# How does interpreting performance correlate with note-taking process, note-taking product and note-reading process?

An eye-tracking and pen-recording study

Huolingxiao Kuang
Durham University
huolingxiao.kuang@durham.ac.uk
https://orcid.org/0000-0003-2662-7904

Binghan Zheng (corresponding author)
Durham University
binghan.zheng@durham.ac.uk
https://orcid.org/0000-0001-5302-4709

**Funding**

This work is supported by National Social Science Fund of China (No.20BYY014). Huolingxiao Kuang is funded by the China Scholarship Council.

**Notes on contributors**

***Miss Huolingxiao Kuang*** is a PhD candidate (with CSC scholarship) at the School of Modern Languages and Cultures, Durham University. She obtained her MA in Translation and Interpreting from Peking University in 2018, and has been working as a freelancer interpreter for eight years. Her research interests include interpreting process research and note-taking in consecutive interpreting. She has published journal article in *Perspectives*, and two of her papers will appear as book chapters published by Cambridge University Press and Routledge.

**Dr Binghan Zheng** is Associate Professor of Translation Studies at Durham University, where he serves as the Director of the Centre for Intercultural Mediation. His research interests include cognitive translation/interpreting studies, and neuroscience of translation. His recent publications appeared in journals such as *Target*, *Across Languages & Cultures*, *Journal of Pragmatics*, *Brain & Cognition*, *Perspectives*, *LANS-TTS*, *Babel*, *Translation & Interpreting Studies*, *Foreign Language Teaching & Research*, and *Journal of Foreign Languages*. He is a guest editor of journals *Translation Studies*, and *Translation & Interpreting Studies*.

Corresponding author: Binghan Zheng
Address: School of Modern Languages and Cultures, Durham University
Elvet Riverside, New Elvet, Durham, DH1 3JT, UK
Email: binghan.zheng@durham.ac.uk

**Abstract:**

This paper explores relationships between consecutive interpreting (CI) performance on the one hand, and interpreters' note-taking effort, note-taking product and note-reading effort, on the other hand. 20 professionals and 29 students consecutively interpreted two easy segments and two difficult segments in an English (L2) speech, with their eye fixations on the notes and handwriting on the digital pad being registered through eye-tracking and pen-recording methods. Both groups' CI performance showed positive but weak correlations with their note quantities in the easy segments, but not in the difficult ones. Almost no significant correlations were found between the students' interpretation quality and effort of note-taking, whereas the professionals' CI performance was negatively correlated with their cognitive effort of note-taking. Significant but weak correlations were observed in both groups between their note-reading effort and interpreting performance, but the students' correlations were mainly found in the difficult segments, and the professionals' correlations were mostly detected in the easy ones. Overall, the interpreters' note-taking behaviour was not closely associated with their interpretation quality, and the associations varied across interpreter groups and task difficulties. These findings suggest that note-taking should be taught more judiciously in interpreter training programs and applied more prudently in interpreting practice.

**Keywords:**

Consecutive interpreting (CI) performance, note-taking, note-reading, features of notes, eye-tracking, pen-recording

## 1. Introduction

In consecutive interpreting (CI), interpreters deliver their rendition only after the speaker finishes a segment of speech, which usually lasts from 45 seconds to 3 minutes (Setton & Dawrant, 2016). Therefore, interpreters generally resort to note-taking to release the pressure on their short-term memory. Serving as a "bridge connecting the gap between the interpreter's memory and the production" (Chuang, 2008, p. 95) of the target speech, note-taking is considered to be an important factor of CI quality (e.g., Gile, 2009; Gillies, 2005). Gile (2020, p. 13) clearly points out that most errors and omissions in CI can be traced back to the comprehension phase because of the "strong cognitive pressure" caused by the "cognitive and mechanical aspects of note-taking during comprehension". Therefore, repeated attempts have been made to explore the secrets of successful note-taking that can contribute to high-quality interpretations in CI.

Previous research has mainly investigated the features of (non)effective notes from three aspects: note quantity (e.g., Dam, 2007), note form (e.g., Cardoen, 2018) and note language (e.g., Dai & Xu, 2007). Overall, a variety of findings have been reported about how these note choices are related to interpretation quality. Recently, researchers (e.g., Chen, 2020b; Chen et al., 2021; Hu, 2008) have shifted their attention from the descriptive features of notes to the underlying cognitive mechanisms during the note-taking and note-reading processes. Only one study has investigated interpretation quality, and it found that the temporal and physical demands of note-taking are significantly correlated with interpreters' CI performance (Chen, 2020b). The fact that no research has yet fully examined the associations between interpretation quality, on the one hand, and the three stages of note-taking activities in CI, namely the process of note-taking, the product of note-taking, and the process of note-reading, on the other hand, leaves the question of note-taking's role in CI underexplored from the "performance" perspective. The present study aims to fill this research gap by adopting process-oriented methods such as eye-tracking and pen-recording to examine interpreters' visual, temporal, cognitive and physical demands during the process of the note activities, and by using product-oriented methods including a descriptive analysis of notes and a summative assessment of interpretation quality, to provide a clearer picture of the associations between interpreters' note-taking behaviour and CI quality.

## 2. Research Background

### 2.1. Note-taking product and CI quality

In previous research on the associations between the product of note-taking and CI quality, some findings are seemingly consistent. For example, a positive relationship between the use of symbols and interpretation quality is reported in Chen (2020b) and Liu (2010) (Table 1). However, Chen's (2020b) result was obtained from professional interpreters in an L2-to-L1 CI task in which the source speech was delivered at a rate of 127 words per minute (wpm), while Liu's (2010) finding was based on undergraduate students who received note-taking training for one semester only in an L1-to-L2 CI task with a delivery rate at 200 wpm. Since the experimental settings and the background of participants varied considerably, these findings cannot be directly compared or generalized to other interpreting scenarios. Cardoen's dissertation is the only study that investigated this issue by asking different interpreter groups (two student groups and one professional group) to interpret both easy and difficult speeches. She found that the students' interpreting fluency improved when they used more symbols, but that the situation was the opposite for the professionals. In the easy tasks, all the participants' interpreting accuracy was higher when they noted fewer symbols, whereas the result was the opposite in the difficult tasks. All these findings suggest that it is important to take interpreting experience, task difficulty and different aspects of interpretation quality into consideration, before we can provide a more accurate account of the relationship between note patterns and CI quality. Moreover, the fact that the "professionals" in Cardoen (2018) mainly taught interpreting at universities and rarely worked as professional interpreters in the market could affect her findings.

Table 1. The relationship between note patterns and CI quality

| Research | Language pair | Interpreting direction | Participant background | Results |
|---|---|---|---|---|
| **Her (2001)** | Chinese-English | Both | 2 groups of students with 8 weeks and 4 weeks of note-taking training | • Note quality + |
| **Dam et al. (2005)** | Danish-Spanish | L2-L1 | 1 professional | • Note quantity +<br>• Abbreviation + |
| **Dai & Xu (2007)** | Chinese-English | L1-L2 | 6 professionally trained (3 months) and 6 non-professionally trained students (1 semester) | • No observed relations |
| **Dam (2007)** | Danish-Spanish | L2-L1 | 5 professionals | • Note quantity +<br>• Abbreviation +<br>• Full word – |
| **Liu (2010)** | Chinese-English | L1-L2 | 62 trained undergraduates (1 semester) | • No statistically significant correlations with the language of notes<br>• Symbol + |
| **Wang et al. (2010)** | Chinese-English | Both | 12 trained undergraduates (15 teaching hours in note-taking) | • No statistically significant correlations |
| **Chen (2017)** | Chinese-English | Both | 5 professionals | • No observed relations |
| **Cardoen (2018)** | Dutch-Spanish | L2-L1 | 5 first-year master students, 5 second-year master students and 5 professionals | Interpreting accuracy:<br>• Note quantity +<br>• Symbol in easy tasks +<br>• Symbol in hard tasks –<br>Interpreting fluency:<br>• Note quantity +<br>• Full word +<br>• Abbreviation – |
| **Chen (2020b)** | Chinese-English | Both | 22 professionals | L1-to-L2 interpreting<br>• Note quantity +<br>• Interpreters' L2 –<br>L2-to-L1 interpreting<br>• Symbols +<br>• Interpreters' L2 – |

*Note.* "+" represents a positive relationship and "–" means a negative relationship.

In summary, our current understanding of the relationship between the product of note-taking and interpretation quality is limited for four major reasons: (1) inadequate consideration of interpreting experience and task difficulty; (2) incomplete assessment of interpretation quality; (3) inappropriate criteria of participant recruitment; and (4) small sample size undermining statistical power. More empirical evidence is needed to show the features of (non)effective notes in CI by recruiting more participants in different interpreter groups, designing CI tasks at different difficulty levels, and assessing interpreting performance from content, delivery and language quality aspects.

### 2.2. Note-taking process and CI quality

Early research into the process of note-taking used video-recording and dual-task paradigm methods. Andres (2002) calculated 14 professionals' and 14 students' ear-pen span (EPS), i.e., the time span between the delivery of the source speech and the note-taking acts, with the help of video recordings. She found that the professionals' EPS was 4 to 6 seconds, while the students' EPS sometimes extended to 10 seconds. Since Andres (2002) found that interpreters generally had listening comprehension problems when their EPS was longer than seven seconds, she suggested that the prolonged EPS among the students indicated a cognitive overload of note-taking. This implies that students' underperformance in CI, when compared with that of professionals, might originate in their still rudimentary note-taking techniques. Similarly, Hu (2008) found that professional interpreters ($N$=10) achieved higher interpretation quality and reacted to a secondary task of sound detection faster than student interpreters ($N$=10) during note-taking, suggesting less cognitive effort of note-taking in the former group than in the latter. Moreover, all the interpreters slowed down in the secondary task completion and performed worse in interpreting when the source material became difficult, again implying that the decrease in interpretation quality could be attributed to the increase of note-taking effort.

Moving forward, Chen (2020b) adopted a digital pen to measure 22 professional interpreters' temporal and physical demands during note-taking, by calculating the distance and duration of note-writing and the time spent on note-planning (EPS). She found that the interpreters' performance was positively correlated with the time they spent on note-taking, which corroborates previous findings that the more notes are made the better the performance (e.g., Dam, 2007). However, the interpreters' performance was negatively correlated with their EPS only in the L2-to-L1 direction. Based on Chen's (2020a)[1] findings, L2 listening and comprehension was more cognitively demanding than that of L1. Therefore, if the interpreters could not quickly transform the L2 source information into written notes, they would have to recall the source contents during the output phase of CI simply based on their short-term memory, which could lead to less accurate and incomplete renditions. Since Chen did not include the participants' eye movements during the note-taking process into the analysis, which is assumed to be important in indicating the cognitive processing involved in the note activities in CI (Chen et al., 2021), the relationship between the cognitive effort of note-taking and interpretation quality remains unexplored. In addition, the professional interpreters in Chen's study had lived in an English-speaking (L2) country for a long time. In that case, their strong L2 competence might affect their note-taking behaviour and interpreting performance. Hence, the reported findings need to be re-examined in other interpreter groups and CI tasks that vary in difficulty levels.

### 2.3. Note-reading process and CI quality

Regarding the process of note-reading, only two papers based on one research project are available. Chen collected 18 professional interpreters' eye-tracking data during the second phase of CI and found that, compared to the L1-to-L2 task, the L2-to-L1 task entailed shorter average fixation duration, i.e., less cognitive load, during note-reading (Chen, 2020b). In addition, interpreters had poorer interpreting accuracy and higher interpreting fluency in the L2-to-L1 interpreting than in the other direction. Although Chen (2020b) did not differentiate the cognitive effort of note-reading from that of other operations in the reformulation phase of CI, such as the production of target speech (Gile, 2009), her results imply that the cognitive effort of note-reading, which constitutes a part of the interpreters' effort at the reformulation stage, could relate to CI quality.

Chen et al. (2021) specifically measured interpreters' cognitive effort of reading notes in different categories with a variety of eye-tracking indicators. The results show that, regardless of interpreting direction, reading language notes was more cognitively demanding than reading symbols[2]. As for abbreviations and full-words, no significant difference was observed between the two when the length of notes was controlled for. In addition, recognizing Chinese (L1) notes and English (L2) notes required similar effort for the Chinese-native interpreters. However, integrating the meaning of English notes at a textual level needed significantly more effort than for the Chinese notes in English-to-Chinese interpreting. Although Chen's study did not focus on CI quality, it demonstrates that interpreters' note choices can affect their cognitive effort of note-reading. Taken together, the process of note-taking, the product of note-taking and the process of note-reading are three integrated note activities in CI that could relate to CI quality in different ways.

### 2.4. The present study

The above review indicates that interpreters' note-taking behaviour has been rarely studied with combined process-oriented and product-oriented methodologies, and that there has been inadequate control for interpreter background and task difficulty when investigating the relationship between note-taking behaviour and CI quality. The present study aims to fill this gap by examining 29 student interpreters' and 20 professional interpreters' pen movements, eye movements and interpreting outputs in an English-to-Chinese CI task that contained both easy and difficult speech segments. It attempts to answer three research questions:

1. How is the product of note-taking related to the quality of CI output? This question looks at the descriptive features of notes, including note quantity, note form and note language.
2. How is the process of note-taking associated with CI quality? This question is answered by measuring interpreters' overt visual attention, cognitive processing,

the time spent on note-planning, and the physical effort during note-taking, through eye-tracking and pen-recording methods.

3. How is the process of note-reading relative to CI quality? A variety of eye-tracking indicators that point to the cognitive processing involved in early and late stages of note-reading is adopted to answer this question.

## 3. Method

### 3.1. Participants

We recruited 24 professional interpreters and 31 student interpreters (Table 2) to participate in the experiment. The participants had normal vision or corrected-to-normal vision by wearing glasses, and all of them have Chinese as their L1 and English as their L2. In the professional group, 21 have a master's degree in interpreting and 3 learned interpreting at professional training courses. In the student group, all had just finished their first year of study in Master of Translation and Interpreting (MTI) in Chinese universities and had no professional interpreting experience. During participant recruitment, all confirmed that they had learned and practiced note-taking in and outside class. Ethical clearance had been obtained before the project (ID: MLAC-2019-06-13T14:42:41-tzcw84) began.

Table 2. Participant details

| **Group** | **Age** | **Gender** | **Training experience in interpreting** | **Work experience in interpreting** |
|---|---|---|---|---|
| **Professional** | 33.58 (*SD*=6.16) | 18 females and 6 males | 1.64 years (*SD*=0.68) | 7.29 years (*SD*=3.24 years) |
| **Student** | 23.22 (*SD*=2.17) | 27 females and 4 males | just finished their first year of MTI program | No |

### 3.2. Material

The stimulus was excerpted from a TED talk that introduces the role of classical music in life in plain language. The talk was divided into four segments and minor revisions were made to the videos to ensure that Segments 2 and 3 were more difficult to interpret than Segments 1 and 4. This sequence arrangement was designed to reduce the possibility of order effects. For the same reason, a 45-second music clip between Segments 2 and 3 from the original talk was kept. We designed the duration of each speech segment to be less than 3 minutes, because the general duration of speech segments in CI is estimated to be between 45 seconds and 3 minutes (Setton & Dawrant, 2016). Moreover, although the total duration of the interpreting task (around 20 minutes) is not typically long in interpreting practice, the reduced task duration helped to ensure high eye-tracking data quality in the experiment, which is closely related to the participants' fatigue level.

Table 3. Speech segment details

| | **Segment 1** | **Segment 2** | **Segment 3** | **Segment 4** |
|---|---|---|---|---|
| **Speech duration** | 2 min 45 s | 2 min 36 s | 2 min 47 s | 2 min 47 s |
| **Delivery speed (wpm)** | 147.63 | 146.15 | 145.51 | 145.51 |
| **Word count** | 406 | 380 | 405 | 405 |
| **Reading Ease score** | 80.70 | 58.00 | 55.00 | 79.20 |
| **Idea density (ppw)** | 0.54 | 0.57 | 0.57 | 0.55 |

*Note:* the unit for "delivery speed" is words per minute (wpm), and for "idea density" is proposition count per word (ppw)

We assessed the difficulty level of interpreting the speech segments through idea density[3] and readability[4] indices, both of which are considered important indicators of source speech difficulty in interpreting (Liu & Chiu, 2009; Liu et al., 2004). Segments 2 and 3 have higher idea densities and lower Reading Ease scores than Segments 1 and 4 (Table 3), indicating more difficulty of interpreting in the first group of segments than in the second. We also asked 6 university interpreting lecturers who were also part-time interpreters (10.5 years of teaching and interpreting experience on average) and 4 freelance interpreters (6 years of professional experience on average) to assess the difficulty level of interpreting the segments from eight aspects: words, syntactic structure, information density, coherence, logic, clarity, abstractness, and required background knowledge (Liu & Chiu, 2009, p. 248). Moreover, 10 MA students from the pilot study rated their mental demand, effort, frustration and performance after interpreting (Sun & Shreve, 2014). In accordance with the objective assessment results, both groups of assessors judged that Segments 2 and 3 were obviously more difficult to interpret than Segments 1 and 4.

### 3.3. Experimental design

The participants used a Wacom CTL 672 digital pen to take notes on a tablet while their eyes were fixating on a 15.6-inch laptop screen. All of their writings on the tablet were synchronized and shown on the screen simultaneously (Figure 1). A Tobii Pro Fusion 250HZ eye tracker, which was attached to the lower part of the laptop, recorded interpreters' eye movements on the screen during interpreting. *Audacity* 2.4.4 software was used for voice recording throughout the experiment.


Figure 1. Experimental settings

*Note*. AOI stands for area of interest.

We decided to use the laptop based remote eye-tracker rather than wearable eye-tracking glasses (Chen et al., 2021) to record the participants' eye movements, because a high percentage (31%) of data loss was reported in Chen et al.'s study. The fact that this Tobii Pro Fusion eye-tracker was fixed to the laptop screen could help us to ensure the quality of eye-tracking data. Considering that this design could cause some hand-eye coordination problems during note-taking, a pilot study was conducted with 10 MA students. We found that all of them adapted to this setting after a time-free (*M*=18 min 27 s, *SD*=5 min 10 s) pen-practice session. In addition, since our experiment was conducted during the Covid-19 pandemic, all participants claimed that they had experience in interpreting a speech presented through a computer screen during work and training, which helped them to adjust to this experimental setting. Meanwhile, using a contact-free eye-tracker could protect our participants from possible cross infections under the pandemic situation.

### 3.4. Procedure

As outlined in Figure 2, the participants firstly conducted an English (L2) listening span task (Cai et al., 2015) to measure their working memory capacity. Thereafter, they copied the gist of the speech and the background information of the speaker with the digital pen. They then familiarised themselves with the prepared vocabularies of the speech by copying, circling or creating symbols with the digital pen on the screen without time limits. After that, they had a nine-point eye calibration by sitting approximately 60cm away from the eye-tracker and did a warm-up exercise of CI. When they claimed to be comfortable with the setting, they had another calibration session and proceeded to the main interpreting task. They were asked to interpret Segments 1 and 2, watch the opera video clip, then interpret Segments 3 and 4. Eye-tracking stopped upon the completion of the interpreting task. Then, they scored the difficulty level of interpreting each segment with a NASA-task load questionnaire (Sun & Shreve, 2014) and explained their notes, including but not limited to the form (language, symbol or number), language (Chinese or English), and the corresponding source and target speech units. Finally, they completed a questionnaire regarding their demographic information and interpreting background. The experiment lasted approximately from 80 to 90 minutes.

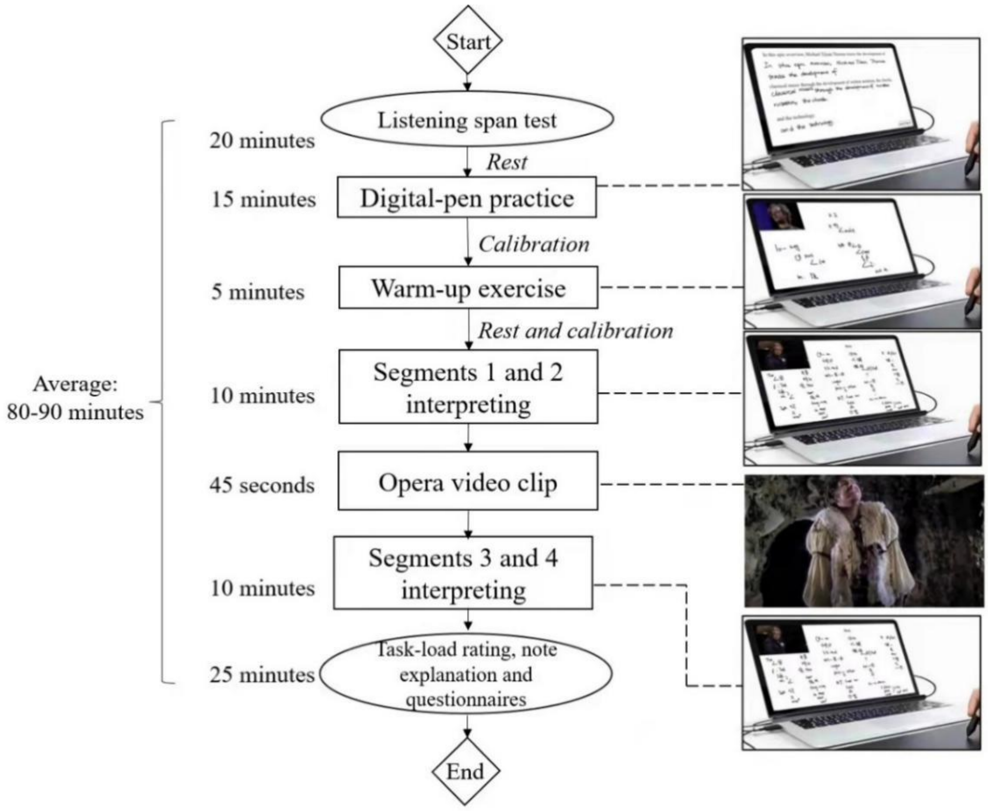


Figure 2. Experimental procedures

3.5. Data and analysis

3.5.1 Data screening

The quality of the eye-tracking data was assessed with three criteria: Mean Fixation Duration, Gaze Time on Screen, and Gaze Sample to Fixation Percentage (Hvelplund, 2014). Datapoints of $P_4$, $P_{15}$, $P_{33}$ and $P_{59}$ were excluded from the analysis because they did not meet at least two of the three criteria (lower than one standard deviation of the sample's mean). In addition, $P_{42}$ and $P_{53}$ were excluded for reporting discomfort with the digital pen. Overall, 10.91% of the data was invalid. Ultimately, we analysed data collected from 20 professionals (17 females and 3 males) and 29 students (25 females and 4 males). The two groups performed similarly ($t(47)=-1.612$, $p>.05$, $d=0.48$) in the English listening span task.

### 3.5.2 Analysis of eye-tracking and pen-recording data

The screen recordings were divided into different Time of Interest (TOI) in the eye-tracking software Tobii Pro Lab according to the two noting processes in CI and the four segments of interpreting. Then, we drew AOI on each note and exported all the fixation data. The adopted measures were as follows.

**Total Fixation Duration** (TFD): referring to Hvelplund (2019), we added the fixations on every note in each TOI to indicate the interpreters' total overt visual attention paid to the note areas during note-taking and note-reading processes.

**Mean Fixation Duration** (MFD): in accordance with other interpreting studies such as Chen et al. (2021), Dragsted and Hansen (2009) and Stachowiak-Szymczak and Korpal (2019), we used MFD to reflect the overall involved cognitive processing during taking or reading one note on average.

**Revisit Count** (RVC): this measure indicates the times that eyes re-enter and leave an AOI (note), and points to the cognitive processing of incomplete note-processing (Chen et al. 2021).

**Click Count** (CC): this measure is the accumulated time of mouse clicks (strokes) in an AOI (note), and was adopted to measure the physical effort of note-writing. For example, in Chinese, both *缺陷* and *不足* mean "shortage", but writing the former word requires more time and effort than writing the second one.

**Ear-Pen Span** (EPS): the onset of note-taking was obtained by exporting the Time to First Click in AOIs (notes) from Tobii Pro Lab, and the onset of source speech units was gained by automatic speech recognition and human correction. Extreme EPS values that were three standard deviations higher or lower than the mean of each participant's EPS in each TOI were excluded (Chen, 2020b).

**First Fixation Duration** (FFD): it represents the length of the first fixation in an AOI (note), indicating the cognitive processing of “early-stage processing such as recognition and identification” (Chen et al., 2021, p. 7) in note-reading.

**First Visit Duration** (FVD) and **Second Visit Duration** (SVD): the former measure indicates the duration of fixations and saccades between when eyes enter and leave an AOI (note) for the first time, and the latter measure represents the total duration that eyes stay in an AOI except for the FVD. According to Irwin (2004), cognitive processing could happen during both fixations and saccades. The visit duration was correlated with sentence length and syntactical complexity (Mishra et al., 2013), and it prolonged as translation difficulty increased (Liu et al., 2019). In the present study, FVD and SVD respectively indicated the cognitive effort of integrating the meaning of notes at a sentence level and a textual level, with the former measure focusing on early-stage note-reading and the latter one on late-stage note-reading.

### 3.5.3 Analysis of note patterns and interpretation quality

We categorized the participants’ notes into symbols, numbers and language notes of different forms and languages. Since the participants’ note quantity varied, we used proportions rather than actual counts to represent the interpreters’ note patterns.

Interpreting performance was evaluated by two experienced native-Chinese interpreter trainers from three aspects: information completeness (InfoCom), fluency of delivery (FluDel) and target language quality (TLQual) (Han, 2019). The total score for each aspect of quality assessment is 8 and it is evenly distributed into 4 bands. The raters first decided which band the interpretation belonged to and then gave a specific score. InfoCom was given a weight of 2, and the other two measures were each given a weight of 1 (Lee, 2015). A Pearson’s correlation test with a 95% confidence interval showed that all the scores given by the two raters were significantly correlated (Table 4). The coefficients for TLQual were noticeably low, but we did not deliberately “manipulate” them because such phenomena have also been observed in interpreting assessment studies (e.g., Han, 2019). Moreover, the correlation coefficients for the total scores were above 0.70, indicating strong associations (Cohen, 1992) between the scores given by the two raters. Therefore, we continued data analysis with the collected scores.

Table 4. Pearson's correlation test results of the two raters' scores

| | **InfoCom** | **FluDel** | **TLQual** | **Total score** |
|---|---|---|---|---|
| **Segment 1** | 0.726*** | 0.736*** | 0.527*** | 0.769*** |
| **Segment 2** | 0.719*** | 0.752*** | 0.545*** | 0.731*** |
| **Segment 3** | 0.724*** | 0.846*** | 0.489*** | 0.736*** |
| **Segment 4** | 0.715*** | 0.747*** | 0.438*** | 0.725*** |

*Note*. * $p<.05$, ** $p<.01$, and *** $p<.001$. This applies to all tables.

### 3.5.4 Correlation tests

Shapiro-Wilk test results showed that a proportion of our data was not normally distributed. However, conducting parametric (Pearson's) and non-parametric (Spearman's) correlation tests at the same time would make the results incomparable. According to Tabachnick and Fidell (2007), when the group size ratio is less than 4:1 and $F_{max}$ is less than 10, non-normally distributed data can be tested with parametric tests. We checked the ratio of group sample size (29:20) and $F_{max}$ of all datasets, finding that all $F_{max}$ values were within 1.00-1.50. Therefore, Pearson's correlation tests were conducted to examine the relationships between interpreters' note-taking behaviour and CI quality (Figure 4). We also performed additional non-parametric tests to ensure that Pearson's and Spearman's tests returned the same results. To ensure the consistency of the reported results, only Pearson's test results are reported.

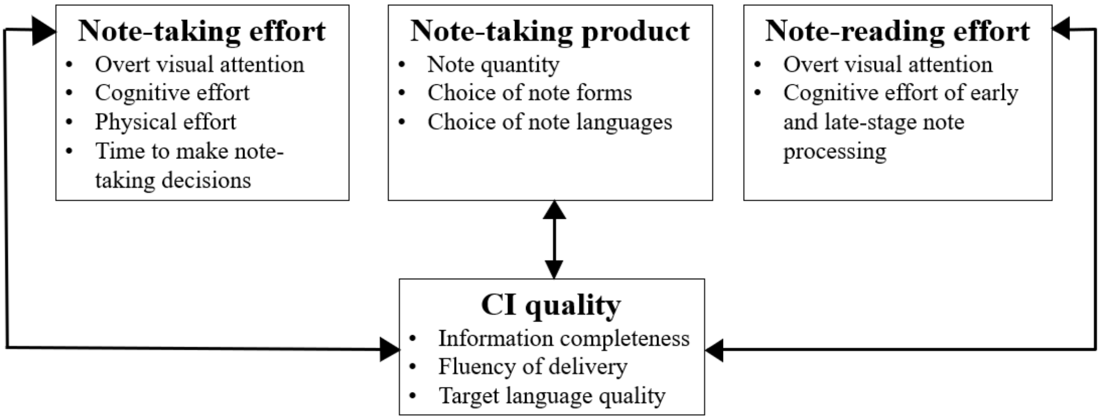


Figure 4. Tested correlations

## 4. Results

### 4.1. Interpreting performance

Paired samples *t*-test results show that the students performed similarly across the easy and difficult segments of interpreting, whereas the professionals performed significantly better in the easy segments than in the difficult ones (Table 5). Moreover, independent *t*-tests show that the professionals had significantly better performance than the students in the easy segments in every aspect (InfoCom: $t(70.752)=-6.388$, $p<.001$, $d=1.29$; FluDel: $t(96)=-3.005$, $p<.01$, $d=0.62$, TLQual: $Z=-3.763$, $p<.001$), but not in the difficult ones.

Table 5. Mean interpreting scores and *t*-test results between easy and difficult tasks

| | Students | | | Professionals | | |
|---|---|---|---|---|---|---|
| | Easy M (SD) | Difficult M (SD) | *t*-test *p* | Easy M (SD) | Difficult M (SD) | *t*-test *p* |
| **InfoCom** | 3.45(1.04) | 3.22(1.82) | >.05 | 4.98(1.32) | 3.63(1.33) | <.001 |
| **FluDel** | 4.92(1.06) | 4.71(1.08) | >.05 | 5.63(1.24) | 5.03(1.32) | <.01 |
| **TLQual** | 4.70(0.82) | 4.55(0.92) | >.05 | 5.48(0.96) | 4.85(1.08) | <.01 |
| **Total score** | 16.52(3.64) | 15.71(4.01) | >.05 | 21.05(4.65) | 17.13(4.70) | <.001 |

### 4.2. Correlations between the note-taking behaviour and interpreting performance

All the significant correlations between interpreters' note-taking behaviour and CI quality observed in the present study are presented in Table 6.

Table 6. Significant correlations between the participants' note-taking behaviour and interpreting performance

| Stages | Aspects | Students | | Professionals | |
|---|---|---|---|---|---|
| | | Easy | Difficult | Easy | Difficult |
| **Note-taking product** | Quantity | All aspects + | | All aspects + | |
| | Symbol | | InfoCom + | | |
| | Full-word | | InfoCom & Total − | | |
| | English | | All aspects − | | |
| **Note-taking process** | Cognitive effort | | | All aspects − | All aspects − |
| | Time spent on note-planning | | | FluDel − | |
| | Physical effort | FluDel + | | | |
| **Note-reading process** | Visual attention | | | All aspects − | |
| | Overall cognitive effort | FluDel − | | FluDel，TLQual & Total − | |
| | Incomplete processing | | FluDel, TLQual & Total + | | |
| | Early processing | | InfoCom & Total − | FluDel − | |
| | Late processing | | TLQual + | All aspects − | FluDel − |

*Note.* "+" represents a positive relationship and "-" means a negative relationship.

### 4.2.1. The product of note-taking and CI quality

Both groups of participants show positive and weak correlations between their note quantity and CI quality in the easy segments, but not in the difficult ones (Table 7). As for the choice of note form and note language, only the students' performance shows weak correlations in the difficult segments (Table 8).

Table 7. Correlations between note quantity and CI quality

| | | InfoCom | FluDel | TLQual | Total score |
|---|---|---|---|---|---|
| **Professionals** | Easy | 0.426** | 0.390* | 0.369* | 0.423** |
| | Difficult | 0.242 | 0.171 | 0.060 | 0.209 |
| **Students** | Easy | 0.267* | 0.410** | 0.287** | 0.337** |
| | Difficult | 0.231 | 0.136 | 0.036 | 0.181 |

Table 8. Correlations between note choices and CI quality among student interpreters in the difficult segments

| | InfoCom | FluDel | TLQual | Total score |
|---|---|---|---|---|
| **Symbol** | 0.280* | 0.139 | 0.095 | 0.224 |

| | | | | |
|---|---|---|---|---|
| **Language** | -0.243 | -0.167 | -0.132 | -0.218 |
| **Full word** | -0.319* | -0.189 | -0.122 | -0.266* |
| **Abbreviation** | 0.116 | 0.043 | 0.001 | 0.080 |
| **Chinese** | 0.156 | 0.189 | 0.138 | 0.174 |
| **English** | -0.413** | -0.384** | -0.289** | -0.412** |

4.2.2. The process of note-taking and CI quality

In the easy segments, the student group's interpreting fluency is positively correlated with their physical effort (CC) of handwriting at a weak level ($r$=0.322, $p$<.05), whereas that of the professionals is weakly and negatively correlated with their EPS ($r$=-0.328, $p$<.05). In addition, the professionals show weak to medium negative correlations between the cognitive effort (MFD) involved in the note-taking process and their interpreting performance in all segments (Table 9).

Table 9. Correlations between the professionals' MFD during note-taking and CI quality

| | **InfoCom** | **FluDel** | **TLQual** | **Total score** |
|---|---|---|---|---|
| **Easy** | -0.357* | -0.543** | -0.357* | -0.422** |
| **Difficult** | -0.320* | -0.345* | -0.293 | -0.341* |

4.2.3. The process of note-reading and CI quality

In the easy segments, the students' interpreting fluency is negatively correlated ($r$=-0.270, $p$<.05) with the involved cognitive effort (MFD). In the difficult segments, their CI quality shows positive correlations with the cognitive effort of incomplete processing (RVC) and negative correlations with the cognitive effort of note recognition (FFD) and late-stage textual processing (SVD) (Table 10).

Table 10. Correlations between note-reading effort and CI quality among student interpreters in the difficult segments

| | **InfoCom** | **FluDel** | **TLQual** | **Total score** |
|---|---|---|---|---|
| **RVC** | 0.182 | 0.280* | 0.354* | 0.263* |
| **FFD** | -0.289* | -0.253 | -0.0128 | -0.267* |

| SVD | 0.097 | 0.149 | 0.260* | 0.157 |
|---|---|---|---|---|

The professional group show various weak correlations between their note-reading effort and CI quality in the easy segments, but only one correlation with the cognitive effort of late-stage processing (SVD) is observed in the difficult segments (Table 11).

Table 11. Correlations between note-reading effort and CI quality among professional interpreters

| | | InfoCom | FluDel | TLQual | Total score |
|---|---|---|---|---|---|
| **Easy** | TFD | -0.394* | -0.447** | -0.346* | -0.415** |
| | MFD | -0.265 | -0.434** | -0.326* | -0.334* |
| | FFD | -0.126 | -0.318* | -0.253 | -0.209 |
| | SVD | -0.425** | -0.436** | -0.431** | -0.447** |
| **Difficult** | SVD | -0.264 | -0.356* | -0.246 | -0.302 |

## 5. Discussion

### 5.1. Weak correlations between note-taking behaviour and interpreting performance

Overall, we observe a few weak ($r$<.50) significant correlations between the interpreters' note-taking behaviour and CI quality, which corroborates previous findings of no close or clear relationships between many aspects of note-taking and interpreting performance (e.g., Chen, 2017; Dai & Xu, 2007; Wang et al., 2010). This suggests that note activities as subtasks in CI (e.g., Gile, 2009; Gillies, 2005) are not closely related to CI quality. After all, "note-taking is just a means, and not the end, of CI" (Viezzi, 2013, as cited in Russel & Takeda, 2015, p.103). This loose connection between note-taking activities and interpreting performance has previously been reported in Chen (2017) where she scored interpreters' notes according to whether they correctly represented source-speech units and were correctly translated in the target speech. Chen (2017) found that high-quality interpretation always appeared with high note-taking scores, but that highly scored notes did not necessarily yield successful interpretation. Such results indicate that good note-taking is only a part of the conditions needed to produce high-quality interpretation. In the following discussion, we explain how this weak association between note-taking and CI quality exists (does not exist) in different interpreter groups and task conditions.

## 5.2. The relationship between CI quality and the production and reception of notes

Many significant correlations observed in the present study are in the note-reading stage rather than in the note-taking stage. One primary reason is that note-reading is directly related to the production of the target speech, as the two tasks are conducted concurrently. According to the limited-capacity resource model (Kahneman, 1973), concurrent tasks usually induce attentional conflicts as more processing capacity devoted to one task will lead to less capacity available for other tasks. Therefore, smooth note-reading can facilitate information recall and leave more cognitive resources for target speech production and monitoring, but arduous note-reading can impede speech organization and hinder speech delivery. For instance, Shen and Liang (2020) discovered that, quite often, student interpreters could not understand the logical relations in their notes. Once they found the sentence structure to be inappropriate in the delivered target speech, they would try another logic to link the meaning of individual notes and re-deliver the interpretation. Therefore, when too much effort was allocated to reading notes, little effort was available to ensure output quality.

Another reason for the observed correlations between the note-reading effort and the participants' interpreting performance could be attributed to the adopted note-taking strategy. It has been proven that the choices interpreters make during note-taking could exert impacts on the effort of note-reading (Chen et al., 2021). According to Craik and Lockhart's (1972) hypothesis of levels of processing, a deeper level of information processing could lead to a stronger memory trace of the processed information in recall. In that case, more SL processing should be involved during note-taking to facilitate interpreters' memory recall during note-reading. We classified interpreters' note-taking strategies into ellipsis and non-ellipsis based on Albl-Mikasa (2006), with the former being a simple reduction of the source information and the latter including more demanding strategies such as syntactical restructuring during note-taking. Results show that over 85% of the two groups' notes in each segment were created through the "ellipsis" strategy. In other words, only a superficial level of language processing was involved during note-taking in most situations. Albl-Mikasa (2006) points out that, although the ellipsis strategy is effort-saving during note-taking, it can be demanding during note-reading because interpreters usually fail to make up for the inadequate comprehension of the source speech and to deliver interpretation within seconds. With more difficulties in note-processing, the effort of note-reading increases and the interpretation quality reduces. Previous research has found a preference for the ellipsis strategy among student interpreters (Albl-Mikasa, 2006), and the present study demonstrates the same strategy preference among professional interpreters. It is also frequently reported that interpreters had problems in note-reading, such as illogical notes that caused restarts in interpreting (Shen & Liang, 2020), and unrecognizable notes that caused omissions and repetitions in interpreting (Arumí Ribas, 2012). Taken together, these results indicate that the strategy of note-taking affects the effort of note-reading; while the effort of note-reading further determines the amount of processing capacity that is available for target speech production and monitoring. The connection between note-reading and CI quality is consequently established.

For the same reason, with an ellipsis-dominated note-taking approach, both groups in the present research only show positive correlations between their note quantity and interpretation quality in the easy segments but not in the difficult ones. Those aspects of note-taking that are linked with increased note quantity in the easy segments, such as shorter note-planning (EPS) among the professionals and more handwriting (CC) among the students, also presented correlations that pointed to the same direction. This is because it was more difficult for the participants to make up for the inadequate comprehension of the source speech in the difficult segments than in the easy segments, which further affected their interpreting performance. This effect of task difficulty also explains why such a positive association between interpreters' note quantity and CI quality has sometimes been observed in previous literature (e.g., Chen, 2020b; Dam, 2007; Her, 2001) and sometimes not (e.g., Dai & Xu, 2007; Liu, 2010; Wang et al., 2010).

## 5.3. The relationship between note-taking and CI quality in students and professionals

Overall, the student group show many correlations between their note-taking behaviour and CI performance in the difficult segments, which mainly involved the product of note-taking and the process of note-reading. In comparison, the professional group present almost all correlations in the easy segments, which concerned both the processes of note-taking and note-reading.

One primary reason for the observed correlations among the student group is their inadequate note-taking expertise. In the difficult segments, the students showed various correlations between their note choices and CI quality. They were troubled by the note form that requires much physical effort of handwriting (full words) and the note language (Chen, 2020b) that demands much effort for note-reading (English as the L2 and TL in our case) (Chen et al., 2021). By comparison, no significant correlations were observed in the easy segments. Nor were these significant correlations observed among the professionals. This finding indicates that, when the demands of comprehending the source speech increased, the students were occupied by source speech processing and could not flexibly retrieve note forms and note languages from their long-term memory. Another reflection of this effect of source speech difficulty on the relationship between their note-taking behaviour and interpretation quality is the change in their interpreting performance in different aspects. In the easy segments, the students' note-reading effort was mainly associated with fluency issues. However, in the difficult segments, the other two aspects of interpretation quality, namely information completeness and TL quality, also show many correlations. This can be caused by the fact that the ellipsis strategy postponed the comprehension and interpreting tasks of the source speech from the process of note-taking to the process of note-reading (Albl-Mikasa, 2006). When it comes to the difficult segments where the cognitive demand for comprehension and language transfer was especially high, if they could not recognize the notes in early processing (FFD) quickly, their information completeness could be directly affected. Extra effort was also needed in late processing (RVC and SVD) to maintain their interpreting quality.

Unlike the student interpreters, the professional interpreters show most significant correlations between their note-reading effort and CI quality in the easy segments. We calculated the professionals' articulation rate in the two task conditions, which is an important determinant of judged fluency, by dividing the total duration of speech (apart from silent and filled pauses) by the number of syllables in the target speeches (Yu & van Heuven, 2017). The results show that the professionals' articulation rates were similar in the two task conditions (easy: $M$=4.95, $SD$=0.67, difficult: $M$=4.85, SD=0.68, $Z$=-1.459, $p$>.05); and their articulation rates were always significantly higher than those of the students ($p$<.01). This means that, compared with the difficult segments, the professionals had to read a greater number of notes at one time in the easy segments to ensure a continuous interpretation. With more processing capacity devoted to note-reading, less was available for target speech production and monitoring (Kahneman, 1973). This could be the reason why negative correlations are observed between the professionals' note-reading effort and CI quality in the easy segments. Despite their high interpreting scores in the easy segments, these negative correlations indicate that a large note quantity can cause note-reading problems, and that the effort of note-reading is associated with the quality of interpretation.

In addition, except for the negative correlations between the professionals' CI quality on the one hand, and their cognitive effort of note-taking and late-stage processing in note-reading on the other hand, no other significant correlations are observed among the professional interpreters in the difficult segments of interpreting. One explanation for the negative relationship concerning the cognitive aspects of note-taking and note-reading could be the professionals' cognitive overload during interpreting. Despite their interpreting experience, the professionals reported even more mental demand (professionals: *M*=6.93, *SD*=1.46, students: *M*=6.59, *SD*=1.24) and a higher level of frustration (professionals: *M*=6.18, *SD*=2.16, students: *M*=5.95, *SD*=1.73) in the difficult segments than those reported by the students. Their performance dropped significantly in the difficult parts and was only slightly better than that of the students. These findings suggest that the professionals were highly sensitive to task difficulty and might experience overload during interpreting. Similar findings have been reported in Hu (2008) where professionals' cognitive effort of note-taking was especially sensitive to source speech difficulty, and in Cardoen (2018) where professionals did not outperform students in difficult CI tasks. Taken together, these results indicate that the role of note-taking can be very limited when professional interpreters perceive much difficulty in CI. However, further investigation is needed to examine how this limitation is established during the process of interpreting.

## 6. Conclusion

This study examined the relationship between professional interpreters' and student interpreters' note-taking behaviour and CI quality in easy and difficult CI tasks through process- and product-oriented methods. Overall, only a few weak correlations are observed between the participants' note-taking behaviour and CI quality. Both groups show positive and weak correlations between their note quantity and CI quality in the easy segments of interpreting but not in the difficult ones. Compared to the process of note-taking, the process of note-reading is shown to be more closely-associated with CI quality because it is concurrent with target speech production. Moreover, the students present most correlations in the difficult segments because of their limited note-taking expertise, whereas the professionals show most correlations in the easy ones which could be attributed to their large note quantity. All these findings suggest that note-taking should be introduced and applied with more discretion in interpreter training and interpreting practice.

It is the combination of eye-tracking and pen-recording measures that helped us to examine the note-taking and note-reading processes in CI in the present study. Eye fixations that point to cognitive processing and click counts that reflect the physical effort of handwriting might be good indices for future research on the production and reception of notes in CI. Nonetheless, this methodology could add a limitation onto our research, as the interpreters were asked to take notes with a smart pen on a digital pad, rather than with a normal pen on a notepad. In addition, the task duration (around 20 minutes) in the present study was not as long as that in some real-life CI scenarios. Therefore, our findings might not be applicable to CI tasks that last for hours or even a whole day. Since interpreters' fatigue level can directly affect their memory span and interpreting performance, the role of note-taking in these situations needs further investigation.

## References


Albl-Mikasa, M. (2006). Reduction and expansion in notation texts. In C. Heine, K. Schubert & H. Gerzymisch-Arbogast (Eds.), *Text and translation: Theory and methodology of translation* (pp.195-214). Narr.

Andres, D. (2002). *Konsekutivdolmetschen und notation* [Consecutive interpreting and note-taking]. Peter Lang.

Arumí Ribas, M. (2012). Problems and strategies in consecutive interpreting: A pilot study at two different stages of interpreter training. *Meta*, *57*(3), 812-835.

Cai, R., Dong, Y., Zhao, N., & Lin, J. (2015). Factors contributing to individual differences in the development of consecutive interpreting competence for beginner student interpreters. *The Interpreter and Translator Trainer*, *9*(1), 104-120.

Cardoen, H. (2018). *Efficient note-taking in consecutive interpreting: Qualitative notes as a polymorph tool* (Unpublished doctoral dissertation). University of Mons.

Chen, S. (2017). Note-taking in consecutive interpreting: New data from pen recording. *Translation & Interpreting*, *9*(1), 4-23.

Chen, S. (2020a). The impact of directionality on the process and product in consecutive interpreting between Chinese and English: Evidence from pen recording and eye tracking. *Journal of Specialised Translation*, *34*, 100-117.

Chen, S. (2020b). The process of note-taking in consecutive interpreting: A digital pen recording approach. *Interpreting*, *22*(1), 117-139.

Chen, S. (2021). The process and product of note-taking and consecutive interpreting: empirical data from professionals and students. *Perspectives*, 1-17. doi: 10.1080/0907676X.2021.1909626.

Chen, S., Kruger, J. L., & Doherty, S. (2021). Reading patterns and cognitive processing in an eye-tracking study of note-reading in consecutive interpreting. *Interpreting*, *23*(1), 76-102.

Chuang, L. (2008). Note-taking know-how: A processing perspective on consecutive interpreting. *Spectrum: Studies in Language, Literature, Translation, and Interpretation 2*, 93-101.

Cohen, J. (1992). A power primer. *Psychological Bulletin*, *112*(1), 155-159.

Craik, F. I. M., & Lockhart, R. S. (1972). Levels of processing: A framework for memory research. *Journal of Verbal Learning and Verbal Behavior, 11*, 671-684.

Dai, W., & Xu, H. (2007). *汉英交替传译过程中译员笔记特征实证研究——以职业受训译员和非职业译员为例* [An empirical study of the features of interpreters' notes in Chinese-English consecutive interpreting: The examples of professionally trained and unprofessional interpreters]. *Foreign Language Teaching and Research*, *39*(2), 136-144.

Dam, H. V. (2007). What makes interpreters' notes efficient? Features of (non-) efficiency in interpreters' notes for consecutive. In Y. Gambier, M. Shlesinger & R. Stolze (Eds.), *Doubts and directions in translation studies* (pp. 183-197). John Benjamins.

Dam, H. V., Engberg, J., & Schjoldager, A. (2005). Modelling semantic networks on source and target texts in consecutive interpreting: A contribution to the study of interpreters' notes. In H. V. Dam, J. Engberg & H. Gerzymisch-Arbogast (Eds.), *Knowledge systems and translation* (pp. 227-254). Mouton de Gruyter.

Dragsted, B., & Hansen, I. (2009). Exploring translation and interpreting hybrids. The case of sight translation. *Meta*, *54*(3), 588-604.

Gile, D. (2009). *Basic concepts and models for interpreter and translator training*. John Benjamins.

Gile, D. (2020). The effort models and gravitational model: Clarifications and update. [PowerPoint slides]. ResearchGate. https://www.cirinandgile.com/powerpoint/The-Effort-Models-and-Gravitational-Model-Clarifications-and-update.pdf.

Gillies, A. (2005). *Note-taking for consecutive interpreting: A short course*. St.

Jerome.

Han, C. (2019). A generalizability theory study of optimal measurement design for a summative assessment of English/Chinese consecutive interpreting. *Language Testing*, *36*(3), 419-438.

Her, H. (2001). Notetaking in basic interpretation class: An initial investigation. *Studies of Translation and Interpretation*, *6*, 53-77.

Hu, L. (2008). *Note-taking in consecutive interpreting: A psycholinguistic perspective* (Unpublished doctoral dissertation). Shanghai International Studies University.

Hvelplund, K. T. (2014). Eye tracking and the translation process: Reflections on the analysis and interpretation of eye-tracking data. In Muñoz Martín, R. (Ed.), *MonTI special issue 1: Con la traducción en mente* [Minding translation], (pp. 201-223). Publicaciones de la Universided de Alicante.

Hvelplund, K. T. (2019). Digital resources in the translation process–attention, cognitive effort and processing flow. *Perspectives*, *27*(4), 510-524.

Irwin, D. E. (2004). Fixation location and fixation duration as indices of cognitive processing. In J. Henderson & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and visual world* (pp. 105-134). Psychology Press.

Kahneman, D. (1973). *Attention and effort.* Prentice-Hall.

Lee, S. B. (2015). Developing an analytic scale for assessing undergraduate students' consecutive interpreting performances. *Interpreting*, *17*(2), 226-254.

Liu, M., & Chiu, Y. H. (2009). Assessing source material difficulty for consecutive interpreting: Quantifiable measures and holistic judgment. *Interpreting*, *11*(2), 244-266.

Liu, M., Schallert, D. L., & Carroll, P. J. (2004). Working memory and expertise in simultaneous interpreting. *Interpreting*, *6*(1), 19-42.

Liu, J. (2010). *英语专业本科学生汉英交传笔记特征及其与口译成绩的关系* [Note-taking characteristics of English majored undergraduates in Chinese-English consecutive interpreting: An empirical study based on students' consecutive interpreting notes]. *Foreign Language World*, *2*, 47-53.

Liu, Y., Zheng, B., & Zhou, H. (2019). Measuring the difficulty of text translation: The combination of text-focused and translator-oriented approaches. *Target*, *31*(1), 125-149.

Mishra, A., Bhattacharyya P., & Carl, M. (2013). Automatically predicting sentence translation difficulty. In *Proceedings of the 51st Annual Meeting of the Association for Computational Linguistics* (pp. 346-351). Sofia.

Russell, D., & Takeda, K. (2015). Consecutive interpreting. In H. Mikkelson & R. Jourdenais (Eds.), *The Routledge handbook of interpreting* (pp. 96-111). Routledge.

Setton, R., & Dawrant, A. (2016). *Conference interpreting–A trainer's guide*. John Benjamins.

Shen, M., & Liang, J. (2020). Self-repair in consecutive interpreting: similarities and differences between professional interpreters and student interpreters. *Perspectives*, *29*(5), 761-777.

Stachowiak-Szymczak, K., & Korpal, P. (2019). Interpreting accuracy and visual processing of numbers in professional and student interpreters: An eye-tracking study. *Across Languages and Cultures*, *20*(2), 235-251.

Sun, S., & Shreve, G. M. (2014). Measuring translation difficulty: An empirical study. *Target*, *26*(1), 98-127.

Tabachnick, B. G., & Fidell, L. S. (2007). *Using multivariate statistics (5th edition)*. Pearson.

Yu, W., & van Heuven, V. J. (2017). Predicting judged fluency of consecutive interpreting from acoustic measures: Potential for automatic assessment and pedagogic implications. *Interpreting*, *19*(1), 47-68.

Wang, W., Zhou, D., & Wang, L. (2010). *口译笔记特征与口译产出质量实证研究* [An empirical study of note taking characteristics and output quality in interpreting]. *Foreign Language World*, *4*, 9-18.

[1] Chen (2020a) and Chen (2020a) are based on the same project, but the former included 22 data points and the latter only had 18 after eye-tracking data screening.

[2] In Chen's series research, symbols included those that were created based on Chinese characters and English words. For example, the Chinese character 心 ("heart") was used to symbolize its associate meaning "love", and the English letter B (an abbreviation of "But") was used to symbolize adversative relations such as "however", "on the other hand" and "although".

[3] This is calculated by dividing the proposition count by the word count in CPIDR 5 (http://ai1.ai.uga.edu/caspr/). The higher the density, the more informative the material.

[4] The readability indices used in this article include the Automated Readability Index (ARI), the Flesch-Kincaid index, the Coleman-Liau index, the Gunning Fog index, the SMOG index, the Flesch Reading Ease Score index, and LIX. As all the indices show the similar results regarding the readability levels of the four segments, we only choose to present Reading Ease Score as one of the representative indices. Details about the readability indices can be retrieved from https://www.webfx.com/tools/read-able/.