Routledge
Taylor & Francis Group

Perspectives

Studies in Translation Theory and Practice

ISSN: (Print) (Online) Journal homepage: https://www.tandfonline.com/loi/rmps20

The effects of training on reading behaviour and performance in sight translation: a longitudinal study using eye-tracking

Jing Fang, Xiaomin Zhang & Haidee Kotze

To cite this article: Jing Fang, Xiaomin Zhang & Haidee Kotze (2023) The effects of training on reading behaviour and performance in sight translation: a longitudinal study using eye-tracking, Perspectives, 31:4, 655-671, DOI: 10.1080/0907676X.2022.2030372

To link to this article: https://doi.org/10.1080/0907676X.2022.2030372

Published online: 27 Jan 2022.

PERSPECTIVES
2023, VOL. 31, NO. 4, 655–671
https://doi.org/10.1080/0907676X.2022.2030372

The effects of training on reading behaviour and performance in sight translation: a longitudinal study using eye-tracking

Jing Fang [a], Xiaomin Zhang [a] and Haidee Kotze [b,c]

[a]Department of Linguistics, Macquarie University, Sydney, Australia; [b]Utrecht University, Utrecht, The Netherlands; [c]Vaal Triangle Campus, North-West University, Vanderbijlpark, South Africa

ABSTRACT

This longitudinal study focuses on the impact of training on trainee interpreters' reading behaviour and performance during sight translation (ST) when dealing with texts of different syntactic complexities, compared to a control group. Participants sight translated four texts, two semesters apart, while their eye movements were captured. Mixed-effects modelling was used to investigate how the independent variables of Complexity, Test Time and Group affect reading behaviour and the ST performance. The findings show that while two semesters of training had limited impact on the reading behaviour of the experimental group, their ST performance improved significantly. It was also found that the effect of training on reading behaviour is mediated by skill level, with greater effects for participants of around or below average performance level.

ARTICLE HISTORY
Received 4 May 2021
Accepted 9 January 2022

KEYWORDS
Sight translation; training effects; reading behaviour; eye-tracking; longitudinal study

1. Introduction

Sight translation (ST) refers to an oral translation of a written text in a source language into a target language, where an interpreter may or may not have preparation time to read the source text. As ST involves real-time processing demands and oral production in the target language, it is widely defined as a mode of interpreting rather than translation (e.g., Cenkova, 2015). Correspondingly, ST is commonly included as an integral part of interpreting training in Translation and Interpreting (T&I) programmes (e.g., T&I programmes in Australia and in the UK). As a recommended interpreting pedagogy, ST is generally considered a good exercise for increasing transfer speed and thus for preparing students for simultaneous interpreting (Song, 2010).

Moreover, ST is a common form of professional practice in the T&I industry in immigration countries, such as Australia and the USA (Hlavac, 2013; Li, 2014). It is frequently used in public service settings, where written documents need to be sight translated on the spot by an interpreter. Such needs in the market have also been reflected in the certification examinations in these countries. For example, ST is a part of the national Certified Interpreter test organised by Australia's National Accreditation Authority for Translators and Interpreters (NAATI).[1]

CONTACT Jing Fang jing.fang@mq.edu.au

Performing an ST task requires specialised skills, including excellent reading skills to be able to analyse the source text rapidly and distinguish key ideas from ancillary ones (see Lee, 2012; Nilsen & Monsrud, 2015), adequate translation skills to be able to chunk translation units and apply condensation strategies to enhance delivery speed and target language quality (Lee, 2012, p. 710), swift transfer skills from one language to another while avoiding literal translation (Angelelli, 1999; Ivars, 2008), effective public speaking techniques (Weber, 1990),and multimodal mediating skills during the face-to-face interactions (Havnen, 2020).

However, despite the wide use of ST in interpreting training, and the suggested benefits of transferable skills from ST to other kinds of translation and interpreting, there has been little research investigating the impact of ST training on interpreters, especially from a longitudinal perspective. In particular, whether training could change trainees' reading behaviour, one of the most fundamental components of the ST process, has not attracted much attention (see Section 2).

Against this background, this study sets out to investigate the impact of two semesters of training (equivalent to half or two-thirds of the length of T&I training at postgraduate level in Australia) on student interpreters' reading behaviour and interpreting performance. Using eye-tracking to record student interpreters' eye movements during ST tasks at two points approximately nine months apart, this study aims to capture changes over time in reading patterns and ST performance in the presence and absence of training.

2. Existing research on ST

ST was not widely investigated until the late 1980s (see Weber, 1990 for early studies), but there has been a notable increase of research in this area (e.g., Agrifoglio, 2004; Jakobsen & Jensen, 2008; Gile, 2009; Su & Li, 2019; Havnen, 2020; Płużyczka, 2015, 2018, 2020).

Reading involves decoding and linguistic comprehension skills (Hoover & Gough, 1990), and reading while providing an oral translation in another language requires more complex skills. As Agrifoglio (2004) points out, to produce a smooth delivery in ST, the interpreter has to read ahead and identify key translation units while planning target-language expressions simultaneously. The recent introduction of eye-tracking as a research tool to study ST processes has opened new avenues in modelling the complex reading processes of ST. Some eye-tracking studies have investigated differences in reading behaviour during ST compared to other settings (e.g., Jakobsen & Jensen, 2008; Dragsted & Hansen, 2009). Other studies on ST using eye-tracking have predominately focused on identifying translation problems and difficulties at syntactic and lexical levels through structural manipulation (e.g., Shreve et al., 2010; Chmiel & Lijewska, 2019), on investigating source language interference (e.g., Płużyczka, 2013; Chmiel et al., 2020) and on testing the credibility of eye-tracking parameters (Płużyczka, 2020). Pertinent to the purposes of this study is a project by Shreve et al. (2010), in which an eye-tracking experiment was conducted to determine how the manipulation of textual complexity disrupts comprehension in ST and written translation. They found that when dealing with sentences of different syntactic complexity, participants invest more cognitive effort in complex syntax, and the syntactic manipulation seems to be more disruptive in ST than in written translation (Shreve et al., 2010).

In general, there has been limited research on the effect of training on ST process and performance. Chmiel and Mazur (2013) find that one year of interpreting training hardly has any significant impact on trainees' ST behaviour. They compared two groups of participants with one year's difference in training duration, and found no significant difference in their eye-movement patterns (and thus their reading behaviour). In exploring the training effect on advanced and beginner interpreting students, Su and Li (2021) use eye-tracking data to compare participants' ST performance in L1 and L2 translation. They find no significant difference between the two groups for L1 translation, but training is more effective for advanced students than beginners in the more challenging L2 translation. Both these studies focus on group differences between two cohorts, but it is not clear if and how training influences developmental changes in the same cohort over a training period. Therefore, more research is needed to examine the effectiveness of training in more rigorous ways (Chmiel & Mazur, 2013), involving longitudinal studies and appropriate control groups.

In this project, we specifically aim to investigate the impact of training on trainee interpreters' reading behaviour and performance in dealing with syntactic structures of different complexity during English-to-Chinese ST. Using a longitudinal research design with both an experimental and a control group, we aim to answer the following two questions:

(1) Do two semesters of training have a significant impact on T&I students' reading behaviour (as indexed by eye-tracking measures), compared to students without training, during English-Chinese ST tasks involving different syntactic complexity?
(2) Do two semesters of training have a significant impact on T&I students' ST performance (as indexed by marks), compared to students without training, during English-Chinese ST tasks involving syntactic structures of different complexity?

We will combine the answers to these two research questions to reflect on potential pedagogical implications.

3. Methodology

3.1. Stimuli

Four texts were used as stimuli for the eye-tracking experiment, adapted from the webpages of three Australian universities. All four texts provide information for new students on topics including campus safety, accommodation, and online enrolment. With all the participants being students from an Australian university, Macquarie University, who were already familiar with these topics, the choice of this subject domain aims to control a potential variable of 'context knowledge', which might impact the participants' reading behaviour if not controlled (see Zheng & Xiang, 2014). All the texts were comparable in terms of length (180–190 words), lexical frequency and syntactic complexity, tested by Coh-Metrix, a computational tool that measures texts in terms of different linguistic criteria (see Graesser et al., 2011).

Subsequently the four texts were manipulated in such a way that two texts have four sentences each containing a nominal group structure of a simplex embedded clause, as shown in Example (1); and two texts have four sentences each containing a nominal group structure of a complex embedded clause, as shown in Example (2):

(1) Also, you should report any lights *[[which are not functioning]] to Security Services.*
(2) Our on-campus housing provides a vibrant community *[[[[where some friendships [[that you will keep lifelong]] are formed]]]].*

As demonstrated by these examples, a nominal group may have an embedded clause modifying the head noun, and such an embedded structure can be either simplex (i.e., an embedded simple clause), or complex (i.e., an embedded clause with a further embedded clause within it). In an English nominal group, an embedded clause typically comes after the head noun and functions as a post-modifier (see Halliday & Matthiessen, 2014). In comparison, in a Chinese nominal group structure, all the modifying elements, including embedded clauses, come before the head noun and function as pre-modifiers. This means that translating a nominal group with an embedded clause from English into Chinese may require a structural shift and/or a shift in terms of grammatical ranks – for example, a shift from a modifier in English to a ranking clause in Chinese. It has been found that such translation shifts are more likely to occur in translating a nominal group with a complex structure than a simplex structure due to the limited affordance of premodifiers in a Chinese nominal group structure (Fang & Wu, 2010). Therefore, by manipulating the nominal group structure in the stimuli, we are able to set the nominal group complexity as an independent variable, and to observe its relationship with various eye-tracking measures. Table 1 gives the details about the complexity of the four texts after manipulation.

In this paper, the two texts featuring a simplex nominal group structure are referred to as 'simplex texts', and the two texts featuring a complex nominal group structure are referred to as 'complex texts'. Within each text stimulus, the four sentences containing a featured nominal group structure were identified as an 'area of interest' (AOI), and therefore a total of 16 AOIs (4 AOIs in each text) are defined. The four AOIs in each stimulus are located in different parts of the text and coded ordinally. An interpreting brief was developed so that the four texts could be used as ST tasks.

The developed stimuli were then trialled on eight second-year T&I students. They were asked to sight translate the texts from English into Chinese, and identify challenging words and sentences, and also rate the difficulty level of the ST task on a scale of 1–8. Based on the results, a glossary was developed listing all the potentially challenging words that had been identified during the trial, with both English and Chinese explanations provided. This glossary later was presented to each participant prior to their eye-tracking experiment so that

Table 1. Text complexity measured by linguistic criteria.

	Simplex A1	Simplex A2	Complex B1	Complex B2
Word count, number of words	179	185	188	185
CELEX word frequency for content words, mean	2.2	2.2	2.2	2.1
Coh-Metrix syntactic simplicity, *z* score (above 0 means easier than mean score)	0.12	0.01	−0.86	−0.21
Nominal group embeddedness, number of modifying words after the head noun, mean	6.5	6.75	11.75	10.2

'difficult words' would not become a variable influencing participants' eye-tracking data. In terms of the translatability of the four texts, an inter-rater reliability test was conducted based on the rating results, using a two-way mixed, consistency, average-measures ICC (McGraw & Wong, 1996). The result of ICC was in a good range (ICC = 0.61) (Cicchetti, 1994), indicating the raters reached a good degree of agreement on the translation difficulty of the four texts. In fact, the four texts all fall into a difficulty range between scale 2 and 4 (mean rating scores: 2.75; 3.125; 3.375; 3.125), and no sentences other than the AOIs were identified as 'highly challenging' in the tasks.

3.2. Participants

Two groups of participants were recruited, including an experimental group (E group) and a control group (C group).

3.2.1. E group participants: Selection criteria and profile

E group participants were required to:

(1) be a native speaker of Chinese
(2) have a most recent IELTS test (or equivalent) overall score at or above 6.5, and IELTS reading score at or above 7.0
(3) have been educated in a Mandarin-speaking country or region till Year 11 or later
(4) be aged between 18 and 40
(5) have just started T&I training as a new student at the time of the first experiment session.

Altogether 17 participants were recruited to form the E group, who were at the beginning of their first semester of T&I training as a postgraduate student when the first eye-tracking experiment session took place. They were doing an introductory level interpreting practice unit, where fundamental skills and knowledge about ST were introduced. None of the E group participants had had professional translation experience at the time of the first session.

Nine months after the first experiment session, 15 participants in the E group returned for the second session. Among them, 2 were male and 13 were female students, and they were between 22 and 36 years old with an average age of 26.

At the time of the second eye-tracking session, participants in the E group were about to complete two semesters of full-time study in T&I. In particular, training in and practice of ST had been ongoing during both semesters, where ST skills such as semantic unit identification and chunking, anticipation, and memory retention training were introduced and developed through regular practice both in and outside the classroom.

3.2.2. C group participants: selection criteria and profile

C group also consisted of 17 participants, who had to meet the same selection criteria 1–4 outlined above. However, they were required to have no T&I background and received no training in ST, and were current students studying at the same university but in a major other than T&I, including accounting, finance, engineering, computer science and psychology.

All 17 participants in the C group returned for the second experiment session. Among them, 2 were male and 15 were female students. Their ages ranged from 19 to 33, with an average age of 24.

3.3. The experiment

In both experiment sessions, the same procedure was followed. The participant completed a consent form and demographic questionnaire, and the interpreting brief and glossary were then presented. The researchers briefed the participant on the eye-tracking experiment, after which the experiment commenced. There was no time limit for a participant to sight translate a text nor restriction on when they should start, though all the participants managed to complete the whole session within an hour. Participants could not access the stimuli prior to the experiment.

During the ST tasks, an SMI RED250Hz eye-tracking system was used to collect eye movement data, and simultaneously a Logitech HD camera was placed above the screen to record the image and voice of the participant. After calibration, experimental stimuli were displayed on a 20″ screen with a resolution of 1680×1050. A complete text was displayed on one screen in a font size suitable for tracking eye movements, which had been tested during the design stage of the experiment.

During each experiment session, the participant would sight translate two texts, one simplex text and one complex text, in an order randomly selected by the computer. Participants were asked to re-calibrate after finishing one text before moving onto the next. In the second experiment session, participants completed the other two texts in the same procedure, and by the completion of the second session, all the participants had sight translated all four texts in various orders.

3.4. Data

Two types of data were collected from the experiment: eye movement data from the eye-tracker and video data from the camera.

3.4.1. Eye data

Following the experiment, the eye data were checked for quality. The data of a participant were used only if a tracking ratio above 90% had been achieved. As a result, the data from 13 participants in the E group and 15 in the C group were ultimately used in the analysis.

Three types of eye measures in the AOIs were used in the analysis: dwell time (the total duration in milliseconds of all fixations and saccades within an AOI), revisits (the total number of glances[2] within an AOI except the first glance), and fixation count (the total number of fixations within an AOI). Reading research has indicated that these eye measures are related to various aspects of language processing during reading and speaking (e.g., Rayner, 1998; Holmqvist et al., 2011). Furthermore, existing ST studies based on eye-tracking have also demonstrated that these eye measures are able to provide reliable data in investigating the ST process (e.g., Chmiel & Mazur, 2013; Płużyczka, 2020). Once the raw data had been exported, these measures were normalised by using the formula: raw value/(AOI coverage in pixels/total coverage in pixels).

3.4.2. Video data

The data of the video recordings captured by the camera were used to assess the ST performance of the participants. Marking criteria were developed based on two equally weighted aspects of ST competence: accuracy (10 marks) and fluency (10 marks). The sight translation assessment criteria used by the NAATI Certified Interpreter Test were adopted as a guideline in assessing 'accuracy' (in transferring the source text message) and 'fluency' (in the manner of delivering the target text).[3] Three independent markers, who were T&I practitioners and educators, were used for assessment following a marking training session and a group marking of three participants' recordings. An inter-rater reliability test was conducted after the marking, based on a two-way mixed, absolute agreement, average-measures ICC (McGraw & Wong, 1996). The result of ICC was in an excellent range (ICC = 0.87) (Cicchetti, 1994), indicating the three markers reached a high degree of agreement on the participants' ST performance.

3.5. Data analysis

To answer the research questions, linear mixed effects modelling in SPSS (IBM, 2017) was used, with three models developed. Focusing on the first research question, the first model aims to determine the effect of different independent variables on each eye-tracking measurement used as dependent variable. The three eye-tracking measures (dwell time, revisits and fixation count) are thus used as dependent variables, whereas Group (E group vs C group), Test Time (1st vs 2nd experiment session), and Complexity (simplex vs complex) are used as independent variables. Marks is also included as an independent variable, in the form of a continuous independent covariate. Participants and AOIs are included as random variables in the model. Three-way interactions of Group, Test Time and Complexity were included in the model to estimate the fixed effects of different combinations of the independent variables with each dependent variable. Following the advice of a research statistician, a Bonferroni-adjusted significance test for pairwise comparisons, which determines whether the simple main effect of one variable is statistically significant at a given level of another variable, is run to better understand the results. The findings from this model are presented and discussed in Section 4.1.

The second and the third mixed effects models aim to answer the second research question by further exploring the relationship between training and performance, also looking in detail at the specific training effect on E group participants' eye data at different performance levels. In the second model, the overall performance marks of participants are used as the dependent variable, whereas Test Time, Group and Complexity are used as independent variables. This is to determine whether the E group demonstrated significant improvement in performance relative to the C group as indexed by Marks. Combining the findings regarding both research questions, we develop a third model to further explore the training effect in relation to reading

Table 2. Fixed effects: dwell time.

Source	Sig. (*p*-value)
Complexity	.58
Test_Time	.01
Group	.75
Complexity*Test_Time	.81
Complexity*Group	.08
Test_Time*Group	.06
Complexity*Test_Time*Group	.15
Dependent Variable: Normalised_Dwell_Time	

behaviour and performance of the E group. More details about this model are provided in Section 4.2.2.

4. Findings and discussion

4.1. Group effects, text complexity and training impact

In this section, the discussion will focus on the first research question. Using a mixed effects model, we examine whether Group (E group vs C group) and Test Time (1st vs 2nd experiment session), which are independent variables indicating training effects, have significant effects on the three dependent variables (dwell time, fixation count, and revisits) in tasks of varying complexity.

4.1.1. Dwell time

Table 2 presents a summary of the fixed effects from the mixed model showing what factors and interactions have significant effects on dwell time.

As Table 2 shows, neither two-factor nor three-factor interactions are significant, but the main effect of Test Time is significant ($p = .01$). It is also worth noting that one of the two-factor interactions, Test_Time*Group, approaches significance ($p = .06$). To further explore the interaction between Group and Test Time, two pairwise comparisons were conducted. Table 3 estimates the simple main effect of Group in the presence of Complexity and Test Time as independent variables ($n = 423$ observations).

As shown in Table 3, in both experiment sessions, the C group spent less dwell time in the complex texts than the E group, although the differences fail to achieve statistical significance ($p = .85$ and .62). However, for the first session, there is a significant Group difference for the simplex texts ($p = .04$), with the C group participants spending much longer time in the AOIs than the T&I students in the E group (5766.38 ms longer).

Table 3. Dwell time: pairwise comparison of Group by Complexity by Test Time.

Test Time	Complexity	(I) Group	(J) Group	Mean difference (I-J)	Sig.[a]
1st	Complex	C_Group	E_Group	−549.49	.85
	Simplex	C_Group	E_Group	5766.38*	.04
2nd	Complex	C_Group	E_Group	−1482.79	.62
	Simplex	C_Group	E_Group	−739.05	.80

Based on estimated marginal means.
*The mean difference is significant at the .05 level.
Dependent variable: Normalised_Dwell_Time.
[a]Adjustment for multiple comparisons: Bonferroni.

Table 4. Dwell time: pairwise comparison of Test Time by Complexity by Group.

Group	Complexity	(I) Test Time	(J) Test Time	Mean difference (I-J)	Sig.[a]
C_Group	Complex	1st	2nd	2875.12	.14
	Simplex	1st	2nd	6132.99*	.00
E_Group	Complex	1st	2nd	1941.82	.32
	Simplex	1st	2nd	−372.44	.85

Based on estimated marginal means.
*The mean difference is significant at the .05 level.
Dependent variable: Normalised_Dwell_Time.
[a]Adjustment for multiple comparisons: Bonferroni.

Table 5. Fixed effects: fixation count.

Source	Sig. (*p*-value)
Complexity	.39
Test_Time	.03
Group	.69
Complexity*Test_Time	.83
Complexity*Group	.08
Test_Time*Group	.03
Complexity*Test_Time*Group	.29
Dependent Variable: Normalised_Fixation_Count	

Notably, this difference diminishes in the second session. The situation is further illustrated in Table 4 which shows the pairwise comparison of Test Time.

As indicated in Table 4, participants in the C group demonstrate, to different extents, decreased dwell time in both simplex and complex tasks in the second experiment session compared to the first, and the decrease is highly significant in the simplex tasks ($p = .001$). In comparison, no significant differences are evident in the E group either in simplex or in complex texts in the second session compared to the first ($p = .32$ and .85).

4.1.2. Fixation count

Table 5 demonstrates the fixed effects from the mixed model showing what factors and interactions have significant effects on fixation count.

As Table 5 shows, again, only the main effect of Test Time ($p = .03$) and the two-factor interaction of Test Time and Group are significant ($p = .03$). Table 6 estimates the simple main effect of Group in the presence of Complexity and Test Time as independent variables ($n = 423$ observations).

As indicated in Table 6, in terms of fixation count, there is a significant difference between the two groups in conducting the simplex tasks in the first experiment

Table 6. Fixation count: pairwise comparison of Group by Complexity by Test Time.

Test Time	Complexity	(I) Group	(J) Group	Mean difference (I-J)	Sig.[a]
1st	Complex	C_Group	E_Group	.49	.96
	Simplex	C_Group	E_Group	18.29*	.03
2nd	Complex	C_Group	E_Group	−6.42	.47
	Simplex	C_Group	E_Group	−1.57	.86

Based on estimated marginal means.
*The mean difference is significant at the .05 level.
Dependent variable: Normalised_Fixation_Count.
[a]Adjustment for multiple comparisons: Bonferroni.

Table 7. Fixation count: pairwise comparison of Test Time by Complexity by Group.

Group	Complexity	(I) Test Time	(J) Test Time	Mean difference (I-J)	Sig.[a]
C_Group	Complex	1st	2nd	11.12	.08
	Simplex	1st	2nd	16.23*	.01
E_Group	Complex	1st	2nd	4.21	.50
	Simplex	1st	2nd	−3.63	.57

Based on estimated marginal means.
*The mean difference is significant at the .05 level.
Dependent variable: Normalised_Fixation_Count.
[a]Adjustment for multiple comparisons: Bonferroni.

Table 8. Fixed effects: revisits.

Source	Sig. (*p*-value)
Complexity	.35
Test_Time	.04
Group	.17
Complexity*Test_Time	.59
Complexity*Group	.17
Test_Time*Group	.31
Complexity*Test_Time*Group	.01

Dependent Variable: Normalised_Revisits.

session, with the C group having many more fixations (18.29 more fixations). However, the situation is reversed in the second experiment session – this time, the C group demonstrate fewer fixations than the E group (1.57 fewer fixations).

The change within each group across the two sessions is more evident in Table 7.

As shown in Table 7, for simplex texts, the C group shows a significant decrease in fixation count in the second session ($p = .01$), and they also have reduced fixations in the complex texts, although the difference is non-significant ($p = .08$). In contrast, for both types of texts, training appears to have little effect on fixation counts for the E group ($p = .50$ and .57).

4.1.3. Revisits

Table 8 summarises the fixed effects from the mixed model showing what factors and interactions have significant effects on revisits. Apart from the main effect of Test Time ($p = .04$), the three-factor interaction (Complexity*Test_Time*Group) is also significant ($p = .01$).

Table 9 shows the pairwise comparison of revisits of both groups in complex and simplex texts in each experiment session ($n = 423$ observations).

Table 9. Revisits: pairwise comparison of Group by Complexity by Test Time.

Test Time	Complexity	(I) Group	(J) Group	Mean difference (I-J)	Sig.[a]
1st	Complex	C_Group	E_Group	−12.38	.68
	Simplex	C_Group	E_Group	63.58*	.03
2nd	Complex	C_Group	E_Group	55.72	.07
	Simplex	C_Group	E_Group	33.16	.28

Based on estimated marginal means.
*The mean difference is significant at the .05 level.
Dependent variable: Normalised_Revisits.
[a]Adjustment for multiple comparisons: Bonferroni.

Table 10. Revisits: pairwise comparison of Test Time by Complexity by Group.

Group	Complexity	(I) Test time	(J) Test time	Mean difference (I-J)	Sig.[a]
C_Group	Complex	1st	2nd	−19.10	.30
	Simplex	1st	2nd	39.97*	.02
E_Group	Complex	1st	2nd	49.00*	.01
	Simplex	1st	2nd	9.55	.62

Based on estimated marginal means.
*The mean difference is significant at the .05 level.
Dependent variable: Normalised_Revisits.
[a]Adjustment for multiple comparisons: Bonferroni.

As shown in Table 9, for the first experiment session, the C group is estimated to have significantly more revisits than the E group in the simplex task ($p = .03$). However, the group difference is reduced to an insignificant level for the second session ($p = .28$). In terms of revisits in the complex texts, the group difference shows the opposite trend: the E group participants demonstrate more revisits than the C group during the first experiment session (12.38 more revisits), but the trend flips in the second experiment session, with the E group having much fewer revisits than the C group (by 55.72 fewer revisits). Although the difference in the second session fails to achieve statistical significance, it is approaching the .05 level ($p = .074$), and given that the direction of the difference has reversed (from more to less) in the second experiment session, the change appears noteworthy.

Table 10 presents a closer look at the change in the number of revisits of each group across the two test times.

Table 10 shows the C group demonstrated significantly reduced numbers of revisits in the simplex text during the second session ($p = .02$), but the number of revisits in complex texts remained relatively stable ($p = .30$). In comparison, participants in the E group have significantly fewer revisits in the complex text in the second session (mean difference = 49.00, $p = .01$), while their revisits in the simplex texts remain relatively stable (mean difference = 9.55).

To sum up, the findings above indicate that, after two semesters of T&I training, T&I students' reading behaviour in terms of dwell time and fixation count have not changed much as far as the participants in the current project are concerned, which echoes the findings of Chmiel and Mazur (2013). However, the situation is different in the case of revisits.

The significant decrease in the number of revisits in complex texts suggests that professional training is likely to have had an impact on the E group participants' way of approaching a complex syntactic structure in ST. It has been argued that readers revisit more often when the text is complex (Rayner & Pollatsek, 1989), as they need to reread the information they have forgotten or are unsure about (see Booth & Weger, 2013). The reduced number of revisits in dealing with complex texts suggests that, through training, the E group participants have likely improved their working memory and/or have developed better reading comprehension and language transfer skills through ongoing practice.

Furthermore, C group demonstrates a decrease in all the eye-tracking measurements in the second experiment session, especially when dealing with simplex texts. This might be related to the fact that they became familiar with the ST procedure by participating in the first session. In other words, in the first experiment session participants in the C

Table 11. Performance scores: pairwise comparison of Test Time by Complexity by Group.

Group	Complexity	(I) Test Time	(J) Test Time	Mean difference (I-J)	Sig.[a]
C_Group	Complex	1st	2nd	−.30	.17
	Simplex	1st	2nd	−1.01*	.00
E_Group	Complex	1st	2nd	−.83*	.00
	Simplex	1st	2nd	−1.13*	.00

Based on estimated marginal means.
*The mean difference is significant at the .05 level.
Dependent variable: Mark_Overall.
[a]Adjustment for multiple comparisons: Bonferroni.

group had no baseline experience of ST and the task was thus completely unfamiliar (while participants in the E group had familiarity with the kind of task); by the second session the C group participants at least had a frame of reference and the task was not wholly unfamiliar.

4.2. The relationship between training, ST performance and reading behaviour

This section focuses on the second research question, and to answer it, we focus on the assessment results of the participants. We first present an analysis of the scores indicating the level of performance in each group at different test times. Following this, we correlate these scores with the results of the eye-tracking analysis, to explore any meaningful implications for training.

4.2.1. Performance

Table 11 provides a detailed analysis by estimating the simple main effect of Test Time on participants' scores in each type of text ($n = 112$ observations).

As shown in Table 11, in the second experiment session, the E group have significantly improved their scores in both simplex and complex texts. In comparison, participants in the C group only show significant improvement in the simplex texts.

When considering these findings alongside those presented in Section 4.1, a few interesting observations arise. Firstly, results indicate that the reading behaviour of the T&I participants did not change much after two semesters of training except for the number of revisits in complex texts. Meanwhile, their performance, as indicated by the significant increase of scores, improved significantly in both simplex and complex tasks. This implies that training did contribute to the improved ST competence, but at the same time, training only has a limited effect on changing participants' reading behaviour. It thus seems that, in order to demonstrate changes in ST competence, it is not sufficient to only use eye-movement behaviour as a measure.

Secondly, participants in the C group, who did not receive any training but demonstrated significant changes in reading behaviour in simplex texts from the first to the second experiment session, at the same time also improved their scores in the simplex tasks. This provides some evidence that the C Group participants had probably adjusted their reading strategies during the second session, most likely since the task was no longer completely unfamiliar to them, or perhaps simply due to increased exposure to reading in English over the nine months between the two experiment sessions, since all of them were full-time university students studying in an English-speaking country during the period when the two sessions took place. However, it is not clear whether this alignment is coincidental or is a result of a correlation between reading behaviour and performance.

Thirdly, given that the C group, without training, also managed to significantly improve their scores in sight translating the simplex texts, the training effect for T&I students in simplex texts becomes uncertain – it seems possible that, with or without training, over time participants improve in sight translating simplex texts after becoming more familiar with the ST process. It may therefore be that training effects truly become evident only under more challenging conditions.

Last but not least, although we cannot be certain that participants' reading behaviour correlates with their ST competence, one particular eye measure, the revisits, is worth a closer look. As discussed in Section 4.1.3, the T&I participants demonstrated a significantly reduced number of revisits in dealing with the complex texts in the second experiment session, and their scores also improved significantly for the same type of texts. Meanwhile, neither reading behaviour nor performance scores in the C Group changed significantly in the complex texts. This suggests that both revisits and performance scores might be able to reflect the effect of training, and are therefore worth further examination.

4.2.2. Revisits and performance scores in T&I participants

To further investigate the relationship between training, the change in performance scores and the change in the number of revisits to AOIs for E Group participants, we develop another linear mixed effect model. Five different bands are identified as indicators of different competence levels: '0' representing the group mean score irrespective of Test Time; '+2' and '−2' representing a result above and below the group mean by 2 marks respectively; '+4' and '−4' representing a result above and below the group mean by 4 marks respectively. These different bands are then used as a categorical variable in the model to estimate the effect of the categorical independent variable Test Time on the dependent variable of revisits for participants at different competence levels. Similar to the other two mixed effects models, Participants and AOIs are included as random variables.

Table 12 presents the estimated effect of Test Time on T&I participants' revisits when the marks fall in the above five bands.

A more accessible impression of the relation between revisits and different bands is presented in Figure 1.

Again, it is clear that training has limited impact on changing participants' revisits in the simplex tasks, irrespective of the participants' performance level. However, training has a more significant impact on participants in the complex tasks, and the effect

Table 12. Pairwise comparison of Test Time by Complexity by Band.

Band	Complexity	(I) Test Time	(J) Test Time	Mean difference (I-J)	Sig.[a]
Mean+4	Complex	1st	2nd	−26.66	.41
	Simplex	1st	2nd	22.90	.47
Mean+2	Complex	1st	2nd	8.32	.69
	Simplex	1st	2nd	10.75	.59
Mean	Complex	1st	2nd	43.31*	.01
	Simplex	1st	2nd	−1.41	.93
Mean-2	Complex	1st	2nd	78.29*	.00
	Simplex	1st	2nd	−13.56	.56
Mean-4	Complex	1st	2nd	113.27*	.00
	Simplex	1st	2nd	−25.71	.47

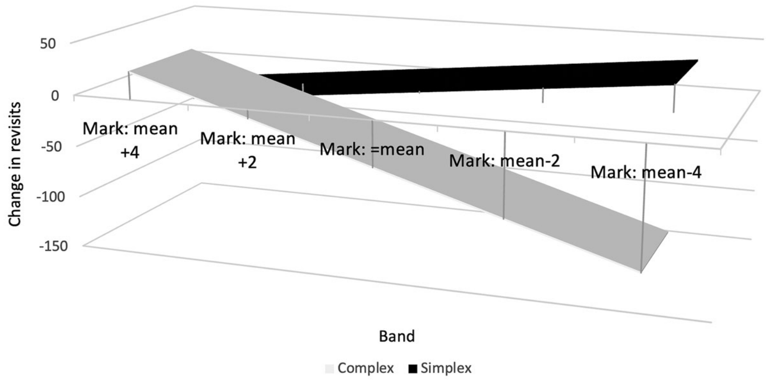

Figure 1. Change of revisits after training in relation to different marks.

increases as the performance or competence of the participants decreases. For those participants whose marks are around the average level or below (mark = mean, = mean-2, and = mean-4), the training effect on reducing revisits is highly significant. In comparison, for the rest of the participants whose marks are above average by 2 and 4 marks, training has only limited impact on reducing revisits. In summary, the training effect is more noticeable for less competent participants than for more competent participants.

5. Conclusion

As far as the first research question is concerned, we have found that, apart from revisits, two semesters of T&I training have in general limited impact on students' reading behaviour in the current project as far as the eye measures are concerned. This finding matches the conclusions of Chmiel and Mazur (2013).

Furthermore, the group difference in terms of eye measures can only be found in revisits, which echoes Płużyczka's (2015) finding that, in ST, the most significant parameter of cognitive effort are revisits. More importantly, after training, T&I students significantly reduced their revisits in reading the complex texts, indicating that training might have helped improve their efficiency in the handling of challenging complex syntactic structures during ST, similar to the findings of Su and Li (2021).

Regarding the second research question, we found that, despite a significant increase in scores in both simplex and complex texts for T&I students, it is not certain if training, in fact, is the factor that has helped improve the performance in the simplex tasks, since non-T&I students also showed an improvement in performance in the simplex tasks – but not in the complex tasks. The findings also raise questions about the relationship between reading behaviour and performance: at least in this project, training did help T&I participants improve their ST performance; however, apart from a decrease in revisits, the reading behaviour of T&I participants did not change much after training. Therefore, we believe that eye-movement data as measure of reading behaviour cannot be used as the only indicator in assessing the training effects on participants.

Nevertheless, although two semesters of training are found to have limited impact on interpreters' reading behaviour, training has enhanced their performance in general, which is consistent with the findings by Hale et al. (2019). It is possible that, notwithstanding limited impact on reading behaviour, training may have impacted other aspects of the ST performance, such as the effective use of strategies in chunking translation units, and the speed of transfer from one language to another, which all contribute to ST competence. In particular, the training effect is very significant for T&I students in dealing with the complex tasks, as reflected by reduced revisits in eye data and also by improved marks. In addition, this training effect is particularly remarkable for T&I participants performing around and below average level, but the training effect is insignificant for participants performing above average.

This study has enabled us to take a step forward in exploring the effect of training on interpreters' reading behaviour and ST performance. The findings provide some evidence that training is effective in improving students' performance in dealing with syntactically challenging texts. However, these gains are evident only under certain conditions. This brings pedagogical implications to T&I trainers and curriculum innovators in understanding the training needs of students with various competence levels. Finally, it must be acknowledged that the current project has limitations. Due to constraints in recruiting eligible participants for this longitudinal study, only 17 people were recruited for each group. It is not certain whether the current findings, derived from an experiment with small groups, reflect the situation in general, though they have contributed to the understanding of the situation. Longitudinal studies based on a larger sample size over longer periods are expected to yield more robust findings. Also, due to the screen size, only 4 sentences could be set as AOIs in each stimulus, which could be a limitation. Obviously, more work is needed to gain further insights into the training effect on T&I students.

Notes

1. https://www.naati.com.au/certification/certification-testing/certified-interpreter/#
2. Glances are defined as each time a fixation hits the AOI if not hit before.
3. Details of the NAATI assessment criteria for ST can be found at: https://www.naati.com.au/wp-content/uploads/2020/10/Certified-Interpreter-Assessment-Rubrics.pdf.

Disclosure statement

No potential conflict of interest was reported by the author(s).

Funding

This work was supported by the Chitra Fernando Fellowships Award granted by the Department of Linguistics, Macquarie University.

Notes on contributors

Dr Jing Fang is a lecturer in the Translation and Interpreting Program at Macquarie University. Her current research interests include systemic functional linguistics, corpus linguistics and sight translation research.

Xiaomin Zhang is a PhD candidate at Macquarie University. Her research area includes translation of children's literature, corpus linguistics and translation process research.

Haidee Kotze is Professor and Chair of Translation Studies in the Department of Languages, Literature and Communication at Utrecht University, and holds an appointment as Extraordinary Professor at the North-West University in South Africa. Her research focuses on language contact, language variation, and language change, with a particular emphasis on translation. She is the co-editor of *Target: International Journal of Translation Studies*, and co-editor of the series Translation, Interpreting and Transfer (Leuven University Press).

ORCID

Jing Fang http://orcid.org/0000-0003-2689-4754
Xiaomin Zhang http://orcid.org/0000-0002-1146-9141
Haidee Kotze http://orcid.org/0000-0002-5721-0733

References

Agrifoglio, M. (2004). Sight translation and interpreting: A comparative analysis of constraints and failures. *Interpreting. International Journal of Research and Practice in Interpreting*, *6*(1), 43–67. https://doi.org/10.1075/intp.6.1.05agr

Angelelli, C. (1999). The role of reading in sight translation. *The ATA Chronicle*, *28*(5), 27–30.

Booth, R. W., & Weger, U. W. (2013). The function of regressions in reading: Backward eye movements allow rereading. *Memory & Cognition*, *41*(1), 82–97. https://doi.org/10.3758/s13421-012-0244-y

Cenkova, I. (2015). Sight interpreting/translation. In F. Pöchhacker (Ed.), *The Routledge encyclopedia of interpreting studies* (pp. 374–375). Routledge.

Chmiel, A., Janikowski, P., & Cieslewicz, A. (2020). The eye or the ear?: Source language interference in sight translation and simultaneous interpreting. *Interpreting. International Journal of Research and Practice in Interpreting*, *22*(2), 187–210. https://doi.org/10.1075/intp.00043.chm

Chmiel, A., & Lijewska, A. (2019). Syntactic processing in sight translation by professional and trainee interpreters: Professionals are more time-efficient while trainees view the source text less. *Target. International Journal of Translation Studies*, *31*(3), 378–397. https://doi.org/10.1075/target.18091.chm

Chmiel, A., & Mazur, I. (2013). Eye tracking sight translation performed by trainee interpreters. In C. Way, S. Vandepitte, R. Meylaerts, & M. Bartlomiejczyk (Eds.), *Tracks and treks in translation studies* (pp. 189–205). John Benjamins.

Cicchetti, D. V. (1994). Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, *6*(4), 284–290. https://doi.org/10.1037/1040-3590.6.4.284

Dragsted, B., & Hansen, I. G. (2009). Études et prospectives. *Meta*, *54*(3), 588–604. https://doi.org/10.7202/038317ar

Fang, J., & Wu, C. (2010). Exploring shifts in translating English nominal groups modified by embedded clauses: A corpus-based approach. In B. Exley (Ed.), *Australian systemic functional linguistics association conference 2009: Proceedings* (pp. 1–20). ASFLA.

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

Graesser, A. C., McNamara, D. S., & Kulikowich, J. M. (2011). Coh-Metrix: providing multilevel analyses of text characteristics. *Educational Researcher*, *40*(5), 223–234. https://doi.org/10.3102/0013189X11413260

Hale, S., Goodman-Delahunty, J., & Martschuk, N. (2019). Interpreter performance in police interviews. Differences between trained interpreters and untrained bilinguals. *The Interpreter and Translator Trainer*, *13*(2), 107–131. https://doi.org/10.1080/1750399X.2018.1541649

Halliday, M. A. K., & Matthiessen, C. M. I. M. (2014). *An introduction to functional grammar* (4th ed.). Arnold.

Havnen, R. (2020). Where did the footnote go? How the change of mode in sight translation affects meaning-making. *The Journal of Specialised Translation*, (34), 78–99.
Hlavac, J. (2013). A cross-national overview of translator and interpreter certification procedures. *The International Journal of Translation and Interpreting Research*, *5*(1), 32–65. https://doi.org/10.12807/ti.105201.2013.a02
Holmqvist, K., Nyström, M., Andersson, R., Dewhurst, R., Jarodzka, H., & Van de Weijer, J. (2011). *Eye tracking: A comprehensive guide to methods and measures*. Oxford University Press.
Hoover, W. A., & Gough, P. B. (1990). The simple view of reading. *Reading and Writing: An Interdisciplinary Journal*, *2*(2), 127–160. https://doi.org/10.1007/BF00401799
IBM Corp. (2017). IBM SPSS Statistics for Macintosh (Version 25.0) [Computer Software]. IBM Corp.
Ivars, A. J. (2008). Sight translation and written translation: A comparative analysis of causes of problems, strategies and translation errors within the PACTE translation competence model. *FORUM. Revue Internationale D'interprétation et de Traduction / International Journal of Interpretation and Translation*, *6*(2), 79–104. https://doi.org/10.1075/forum.6.2.05iva
Jakobsen, A. L., & Jensen, K. T. H. (2008). Eye movement behaviour across four different types of reading task. In S. Göpferich, A. L. Jakobsen, & I. Mees (Eds.), *Looking at eyes – eye tracking studies of reading and translation processing* (pp. 103–124). Samfundslitteratur.
Lee, J. (2012). What skills do student interpreters need to learn in sight translation training? *Meta*, *57*(3), 694–714. https://doi.org/10.7202/1017087ar
Li, X. (2014). Sight translation as a topic in interpreting research: Progress, problems, and prospects. *Across Languages and Cultures*, *15*(1), 67–89. https://doi.org/10.1556/Acr.15.2014.1.4
McGraw, K. O., & Wong, S. P. (1996). Forming inferences about some intraclass correlation coefficients. *Psychological Methods*, *1*(1), 30–46. https://doi.org/10.1037/1082-989X.1.1.30
Nilsen, A. B., & Monsrud, M. B. (2015). Reading skills for sight translation in public sector services. *Translation & Interpreting*, *7*(3), 10–20.
Płużyczka, M. (2013). Eye-tracking supported research into sight translation: Lapsological conclusions. In S. Grucza, M. Plużyczka, & J. Zajac (Eds.), *Translation studies and eye-tracking analysis* (pp. 105–138). Peter Lang.
Płużyczka, M. (2015). Tłumaczenie a vista. Rozważania teoretyczne i badania eyetrackingowe (Sight translation. Theoretical considerations and eye tracking research). Warsaw.
Płużyczka, M. (2018). The first hundred years: A history of eye tracking as a research method. *Applied Linguistics Papers*, *4/2018*(25), 101–116. https://doi.org/10.32612/uw.25449354.2018.4.pp.101-116
Płużyczka, M. (2020). Tracking mental processes in sight translation: Neurobiological determinants of selected eyetracking parameters. *Translation, Cognition & Behavior*, *3*(2), 209–232. https://doi.org/10.1075/tcb.00041.plu
Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. https://doi.org/10.1037/0033-2909.124.3.372
Rayner, K., & Pollatsek, A. (1989). *The psychology of reading*. Prentice Hall.
Shreve, G. M., Lacruz, I., & Angelone, E. (2010). Cognitive effort, syntactic disruption and visual interference in a sight translation task. In G. M. Shreve & E. Angelone (Eds.), *Translation and cognition* (pp. 63–84). John Benjamins.
Song, S. Z. (2010). Skill transfer from sight translation to simultaneous interpreting. *International Journal of Interpreter Education*, *2*(1), 120–134.
Su, W., & Li, D. (2019). Identifying translation problems in English-Chinese sight translation. *Translation and Interpreting Studies*, *14*(1), 110–134. https://doi.org/10.1075/tis.00033.su
Su, W., & Li, D. (2021). Exploring the effect of interpreting training: Eye-tracking English-Chinese sight interpreting. *Lingua*, *256*, 103094. https://doi.org/10.1016/j.lingua.2021.103094
Weber, W. K. (1990). The importance of sight translation in an interpreter training program. In D. Bowen & M. Bowen (Eds.), *American translators association scholarly monograph series 4. Interpreting – yesterday, today and tomorrow* (pp. 44–52). State University of New York.
Zheng, B., & Xiang, X. (2014). The impact of cultural background knowledge in the processing of metaphorical expressions: An empirical study of English-Chinese sight translation. *Translation and Interpreting Studies*, *9*(1), 5–24. doi:10.1075/tis.9.1.01zhe