list.files("figures/sample-content/pdf\_embed\_example/split", full.names = TRUE)  
  
# set how wide you want the inserted PDFs to be:   
# 1.0 is 100 per cent of the oxforddown PDF page width;  
# you may want to make it a bit bigger  
pdf\_width <- 1.2  
  
# for each PDF page, insert it nicely and  
# end with a page break  
cat(stringr::str\_c("\\newpage \\begin{center} \\makebox[\\linewidth][c]{\\includegraphics[width=", pdf\_width, "\\linewidth]{", pages, "}} \\end{center}"))

# Appendix

## More info

And here’s some other random info: the first paragraph after a chapter title or section head *shouldn’t be* indented, because indents are to tell the reader that you’re starting a new paragraph. Since that’s obvious after a chapter or section title, proper typesetting doesn’t add an indent there.

This paragraph, by contrast, *will* be indented as it should because it is not the first one after the ‘More info’ heading. All hail LaTeX. (If you’re reading the HTML version, you won’t see any indentation - have a look at the PDF version to understand what in the earth this section is babbling on about).

# (APPENDIX) Appendix

# 8 The First Appendix

This first appendix includes an R chunk that was hidden in the document (using echo = FALSE) to help with readibility:

**In 02-rmd-basics-code.Rmd**

library(tidyverse)  
knitr::include\_graphics("figures/sample-content/chunk-parts.png")

**And here’s another one from the same chapter, i.e. Chapter ??:**

knitr::include\_graphics("templates/download.png")

# 9 The Second Appendix, for Fun

# References

Lottridge, D., Marschner, E., Wang, E., Romanovsky, M., & Nass, C. (2012). Browser design impacts multitasking. *Proceedings of the Human Factors and Ergonomics Society 56th Annual Meeting*. <https://doi.org/10.1177/1071181312561289>

Mill, J. S. (1965 [1843]). *A system of logic, ratiocinative and inductive: Being a connected view of the principles of evidence and the methods of scientific investigation*. Longmans.

Shea, N., Boldt, A., Bang, D., Yeung, N., Heyes, C., & Frith, C. D. (2014). Supra-personal cognitive control and metacognition. *Trends in Cognitive Sciences*, *18*(4), 186–193. <https://doi.org/10.1016/j.tics.2014.01.006>

Wu, T. (2016). *The Attention Merchants: The Epic Scramble to Get Inside Our Heads*. Knopf Publishing Group.

1. my footnote text [↑](#footnote-ref-52)
2. This is a random test. [↑](#footnote-ref-53)
3. The bibliography can be in other formats as well, including EndNote (**.enl**) and RIS (**.ris**), see [rmarkdown.rstudio.com/authoring\_bibliographies\_and\_citations](https://rmarkdown.rstudio.com/authoring_bibliographies_and_citations.html). [↑](#footnote-ref-84)