Author/s:
Khan, AA;Nawaz, S;Newn, J;Kelly, R;Lodge, J;Bailey, J;Velloso, E

Title:
To Type or to Speak? The Effect of Input Modality on Text Understanding During Note-Taking

Date:
2022-04-30

Citation:
Khan, A. A., Nawaz, S., Newn, J., Kelly, R., Lodge, J., Bailey, J. & Velloso, E. (2022). To Type or to Speak? The Effect of Input Modality on Text Understanding During Note-Taking. Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems, Association for Computing Machinery. https://doi.org/10.1145/3491102.3501974.

Persistent Link:
https://hdl.handle.net/11343/299907

To type or to speak? The effect of input modality on text understanding during note-taking

Anam Ahmad Khan
The University of Melbourne
Melbourne, VIC, Australia
anamk@student.unimelb.edu.au

Sadia Nawaz
The University of Melbourne
Melbourne, VIC, Australia
snawaz@unimelb.edu.au

Joshua Newn
The University of Melbourne
Melbourne, VIC, Australia
joshua.newn@unimelb.edu.au

Ryan M. Kelly
The University of Melbourne
Melbourne, VIC, Australia
ryan.kelly@unimelb.edu.au

Jason M. Lodge
The University of Queensland
Brisbane, QLD, Australia
jason.lodge@uq.edu.au

James Bailey
The University of Melbourne
Melbourne, VIC, Australia
baileyj@unimelb.edu.au

Eduardo Velloso
The University of Melbourne
Melbourne, VIC, Australia
eduardo.velloso@unimelb.edu.au

ABSTRACT

Though recent technological advances have enabled note-taking through different modalities (e.g., keyboard, digital ink, voice), there is still a lack of understanding of the effect of the modality choice on learning. In this paper, we compared two note-taking input modalities—*keyboard* and *voice*—to study their effects on participants' understanding of learning content. We conducted a study with 60 participants in which they were asked to take notes using voice or keyboard on two independent digital text passages while also making a judgment about their performance on an upcoming test. We built mixed-effects models to examine the effect of the note-taking modality on learners' text comprehension, the content of notes and their meta-comprehension judgement. Our findings suggest that taking notes using voice leads to a higher conceptual understanding of the text when compared to typing the notes. We also found that using voice triggers generative processes that result in learners taking more elaborate and comprehensive notes. The findings of the study imply that note-taking tools designed for digital learning environments could incorporate voice as an input modality to promote effective note-taking and higher conceptual understanding of the text.

CCS CONCEPTS

• **Human-centered computing** → **Empirical studies in HCI**; **Empirical studies in interaction design**.

CHI '22, April 29-May 5, 2022, New Orleans, LA, USA

ACM ISBN 978-1-4503-9157-3/22/04…$15.00
https://doi.org/10.1145/3491102.3501974

KEYWORDS

Digital learning, Input Modality, Voice notes, Note-taking tools

ACM Reference Format:
Anam Ahmad Khan, Sadia Nawaz, Joshua Newn, Ryan M. Kelly, Jason M. Lodge, James Bailey, and Eduardo Velloso. 2022. To type or to speak? The effect of input modality on text understanding during note-taking. In *CHI Conference on Human Factors in Computing Systems (CHI '22), April 29-May 5, 2022, New Orleans, LA, USA.* ACM, New York, NY, USA, 15 pages. https://doi.org/10.1145/3491102.3501974

1 INTRODUCTION

In a world where distractions abound, capturing learners' attention and encouraging deliberate reading is a significant challenge. Deliberate reading refers to the process of deeply engaging with the text to decode, understand, and construct meaning [1]. Such reading practices can enhance text comprehension and support a better understanding of the reading material [2]. One reading strategy that is effective in supporting learners' text comprehension is to encourage them to take notes as they read [37, 48, 52]. Note-taking is a means for learners to extract useful information from the content and reflect their understanding regarding the reading material [65]. As such, note-taking helps learners to focus their attention on the text, which in turn can enable them to effectively understand and encode the content into long-term memory [27].

Although traditional longhand (i.e., paper and pen) note-taking remains a common practice, the prevalence of digital text and the increasing availability of portable electronic devices have made digital note-taking more popular among learners. Studies have found digital note-taking to be ubiquitous, faster, and beneficial for learners' performance [38]. As compared to hand-written notes, digital notes offer several advantages, including the fact that they are easily searchable, editable, sharable, and more malleable [24]. Recognising the utility of digital note-taking, a wide range of tools have emerged that enable learners to record notes while reading digital texts [7, 63, 67]. Although the functionality available in these tools are relatively similar, they offer variety in terms of their input modality. For instance, some tools allow learners to compose

written notes using the keyboard or digital ink (i.e., pen based digital written input [63]), while others enable them to verbalise their note content using voice [67]. As the expressiveness of the input modality shapes what learners externalise [30], the use of different input modalities for recording notes may have a varying impact on learners' understanding of the digital text. Thus, to design tools that are conducive to the optimal note-taking experience, it is imperative to understand how the input modality for recording notes shapes learners' note-taking behaviour and impacts their overall understanding of the learning content.

Recent studies have investigated the effectiveness of the keyboard—which is the dominant note-taking modality in most current digital note-taking applications—on learners' understanding of digital content [18, 38, 44]. The findings of these studies suggest that compared to longhand note-taking, learners record more information in their notes when typing them [18]. However, typing notes tends to lead to transcription-oriented or verbatim copying of content which may not always boost learning performance [18, 38]. This finding emphasises the need to explore other input modalities for recording digital notes that can better instigate learners to practice effective note-taking practices, which, in turn, can help them better understand the learning material.

With the increased acceptance of voice as an input modality and advances in speech recognition technology, voice has emerged as a promising alternative note-taking modality. Speaking has been found to consume fewer cognitive resources than writing, thus permitting more attention to be directed to the text itself [10, 23]. Compared to writing, speaking to record information involves a higher level of social presence, which encourages learners to tailor their explanations for a potential audience [9, 34]. This behaviour could be particularly fruitful for learners to better understand the content, as it may encourage them to illustrate the subject matter by providing real-life examples or analogies [17, 26]. Lastly, the act of speaking is faster and requires less effort than writing, allowing momentary thoughts to be recorded before they are forgotten [55].

Recognising the potential of voice notes, a considerable number of note-taking tools such as *Sononcent*[1], *Luminant*[2], and *Notability*[3] have entered the market, enabling learners to not only record and index voice notes, but also to access voice recordings as transcripts at a later stage. The wide availability and adoption of these tools then raise questions regarding the impact of the input modality used for note-taking on the learning experience. In particular, how does note-taking using voice impacts learners' comprehension of text passages? Does a change in input modality from keyboard to voice influence learners' note-taking behaviours during digital reading? Answering these questions can help in understanding the impact of note-taking modalities on learners' understanding and inform the design of digital note-taking tools that can trigger higher-order comprehension while reading.

In this paper, we present an experiment in which 60 participants took notes using voice or keyboard on two independent pieces of digital text of comparable complexity. After the note-taking task, participants reported on their meta-comprehension level and completed a post-test, which included both factual and inference-based questions. This was followed by an interview that investigated participants' experience in adopting voice as an input modality for taking notes. We then built mixed-effects models to explore the effect of the note-taking modality on (1) learners' text comprehension, (2) the content of the notes, and (3) learners' ability to make an accurate judgement regarding their text comprehension. We further analysed the interview transcripts to identify the possible benefits and challenges of using voice notes for conducting learning activities.

Our findings highlight the potential benefits of using voice for note-taking. Though participants taking notes with voice performed similarly to those taking notes with the keyboard on factual questions, they scored significantly higher on inference-based questions, which require them to make inferences regarding the text and tested their conceptual understanding of the learning material. This finding can be explained by the substantial differences in the content of the notes—voice notes tended to include more elaborations and more idea units than typed notes. However, this advantage was not reflected in participants' meta-comprehension judgement, as we found no significant effect of the note-taking modality on their ability to predict their learning performance on a future test. Our findings suggest that though voice notes may not be suitable for every context (e.g. public spaces, traditional classrooms), private digital learning environments may incorporate voice as a complementary input modality for effective note-taking, which can assist learners in gaining a better conceptual understanding of the learning content. Given the fast speed and ubiquitous availability of voice input on contemporary electronic devices, our findings overall highlight a promising use case for speech interfaces in educational contexts.

2 RELATED WORK

Our research draws on prior work related to the significance of note-taking for text comprehension, its relation with meta-comprehension and the current rise and usage of voice note-taking.

2.1 Text Comprehension and Note-taking

Text comprehension refers to the process of constructing the meaning of the text being read. It requires learners to construct a mental representation of the text by integrating different parts of the text, making correct inferences regarding the text and linking the read information to their prior knowledge [33]. Such coherent representation of the text may aid learning as it expands the learners' existing knowledge and corrects any misconceptions in their knowledge base [60]. Previous research has suggested that the processes involved in comprehending while reading have a direct bearing on learning from the text [28, 33, 60]. This is because when learners process text, the informational elements and their relations get embedded in learners' mental representation of the text and update their prior knowledge, which may deepen learners' understanding of the text.

Prior research has identified note-taking as an effective strategy that can assist learners to effectively comprehend text and consequently aid learning [50]. From a cognitive psychology perspective, note-taking relates to information management where learners

[1]https://sonocent.com/
[2]https://luminantsoftware.com/apps/audionote-notepad-and-voice-recorder/
[3]https://www.gingerlabs.com/

need to comprehend and write down the information that is personally meaningful to them [51]. Research on note-taking suggests that the reasons for note-taking may differ between learners. Some engage in note-taking as a "process" that can promote their recall and aid their concentration—known as the *encoding effect*—while others take notes because of the resulting "product", which can help them in revising the material and keeping track of what was covered—known as the *external storage effect* [61]. Prior research has investigated the effects of note-taking on text comprehension [4, 43, 50] and found that note-taking encourages learners to select, organise, and integrate ideas presented in the text. For instance, Bahrami and Nozratzadeh investigated whether note-taking during reading improved learners' text comprehension. Their findings suggest that learners who take notes while reading were better able to integrate ideas between separate texts and performed better in post-reading tests than learners who did not take notes [4].

As digital technologies become more prevalent, learners are increasingly taking notes using portable digital devices, such as their laptops and tablets. To facilitate the easy recording and archiving of digital notes, many note-taking applications exist, using a wide range of input modalities such as keyboard, digital ink, and voice. However, the affordances of different input modalities vary substantially, which may directly impact learners' note-taking behaviour and, subsequently, their overall comprehension of learning material. Therefore, recent research efforts have investigated how the input modality used for recording notes impacts the content of the notes and learners' understanding of the text content. For instance, studies have investigated how notes typed using the keyboard impacts learners' performance by comparing them with handwritten notes [18, 38, 45]. The results of these studies showed that keyboard note-takers recorded more content in their notes in the form of words as they can capitalise on digital affordances such as spell checking and copying-and-pasting. However, longhand note-takers achieved higher performance on a post-reading test consisting of conceptual questions than keyboard note-takers. Overall, these findings suggest that note-taking input modality can impact both the content of notes generated by learners and their understanding of the text. Therefore, in this work, we investigate how voice as an note-taking input modality affects learners' content of notes and text comprehension.

2.2 Meta-comprehension and Note-taking

Meta-comprehension is defined as the learners' ability to judge their own understanding of the reading materials [14]. Prior research suggests that when learners judge their own text comprehension, they optimally decide how much time and cognitive resources are to be spent on revising segments of the text [13, 57]. For instance, during exam preparation, if learners can distinguish between well-learned material and less-learned material, they are more likely to spend time restudying less-learned material in order to compensate for the difference in learning. Therefore, accurate meta-comprehension judgment is critical to effective self-regulation of the reading task, which can ultimately result in a higher text comprehension [49, 64].

In order to assist learners in making an accurate judgment regarding their comprehension, previous research has suggested various learning strategies such as generating explanations and note-taking [49, 66]. For instance, studies suggest that learners who take notes during reading can more accurately monitor their understanding of the learning material and regulate their study tasks than those who did not indulge in note-taking [8, 49, 58]. This is because note-taking while text comprehension allows learners to evaluate the quality of cues used to retrieve and infer information about the text. For instance, while judging comprehension, learners may consider how successfully they recalled and retrieved information during note-taking. Resultantly, a text may receive a high rating of meta comprehension if the learner had been able to quickly retrieve information regarding the text during note-taking. However, if the learner struggled to retrieve information about the text during note-taking, they may give a low rating of comprehension to the text. These findings suggest that note-taking during reading can assist learners in judging their understanding of text more accurately.

Given the critical role of note-taking in regulating the meta-comprehension judgement of learners, it is important to examine how meta-comprehension is affected by the input modality used for note-taking during text comprehension tasks.

2.3 Voice Note-taking

With advances in multimedia sensing capabilities, researchers have been exploring modalities beyond the keyboard for digital note-taking. For instance, reading devices that support touch and pen input (e.g., eWriters) allow learners to leverage digital ink to engage in free-form note-taking. The use of digital ink to input notes enables learners to include illustrations (e.g., concept maps and drawings) in their notes which, in turn, is beneficial for learning [16]. Voice is another modality that has been gaining popularity. When compared to typing using the keyboard, voice can be faster to record and be more expressive [68]. Further, speaking involves a higher level of social presence which can instigate learners to address a potential audience, thereby elaborating more on the subject matter by providing real-life examples or analogies [26]. Such elaboration of ideas can, in turn, be helpful for learners to attain a better understanding of the subject matter [17]. Moreover, voice notes can also be transcribed to textual notes that, at a later stage, can make the notes review process easier.

With the current development of voice note-taking tools and extensions for cloud-based document viewers such as *Mote*[4] and *Kaizena*[5], learners can now easily record voice notes. For example, *Mote* can assist learners in taking voice notes on digital documents and in labelling the voice recordings. Further, the recorded voice can automatically be transcribed to text and later be shared with peers. Although tools supporting voice notes are starting to enter the market, there is a lack of evidence whether the experiences provided by voice note-taking complement the cognitive and comprehension goals of learners. To build applications that help learners practice effective note-taking, there is a need to gain insight into learners' perception of composing and reviewing voice notes. The findings from their perception could inform the design of voice note-taking tools that can better complement learners' note-taking experiences and help them understand the learning material. Therefore, in this paper, we investigate learners' perception of voice note-taking in

[4]https://www.justmote.me/
[5] https://www.kaizena.com/

order to highlight the potential benefits and limitations of adopting voice notes in educational settings.

In summary, in this research, we focused on comparing voice with the most commonly used digital note-taking modality—*keyboard*. We analyse the effects of these note-taking modalities on learners' text comprehension and meta-comprehension. Specifically, we aim to answer the following research questions:

- RQ1: How does the note-taking modality affect learners' text comprehension?
- RQ2: How does the note-taking modality impact the content of the notes made by the learners?
- RQ3: How does the note-taking modality influence learners' meta-comprehension judgements?
- RQ4: What are the potential benefits and obstacles of using voice notes in learning activities?

3 EXPERIMENT

To investigate the effect of the modality of note-taking on learners' understanding of leaning content, we conducted a within-subjects study in which learners practised note-taking with two different modalities while reading two digital articles. Through the study, we analyse (1) the effects of modality on text comprehension, (2) the content of the notes and (3) the meta-comprehension judgement in a reading task.

3.1 Participants

We recruited 60 participants aged between 21–35 (M = 27.04, SD = 4.39) via an online notice board from a large Australian university. 31 participants identified as men and 29 identified themselves as women. Participants were diverse in terms of their native language; with 16 English speakers, 3 Italian, 15 Mandarin, 4 Urdu, 5 Hindi, 7 Persian, 6 Sinhala, 2 Korean and 2 Tamil speakers. In terms of their education level: 29 participants were Masters students, and 31 were PhD students. As the reading task involved in the study related to the field of engineering, we ensured that none of the recruited participants were from an engineering discipline, reducing the potential familiarity with the subject matter.

3.2 Study Text

We used two scientific text passages from a study by Mayer et al. [41]. These passages have been widely used in various studies for measuring the impact of note-taking and accurate meta-comprehension judgment on learners' performance [49]. The first text passage comprised 792 words and was about different kinds of brakes and their functionality. The second passage discussed the mechanics of different types of pumps and had 850 words. We slightly modified the text in these passages so that both topics had a comparable reading difficulty level of 6.8, as measured by the Flesch-Kincaid Grade Level [31].

3.3 Pre-test and Post-test

For each study text passage, we conducted a pre-test to measure participants' prior knowledge, and a post-test to measure their text comprehension. The pre-test consisted of six multiple-choice questions (MCQs). There were four candidate answers and one "I don't know" (IDK) option for each question. Following Jacob et al.'s approach, participants' text comprehension was assessed on a post-test that included *factual* as well as *inference*-based MCQs [26]. For the post-test of each text passage, we developed 12 factual questions that required the knowledge of one idea unit presented in the passage (e.g., *Light-weight bicycles use which kind of brakes?*). Further, for the post-test of each text passage, we developed six inference-based questions that required the integration of two or more idea units and tested the conceptual understanding of the text (e.g., *A young lady in a village in India wants to pump water for cooking using a hand pump situated in her front yard. The water level in that area is 30 feet below the ground. What will happen when she moves the handle up and down?*). Similar to the pre-test, each post-test question had four candidate answers and one "I don't know" (IDK) option. All MCQs of the pre-test and post-test either received one point for the correct option or received zero points for incorrect or "I don't know" options.

3.4 Meta-comprehension Judgement

To investigate the impact of the note-taking modality on participants' meta-comprehension judgement, we asked participants to report their judgement regarding their perceived performance on a knowledge test after each study phase. For this, in line with prior literature in Educational Psychology [54], after completing each reading task, we asked participants to report their confidence level to correctly answer questions regarding the read text passage in an upcoming knowledge test. Specifically, participants were asked to answer the following question: *Please indicate how confident you are that you can correctly answer questions about the read text (0: no confidence; 100: absolute confidence).* A value of 0 indicated that the participants judged they definitely would not be able to answer questions correctly and a value of 100 indicated that the participants judged they definitely would be able to answer questions regarding the read text passage correctly.

3.5 Subjective Interest

Note-taking behaviour may be impacted by learners' interest in the topic as well as their perceived difficulty [3, 5]. For this reason, we asked participants to rate their interest and perceived difficulty of the text on a 7-point scale from (1) "*Not at all interesting*" to (7) "*Extremely interesting*".

3.6 Distraction Task

We further included a distraction task in our study to prevent against the *recency effect* of working memory [47]. The recency effect refers to learners' ability to better recall the most recently presented material. To control for such an effect, we selected a distraction task that was unrelated to the research questions being investigated and was to be completed between each note-taking task and its corresponding post-test. Following Mundt et al., we asked participants to count backwards from 200 to 0 in multiples of seven for two minutes [46]. In the second condition, participants performed the same distraction task but started counting from a different number (203).

3.7 Study Design and Experimental Setup

To explore the impact of note-taking modality, we designed a within-subjects study with note-taking modality as a within-subjects factor (keyboard vs. voice)—all participants were exposed to both conditions of note-taking modality while reading a different text passage (i.e. brakes or pumps) in each condition. The combinations of text passage and note-taking modality were counter-balanced across the sample. We facilitated the study entirely online through *Zoom* video-conferencing with *Qualtrics*[6] as an online survey tool alongside.

To collect the data for both conditions, we used *Google Docs* with the *Kaizena*[7] add-on enabled as an online platform for participants to read text passages and to take notes using keyboard or voice. *Kaizena* is available as a plug-in inside Google Docs and enables participants to highlight a text segment and click a button to record a voice note. Further, all participants also reported on their meta-comprehension judgment and answered the post-study test after each condition. Each study session was recorded with the written consent of participants and lasted approximately 90 minutes.

3.8 Procedure

Figure 1 illustrates the study procedure. Once the participant connected to the session, the researcher gave a brief introduction to the study and the URL link to the Qualtrics page. The participant was then asked to read through the *Plain Language Statement* and fill in the *Consent Form* on the page. Upon obtaining the participant's written consent, the researcher shared their screen and demonstrated the note-taking task using Google Docs and Kaizena for both typed and voice notes. The researcher then instructed the participant to share their screen and to try to create a few notes using both modalities in the same environment. The participant then shared their screen for the remainder of the session.

In each condition, participants first completed a pre-test that consisted of six MCQs. They then read the text passage and were asked to take notes that would help them in understanding the passage content and in answering the post-test questions using voice or keyboard, depending on the condition assigned. Before commencing the note-taking task, we informed participants that they would not be allowed to access or review the notes after the task or during the post-test. Participants were free to delete or re-record notes whenever they wanted. In line with prior studies investigating the effects of note-taking on text comprehension [4, 38, 45], we did not impose any time restrictions for the note-taking task. This design rationale was also instilled from the underlying aim of the study that focuses on investigating which note-taking modality allows learners to understand text more deeply. However, when time constrain is imposed, participants can drift from their natural behaviour of reading and note-taking to quickly complete the task in the given time frame. Therefore, to capture the naturalistic behaviour of learners while they read, comprehend and record notes without being pressured by any constraints, we allowed the participants to complete the task at their own pace.

The researcher stayed online with the participant during the entire study session to respond to any questions or provide any technical assistance. However, during the note-taking task, both the researcher and the participant had their cameras and microphones turned off to minimise any experimenter bias. Immediately after completing the note-taking task, we asked the participants to complete a short survey segment (meta-comprehension and subjective opinion). Then, participants performed the distraction task for two minutes before answering the post-test. Between the two conditions, participants were given the option to take a short break (3 minutes maximum).

The order of the modality and the text passage was counterbalanced between participants, such that each participant took notes with both note-taking modalities but on different text passages. For instance, if in the first learning condition, a participant read the text passage on *brakes* while taking notes using voice, then in the second learning condition, the same participant would read the text passage regarding *pumps* while taking notes using the keyboard. Before concluding the session, we conducted a short audio-recorded interview in which the participants were asked to comment on the usefulness and limitations of using voice note-taking for learning activities. We rewarded each participant with a $30 gift card for their time and contribution.

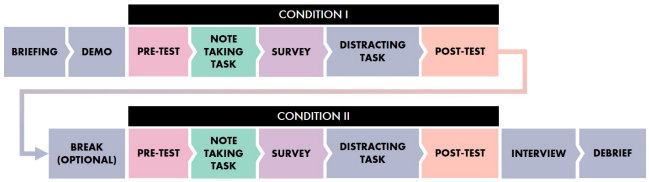

Figure 1: Study procedure adopted for exploring the effects of note-taking modality on participants' learning. The order of the text passage and modality was counterbalanced across the sample, such that each participant took notes with both modalities but with different text passages.

4 MEASURES

We investigate the effects of note-taking modality (voice vs. keyboard) on participants' learning based on their text comprehension, the content of notes generated, and their ability to make a meta-comprehension judgement on the post-test.

4.1 Text Comprehension

We operationalised learners' text comprehension in terms of their post-reading test score. The post-test consisted of 12 factual and 6 inference questions. The factual questions tested learners' ability to recall single idea units from the passage. In contrast, inference questions required participants to integrate multiple idea units across the text passage and tested their conceptual understanding of the passage.

4.2 Note Content

We analysed participants' notes based on the level of comprehensiveness and elaboration. For the analysis of voice notes, we hired a professional service to manually transcribe the notes.

[6] https://www.qualtrics.com/
[7] https://www.kaizena.com/

4.2.1 Note Comprehensiveness. Note comprehensiveness is measured by the *number of idea units* stated in the notes. An idea unit is a concept consisting of an argument and its relations [32]. For instance, the note: *"Electric brakes are used as an electromagnet"* consists of one idea unit, in which *"Electric brakes"* is an argument and *"are used as an electromagnet"* is a relation. To measure note comprehensiveness, we first split each of the notes into individual statements and then counted the number of idea units. Two independent raters marked the number of idea units for 25% of the notes. The overall inter-rater reliability as measured by Cohen's Kappa [42] was 0.95 – suggesting a good agreement between the two raters. Hence, the remaining notes were coded by only one of the raters.

4.2.2 Elaboration Level of Notes. Elaboration Level is indicated by the *number of elaborations* discussed in notes. Following Jacob et al., we considered an elaboration to be an idea unit that was not explicitly stated in the text passage but was either an example or an analogy from participants' personal experience [26]. For example, the following note contains two elaborations marked in italics as they were not stated in the text passage "Seems like there are two kinds of pumps, one is the dynamic pump, and other is positive displacement pump. *I would compare positive displacement to a dam*, and *I would compare dynamic displacement to something driven by a motor*". Two independent coders marked the number of elaborations for 25% of the notes. The overall inter-reliability rate was 0.90, indicating good agreement between the two raters. Hence, only one of the raters marked the remaining notes.

4.3 Meta-Comprehension

We operationalised meta-comprehension judgement in terms of meta-comprehension accuracy. This measure is defined as the absolute difference between participants' meta-comprehension judgment and their performance on the post-test [39]. A score of zero on meta-comprehension accuracy indicates an absolutely accurate judgment. The lower the value of this measure, the higher is the accuracy. The accuracy was calculated separately for each note-taking modality and for each participant.

5 RESULTS

We first report the effects of note-taking modality on learners' text comprehension. Then, we report the analysis of note content when different note-taking modalities are used. Lastly, we investigate the effects of note-taking modality on learners' meta-comprehension judgement.

5.1 Preliminary Analysis

To better understand how the choice of text passages might have affected our results, we analysed the perceived difficulty and participant interest about them. We conducted separate paired t-tests for this analysis, and report effect sizes with Cohen's d [12]. Table 1 shows the mean perceived difficulty and interest across two text passages reported by participants. We did not find a significant difference in the perceived difficulty of the text passages, $t_{59} = -.77, p = 0.44, CI = [-.25, .5], d = .09$. Similarly, we did not find a significant difference in participants' interest in the text passages, $t_{59} = 1.35, p = 0.17, CI = [-.13, .70], d = .17$. As such, we assume the two passages to be of comparable difficulty and perceived interest. Further, to observe whether the time taken for the note-taking task could have affected participants' post-study test performance, we calculated the Pearson correlation [19] between the time taken by the participants to complete the note-taking task and their score in the post-study test. Table 1 also reports the average time taken by the participants to complete the note-taking task for the two text passages. We observed that the participants' post-study test performance did not correlate with the time taken to read and record notes, as the correlation score between these two variables was week, $r = .001$ for brakes and $r = .01$ for pumps text passage.

Table 1: Mean with standard deviation (in parenthesis) of the perceived difficulty, interest and time taken (reported in minutes) by the participants to complete the note-taking task for the two text passages.

Measure	Brakes	Pumps
Difficulty (0-7)	4.32 (1.76)	4.04 (1.74)
Interest (0-7)	3.85 (1.64)	4.23 (1.76)
Time taken for the note-taking task	16.1 (3.8)	17 (2.34)

5.2 Effect on Text Comprehension

To investigate whether the note-taking modality affects participants' text comprehension, we built a linear mixed-effects model in *R* using the `lme4` package [6]. Linear mixed-effects models are used to represent dependent or clustered data which arise, for instance, when observations are collected over time for the same subject [21]. It is an extension of the general linear model and includes both fixed effects and random effects. Variables that are expected to have an influence on the dependent variable are specified as *fixed effects* in the model. In contrast, variables that are expected to have random influence on the dependent variable are specified as *random effects*. The resulting model estimates the impact of the fixed effects on the dependent variable, after accounting for any influence of random effects. The rationale for using a linear mixed-effects model over a simple linear regression model was that the former can handle correlated data and unequal variances more effectively [62]. As our design included several observations for the same participants, we built a mixed-effect model for the data to indicate both the random and fixed effects.

In this model, following prior studies on note-taking [29], we treated the participant ID as a random effect to control for participants' idiosyncrasies. Participants in our study read a separate text passage in each of the two experimental tasks. To avoid any possible learning effects caused due to the differences between the text passage read in the two experimental conditions, we also included the passage ID as a random effect in the model. Participants' test scores were treated as the dependent variable. We treated the note-taking modality (Voice vs. Keyboard), the question type (Factual vs. Inference), and the interaction between note-taking modality and question type as fixed effects. Further, we added the participants' pre-test score for each passage as a covariate in the model.

Table 2: Effects of model factors on predicting participant's test score. Model Formula = $Score \sim Modality * QuestionType + PriorKnowledge + (1|Participant) + (1|Passage)$, **where** *Score* = **post-test score,** *Modality* = **note-taking modality (Voice vs. Keyboard),** *QuestionType* = **type of question (Factual vs. Inference),** *PriorKnowledge* = **pre-test score,** *Participant* = **identification number of each participant and** *Passage* = **identification number of each passage, ***p < 0.001, *p < 0.05**

Variable	Estimate	SE	t	p
(Intercept)	16.52	4.3	5.15	2e-16***
Modality (Keyboard)	-2.46	3.09	-0.79	0.42
QuestionType (Inference)	-3.05	3.09	-0.98	0.32
PriorKnowledge	-0.18	0.10	-1.7	0.2
Modality:QuestionType	-3.40	4.01	-2.40	0.04*

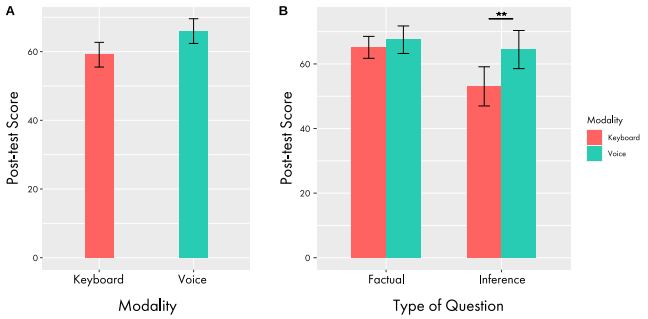

Figure 2: A: Overall Post-test scores across two conditions of note-taking modality. B: Post-test scores across two conditions of note-taking modality split on the question type. Error bars represent 95% CI, **p < 0.05

The final model, with its parameters, is shown in Table 2. To test for the assumptions of the linear mixed effect model, we visually inspected the residual plots, which did not reveal any apparent deviation from normality. Further, we checked for the presence of multi-collinearity between the fixed effects. All the predictors for the model had a variance inflation factor between 1 and 3, which is below the threshold to suggest multi-collinearity.

We measured the effect of note-taking modality on participants' text comprehension. We report this in terms of their overall post-test scores. The t-values and p-values of test results were reported on Satterthwaite's approximation of the effective degree of freedom [53]. Figure 2 presents a bar chart of participants' post-test scores across the two note-taking modalities.

The model for text comprehension was statistically significant ($\chi^2_4 = 27.29, p < 0.01$) and describes 9% of variance of participants' test scores (Marginal $R^2 = 0.09$, Conditional $R^2 = 0.29$). The results indicate that although the mean post-test score was higher for voice note-taking (mean = 65.2%, SD = 19.9) than keyboard note-taking (mean = 59.1%, SD = 19.1), the effect of modality in predicting learners' score was not significant across all questions ($p > 0.05$).

As expected, participants' scores in the inference-based questions (mean = 58.8%, SD= 23.8) were lower than in the factual questions (mean = 66.31%, SD = 14.9). However, this difference was not statistically significant ($p > 0.05$).

Finally, we observed a significant interaction effect between modality and question type ($t_{178} = -2.40, p = 0.04$). This result suggests that the note-taking modality has different effects depending on the type of question being asked. To explore the interaction between the modality and question type, we conducted pairwise comparisons from the contrasts between factors using the `lsmeans` package in R [36]. Table 3 shows the results of differences in the least square (LS) means with the associated confidence intervals (CI) and p-values for pairwise comparison between the two factors.

The results in Table 3 highlight two interesting insights. First, whereas we found no significant effect of the note-taking modality on test scores in factual questions, ($CI = [-5.6, 10.5], p > 0.05$), we did find a significant effect in inference-based questions ($CI = [3.4, 19.6], p < 0.01$). Second, whereas we did not observe a significant difference between factual and inference-based question scores when using voice notes ($CI = [-5.0, 11.1], p = 0.76$), we found a significant drop in performance in inference-based questions when using typed notes ($CI = [4.0, 20.1], p < 0.01$).

5.3 Effect on Note Content

Given that the affordances of the input modality can shape user behaviour [22], we expected to see differences in the content of the notes depending on the choice of modality. To explore the effect of the modality on the note content, we analysed the notes in terms of their comprehensiveness and and level of elaboration. *Comprehensiveness* reflects the number of idea units discussed in notes, while the *level of elaboration* refers to the number of elaborations of an example, analogies, and personal experiences discussed in notes. We built two separate linear mixed-effects models to investigate this effect. In both models, to control for the idiosyncrasies of the participants and the text passages, we treated the participants and passage IDs as random effects. We treated the note-taking modality (voice vs. keyboard) as a fixed effect for both models. The number of elaborations and the number of idea units were treated as the dependent variables. Figure 5 shows representative examples of the kinds of notes made by participants using each of the two note-taking modalities.

The model for note comprehensiveness was statistically significant ($\chi^2_1 = 24.9, p < 0.001$) and described 13% of variance (marginal $R^2 = .13$, conditional $R^2 = .45$). The main effect of modality on note comprehensiveness was significant ($d = 1.63, p < 0.001$). The model is shown in Table 4. The model suggests that when taking voice notes, learners include more idea units than when typing notes by using a keyboard (see Figure 3).

The model for the level of elaboration was also statistically significant ($\chi^2_1 = 14.28, p < 0.001$) and described 11% of variance (marginal $R^2 = .11$, conditional $R^2 = .13$). The main effect of modality on elaborations was significant ($d = 0.69, p < 0.001$). The model is shown in Table 4. These results suggest that learners tend to elaborate more on the content when taking voice notes than when typing notes (see Figure 3).

These findings suggest a potential explanation for the positive effect of the use of voice notes on participants' performance on inference-based questions—because voice notes lead learners to

Table 3: Pairwise contrast of interaction between note-taking modality and question type, $^{**}p < 0.01$, $^{***}p < 0.001$

Contrast	Estimate	SE	*t*	lower CI	upper CI	*p*
Voice Factual - Keyboard Factual	2.46	3.12	0.78	-5.63	10.56	0.85
Voice Factual - Voice Inference	3.05	3.12	0.98	-5.04	11.15	0.76
Keyboard Factual - Keyboard Inference	12.08	3.12	3.87	3.99	20.18	0.000***
Voice Inference - Keyboard Inference	11.49	3.12	3.68	3.40	19.59	0.001**

Table 4: Effect of modality on predicting note comprehensiveness, indicated by number of idea units and level of elaboration indicated by the number of elaborations, $^{**}p< 0.01$, $^{***}p < 0.001$

Note Comprehensiveness				
Variable	Estimate	SE	*t*	*p*
Intercept	23.76	3.0	7.89	2.84e-12***
Modality	18.6	3.38	5.5	0.000***
Elaboration level of notes				
Intercept	1.45	0.46	3.14	0.002**
Modality	2.53	0.64	3.93	0.000***

increased elaboration and comprehensiveness in their notes by discussing more idea units together, they lead to improved comprehension of the text. To test this hypothesis, we conducted a mediation analysis by taking the number of elaborations and the number of idea units as mediators. Note-taking modality was the dummy-coded predictor variable, and learners' performance on the post-test inference-based questions was taken as a dependent variable. To conduct the mediation analysis, we applied Hayes's bootstrapping methodology [25] with 10,000 simulations. The result of the mediation analysis is shown in Figure 4. We report the results of the significant effect of the mediation analysis using the estimates, SE, and 95% CI of the estimates. The CI estimates that do not include zero are considered significant.

The findings of the mediation analysis revealed a significant direct effect of the note-taking modality on the learning performance (estimate = 9.03, 95% CI [.05, 18.01]). Additionally, we found that modality had a significant indirect effect of 4.77, SE= 2.21, 95% CI [.82, 9.95], on learners' learning performance mediated via the number of idea units. On the other hand, regarding the number of elaborations, the analysis revealed a non-significant indirect effect of 2.41, SE= 1.75, 95% CI [-6.06, 0.88] of modality on learners' learning performance. The results of the mediation analysis suggest that voice leads learners to elaborate more on the information read in the text passage and to make more comprehensive notes, indicated by the number of idea units. However, only the note comprehensiveness mediated the effect of note-taking modality on learning performance for inference-based questions.

5.4 Effect on Meta-comprehension Accuracy

The meta-comprehension accuracy was operationalised as the absolute deviation between learners' judgement and their actual test performance on the entire post-test. The mean meta-comprehension for the voice note-taking condition was 17.25 (SD = 11.93), and for keyboard note-taking condition was 15.09 (SD=11.9). To measure the effect of note-taking modality on learner's meta-comprehension judgement, we again built a linear mixed effect with the absolute meta-comprehension accuracy as a predictive variable, the participant and passage IDs as random effects, and the note-taking modality as a fixed effect. The model for meta-comprehension accuracy

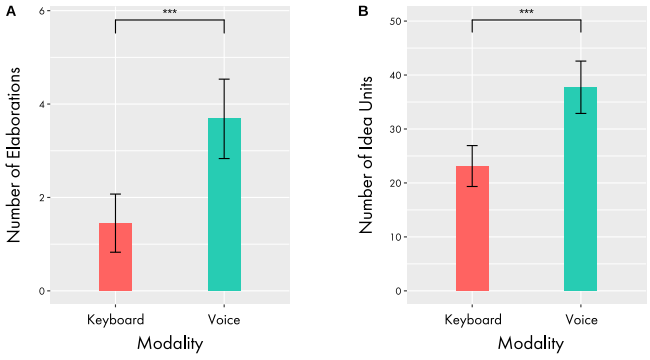

Figure 3: A: Total number of elaborations made by participants using voice and keyboard note-taking modality. B: Total number of idea units discussed by participants using voice and the keyboard note-taking modality. Error bars represent 95% CI, $^{***}p < 0.001$

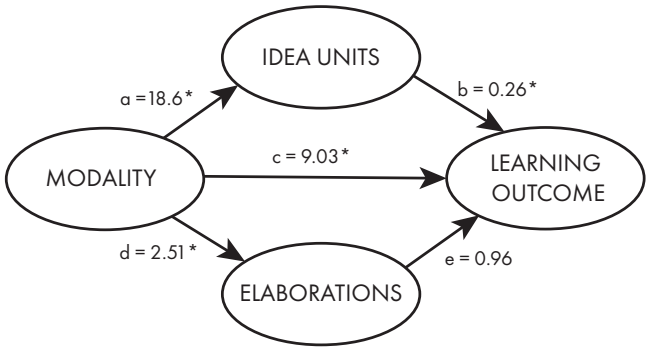

Figure 4: The model showing the effects of note taking modality on learners' learning performance on inference based question, with the number of idea units and number of elaborations as mediator. Values on paths are regression coefficients. * p < 0.05

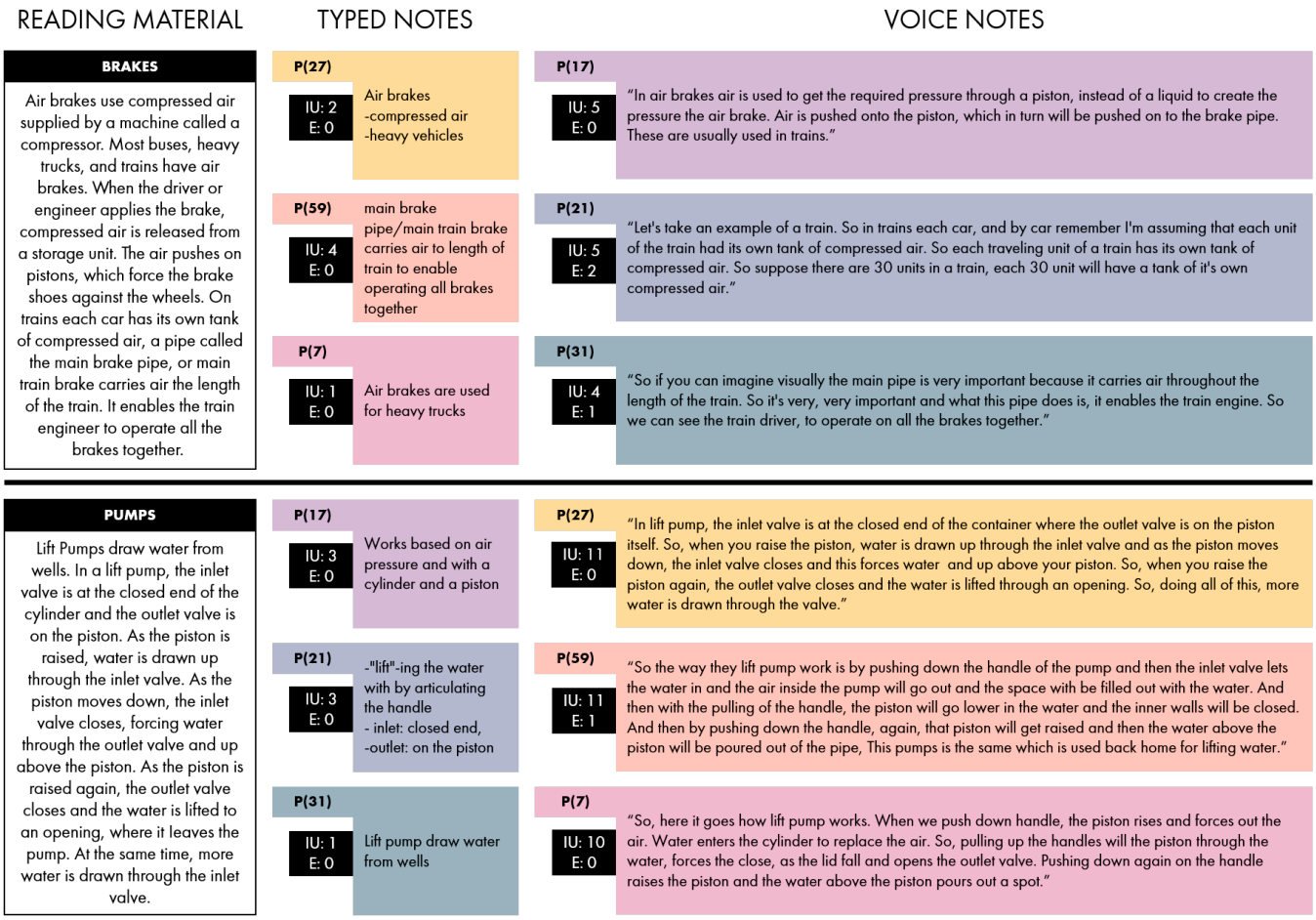

Figure 5: Typed and voice notes made by participants (ID given in parenthesis) for a paragraph of the two text passages. IU indicates the number of idea units discussed in each note. E indicates the number of elaborations made in each note. Notes are color coded based on participants' ID

was not statistically significant ($\chi^2_1 = 1.07, p = 0.29$) and the main effect of modality on learner's meta-comprehension judgement was not significant ($t_{60} = -1.43, p = 0.30$). The results indicate that even though the mean meta-comprehension accuracy reported by learners using voice as a note-taking modality was higher than the keyboard, we found no significant effect of the note-taking modality on participants' ability to predict their performance on a future test.

5.5 Learners' Perception Toward Voice Note-taking

To identify potential opportunities and challenges in adopting voice note-taking in educational settings, we analyzed participants' interview transcripts using a general inductive approach [59]. For this purpose, two authors from the research team independently read the data and developed codes, focusing on the potential benefits and limitations of voice notes. An initial set of twelve thematic labels were developed by the first coder. The developed themes were further reviewed by the second coder, which resulted in the removal of one theme. A final set of eleven codes were agreed upon, out of which six focused on the benefits and five on the limitation of voice notes. Lastly, a coding consistency check was conducted in which the two coders independently applied the developed codes back to the raw data deductively. We observed an overall inter-reliability rate of 0.86, measured by Cohen's Kappa [42], suggesting a good agreement between the two coders. All disagreements between coders were resolved through discussion.

We split the findings into two high-level categories which focused on the potential benefits and limitations of adopting voice note-taking.

5.5.1 Benefits of Voice Notes. Participants were generally receptive towards using voice notes. The themes constructed in this category were as follows.

(1) **Ease of Input:** Twenty-seven participants acknowledged that recording notes via voice was more convenient than typing them, simply because speaking is much *"faster"* (P7) and *"easier"* (P8) than typing. Participants further commented that compared to typed notes, voice notes assisted them in recording more content in less time: *"I feel from recording the audio is definitely much easier than text because I just speak and, um, they're recorded, but the text one has I have to type and this is really time consuming compared with the audio one."* (P28).

(2) **Capturing Momentary Thoughts:** Eight participants commented that recording notes via voice was more spontaneous and assisted them in capturing momentary thoughts and ideas before they are forgotten: *"With voice notes you can just say out loud and it gets recorded the thought process instantly without forgetting any idea"* (P35). Participants further elaborated that typing notes using a keyboard is a time-consuming activity, hence there exists a high possibility that essential ideas might be forgotten during the process of typing content: *"Sometimes, I would have an idea in my mind, that I want to note down, but when I type it, maybe if it's too long and halfway, I might forget what I was thinking before, but with audio it will be faster so that I can just directly say what I had in mind"* (P16).

(3) **Collaborative Note-taking:** Two participants recognized the benefits of voice notes for collaborative learning tasks (P23, P51). This was mainly because voice allows them to easily emphasize essential ideas by changing their *"tone"* (P23), which can help them to better communicate with fellow peers: *"In case, if I'm using audio on a project with my other team members, like, telling ideas in audio could be in our own slang, then this will actually give them the exact meaning of what I'm trying to tell but when it text, that could be any different kind of meaning they could have understood from what I am trying to say"* (P51).

(4) **Rehearsing and Encoding:** Twelve participants reported that while practising voice note-taking, they were *"intentionally thinking"* (P26) about the note content before recording it. This encouraged them to *"re-read"* (P33), *"synthesize"* (P7) and *"memorize"* (P10) the learning content before speaking aloud, which they suggested was helpful for developing a deeper understanding and encoding of the learning material. Compared to this, with typed notes, participants commented that they were more inclined to copy content from the reading material without much understanding: *"With text ... because I am thinking internally and some time even if I don't understand I made a note just because for the sake of taking notes. But with audio if I make a note and I don't actually have not understood the idea while I am making note, I will stop and re-read it and then explain to my self which I don't know helped me get better idea of the text passage"* (P52).

(5) **More Attention Towards the Learning Content:** Eight participants felt that recording voice notes prompted them to pay more attention to the learning content. This was mainly due to two reasons. Firstly, with voice note-taking participants were able to simultaneously inspect the learning content as well as speak out loud. This notion of recording notes while still visualizing the text helped them focus more attention on the learning content: *"For note composing purpose ... I think somehow I read the text more when I was recording an audio note, even while I am recording my eyes were on the text rather then you know focused on the keyboard or the note. So your visual memory is also free for you to read the text while speaking as well"* (P50). Compared to voice, participants commented that with typed notes, they had to switch their visual attention from the text towards the keyboard modality when manually typing content. This broke their thought process and diverted their attention from the primary learning task: *" I paid more attention toward the reading task with audio notes as you can easily record audio notes without much work but with text the way I did it is that it requires a lot of effort. I have to stop reading and then use a keyboard for typing a concise note"* (P34). Secondly, participants commented that while recording voice notes, they did not have to take care of grammatical or typing errors commonly associated with typing text. This further allowed them to focus more attention on the learning content rather than being distracted by these errors: *"With typing I was more concerned about the grammar and spelling mistakes I can make rather then the actual content which I tended to write"* (P48).

(6) **Active Explanation and Reflection:** Twenty-one participants agreed that taking voice notes compelled them to explain learning concepts to themselves or to an imaginary audience. These individuals felt that they subsequently *"reflected"* (P25) and *"elaborated"* (P18) more on the learning content by giving real-life examples: *"with audio notes, I was speaking so its like I am explaining things to some one so I was like saying more content and giving example."* (P42). They further felt that the expression of content in their own words helped them to gain a deeper understanding of the subject matter: *"With audio... like you could tell it in your own way and expression was better and understanding was better"* (P52). Contrarily, participants felt that while typing notes, their thoughts were more *"internalized"* (P52) and they simply wrote minimal content in their notes without much reflection: *"With text I just tried to highlight, what are the important parts of the passage and referring them in notes. But for the audio part, I was trying to express the message of the content rather than just highlighting the important things"* (P6).

5.5.2 Limitations of Voice Notes. Participants expressed some concerns regarding the adoption of voice notes. We have split their feedback into five themes.

(1) **Recording and Replaying Voice:** Eight participants were less enthusiastic about voice notes as they self-counscious while recording their voice: *"When you're speaking, you have some sense of kind of shyness. Also, you kind may not be, feel totally confident in saying things."* (P1). Similarly, participants perceived that when reviewing voice notes, they would not be comfortable while listening to their own voice: *"The only thing that I'm considering is, so suppose I'm learning it again, like I moved to the concepts again, would I listen to my own audio? Cause you know, sometimes listening to your own voice*

is sometimes weird but if you can change it... to smart voice that would be best" (P31).

(2) **Editing Voice Notes:** Four participants were concerned with the difficulty in editing their voice notes. They commented that to edit content in a recorded voice note, they had to record a new voice note which can become cumbersome: *"the issue is, when we did something wrong in the audio, it's very hard to edit ... we need to hear it again to find the bit and if we recorded wrong then I cannot directly edit it rather, I have to make a new note"* (P36).

(3) **Search and Review:** Seventeen participants expressed their concern that reviewing voice notes would require *"time and effort"* (P56, P60) as they had to serially listen to the recording to understand the note content: *"If you want to review your audio comments, you have to repeatedly listen to it. And for example, if you made an audio comment for 10 seconds then at least you will need to spend 10 seconds to review it"* (P34). Moreover, participants noted that as voice notes could not be visually skimmed, searching for the desired content in a voice note would be challenging: *"With text you can categorize content and then you can see what you have written, but in audio, like you can't really have these things ... let's say, like, if you see how I took the text now there were like categories and like, could you see it overall? So for reviewing I could have an overview summary of the things without even looking at in detail and it's easy to find what you want to find."* (P31).

(4) **Structuring and Sketching:** Eight participants reported that with voice note-taking, they had difficulty in organizing their thoughts as they could not add structure to their note content: *"With text you have the feasibility, like, you can better organize ideas, like for audio, I can say, okay, these are the bullet points or anything. I have to speak. But with text I can better organize my notes. Like I can highlight the main statement and then break it down and write a bullet points or anything like that."* (P44). Participants further noted with voice notes, it would be impossible to draw a graph or image which may be necessary for some learning content: *"I still think there's value when I have to hand write things down and especially diagrams and drawing pictures and stuff, I think that's, something that audio cannot achieve"* (P30).

(5) **Shared workplace:** Seven participants expressed their concern about using voice notes in certain settings like classrooms, libraries or quiet shared workspaces as fellow occupants might be disturbed while audio is being recorded: *"Another issue is, cause now I totally study online and stay at home so audio is okay. But instead, I mean, before the semester, I always study in library, which is a quiet, convenient environment. So if I speak loud to take notes it may affect others"* (P37).

6 DISCUSSION

In this paper, we investigated whether note-taking modality could affect learners' text comprehension, their note contents, and their ability to make an accurate judgement regarding text comprehension. By analysing these behaviours, this research provides evidence that support the use of speech as an input modality for taking notes on digital documents and thus motivates the design of note-taking tools that assist learners to gain a better conceptual understanding of digital texts. In this section, we discuss the findings of our research questions.

6.1 Towards a Better Conceptual Understanding of Text

Our first research question examined how note-taking modality can affect learners' text comprehension. To measure this, we assessed learners' performance on the post-test, which consisted of factual and inference-based questions. Our findings suggest that typed notes and voice notes may offer similar benefits when answering factual questions. These findings are in line with prior research that investigated the effect of explanatory modality on learners' text comprehension [26, 34]. Jacob et al. [26], for instance, found that if the goal of the learning activity is to acquire basic knowledge (as opposed to deeper knowledge), then the modality used for articulating the explanation may not affect participants' performance.

However, we found that voice notes seemed to offer an advantage in answering inference-based questions. A possible explanation for this may be that inference-based questions required learners to connect multiple idea units across the text passage and tested their conceptual understanding of the passage. Based on learners' notes, we found that the number of idea units in voice notes (mean idea units/voice note = 5.14 ± 3.7) far exceeded the number of idea units in typed notes (mean idea units/text note = 2.36 ± 1.93). Thus, it appears that voice note-taking may assist learners in generating inferences based on the information presented in the text passages, where learners seemed to connect the ideas and concepts across those passages. The act of integrating idea units across multiple sentences may lead to the construction of new knowledge [11, 15]. Learners who are able to discuss more idea units in a single note are likely to make better inferences about the text, and are also likely to develop a better conceptual understanding, which is reflected in their performance in the post-test. These findings were also in line with prior studies that investigated the effect of explanatory modality on inference-based questions [26, 34]. These studies found that while generating explanations using voice, learners elaborated more on the text passage, which increased their score in inference based questions [26].

We also found that learners' performance was significantly lower for inference-based questions than the factual questions when using keyboard. We believe that a possible reason for this could be that learners generally used fewer idea units in their typed notes (see Figure 3). Moreover, the number of elaborations in typed notes were either very few or none at all (see Figure 3). Therefore, after typing notes using a keyboard, while learners could recall specific idea units in factual questions, they struggled to connect multiple idea units when they were required to answer inference-based questions.

6.2 Towards Generating Better Notes

Our second research question examined the impact of the note-taking modality on the content of the notes generated by our participants. We analysed the content of the notes in terms of their

comprehensiveness and level of elaboration. We defined comprehensiveness in terms of the number of idea units discussed in a note. Further, we measured the level of elaboration in terms of the number of elaborations made by the learner. We found that learners who used voice as a modality for their notes discussed more idea units than those who used the keyboard. Similarly, learners elaborated more on the subject matter when they made notes using voice modality than when they used the keyboard. A possible reason for this behaviour could be the presence of social involvement. As suggested in prior research [26, 34], voice can encourage learners to address a potential audience by making more person-deictic references (e.g., "me", "you") in their notes [35, 56]. This higher levels of social involvement could potentially lead learners to discuss more idea units and elaborate more on the subject matter.

Further, we observed that note comprehensiveness (indicated by the number of idea units) mediated the effect of note-taking modality on text comprehension. These findings suggest that voice notes led learners to take more comprehensive notes, which, in turn, improved inferences about the text, which ultimately resulted in a higher score for inference-based questions.

6.3 Comparable Effects of Modality on Meta-comprehension

The third research question investigated the effect of note-taking modality on learners' meta-comprehension judgement. The results of our analysis revealed no significant effect of note-taking modality on learners' ability to judge their performance for an upcoming test. From the results of this study, we suggest that even though a change in note-taking modality from keyboard to voice increases learners conceptual understanding of the text, the perception of their learning performance might remain unaffected. These results are consistent with the findings of prior research [20, 40], which revealed no significant correlation between learning performance of participants on immediate post-reading test and their meta-comprehension judgement. The findings of these studies suggested that even if learners engage in active elaborations which leads them to better comprehend the text and to provide comprehension-level cues, learners may not use these cues when judging their own text comprehension.

6.4 Potential Benefits and Challenges for the Adoption of Voice notes

The last research question aimed to identify the opportunities of voice notes and challenges learners might face when using voice note-taking for learning activities. We conducted a thematic analysis on participants' interview responses and identified eleven themes that focused on the identified aims. The findings of our analysis revealed that participants were generally receptive to voice notes. This was primarily because the act of speaking content was a more convenient way of composing notes and allowed them to capture instant ideas. Moreover, participants also reported that they actively paraphrased and reflected on the learning content while taking voice notes. This rich expression of content can be explained by the higher level of social presence associated with the oral discourse, which can prompt learners to address a potential audience and elaborate on the subject matter by giving examples or analogies [26]. Additionally, as voice is a gaze-free modality, participants were able to simultaneously read and compose voice notes, thus paying more attention to the learning content itself rather than shifting their attention to type up notes manually.

Apart from the benefits of voice notes, our findings revealed that using voice for recording notes may pose some challenges. Participants were hesitant to record voice notes in classrooms or shared workspaces as speaking aloud to record voice notes may pose privacy concerns and disturb surrounding peers. Moreover, participants raised concerns regarding reviewing and searching voice notes. This was because, compared to typed notes where visual cues (e.g., spatial structuring) could be capitalised, searching and reviewing voice notes is a relatively time-consuming and cumbersome activity. This finding could inform the design of voice note-taking applications to provide text transcriptions of recorded voice notes. This will facilitate not only the note review process but also help learners to edit and search content in notes. Further, some participants were hesitant regarding listening to their own voice while reviewing voice notes, suggesting that they might not prefer to review votes notes. This suggests that the benefits of voice note-taking may be more apparent during note recording as it allows learners to actively reflect on the learning content than during the note-reviewing process. As such, future research can conduct studies on how output modality (audio vs written) can affect the task of note-reviewing.

Overall, the mobility afforded by voice notes suggests that learners could practice this form of note-taking in a digital learning environment, where they are more prone to consume material in a relatively quiet place, and still support comprehension goals. Additionally, learners could also use voice notes in collaborative learning activities (e.g., peer-reviewing and peer-discussion) to provide meaningful and semantically rich feedback to their peers.

7 LIMITATIONS AND FUTURE WORK

We acknowledge five limitations in this study. Previous research indicates that the effects of modality used for learning tasks based on text comprehension depend on the complexity of the text passage. For instance, Jacob et al. [26] suggest that when generating explanations, the effect of modality on text comprehension depends on the text complexity. Whereas the effect of the modality is small in texts of lower complexity, the effect is more pronounced in texts of higher complexity. In the present study, the reading material used for note-taking was perceived to have high levels of text complexity by the novice learners (mean subjective reading difficulty reported = 4.18/7, SD = 1.76). Hence our findings regarding the effects of voice notes for the conceptual understanding are suitable for text passages that are perceived to have a higher level of text complexity. Future research should investigate whether these findings could be generalized with reading materials with a lower level of text complexity.

Further, to observe the effects of note-taking modality on learners' text comprehension, we conducted the post-test immediately after reading and distraction task. As a consequence, we only examined the short-term effects of the note-taking modality on text comprehension. The question of whether these effects translate to

a longer-term remains open for future work to explore, possibly by conducting a delayed knowledge test.

A third limitation is that in this study participants were not allowed to review their notes before the post-reading test. The rationale for this design decision was that reviewing typed and voice notes could demand varying time which could potentially be a confounding factor in this study. Nevertheless, future research could be conducted to allow learners to review their notes during the reading task to investigate the impact of the review of voice notes on text comprehension. In particular, we note the distinction between the input modality used to create the note and the output modality used to consult it. Through the use of text-to-speech and speech-to-text technologies, it is feasible to separate the creation and consumption of notes. As such, future work can explore the different combinations of typed/spoken notes as input modalities and written/audio notes as output modalities.

Further, we conducted our study in a controlled setting with university students. Therefore, even though our findings indicate that compared to keyboard, voice note-taking modality assists learners to comprehend the text better, the external validity of our findings is limited. Therefore, future research is encouraged to replicate our study in a more realistic setting (e.g., in physical and online classrooms) with different texts and study populations to validate the robustness of our findings.

Lastly, we observed that while taking voice notes, learners recorded information that did not require spatial structuring. However, many areas of knowledge can benefit from a visual representation, which includes anything from equations to mind maps. In these cases, the voice modality may not be as effective as compared to other modalities. Therefore, our findings regarding the superior effects of voice notes on text comprehension are limited to the reading content regarding which learners can easily record the information using the voice modality without needing to draw diagram or concepts maps in their notes. This suggests that note-taking applications should provide users with an alternative or complementary input modalities, enabling users to decide how to best leverage the affordances of each of them.

8 CONCLUSION

Although voice note-taking is becoming increasingly popular, there is a lack of evidence about how this modality may support comprehension goals. The present study bridges this gap by providing insights into the effects of voice note-taking on learners' text comprehension, their meta-comprehension judgements, and the contents of their notes. Our findings suggest that a change of modality from keyboard to voice may enable learners to make more elaborations and discuss more idea units in their notes. This can assist learners in text comprehension, as reflected in their post-test scores. Further, the findings indicate that if the goal is to gain basic knowledge of the study content, then both modalities (keyboard vs. voice) perform similarly, but if the goal is to enable learners to gain a deeper conceptual understanding by making inferences, voice note-taking might be more effective than typing notes using keyboard. Lastly, the findings suggest that both note-taking modalities have comparable effects on learners' meta-comprehension judgement of text comprehension. Overall, our findings suggest that voice modality could be incorporated in digital learning environments for taking notes as it could trigger higher-order comprehension and enable learners to get a better conceptual understanding of digital texts.

ACKNOWLEDGMENTS

We wish to thank Prof. Simon Buckingham Shum and Dr Paula G. de Barba for the useful discussion on this work. Dr Eduardo Velloso is the recipient of an Australian Research Council Discovery Early Career Researcher Award (Project Number: DE180100315) funded by the Australian Government. Anam Ahmad Khan is supported under the Melbourne Graduate Research Scholarship.

REFERENCES

[1] Peter Afflerbach, P. David Pearson, and S. Paris. 2008. Clarifying Differences Between Reading Skills and Reading Strategies. *The Reading Teacher* 61 (2008), 364–373. https://doi.org/10.1598/RT.61.5.1

[2] Samah Zakareya Ahmad. 2019. Impact of Cornell Notes vs. REAP on EFL Secondary School Students' Critical Reading Skills. *International Education Studies* 12, 10 (2019), 60–74. https://doi.org/10.5539/ies.v12n10p60

[3] Steven R Asher, Shelley Hymel, and Allan Wigfield. 1978. Influence of topic interest on children's reading comprehension. *Journal of Reading Behavior* 10, 1 (1978), 35–47. https://doi.org/10.1080/10862967809547253

[4] Farnoush Bahrami and Heidar Nosratzadeh. 2017. The effectiveness of note-taking on reading comprehension of Iranian EFL Learners. *International Journal of Applied Linguistics and English Literature* 6, 7 (2017), 308–317. https://doi.org/10.7575/aiac.ijalel.v.6n.7p.308

[5] Marie-Laure Barbier, Jean-Yves Roussey, Annie Piolat, and Thierry Olive. 2006. Note-taking in second language: Language procedures and self evaluation of the difficulties. *Current psychology letters. Behaviour, brain & cognition* 20, Vol. 3, 2006 (2006). https://doi.org/10.4000/cpl.1283

[6] Douglas Bates, Martin Mächler, Ben Bolker, and Steve Walker. 2015. Fitting Linear Mixed-Effects Models Using lme4. *Journal of Statistical Software* 67, 1 (2015), 1–48. https://doi.org/10.18637/jss.v067.i01

[7] Peter Brandl, Christoph Richter, and Michael Haller. 2010. *NiCEBook: Supporting Natural Note Taking*. Association for Computing Machinery, New York, NY, USA, 599–608. https://doi.org/10.1145/1753326.1753417

[8] William G Brozo, Norman A Stahl, and James R King. 1996. Tracking thinking while reading: A strategy for improving metacomprehension. *Research and Teaching in Developmental Education* (1996), 19–31.

[9] Ying-Chih Chen, Soonhye Park, and Brian Hand. 2016. Examining the use of talk and writing for students' development of scientific conceptual knowledge through constructing and critiquing arguments. *Cognition and Instruction* 34, 2 (2016), 100–147.

[10] Minyoung Cho. 2018. Task complexity, modality, and working memory in L2 task performance. *System* 72 (02 2018), 85–98. https://doi.org/10.1016/j.system.2017.10.010

[11] Virginia Clinton, Ben Seipel, Paul van den Broek, Kristen L McMaster, Panayiota Kendeou, Sarah E Carlson, and David N Rapp. 2014. Gender differences in inference generation by fourth-grade students. *Journal of Research in Reading* 37, 4 (2014), 356–374. https://doi.org/10.1111/j.1467-9817.2012.01531.x

[12] Jacob Cohen. 2013. *Statistical power analysis for the behavioral sciences*. Academic press.

[13] John Dunlosky, Christopher Hertzog, Mary RF Kennedy, and Keith W Thiede. 2005. The self-monitoring approach for effective learning. *Cognitive Technology* (2005).

[14] John Dunlosky and Amanda R Lipko. 2007. Metacomprehension: A brief history and how to improve its accuracy. *Current Directions in Psychological Science* 16, 4 (2007), 228–232.

[15] Luke G Eglington and Sean HK Kang. 2018. Retrieval practice benefits deductive inference. *Educational Psychology Review* 30, 1 (2018), 215–228. https://doi.org/10.1007/s10648-016-9386-y

[16] Logan Fiorella and Richard Mayer. 2015. *Learning as a Generative Activity: Eight Learning Strategies that Promote Understanding*. https://doi.org/10.1017/CBO9781107707085

[17] Logan Fiorella and Richard E Mayer. 2016. Eight ways to promote generative learning. *Educational Psychology Review* 28, 4 (2016), 717–741. https://doi.org/10.1007/s10648-015-9348-9

[18] Logan Fiorella and Richard E Mayer. 2017. Spontaneous spatial strategy use in learning from scientific text. *Contemporary Educational Psychology* 49 (2017), 66–79. https://doi.org/10.1016/j.cedpsych.2017.01.002

[19] David Freedman, Robert Pisani, and Roger Purves. 2007. Statistics (international student edition). *Pisani, R. Purves, 4th edn. WW Norton & Company, New York* (2007).

[20] Tatsushi Fukaya. 2013. Explanation generation, not explanation expectancy, improves metacomprehension accuracy. *Metacognition and learning* 8, 1 (2013), 1–18. https://doi.org/10.1007/s11409-012-9093-0

[21] Andrzej Gałecki and Tomasz Burzykowski. 2013. Linear mixed-effects model. In *Linear Mixed-Effects Models Using R*. Springer, 245–273.

[22] James J Gibson. 2014. *The ecological approach to visual perception: classic edition*. Psychology Press.

[23] Joachim Grabowski. 2010. Speaking, writing, and memory span in children: Output modality affects cognitive performance. *International Journal of Psychology* 45, 1 (2010), 28–39. https://doi.org/10.1080/00207590902914051

[24] Jason A Grahame. 2016. Digital note-taking: Discussion of evidence and best practices. *The Journal of Physician Assistant Education* 27, 1 (2016), 47–50.

[25] Andrew F Hayes. 2017. *Introduction to mediation, moderation, and conditional process analysis: A regression-based approach*. Guilford publications.

[26] Leonie Jacob, Andreas Lachner, and Katharina Scheiter. 2020. Learning by explaining orally or in written form? Text complexity matters. *Learning and Instruction* 68 (2020), 101344.

[27] Renée S Jansen, Daniel Lakens, and Wijnand A IJsselsteijn. 2017. An integrative review of the cognitive costs and benefits of note-taking. *Educational Research Review* 22 (2017), 223–233. https://doi.org/10.31234/osf.io/ty4nq

[28] Panayiota Kendeou, David N Rapp, and Paul van den Broek. 2003. The influence of reader's prior knowledge on text comprehension and learning from text. *Progress in education* 13 (2003), 189–209.

[29] Hyeyoun Kim. 2018. Impact of slide-based lectures on undergraduate students' learning: Mixed effects of accessibility to slides, differences in note-taking, and memory term. *Computers & Education* 123 (2018), 13–25. https://doi.org/10.1016/j.compedu.2018.04.004

[30] Yea-Seul Kim, Nathalie Henry Riche, Bongshin Lee, Matthew Brehmer, Michel Pahud, Ken Hinckley, and Jessica Hullman. 2019. Inking your insights: Investigating digital externalization behaviors during data analysis. In *Proceedings of the 2019 ACM International Conference on Interactive Surfaces and Spaces*. 255–267. https://doi.org/10.1145/3343055.3359714

[31] J Peter Kincaid, Robert P Fishburne Jr, Richard L Rogers, and Brad S Chissom. 1975. *Derivation of new readability formulas (automated readability index, fog count and flesch reading ease formula) for navy enlisted personnel*. Technical Report. Naval Technical Training Command Millington TN Research Branch.

[32] Walter Kintsch. 1988. The role of knowledge in discourse comprehension: A construction-integration model. *Psychological review* 95, 2 (1988), 163. https://doi.org/10.1037/0033-295X.95.2.163

[33] Walter Kintsch and CBEMAFRS Walter Kintsch. 1998. *Comprehension: A paradigm for cognition*. Cambridge university press.

[34] Andreas Lachner, Kim-Tek Ly, and Matthias Nückles. 2018. Providing written or oral explanations? Differential effects of the modality of explaining on students' conceptual learning and transfer. *The Journal of Experimental Education* 86, 3 (2018), 344–361. https://doi.org/10.1080/00220973.2017.1363691

[35] Robin Tolmach Lakoff. 1982. Some of my favorite writers are literate: The mingling of oral and literate strategies in written communication. *Spoken and written language: Exploring orality and literacy* 9 (1982), 239–60.

[36] Russell V. Lenth. 2016. Least-Squares Means: The R Package lsmeans. *Journal of Statistical Software* 69, 1 (2016), 1–33. https://doi.org/10.18637/jss.v069.i01

[37] Ching (Jean) Liu, Chi-Lan Yang, Joseph Jay Williams, and Hao-Chuan Wang. 2019. NoteStruct: Scaffolding Note-taking While Learning from Online Videos. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems* (Glasgow, Scotland Uk) *(CHI EA '19)*. ACM, Article LBW1116, 6 pages. https://doi.org/10.1145/3290607.3312878

[38] Linlin Luo, Kenneth A Kiewra, Abraham E Flanigan, and Markeya S Peteranetz. 2018. Laptop versus longhand note taking: effects on lecture notes and achievement. *Instructional Science* 46, 6 (2018), 947–971. https://doi.org/10.1007/s11251-018-9458-0

[39] Ruth H Maki. 1998. Test predictions over text material. In *Metacognition in educational theory and practice*. Routledge, 131–158.

[40] Ruth H Maki, Micheal Shields, Amanda Easton Wheeler, and Tammy Lowery Zacchilli. 2005. Individual Differences in Absolute and Relative Metacomprehension Accuracy. *Journal of Educational Psychology* 97, 4 (2005), 723. https://doi.org/10.1037/0022-0663.97.4.723

[41] Richard E Mayer and Joan K Gallini. 1990. When is an illustration worth ten thousand words? *Journal of educational psychology* 82, 4 (1990), 715. https://doi.org/10.1037/0022-0663.82.4.715

[42] Mary L McHugh. 2012. Interrater reliability: the kappa statistic. *Biochemia medica: Biochemia medica* 22, 3 (2012), 276–282. https://doi.org/10.11613/BM.2012.031

[43] Sholeh Moradi, Shima Ghahari, and Mohammad Abbas Nejad. 2020. Learner-vs. Expert-Constructed Outlines: Testing the Associations with L2 Text Comprehension and Multiple Intelligences. *Studies in Second Language Learning and Teaching* 10, 2 (2020), 359–384. https://doi.org/10.14746/ssllt.2020.10.2.7

[44] Kayla Morehead, John Dunlosky, and Katherine A Rawson. 2019. How much mightier is the pen than the keyboard for note-taking? A replication and extension of Mueller and Oppenheimer (2014). *Educational Psychology Review* 31, 3 (2019), 753–780. https://doi.org/10.1007/s10648-019-09468-2

[45] Kayla Morehead, John Dunlosky, Katherine A Rawson, Rachael Blasiman, and R Benjamin Hollis. 2019. Note-taking habits of 21st century college students: implications for student learning, memory, and achievement. *Memory* 27, 6 (2019), 807–819. https://doi.org/10.1080/09658211.2019.1569694

[46] Daria Mundt, Roman Abel, and Martin Hänze. 2020. Exploring the effect of testing on forgetting in vocabulary learning: an examination of the bifurcation model. *Journal of Cognitive Psychology* 32, 2 (2020), 214–228. https://doi.org/10.1080/20445911.2020.1733584

[47] Bennet B Murdock Jr. 1962. The serial position effect of free recall. *Journal of experimental psychology* 64, 5 (1962), 482. https://doi.org/10.1037/h0045106

[48] Cuong Nguyen and Feng Liu. 2016. Gaze-based Notetaking for Learning from Lecture Videos. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. ACM, 2093–2097. https://doi.org/10.1145/2858036.2858137

[49] Khuyen Nguyen and Mark A McDaniel. 2016. The JOIs of text comprehension: Supplementing retrieval practice to enhance inference performance. *Journal of Experimental Psychology: Applied* 22, 1 (2016), 59. https://doi.org/10.1037/xap0000066

[50] Hüseyin Özçakmak. 2019. Impact of Note Taking during Reading and during Listening on Comprehension. *Educational Research and Reviews* 14, 16 (2019), 580–589. https://doi.org/10.5897/ERR2019.3812

[51] Annie Piolat, Thierry Olive, and Ronald T Kellogg. 2005. Cognitive effort during note taking. *Applied cognitive psychology* 19, 3 (2005), 291–312. https://doi.org/10.1002/acp.1086

[52] Yi Ren, Yang Li, and Edward Lank. 2014. InkAnchor: Enhancing Informal Ink-based Note Taking on Touchscreen Mobile Phones. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (Toronto, Ontario, Canada) *(CHI '14)*. ACM, New York, NY, USA, 1123–1132. https://doi.org/10.1145/2556288.2557302

[53] Franklin E Satterthwaite. 1946. An approximate distribution of estimates of variance components. *Biometrics bulletin* 2, 6 (1946), 110–114. https://doi.org/10.2307/3002019

[54] Michael J Serra and John Dunlosky. 2010. Metacomprehension judgements reflect the belief that diagrams improve learning from text. *Memory* 18, 7 (2010), 698–711.

[55] Hao Shen and Jaideep Sengupta. 2018. Word of mouth versus word of mouse: Speaking about a brand connects you to it more than writing does. *Journal of Consumer Research* 45, 3 (2018), 595–614. https://doi.org/10.1093/jcr/ucy011

[56] Maria Grazia Sindoni. 2014. *Spoken and written discourse in online interactions: A multimodal approach*. Routledge.

[57] Keith W Thiede, Mary Anderson, and David Therriault. 2003. Accuracy of metacognitive monitoring affects learning of texts. *Journal of educational psychology* 95, 1 (2003), 66.

[58] Keith W Thiede and Mary CM Anderson. 2003. Summarizing can improve metacomprehension accuracy. *Contemporary Educational Psychology* 28, 2 (2003), 129–160.

[59] David R. Thomas. 2006. A General Inductive Approach for Analyzing Qualitative Evaluation Data. *American Journal of Evaluation* 27, 2 (2006), 237–246. https://doi.org/10.1177/1098214005283748

[60] Paul Van den Broek. 2010. Using texts in science education: Cognitive processes and knowledge representation. *Science* 328, 5977 (2010), 453–456.

[61] Francis J. Di Vesta and G. Susan Gray. 1972. Listening and note taking. *Journal of Educational Psychology* 63, 1 (1972), 8–14. https://doi.org/10.1037/h0032243

[62] Xu Wang, Miaomiao Wen, and Carolyn P Rosé. 2016. Towards triggering higher-order thinking behaviors in MOOCs. In *Proceedings of the Sixth International Conference on Learning Analytics & Knowledge*. 398–407. https://doi.org/10.1145/2883851.2883964

[63] Lynn D. Wilcox, Bill N. Schilit, and Nitin Sawhney. 1997. Dynomite: A Dynamically Organized Ink and Audio Notebook. In *Proceedings of the ACM SIGCHI Conference on Human Factors in Computing Systems* (Atlanta, Georgia, USA) *(CHI '97)*. Association for Computing Machinery, New York, NY, USA, 186–193. https://doi.org/10.1145/258549.258700

[64] Philip H Winne and Nancy E Perry. 2000. Measuring self-regulated learning. In *Handbook of self-regulation*. Elsevier, 531–566.

[65] Amber Witherby and Sarah Tauber. 2019. The Current Status of Students' Note-Taking: Why and How Do Students Take Notes? *Journal of Applied Research in Memory and Cognition* 8 (05 2019), 139–153. https://doi.org/10.1016/j.jarmac.2019.04.002

[66] Merlin C Wittrock. 1989. Generative processes of comprehension. *Educational psychologist* 24, 4 (1989), 345–376.

[67] Dongwook Yoon, Nicholas Chen, François Guimbretière, and Abigail Sellen. 2014. RichReview: Blending Ink, Speech, and Gesture to Support Collaborative Document Review. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology* (Honolulu, Hawaii, USA) *(UIST '14)*. Association

for Computing Machinery, New York, NY, USA, 481–490. https://doi.org/10.1145/2642918.2647390

[68] Qian Yu, Tonya Nguyen, Soravis Prakkamakul, and Niloufar Salehi. 2019. " I Almost Fell in Love with a Machine" Speaking with Computers Affects Self-disclosure. In *Extended Abstracts of the 2019 CHI Conference on Human Factors in Computing Systems*. 1–6. https://doi.org/10.1145/3290607.3312918