Supporting computer science student reading through multimodal engagement interfaces

Matthew Pike, Kejia Shen, Dave Towey

University of Nottingham

UK | CHINA | MALAYSIA

University of Nottingham Ningbo China, 199 Taikang East Road, Ningbo, 315100, Zhejiang, China.

First published 2020

Supporting Computer Science Student Reading through Multimodal Engagement Interfaces

Matthew Pike, Kejia Shen, Dave Towey
School of Computer Science
University of Nottingham Ningbo China
Ningbo, China
{matthew.pike, zy18722, dave.towey}@nottingham.edu.cn

Abstract— While many computer science (CS) curricula are increasingly addressing a demand for more communicative and ethical graduates, reports of CS student difficulties with non-technical subjects, such as Professional Ethics, persist. These seem compounded for students learning through a second or foreign language. This paper explores the impact that multi-modal engagement interfaces can have on content comprehension. 30 participants of varying English language ability were asked to engage with four unrelated articles under four different conditions: baseline reading (C1); guided reading (sentence-by-sentence) (C2); audio/listening only (C3); and concurrent (multi-modal) presentation of C2 & C3 (C4). After each engagement, participants were asked to complete a comprehension test on the material that they had just encountered. A subjective survey evaluating the "comfort" and "engagement quality" of each interface was also completed after each interaction. Our results paint a complex picture with the guided reading interface (C2) producing both the best performance, and the poorest subjective evaluation from participants. This result aligns with existing findings identified in the field of reading education. The results highlight how varying language levels in participants impact subjective and performance metrics, suggesting how future interfaces may better support readers, according to their language ability or intended outcomes of reading.

Keywords— article comprehension, multi-modal interfaces, Simultaneous listening and reading, listening comprehension, L2 readers

I. INTRODUCTION

Many educators and employers agree that strong communication ability in computer science (CS) students is a crucial skill for graduates [1], [2], [3], [4]. However, most CS programs tend to focus on technical aspects of the subject, covering details such as software engineering, programming, algorithms and machine learning [5], [6]. Increasingly, however, professional bodies and accreditation boards are requiring that programs provide coverage of the social and ethical implications of computing [7]. At the University of Nottingham Ningbo China (UNNC), we are in a unique position to address these two different requirements. As the first Sino-foreign higher education institution (SfHEI), delivering a British education (including degree), through English, in China, we face the challenge of balancing the tensions between the UK accredited and designed curriculum in the context of predominantly Chinese students studying in China through the medium of English.

The current work is partly motivated by one of the author's experiences as module convenor teaching the module "Professional Ethics in Computing" (PEC) to final year CS students at UNNC. The stated education aims of the module are [8]:

> "To acquire the ability to recognise the professional, ethical, social and legal issues involved in the exploitation of computer technology, and be guided by the adoption of appropriate professional, ethical and legal practices. To apply these professional ethics perspectives to contemporary situations and to reflect on one's own experience and practice. To understand and be able to participate within the professional, social and legal framework within which one would have to operate as a professional."

The PEC module structure and contents are unlike any other that CS students at UNNC will have previously encountered in the CS program. PEC requires students to comprehend, articulate and develop arguments from long-form written articles (such as academic papers and textbooks). In delivering the module, the convener identified that students were challenged by the application of these unfamiliar skills, observing them struggling to articulate arguments, and lacking the necessary skills to identify key details and develop convincing presentations. A reflection on these conditions led to identification of the UNNC CS students not having English as their first language (L1) being a possible significant factor contributing to the challenges.

The need for our students to be able to comprehend long-form material led us to investigate the opportunity to develop novel engagement interfaces. While a body of work investigating the negative impact of digital interfaces on reading comprehension exists, less work has examined how technology can be used to improve content comprehension (CC). In this paper, we report on a preliminary investigation into the impact that different digital interfaces can have on CC. We evaluated four different web-based interfaces using the following different modalities: two visual-only; one audio-only; and one audio-visual. The aim of the study was to identify the impact these different modalities could have on CC. The results of this study indicate an interesting dynamic between learner preference and comprehension efficiency. We anticipate that the findings from this study will help inform the development of future engagement interfaces, especially in the context of non L1 English users, including our UNNC CS students.
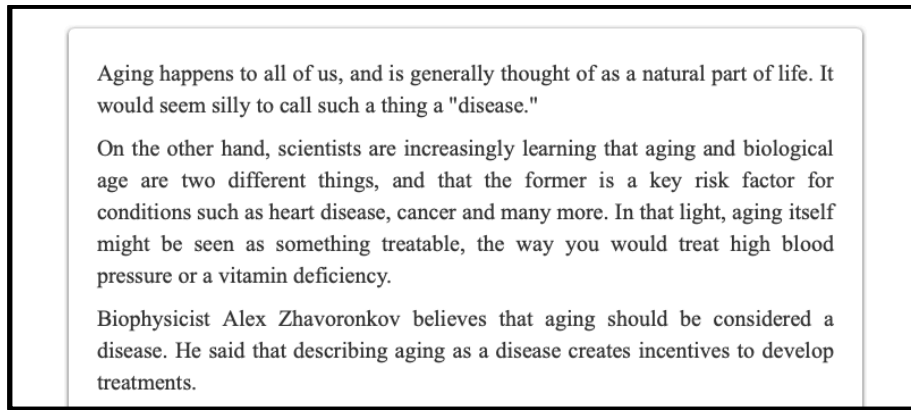
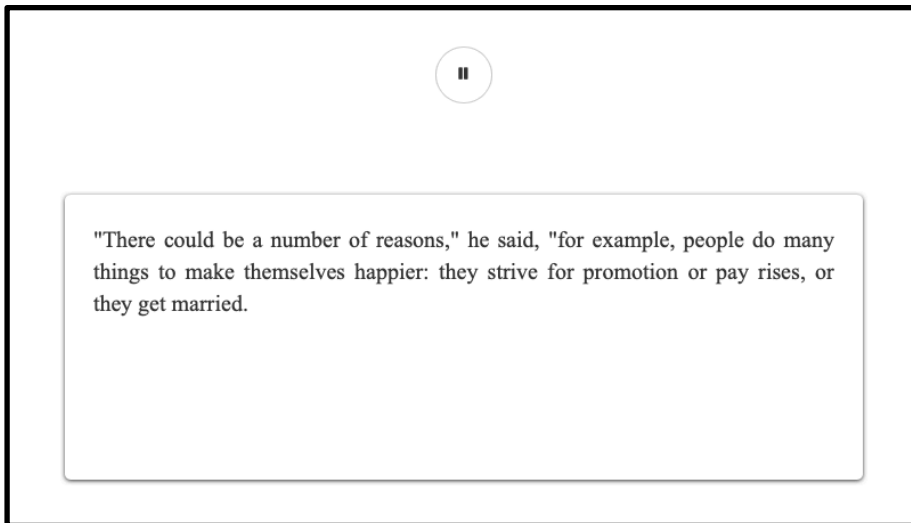Fig. 1. Baseline Reading Interface (C1)
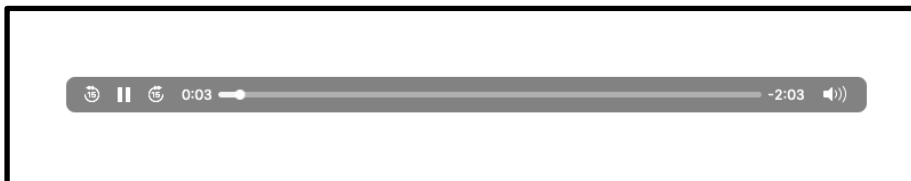

Fig. 2. Guided-Reading Interface (C2)
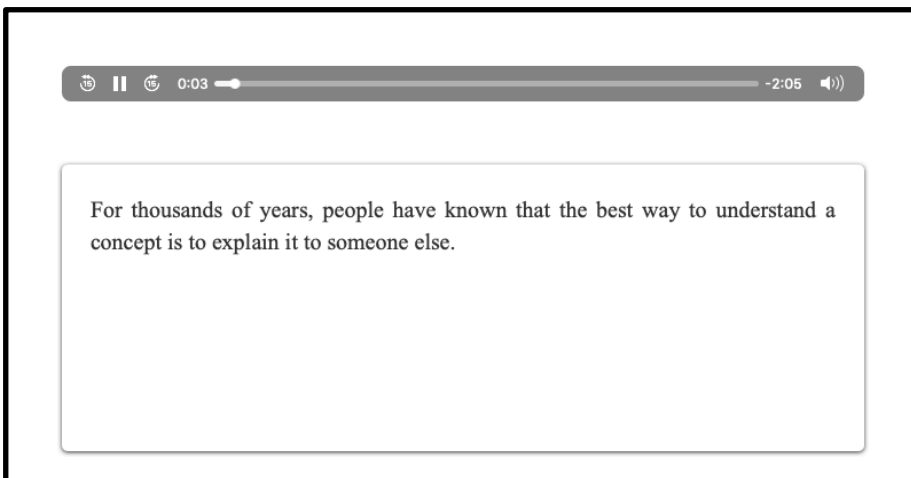

Fig. 3. Audio Only Interface (C3)

Fig. 4. Multi-Modal Engagement Interface (C4)

II. BACKGROUND

Several studies have investigated the effects of multi-modal interfaces in educational contexts, including early stage L1 education, and second-language (L2) acquisition [9], [10]. Hibbing & Rankin-Erickson have highlighted the impact that visual graphics can have in supporting developing or weak readers, with images assisting or confirming understanding of the written material [11]. Similarly, reading-while-listening (RWL) interfaces have been shown to support L2 language learners in reading skills development. Pellicer-Sanchez et al. conducted an eye-tracking based study to quantify the impact RWL interfaces have on L2 learners' processing of two multi-modal interfaces: reading & illustrations; and illustrations & RWL [9]. While results of their study found no significant differences in reading comprehension between the two conditions, the authors noted a behavioural change with the RWL interface: Participants were afforded additional time to process the presented illustrations as a result of the included auditory modality. In a study comparing RWL with listening only (LO) in L2 college students, Chang identified a 10% increase in reading comprehension with the RWL mode, with students reporting a strong preference for RWL [12]. Similarly, Brown et al. examined the acquisition of English vocabulary in L2 learners using three different input modes — reading, RWL, and LO — and found that while acquisition rates with reading and RWL were similar (45% and 48%), LO resulted in only 29% of the examined vocabulary being acquired [13].

III. STUDY DESIGN

A. Research Questions

The aim of this study was to investigate how different engagement interfaces affect CC in participants of varying English language abilities. The study was guided by the following research questions:

RQ1 How do different engagement interfaces impact on comprehension?

RQ2 How does English language proficiency affect preference and performance under different engagement modalities?

To answer these questions, we designed the study to examine the impact that interfaces with different modalities can have on participants of varying English language proficiency levels (specifically, either with English as a first language, L1; or without, L2). Table 1 summarizes details of the four engagement interfaces.

B. Hypotheses

We developed the following hypotheses to inform the investigation of the research questions:

HPMM (Multi-modal hypothesis) There will be no significant difference in comprehension between the baseline (C1) and the multi-modal interface (C4), for L2 participants.

HPAM (Audio modality hypothesis) There will be a significant difference in comprehension between audio-only (C3) and the multi-modal interface (C4), for L2 participants.

HPMM is based on findings by Brown et al. [12], who identified similar rates of vocabulary acquisition in reading and RWL conditions among L2 participants. HPAM is based on findings by Chang and Brown, who identified significant (negative) differences between LO and RWL conditions [12], [13].

C. Task

While seated in a comfortable, classroom environment, participants were asked to complete four engagement assignments using the four different interfaces (Table 1). Each interface presented the article using either different modalities or differing forms of visual presentation. The article difficulty and comprehension tests were based on materials from a standard undergraduate English test (described below).

TABLE I.   DETAILS OF EACH STUDY CONDITION

Condition ID	Name	Description	Modalities	Screenshot	Duration (s)
C1	Reading (Baseline)	Styled as a typical webpage, with all text being displayed as a block (wall-of-text).	Visual	Figure 1	Unlimited
C2	Guided Reading	An engagement interface were sentences are presented one-by-one. Sentences automatically advance.	Visual	Figure 2	120
C3	Audio	An audio recording of the article is played to the participant. No supporting text is provided.	Auditory	Figure 3	127
C4	Multi-Modal Engagement Interface	A novel combination of the interfaces provided in C2 & C3. Article content is concurrently presented using text and audio on a sentence-by-sentence basis.	Visual + Auditory	Figure 4	128

D. Participants

30 participants (17 male, 13 female) aged between 20 and 69 were recruited for the study. All participants were from UNNC, an SfHEI in the People's Republic of China (PRC), and represented both students and staff from the Faculty of Science and Engineering (FoSE). The study was approved by the University's Ethics Committee before any participants were recruited. Participants provided informed consent and received no compensation for their participation. All participants had normal or corrected vision. Given UNNC's unique situation as an SfHEI, the participants presented a broad range of English language proficiency levels. We categorised participants according to whether or not English was their L1 or L2, with seven being classified as L1, and 23 as L2.

E. Procedure

Participants were first introduced to the task that they would be completing during the study, but received no practice runs (for any of the interfaces). Participants used all four interfaces, counterbalanced using a Latin-Square rotation [17]. This approach ensures that each interface is experienced at each stage of the study by an equal number of users, reducing the impact of an ordering effect. Each engagement began with the researcher explaining the particular modality or presentation of the material. Participants were also notified as to the degree of control they had over each interface: control was limited to scrolling in the baseline (reading, C1); but all other interfaces were presented in an identical manner to all participants, with no personalization or customization permitted (of font size/style or audio pitch, for example).

Although the participants were given unlimited time to complete the article under the baseline condition (C1), they were encouraged to read at a natural pace. Because the delivery of C2, C3, and C4 was controlled through the interface, these all had a fixed time. Table 1 summarises the duration of each interface.

After completing all four engagements, participants were then asked to complete a subjective questionnaire (described below) evaluating the interfaces used in the study. They were also invited to rank each interface according to their personal preference.

F. Measurements

We collected two primary measurements during this study.

1) Reading Comprehension (Performance)

To measure the impact of each interface on comprehension, suitable, standardised English articles and comprehension tests were required. Our study used materials from the College English Test (CET), a national examination for assessing PRC undergraduate and postgraduate students' English language level [10]. The test consists of two levels: CET4 and CET6, with CET4 materials being used in this study. Each article was accompanied by five multiple choice comprehension questions. Participants' correct responses in the comprehension test served as the performance measure in the study.

2) Interface Survey (Evaluation)

Understanding participant preference and attitude towards each interface was the second metric used to evaluate the impact of the four interfaces. After the experiment, participants were asked to complete a questionnaire to subjectively evaluate each of the interfaces. In particular, participants were asked to rate each interface according to their "engagement quality," "ease & comfort," and "(perceived) impact on comprehension." Finally, participants were asked to rank each of the interfaces according to their overall preference.

G. Software

The interfaces used in the study were developed using standard web-technologies (HTML5, CSS, JavaScript) and were displayed using the latest version of Google Chrome (Version 74.0.3729+) 4.

The audio for the C3 and C4 interfaces was generated using high-quality text-to-speech (TTS) technology provided by IBM's Watson online service 5. C4 required that the text and audio be synchronised — because they were to be presented simultaneously: this was achieved using the Sakoe-Chiba Band Dynamic Time Warping (DTW) algorithm [18], and was automated using the Aeneas software tool [1].

IV. RESULTS

A. Performance

To determine which statistical functions to utilize, we first established if the data was normally distributed. This was achieved using a Shapiro-Wilk normality test which revealed that the study's data was not normally distributed (W = 0.92462, p = 0.00005) [14].
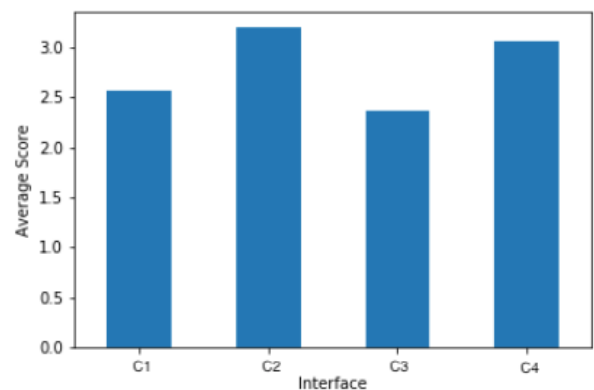

Fig. 5. Reading Comprehension (performance) across all participants.

To compare performance between conditions for non-normally distributed data, a Wilcoxon-Mann-Whitney rank-sum test was used [15]. Analysis of the performance results, for all participants, identified a significant difference between C1 and C2 (p = 0.05), indicating that guided reading significantly increases reading comprehension. No significant differences were found between C1 and C3 (p = 0.4) or between C1 and C4 (p = 0.1).

Analysis of L2 participant performance data identified a significant difference between conditions C3 and C4 (p < 0.05), in agreement with HPAM. Similarly, in agreement with HPMM, no significant differences were found between conditions C1 and C4 (p > 0.05).
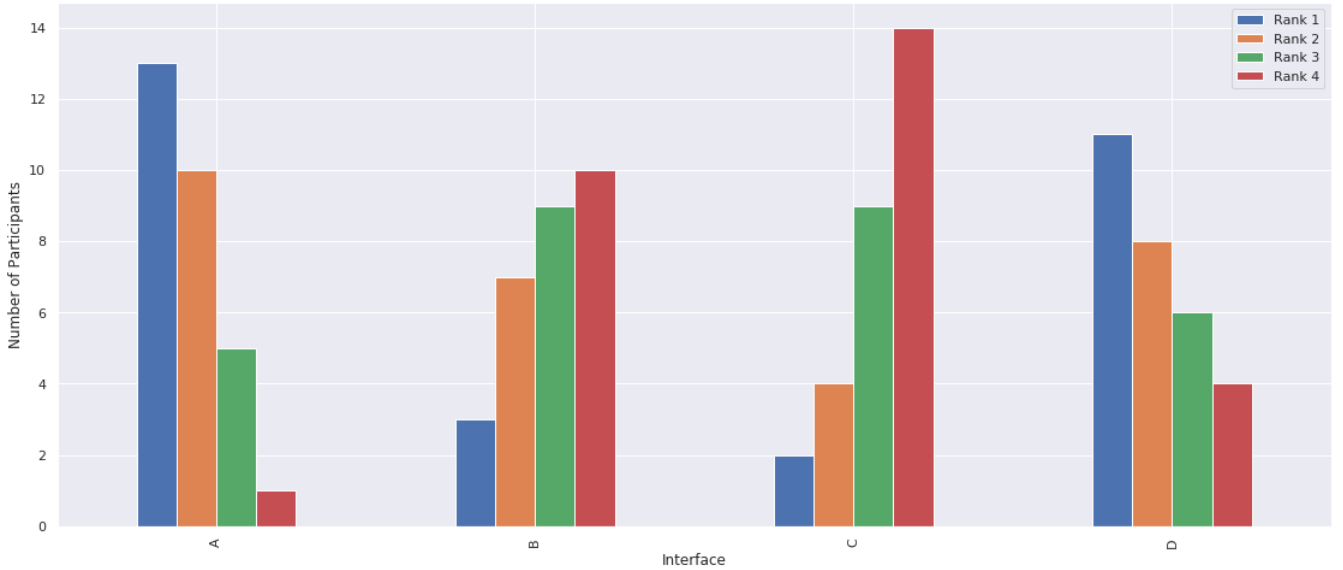
[1] Aeneas - https://www.readbeyond.it/aeneas/

Fig. 6. Subjective Rankings of each Engagement Interface.

B. Subjective Evaluation

Overall, the reading interface C1 was the most highly ranked among participants, with post-study feedback identifying the familiarity and feeling of control provided by the interface as being important factors. Participant10, for whom English was a second language ([P10, L2]), indicated a preference for the interface C1, stating, "I can read at my own pace." Similarly, [P12, L1] reported that, "It felt most natural and I could easily reread."

C2 and C3 were evaluated poorly by participants, despite the fact that C2 produced the best overall comprehension performance. In both C2 and C3, participants highlighted the lack of control over the presentation of article content: [P28, L2] said, "(I) dislike C2, a little bit too fast and (I feel that I) cannot control." Participants noted that reading progress was not displayed to readers in C2, leaving them uncertain as to how long they were required to concentrate for: [P23, L2] reported, "It would be better if the remaining time can be shown on interface C2." Equally, participants expressed unease in conditions C2 and C4, with [P12, L1] stating, "The feeling of not knowing when the sentence will disappear is unnerving if it was easier to return to previous sentences it would be better." Conversely, [P8, L2] stated that "...(C2) can keep my focus because it keeps showing (new) text". [P13, L2] highlighted the issue of de-synchronisation between an individual's natural pace of reading and the audio playback provided in interface C4: "The audio speed of C4 is slower than my reading speed. For example, when I finished reading a sentence, the audio is still playing, which might interrupt me." The issue of control of the pace of delivery for interfaces C2, C3 and C4 was highlighted by a number of participants, across English levels.

L2 participants rated C3 the lowest, mirroring the findings of Chang [2], with the pace of delivery being highlighted as a major factor: [P5, L2] suggests that "Speaking may need to be slower."

Participants were generally in favour of C4, which received the second highest ranking of all four interfaces, and was preferred to C3 — again mirroring the findings of Chang

[2]. Participants reported finding the combination of modalities to be "engaging": [P15, L2] said, "It engages both eyes and ears. This combination makes it easy to stay focused"; and [P27, L2] reported that "C4 helps to stay focused."

V. THREATS TO VALIDITY

Although our study included rotation of the tasks (to reduce the possibility of an ordering effect [16]), we did not rotate the articles between conditions. This could lead to a possibility that our findings were influenced by the relative complexity of the presented article, rather than solely the interface modality. We believe the impact of this threat is mitigated by the readings having been selected from a similar grading (CET-4). Future work will involve rotation of both the articles and conditions, and examining any observed impact this may have.

Additionally, participants received no renumeration for participation. The impact of this is unknown, but, given the lack of article relevance to the individual participants' context, this may represent a threat to the study's validity. This issue was highlighted by [P10, L2], who said, "The topics are not [very] interesting. I [lost] interest halfway through."

A small sample size, especially for L1 participants, prevented us from drawing larger conclusions. Nevertheless, as a preliminary study, we believe the results offer a clear direction for future work, which will seek to replicate this study, but on a larger scale.

VI. CONCLUSION

Our results paint a complex picture, with interface C2 delivering both the best performance, and receiving the poorest subjective evaluation from participants. One possible explanation for this result may be the anxiety caused by the lack of control in this condition, which may have caused participants to engage and focus more. Identifying the relationship between anxiety in engagement in interfaces and impact upon CC are avenues of future work that we intend to explore.

Interface C4 received favourable feedback from participants and resulted in the second highest comprehension in the study. This study did not identify why C4 was preferred to C2, with both conditions sharing the same visual appearance — it did, however, reveal that the inclusion of audio was important to participant comprehension, and appeared to impact on the degree to which they reported liking the interface. This is in contrast to C3, which had both the lowest comprehension performance, and was the least liked interface overall. We may (tentatively) conclude that audio does have a role to play in multi-modal comprehension interfaces — but not without a related visual modality.

In conclusion, our results indicate that language levels impact subjective and performance metrics. This provides some insight into how future engagement interfaces can be designed to support CC. Future work will attempt to further quantify the degree of this impact relative to individual learners' contexts (language proficiency, styles of learning, skillsets, etc.).

The exploration of different engagement interfaces performed in this study was motivated by an author's experience of teaching an ethics module to CS students. The application of a multi-modal engagement interface in long-form reading assignments, and its impact upon content comprehension in CS students, is an interesting avenue of future work that we plan to explore more deeply. This work has the potential to inform and augment future engagement interfaces for CS students.

Acknowledgment

References

[1] J.T. Havill and L.D. Ludwig, "Technically speaking: fostering the communication skills of computer science and mathematics students," in *Proceedings of the 38th SIGCSE technical symposium on Computer Science education - SIGCSE 07*, pp. 185–189, 2007.

[2] G. McDonald and M. McDonald, "Developing oral communication skills of computer science undergraduates," *ACM SIGCSE Bulletin*, vol. 25, no. 1, pp. 279–282, Jan. 1993.

[3] M.A. Fox and N. Hackerman (Eds), *Evaluating and Improving Undergraduate Teaching in Science, Technology, Engineering, and Mathematics,* The National Academies Press, Dec. 2001.

[4] L. Blume, R. Baecker, C. Collins, and A. Donohue, "A 'communication skills for computer scientists' course," *ACM SIGCSE Bulletin*, vol. 41, no. 3, p. 65, 2009.

[5] R.D. De Veaux et al., "Curriculum Guidelines for Undergraduate Programs in Data Science," *The Annual Review of Statistics and Its Application*, pp. 2.1–2.16, 2017.

[6] S.D. Brookes et al, *The Carnegie-Mellon Curriculum for Undergraduate Computer Science,* Springer Science & Business Media, 1985.

[7] M.J. Quinn, "On teaching computer ethics within a computer science department," *Science and Engineering Ethics*, vol. 12, no. 2, pp. 335–343, 2006.

[8] Nottingham, U. of. (n.d.). *University of Nottingham Module Catalogue*. Retrieved October 14, 2019, from http://modulecatalogue.nottingham.ac.uk.

[9] A. Pellicer-Sanchez, E. Tragant, K. Conklin, M. Rodgers, A. Llanes, and R. Serrano, "L2 reading and reading-while-listening in multimodal learning conditions: An eye-tracking study," *ELT Research Papers*, vol. 18, no.1, pp. 1-28, 2018.

[10] F. Serafini, "Reading Multimodal Texts: Perceptual, Structural and Ideological Perspectives," *Children's Literature in Education*, vol. 41, no. 2, pp. 85–104, Feb. 2010.

[11] A.N. Hibbing and J.L. Rankin-Erickson, "A picture is worth a thousand words: Using visual images to improve comprehension for middle school struggling readers.," *The Reading Teacher*, vol. 56, no. 8, pp. 758–770, May 2003.

[12] A.C.-S. Chang, "Gains to L2 listeners from reading while listening vs. listening only in comprehending short stories," *System*, vol. 37, no. 4, pp. 652–663, 2009.

[13] R. Brown, R. Waring, and S. Donkaewbua, "Incidental vocabulary acquisition from reading, reading-while-listening, and listening to stories," *Reading in a Foreign Language*, vol. 20, no. 2, pp. 136–163, Oct. 2008.

[14] S.S. Shapiro and M.B. Wilk, "An Analysis of Variance Test for Normality (Complete Samples)," *Biometrika*, vol. 52, no. 3/4, p. 591, 1965.

[15] H.B. Mann and D.R. Whitney, "On a Test of Whether one of Two Random Variables is Stochastically Larger than the Other," *The Annals of Mathematical Statistics*, vol. 18, no. 1, pp. 50–60, 1947.

[16] P.W.M. John, *Statistical design and analysis of experiments*. Philadelphia: Society for Industrial and Applied Mathematics, 1998.

[17] G.R. Loftus and M.E.J. Masson, "Using confidence intervals in within-subject designs," *Psychonomic Bulletin & Review*, vol. 1, no. 4, pp. 476–490, 1994.

[18] H. Sakoe and S. Chiba, "Dynamic Programming Algorithm Optimization for Spoken Word Recognition," *Readings in Speech Recognition*, pp. 159–165, 1990.