Language Learning & Technology
ISSN 1094-3501 CC BY-NC-ND

2023, Volume 27, Issue 1
pp. 1–22

ARTICLE

How do learners use a CALL environment? An eye-tracking study

Isabeau Fievez, KU Leuven, ITEC, imec Research group at KU Leuven

Maribel Montero Perez, Ghent University

Frederik Cornillie, KU Leuven, ITEC, imec Research group at KU Leuven

Piet Desmet, KU Leuven, ITEC, imec Research group at KU Leuven

Abstract

This study investigates how learners use a freely available web-based multimodal language learning environment based on videos and newspaper articles with corresponding exercises and learning support in the form of help options, NedBox*, for learning L2 Dutch. In order to measure learners' use, a combination of eye movement and screen recording data were analysed. The study also examines how learners' eye movements and use of the environment are related to their proficiency and performance in the environment. Twenty-one learners of Dutch used NedBox for 40 minutes, while their eye movements were tracked and their screen was recorded. Results demonstrated that learners used the environment in different ways in order to deal with the simultaneous presentation of input and exercises, and that this was dependent on the type of input. Eye movements were, to some extent, related to proficiency and performance in the environment. Finally, learners' use of the help options can be defined as idiosyncratic. In sum, the findings of this study suggest that the simultaneous presentation of input, be it video or text, does not overwhelm learners and that learners do not always seem to use help options effectively.*

Keywords*: *Eye-tracking, Multimodality, Help Options, Learning Environment

Language(s) Learned in This Study: *Dutch*

APA Citation: Fievez, I., Montero Perez, M., Cornillie, F., & Desmet, P. (2023). How do learners use a CALL environment? An eye-tracking study. *Language Learning & Technology*, *27*(1), 1–22. https://hdl.handle.net/10125/73514

Introduction

The enormous number of second language (L2) materials that are currently available online, such as videos and CALL environments, are transforming the way learners can engage with the L2 outside the classroom (Godwin-Jones, 2019). These types of materials have the potential to promote informal L2 learning as well as learner autonomy (Godwin-Jones, 2019; Reinders & White, 2016). In the last decade, a considerable number of studies have investigated how videos (e.g., Fievez et al., 2020) and CALL environments have been used to stimulate L2 learning (e.g., Cárdenas-Claros, 2020; Hegelheimer & Tower, 2004; Heift, 2006). However, these studies mainly focused on formal learning activities, such as those linked to a curriculum (i.e., Benson, 2011; Reinders & White, 2016) that took place inside the classroom/laboratory. Little is known about how learners interact with and learn from self-instructional L2 materials such as a CALL environment intended for autonomous use in informal settings outside the classroom (e.g., Pujolà, 2002). In addition, research that investigates the relationship between learners' interaction with learning materials and their proficiency and performance in such a CALL environment is scarce.

In order to address these gaps, this study uses eye-tracking and screen capture technology to examine how

L2 learners of Dutch work with a freely available self-instructional and informal web-based multimodal language learning environment in a relatively uncontrolled context. Further, it investigates the relationship between how learners use the environment, their proficiency, and their performance on the platform.

Literature Review

Understanding how learners use multimodal CALL environments is crucial for both theoretical and practical reasons. First, it helps determine to what extent learners interact with these environments and whether they use the tools as intended by the instructional design. These insights can help us understand the strategies and learning processes that learners employ as they work with CALL tools. These can then serve to improve the instructional design of the program and increase learner autonomy (Fischer, 2007).

In that respect, tracking technologies, such as log files and screen recordings, have been widely used and valued to research learners' interactions in multimedia environments (e.g., Cárdenas-Claros, 2005; Fischer, 2007). Indeed, there has been a substantial amount of research on the use of generic technologies used for language learning such as instant messenger, computer games, and internet chats (e.g., Isbell, 2018); CALL environments developed for research or instructional purposes and designed for specific language skills (e.g., Brandl, 1995; Hegelheimer & Tower, 2004; Heift, 2006); and Computer-Mediated Communication (e.g., O'Rourke et al., 2017). Most of these CALL environments "follow a lock-step scope and sequence" (Beatty, 2010, pp. 12–13) by following the concept of *programmed instruction* (Beatty, 2010). However, over the past few years, a considerable number of CALL environments intended for autonomous use have been created (Danjou, 2020). In contexts where L2 learners have only limited out-of-classroom exposure to authentic L2 input, such CALL environments "extend and change the breadth and depth of exposure that learners can have with the target language" (Chapelle, 2009, p. 750).

Nonetheless, studies on the use of a CALL / multimodal environment actually developed for autonomous L2 learning outside the classroom are scarce. One exception is the study by Pujolà (2002), who researched the use of a web-based multimedia program for English learners. This environment was created for research purposes but learners had to use it autonomously. Pujolà's web-based multimedia program consisted of authentic videos, excerpts from radio shows, and newspaper articles. The program also offered numerous help options which he classified as *assistance* (e.g., dictionary, cultural notes, feedback) and *guidance* (e.g., information on how to use the program, instructions), and includes comprehension questions based on the multimedia input. Participants navigated freely in this environment over four sessions for 90 minutes each while their screens were recorded. Analyses of the data allowed Pujolà to define two groups of users. The first group of users adopted a global approach and did not consult the help options frequently, but relied on their prior knowledge. The second group, on the other hand, were classified as *compulsive consultors* because they called up the help facilities more intensively in order to get more details on the input.

Although the aforementioned studies provide some understanding of how learners use a CALL / multimodal environment, more data is required in order to find out how learners engage with an informal multimodal language learning environment intended for autonomous use that presents both input and exercises simultaneously. Moreover, while most studies have made use of log files or screen recording in order to collect data, none of these instruments provide exact information on whether or not learners actually paid attention to what they have done. For example, while log files and screen recordings enable the researcher to see when the learner clicked on "play," they do not inform us whether the learner actually watched the video or not and how they divided their attention on the screen. This kind of information can be obtained by using eye-tracking technology.

Eye tracking provides information on the user's eye movements. The most frequently used eye-tracking measures are the following: (a) fixations, which are the moments during which the eyes remain relatively still, (b) saccades, which are the eyes' forward movements or so-called twitches in-between the fixations

(Conklin et al., 2018) and (c) regressions, which are the eyes' backward movements, mostly executed when reading. Eye tracking also informs us on *how many* fixations the user has made (i.e., fixation count), *how long* the fixation lasted (i.e., fixation time) as well as on the location of the fixations and the direction of learners' fixations (Conklin et al., 2018).

In the SLA literature, research using eye-tracking technology is expanding (Conklin et al., 2018; Godfroid, 2019). Studies have focused on how on-screen text is processed in audio-visual input (Bisson et al., 2014; Montero Perez, 2019; Tragant & Pellicer-Sánchez, 2019; Winke et al., 2013) and how the potential differences in word processing in captions affects vocabulary uptake (Montero Perez et al., 2015). However, within research on the use of CALL / multimodal language learning environments, eye-tracking technology is relatively new (e.g., Pellicer-Sánchez et al., 2020) and has mainly been exploited in studies on synchronous computer-mediated communication, (e.g., Latif, 2019; O'Rourke et al., 2017; for a review, see Michel & Smith, 2017).

To our knowledge, no study has looked at learners' use of multimedia input and exercises in a multimodal language learning environment. Yet, understanding how learners integrate the different sources of information presented simultaneously (i.e., input and exercises) might not only be beneficial for developers but also further support learners' learning process when interacting with the environment. Particularly relevant for the present article are investigations of learners' eye movements while completing a computerised reading task (e.g., Brunfaut & McCray, 2015). For example, Bax (2013) investigated the reading processes of successful and unsuccessful readers while taking a computerised IELTS test. He argues that reading in order to answer questions is different from 'default' reading in the L1, that is when readers are assumed to comprehend the text smoothly and are mostly reading along a line of text (Bax, 2013). Findings demonstrated that unsuccessful readers spent much more time reading and finding the appropriate extract to answer the question than the successful readers. Although the present study does not aim to discover learners' cognitive processing while taking a test, the simultaneous presentation of multimodal input and exercises resembles such computerised tasks. Findings from these studies might therefore prove useful for interpretation of our results (Jacob & Karn, 2003).

In sum, there is a need for more research on how learners engage with multimodal learning environments developed for self-study that consist of multimedia input and exercises. In addition, few studies have used eye-tracking technology even though this technique makes it possible to study these questions in an unobtrusive (i.e., with a remote eye-tracker) and real-time fashion. This paper aims to fill these gaps by investigating learners' use of a multimodal environment comprised of multimedia input, exercises, and help options by means of eye-tracking and screen capture technology. It also investigates the relationship between learners' proficiency, performance, eye movements, and use of the environment's content and features. The research questions are:

1. How do learners engage with the simultaneous presentation of input and exercises?
2. Are eye-tracking measures (i.e., transitions, proportion of time, average fixation duration) related to participants' proficiency and performance in the environment?
3. How do learners use the different help options available (e.g., control, hint, feedback)?
4. Is the learners' use of help options related to their proficiency and performance in the environment?

Methodology

Participants

Twenty-one French-speaking low to intermediate learners of Dutch (M_{age} = 32.1, *SD* = 11.91) volunteered to participate in the study. They were all jobseekers taking L2 Dutch lessons in different job centres at the time of the study. By focusing on these learners, we address a population that has received little to no attention in the SLA and CALL literature. However, effective L2 learning experiences are extremely important in order to increase learners' chances at the job market. Their scores on a vocabulary size test varied from 22 to 109 out of 120 (M = 71.90, SD = 26.54). As regards their educational background, 10 participants had a degree of secondary school, six a bachelor's degree and five a master's degree. Among them, 13 participants had normal, uncorrected eyesight, whereas eight had to wear glasses. While 15 of them reported never using NedBox (the online language learning environment of this study) before, six had already used it once prior to the study.

Instruments

Vocabulary Size Test

We developed a frequency-based multiple-choice vocabulary size test following Peters et al.'s (2019) criteria. The item selection was based on the *SUBTLEX-NL* corpora (Keuleers et al., 2010). The test consists of four different frequency bands (0–4000 most frequent words) with 30 items per frequency band. Each multiple-choice item contains the correct translation, three distractors, and an 'I don't know' option (see Figure 1). The internal reliability of the test was high (Cronbach's alpha =.95, N = 21). It is strongly assumed that this type of test gives a good estimate of learners' proficiency level in the L2 (Laufer et al., 2004).

Figure 1

Examples of Vocabulary Size Test Items

35. Ingewikkeld

- Horrible
- Sec
- Brillant
- Compliqué
- Je ne sais pas

36. Lijden

- Souffrir
- Plier
- Arrêter
- Rechercher
- Je ne sais pas

NedBox

NedBox, the multimodal language learning environment used in this study, is an online environment intended for self-study of L2 Dutch. The development was coordinated by KU Leuven (Centrum voor Taal en Onderwijs) and carried out in collaboration with other organisations such as Televic Education and the Flemish public broadcasting company VRT. The learning environment is based on news items in the form of newspaper articles and videos, accompanied by exercises (for more information on the development and choice of input, see Schiepers et al., 2016) and targets low to high proficiency learners.

Users can navigate freely through the website. For each news item, they can choose what skill they want to practise, namely viewing / listening, reading, writing, and speaking. The platform also contains vocabulary and grammar exercises. They can also choose between three difficulty levels indicated by one, two, or three stars (equivalent of CEFR-A1, A2, and B1 level, respectively). Appendix A depicts typical screens from news items at the time of the experiment (in 2019). In addition, NedBox also provides several help options that can help users complete the exercises (see Appendix B for an exhaustive list of the help options and their description).

Design of the Experiment

The experiment was designed with the research lab's eye-tracker software, *Experiment Centre of SMI*. This software allows users to select webpages and launches these automatically during the experiment. It also allows the insertion of eye gaze calibrations in-between web pages. In order to compare the data of the participants, we ensured that they would navigate to the same news items. To this end, 13 news items were selected and outlined in *Experiment Centre*. In total, exercises of nine different news items were completed by the participants. In the remainder of this article, we will refer to these completed news items as *trials* (Appendix C).

It is important to note that the trials were not comparable in terms of the types and number of exercises offered. For example, the first trial offered five listening exercises, two reading exercises, and one vocabulary exercise, whereas the second one had eleven listening exercises, three reading exercises, and one vocabulary exercise. However, since the goal of the study was to examine the actual use of a free-to-use online language learning platform in a maximally ecologically valid way, participants were allowed to navigate freely on the preselected webpages. If they had finished the exercises of a trial or did not want to complete them, they could simply close the webpage so that another one would open automatically.

Between each trial, a short explanation of the upcoming one appeared on the screen, followed by a new calibration. In order to get precise data of participants' behaviour in the environment, the screen of the computer was also recorded by *Experiment Centre*.

Procedure

One month before the actual experiment, a pilot study was conducted with five French-speaking low proficiency learners of Dutch who took Dutch lessons in a job centre. Participants navigated on NedBox for 30 minutes and also took the vocabulary size test. The results allowed the researchers to select the most suitable news items based on the time participants took to complete exercises and on the quality and degree of difficulty of the exercises (e.g., one item was removed because participants did not need to read the newspaper article in order to complete the exercises successfully).

Approximately two weeks before the experiment started, the first author visited Dutch classes in job centres in order to recruit participants for the experiment. If they agreed to participate, a schedule was established. One session lasted approximately 1 hour (40 minutes for using NedBox and 10 to 20 minutes for completing the vocabulary size test) and was held during their Dutch lesson.

The experimental session was conducted individually in a quiet room in the job centres' premises. Participants sat on a non-adjustable, non-swivel chair to reduce movement. They were instructed that they had to use NedBox in the same way as they would do at home and were informed that their eye movements would be tracked. They received a short explanation of the website and its help options. They were also notified that they could close the webpage, but not change it themselves since the software would open another one automatically for them. Before or after using NedBox, they completed the vocabulary size test. While it is strongly recommended to supplement the data with interviews or stimulated recall when using eye-tracking technology (O'Rourke et al., 2017), time constraints prevented us from including these. However, the researcher observed the participants during the whole learning session and could therefore ask ad hoc questions at the end of the session if necessary.

Before they started using NedBox, a 9-point calibration was conducted and a 5-point calibration was executed between each new trial. Participants' eye movements were recorded using a remote SMI RED250 mobile eye-tracker with a sampling rate of 250 Hz. The eye-tracker records binocular eye movements through two hidden cameras in the monitor. Its accuracy and resolution are typically of 0.4° and 0.03° respectively (as measured in ideal conditions). Data of both eyes were included in the analyses.

Analyses

Performance on the Platform

In order to be able to link performance on the platform with attention allocation and actual use, performance on exercises was calculated for each participant. To do so, the first author checked the answers to the exercises. If the exercise was answered correctly the first time, the learner would get 100%, and if it was answered correctly the second time, they would get 50%. If the help option *solution* was consulted or the exercise was not completed, they would get 0. As the total number of exercises differed across participants, the scores were averaged by trial and participant to get a percentage.

Learners' Use of Help Options

Because learners' use of the website NedBox was not logged, data about the use of the help options were manually encoded by watching the screen recordings. The data account for each time a help option was consulted. To answer research question 3, which analyses the use of the help options, frequency counts as well as descriptive statistics for the consultation of the *check*, *solution*, *hint*, and *feedback* buttons and of the *subtitles* and *glossary* were obtained.

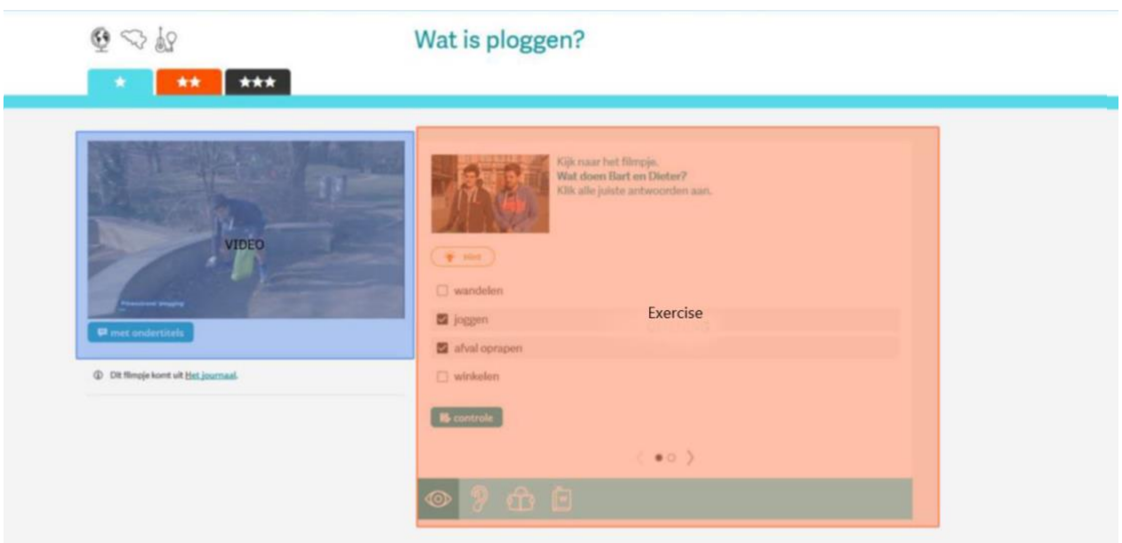
To address the fourth research question, exploring the relationship between the use of the help options and the participants' proficiency and performance on the platform, subtotals of the use of help options were consolidated per trial for each participant by dividing the number of times they requested the help option during one trial by the number of times the help option was available during that trial.

Eye-Tracking Data

Participants' eye movements were measured in specific parts of the screen of interest, which are called areas of interest (AOIs) and had to be defined manually for each trial. There were two different AOIs per trial that are present on the screen at the same time: (a) the text or video and (b) the exercises (Figure 2).

Figure 2

Example of AOI Video and Exercises



The size and position of the AOIs changed during trials because of the variable layout of each page and the participants' scrolling behaviour. The eye-tracking measures used in this study are based on fixations. These are periods of time during which the eyes remain relatively still and the gaze is maintained on a single location inside an AOI. We examined three different measures:

1. the *transitions*, which correspond to the number of times learners switched between the AOIs
2. the *proportion of time*, which corresponds to the total fixation time inside the AOI during a trial, divided by the time the AOI was visible during the trial (total fixation time/visible time of AOI)
3. the *average fixation duration*, which is the sum of fixation times divided by the number of fixations inside an AOI

The choice to analyse the proportion of time instead of total fixation time was motivated by three main reasons: (a) the duration of the trials differed, (b) it has already been used in studies comparing attention allocation to text and pictures, and (c) our AOIs conveyed different types of information (i.e., textual and visual; Conklin et al., 2018; Pellicer-Sánchez et al., 2020). This measure gives information on the relative time spent on the different AOIs, which is what interested us in this experiment. The average fixation duration, on the other hand, gives information on the cognitive processes at hand when looking at the AOIs, since it is assumed that what is fixated on longer is harder to process (Conklin et al., 2018).

Prior to analysis, eye movement data had to be cleaned. For sake of data quality, eye movements with a recording accuracy lower than 70%, as determined by the eye-tracker software, were discarded and fixations longer than 800ms were removed (Conklin et al., 2018). This corresponded to 16.7% of the data points (23.945 out of 143.041 fixations, corresponding to 9 out of 68 trials). This amount of data loss is quite common in eye-tracking research (Conklin et al., 2018).

Results

Vocabulary Size Test and Performance on the Platform

The descriptive statistics of the vocabulary size test and the performance on the exercises of the platform are displayed in Table 1.

Table 1

*Descriptive Statistics of Vocabulary Size Test and Performance on the Platform (*N*= 21)*

	Vocabulary Size Test	**Performance on the Platform**
M (*SD*)	71.90 / 120 (26.54)	72.70% (14.38)

RQ1: How do Learners Engage with the Simultaneous Presentation of Input and Exercises?

Visual Analysis

In order to answer the first research question, we first created sequence charts for each participant and trial. The examination of the sequence charts enabled us to distinguish between three strategies of use: (a) the *switching approach* (see Figure 3, pattern 1), when learners switched very often between the input and the exercises during the whole trial (*Number of trials* = 21); (b) the *systematic approach* (see Figure 3, pattern 2), when learners first focused on the input then completed the exercises without going back to the input often (*Number of trials* = 21); and (c) the *combined approach* (see Figure 3, pattern 3), whereby learners mixed both switching approach and systematic approach patterns (*Number of trials* = 17). Figure 3 represents two examples of each pattern.

To investigate this further and establish whether learners always adopted the same strategy or not, we analysed the number of transitions they made between the two AOIs during the different trials (see Table 2). In order to distinguish between the strategies adopted by participants, we use the top (i.e., *switching approach*), middle (i.e., *combined approach*), and bottom third (i.e., *systematic approach*) of the number of transitions. It turned out that while some participants always used the same strategy across trials (n = 8), others had a different approach depending on the type of input (i.e., text or video, n = 5), while the rest mixed the patterns (n = 8). As can be seen in Table 2, learners had far more transitions between the exercises and the input when the input was a video than when it was a text. This difference was significant (t (4.11) = 53.01, p < .001, d = .97).

Figure 3

Example of Sequence Chart Patterns

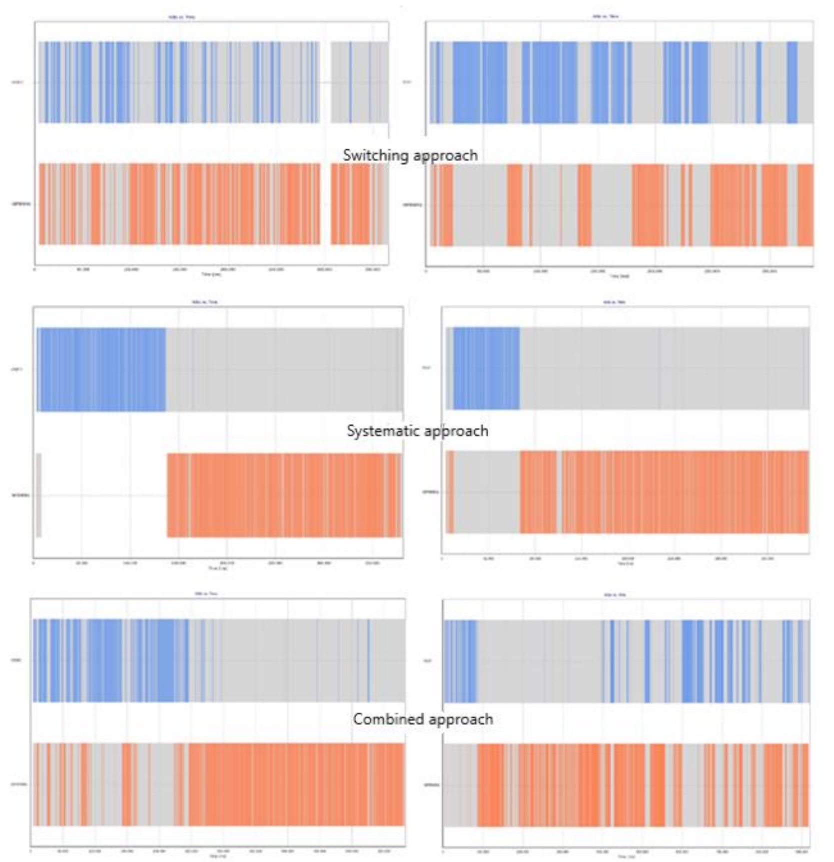


Table 2

Number of Transitions between Exercises and Input

Input Type	**Number of Transitions Between AOI Exercises and AOI Video or Text** ***M*** **(*SD*)**
Video (*n* = 34)	93.66 (*70.37*)
Text (*n* = 25)	35.76 (*35.99*)
Total	69.53 (*64.87*)

Finally, we were interested in discovering whether these strategies of use would influence the way in which learners exploited the input. We therefore analysed this by watching the gaze replays and looking at scan paths in order to check whether the video was watched until the end and the text read in its entirety. The analysis revealed that when adopting a systematic approach, learners would watch/read the entire input in 95% of cases, whereas when adopting the switching approach or combined approach, this would add up to 61% and 79% of cases respectively. More specifically, learners adopting the systematic approach would watch / read the entire input before starting to complete the exercises in 87.5% of cases (100% and 75% when the input was a video and a text respectively), against 30.5% of cases when learners used the switching approach (18% and 42% when the input was a video and a text respectively), and 69% of cases when adopting the combined approach (50% and 88% when the input was a video and a text respectively).

Statistical Analysis

First, it should be mentioned that because we analysed eye movements on types of multimodal materials that are different in nature (e.g., video, text) and that were not controlled for research purposes, these results should be interpreted cautiously. We analysed the proportion of time spent fixating each AOI (see Table 3) as well as the average fixation duration. As can be observed, and as the sequence charts also demonstrated, more time was spent fixating on the exercises, independent of whether the input was a video or a text. Independent sample *t*-tests revealed that the differences of time spent on the AOIs were significant between *Exercises based on video* and *Video* ($t(112.30) = 15.78$, $p < .001$, $d = 2.71$) as well as between *Exercises based on text* and *Text* ($t(90.01) = 9.97$, $p < .001$, $d = 1.99$). The *average fixation duration* on the *Exercises* was slightly shorter than on the *Video* or *Text*. These differences were not significant.

Table 3

Eye-Tracking Measures per AOI

AOI Name	Proportion of Time *M* (*SD*)	Average Fixation Duration (ms) *M* (*SD*)
Exercises based on video (*n* = 68)	.585 (.1541)	302.02 (38.41)
Video (*n* = 68)	.238 (.0978)	310.13 (41.99)
Exercises based on text (*n* = 50)	.520 (.1733)	294.97 (64.98)
Text (*n* = 50)	.216 (.1274)	324.27 (85.87)

Note. The number of cases was 68 and 50 respectively because each AOI has two data entries per trial (one for right and one for left eye) – there were 34 completed trials with video and 25 completed trials with text (after data cleaning).

RQ2: Are Eye-Tracking Measures Related to Participants' Proficiency and Performance in the Environment?

In order to answer the second research question, we checked whether the strategy of use adopted by the participants was affected by their proficiency, and whether strategy of use would affect their performance in the environment. No significant relationship between strategy of use, proficiency, and performance in the environment could be established. Yet, it still appeared that adopting a switching approach led to worse scores ($M = 70.19\%$, $SD = 18.79$) than using a combined ($M = 79.48\%$, $SD = 14.22$) or a systematic approach ($M = 81.26\%$, $SD = 15.29$).

Second, we conducted a Kendall Tau correlation analysis (because the vocabulary size test scores were not normally distributed) in order to explore whether participants' vocabulary size test scores are related to their eye movements on the different AOIs. The analysis revealed three relationships. First, the correlations between vocabulary size test scores and proportion of time were positive for both *Exercises* *based on video* (r_τ = .198, p = .021) and *Exercises based on text* (r_τ = .411, p < .001). In addition, the relationship between vocabulary size test scores and average fixation duration on the *Video* was negative (r_τ = -.412, p < .001). No other significant relationship was found.

Finally, the same analysis was run with average score on the platform and eye movement in the different AOIs in order to ascertain if the performance on the platform was related to eye-tracking measures. Only two significant relationships emerged from the analysis, namely (a) the more time spent fixating the *Exercises based on text*, the higher the scores (r_τ = .355, p < .001) and (b) the longer the average fixation duration on the *Video*, the lower the scores (r_τ = -.393, p < .001).

RQ3: How do Learners Use the Different Help Options Available?

Data of the 68 trials were included in this analysis. The descriptive statistics illustrate that the help options were accessed in various ways (see Table 4). Overall, participants relied on the *check* button the most, with participants accessing it an average of 47 times during the treatment. The *hint* and *solution* were consulted approximately two times per session. Only the *feedback* and the *glossary* were almost never used. We also analysed the use of the help options depending on the strategy of use adopted (see Table 4 for frequency counts). No specific pattern emerged, except that when participants used a combined approach, they consulted the help options less frequently.

Table 4

Frequency of Use of the Help Options

Help Option	**Total (count)**	***M* (*SD*)**	**Switching Approach (count)**	**Systematic Approach (count)**	**Combined Approach (count)**
Check	988	46.9 (*24.57*)	345	418	225
Hint	50	2.38 (*2.82*)	20	22	8
Solution	46	2.19 (*2.84*)	22	19	5
Feedback	12	0.57 (*1.99*)	0	11	1
Subtitles	12	0.17 (0.*38*)	7	3	2
Glossary	4	0.20 (0.*12*)	0	1	3
Nr of exercises	864	41	304	317	243

Note. The *check*, *solution*, and *feedback* options were available a total of 868 times, the *hint* only 271 times. *Subtitles* were available in 42 trials and *glossary* in 26 trials.

RQ4: Is Learners' Use of the Help Options Related to their Proficiency and Performance in the Environment?

Previous research findings (e.g., Hegelheimer & Tower, 2004; Heift, 2006) suggest that proficiency can have an influence on participants' behaviour with the help options of a language learning environment. The frequency with which participants invoked the help options compared to the participants' vocabulary size test scores (i.e., top, middle, and bottom third) was investigated in Table 5.

To explore whether there was a relationship between vocabulary size test scores and use of the help options, we conducted Kendall Tau correlations. A significant positive relationship between vocabulary size test scores and use of *check* (r_τ = .083, p = .002) as well as use of the media in its entirety (reading / viewing the input from start to finish; r_τ = .208, p < .001) was established.

Table 5

Frequency of Use of the Help Options per Proficiency Profile

Help Options	**Bottom Third (count, *N* = 7)**	**Middle (count, *N* = 7)**	**Top Third (count, *N* = 7)**
Check	309	242	437
Hint	13	24	13
Solution	21	5	20
Feedback	2	10	0
Subtitles	5	6	1
Glossary	1	1	2
Read / Viewed the entire media	13 out of 24	16 out of 18	21 out of 26
Total exercises completed	278	225	361

Further investigation of the relationship between performance on the environment and use of the help options was conducted with Kendall Tau correlations. The results revealed two relationships (see Table 6). Using *solution* more often was significantly and negatively correlated with performance, whereas using the media in its entirety was significantly and positively correlated with scores in the environment.

Table 6

Correlations Between the Use of Help Options and Performance

	Performance in the Environment	
	r_τ	p
Solution	-.405	.001
Entire media used	.286	.018

Discussion

This paper investigates how learners work with an informal L2 learning environment based on multimodal input and exercises designed for self-study, and how they engage with the available help options in order to support their learning process. It also attempts to study if and in what ways learner behaviour is related to vocabulary size and performance in the environment. This was done by using eye-tracking technology as well as screen recordings to track learners' eye movements and actions in the environment.

RQ1: How Learners Engage with the Simultaneous Presentation of Input and Exercises

With regard to the first research question investigating how learners engage with the simultaneous presentation of exercises and input, the visual analysis reveals that across trials, learners show three strategies of use. Not all learners use the same strategy, which is not surprising since a strategy may be useful for one and not useful for another (Godwin-Jones, 2019). The fact that all learners do not always use the same strategy across trials is interesting. Particularly, when the input is a video, we observed that learners adopt a switching approach far more often than when the input is a text. Since the double modality (i.e., visual and aural) of a video allows learners to listen without having to watch the video while completing the exercises, this phenomenon may seem unexpected. We can hypothesise that the dynamic nature of the video incites the learners to make more transitions between the exercises and the input. The other visual and statistical analyses of eye-tracking measures seem to support this hypothesis. First, in 40% of cases, learners have started completing the exercises before the end of the video, whereas when the input is a text, learners will more often read the entire text before starting to complete the exercises. In addition, the mean average fixation duration on video (i.e., 310 ms) is in concordance with other studies investigating eye movements on L2 videos (e.g., 450 ms in Bisson et al., 2014 and 306 ms in Tragant & Pellicer-Sánchez, 2019)[1] as well as with research on visual perception (Rayner, 2009). Therefore, what these analyses reveal is that the dynamic nature of the video does not hamper learners' adequate processing of the content, since they are able to answer the exercises and the average fixation duration does not reflect abnormal processing. However, it may have encouraged them to adopt a dynamic strategy of switching between input and exercises (Tragant & Pellicer-Sánchez, 2019).

On the other hand, the average fixation duration on text (i.e., 324 ms) may seem abnormally larger than the mean average fixation duration of 240 ms to 275 ms found in eye-tracking studies looking at L2 reading (e.g., Pellicer-Sánchez et al., 2020; Tragant & Pellicer-Sánchez, 2019). This may lead to the assumption that learners encountered processing difficulties while reading the texts. Yet, the larger average fixation duration in the present study can be explained by different factors. First of all, it can be a consequence of the fonts of the texts (i.e., 9 pt, 13 px) and line spacing which may have been too small for some participants, making some of them lean forward. Since we wanted to keep the environment as ecologically valid as possible by using the actual learning environment, we could not control for that. This is likely to have influenced their eye movements and to have generated longer fixations (Rayner et al., 2006). In addition, it should also be noted that the standard deviation of the average fixation duration on text also corroborates other L2 reading studies (e.g., Pellicer-Sánchez et al., 2020; Tragant & Pellicer-Sánchez, 2019), and as Rayner (2009) argues, "fixation duration can be as short as 50-75 ms and as long as 500-600 ms" (p. 1460). Therefore, while the average fixation duration on text may at first seem abnormally large, it indicates that learners took sufficient time to process the text and its content, hence, sufficiently exploiting the input in order to complete the exercises. The fact that the average score on the exercises is of 72.7% tends to corroborate the hypothesis that learners did not encounter specific difficulties processing the text.

We should also mention the *proportion of time* measure, which indicates that the exercises were fixated on for a longer period of time than the videos or the texts. This is unsurprising since the AOI Exercises is the place where the exercises have to be completed. It can therefore be reasonable to expect learners to spend more time fixating on the exercises than the input.

RQ2: Eye-Tracking Measures and Participants' Proficiency and Performance in the Environment

We were unable to establish any relationship between learners' proficiency (as estimated by the vocabulary size test scores), performance in the environment, and different strategies of use. The small sample and the relatively short time of the trial may explain this. Yet, the analyses of the correlations between learners' proficiency and their eye movements on the AOIs, together with the analysis of the relationship between the performance in the environment and the eye movements on the AOIs, reveal interesting results. The average fixation duration on the video is negatively correlated with the proficiency and performance on the environment. We can hypothesise that the longer average fixation duration of less proficient learners may arise from the fact that they activate the subtitles more often than more proficient learners. They will therefore need more time to process the subtitles (Muñoz, 2017) and this will reflect higher engagement and index higher difficulty in processing the information (Poole & Ball, 2006).

Next, the positive correlations between vocabulary size and proportion of time spent on exercises based on text, as well as between performance and proportion of time spent on exercises based on text, may be explained by the analyses of the use of the help options. We have also found that using the media in its entirety correlated positively both with vocabulary size test scores and performance in the environment. We can therefore argue that learners with higher scores on the vocabulary size test can solve most of the exercises by reading the text in its entirety and will not have needed to go back to it as often as less proficient learners. Consequently, they spend more time looking at the exercises. This may indicate that more proficient learners have a better grasp of what is called *careful local reading* (Bax, 2013), or *local comprehension*, which can be defined as the thorough processing of a text or part of text in order to extract complete meanings from it and which seems associated with linguistic knowledge (Bax, 2013; Weir et al., 2009). In other words, in response to an explicit question (i.e., that does not need inferential skills), more proficient learners can understand and find explicit information from the text more easily than less proficient learners.

In sum, the eye-tracking data show that (a) learners adopt different strategies in order to complete the exercises, (b) the same strategies are not always used across trials and participants, and (c) the simultaneous presentation of a video with exercises tends to provoke a more dynamic switching strategy than the simultaneous presentation of a text with exercises. However, more controlled studies investigating how learners deal with such an environment using eye-tracking are still needed in order to shed light on this further (Jacob & Karn, 2003).

RQ3: Use of the Help Options

As for the third research question, focusing on the actual use of the help options, findings reveal some variation, which corroborates results of studies investigating the use of different help options in a CALL-activity (Fischer, 2012; Pujolà, 2002). In the present study, *check*, which allows learners to check their answers, was used by all but two participants. The frequent use of this help option can be explained by its role, since clicking on *check* was the only way learners could know whether their answers were correct or not. Two participants decided not to use it at all. After the session, the researcher asked them why they chose not to check their answers. One participant explained to the researcher that she thought it would give her a score at the end of the series of exercises, the other one said he was confident in his answers and therefore did not want to check them.

On the other hand, help options such as *hint*, *subtitles*, *glossary*, and *feedback* were accessed a limited number of times. More specifically, the *hint* option was used only 18% of the time it was offered. The fact that it was not exploited as much as the *check* option may originate in learners' will to find the answers by themselves. Interestingly, learners pertaining to the middle third of proficiency used it much more than the bottom third group. We may hypothesise that they are more able to realise they lack the language knowledge or understanding to solve the exercise and therefore sought help, as opposed to less proficient learners. Regarding *solution*, the top third proficiency group consulted it as frequently as the bottom third. In order to understand this, we checked who and how they used it. One participant in the top third group can be considered an 'outlier.' They used the *solution* option eight times in a row. When completing the exercises and getting it wrong, instead of watching the video again in order to find the right answer or consulting *hint* and *feedback* to get help, they directly checked the solution. This may come from a lack of motivation or interest in the exercises.

Similarly, the almost non-use of the *feedback* and *glossary* options may be due to participants not noticing them or thinking they will be helpful, as reported in some of the ad hoc questions. One can infer from this that the learning tasks may have been too easy for the learners. Consequently, they did not feel the need to consult the *feedback* or *glossary*. Another hypothesis can be that the help options were not clear and visible enough. To get feedback, learners have to hover the mouse on the icons, whereas letting the feedback appear directly may be more effective and useful. The example of the participant using *solution* instead of consulting *feedback* may support this hypothesis. In the same way, although the glossed words in the texts are underlined, this is probably not salient enough for learners to notice them and to remind them that they can get a definition by clicking on it.

RQ4: Use of the Help Options, Participants' Proficiency, and Performance in the Environment

Although differences in behaviour between more and less proficient learners have been found in various studies examining learners' behaviour in a CALL-activity (Fischer, 2012), our more detailed inspection of the learners' actual use in relation to their proficiency does not reveal any strong relationship. The positive relationship between use of *check* and vocabulary size is probably due to the higher number of exercises performed by the top third group. In general, less proficient learners mainly require help that will assist them (Pujolà, 2002) in understanding the content, such as *subtitles* and *glossary*.

While no learner profiles can be drawn from this experiment, in contrast with Pujolà (2002) for example, this study still demonstrates that learners use the multimodal environment and its help options in different ways. Although neither extensively trained nor guided, learners were still able to process the necessary information and to find their own way to complete the exercises. However, it appears that some of them did not know when they needed help and which help option to consult.

Limitations

While the present study offers insights into learners' behaviour in a CALL / multimodal language learning environment designed for self-study and use in informal contexts, it has some limitations. First of all, because one of the most important criteria to the design of this study was the ecological validity, we decided not to alter the content of the environment. Hence, the input and exercises of the news items were very distinct (e.g., number of exercises per news item, length of the text and video). The pilot study still enabled us to select items participants found interesting. Special care was also taken to offer an equivalent amount of news items based on video or text. However, since the participants were allowed to navigate freely on those items, it was impossible to predict what they would do. In turn, the difference between the content and the users' freedom to complete the exercises precluded organising any kind of language post-test. Therefore, while eye movements on and use of the environment were examined, this study did not focus on whether the processing and behaviour differences impacted any learning gains.

Secondly, the participants in this study were not very familiar with the environment. The use of help options would probably have been different if participants had already used the environment a certain number of times before the experiment took place. Finally, although the environment is designed for autonomous use outside the classroom, the experiment was conducted in a learning centre—not at learners' home—which might reduce the ecological validity of our study. However, by opting for a mobile eye-tracker, we made the data collection as unobtrusive as possible.

Conclusion and Implications

The present study focuses on investigating learners' behaviour with a multimodal environment designed for self-study purposes. The results provide evidence that learners use the input and exercises in different ways. Although the simultaneous presentation of input and exercises does not overwhelm them, as the eye-tracking measures reveal, learners are still influenced by the modality of the input that is presented. Also, the findings demonstrate that some learners do not know how nor when to seek support by using the help options. While our data does not show a significant relationship between learner proficiency and different uses of the help options, it did demonstrate that more proficient learners exploit the media more often in its entirety than less proficient learners. These data, together with the data of help option use, may be relevant for adding adaptivity and personalisation to such learning environments. More research and data are still needed, though, before being able to draw any firm conclusion.

Further, the observed variety in the use of the help options can also be linked to the learners' performance on the platform. The non-use of certain learning supports point to the learners' willingness to find the answers by themselves or to their lack of self-awareness when it comes to needing help. Finally, the lack of salience of these help options in the environment may also have caused them to be used less frequently.

This paper offers some implications for instructional designers. Since learner training (Fischer, 2012) is inconvenient for web-based multimodal environments such as the one used in this study, more guidance should be offered for learners (Pujolà, 2002) within the environment. A demo on how to use the help options can be added (Fischer, 2012) so that learners get a good idea of their utility before engaging with the exercises. Also, *feedback* may have been accessed more often if it was displayed directly instead of requiring the learners to click on it to get it. In the same way, the words for which a definition is available via *glossary* should be more salient. Furthermore, collecting more data on the behaviour and attention allocation of users should enable instructional designers to guide users towards adequate exercises, but also to help them solve exercises when they struggle by, for example, highlighting a certain part of the text where the answer is found. Finally, it may be beneficial to keep track of learners' scores in CALL environments, which it is currently not offered in *NedBox*. In this way, the environment can lead learners to specific exercises based on their needs in different skills. This may, in turn, increase their motivation to complete exercises on the platform to improve their performance.

Notes

1. Because the subtitles were only activated 11 times, it was not possible to compare average fixation duration on video with and without subtitles.

References

Bax, S. (2013). The cognitive processing of candidates during reading tests: Evidence from eye-tracking. *Language Testing*, *30*(4), 441–465. https://doi.org/10.1177/0265532212473244

Beatty, K. (2010). *Teaching and researching: Computer assisted language learning* (2nd ed.). Routledge.

Benson, P. (2011). Language learning and teaching beyond the classroom: An introduction to the field. In P. Benson & H. Reinders (Eds.), *Beyond the language classroom* (pp. 7–11). Palgrave Macmillan.

Bisson, M.-J., van Heuven, W. J. B., Conklin, K., & Tunney, R. J. (2014). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics*, *35*, 1–20. https://doi.org/10.1017/S0142716412000434

Brandl, K. K. (1995). Strong and weak students' preferences for error feedback options and responses. *The Modern Language Journal*, *79*(2), 194–211. https://www.jstor.org/stable/329619

Brunfaut, T., & McCray, G. (2015). *Looking into test-takers' cognitive processes while completing reading tasks: A mixed-method eye-tracking and stimulated recall study*. The British Council. http://eprints.lancs.ac.uk/72005/1/Brunfaut_and_McCray_final_report_FINAL.pdf

Cárdenas-Claros, M. S. (2005). *Field dependence/field independence: How do students perform in CALL-based listening activities?* [Unpublished master thesis]. Iowa State University.

Cárdenas-Claros, M. S. (2020). Conceptualizing feedback in computer-based L2 language listening. *Computer Assisted Language Learning*, *0*(0), 1–26. https://doi.org/10.1080/09588221.2020.1774615

Chapelle, C. (2009). The relationship between second language acquisition theory and computer-assisted language learning. *The Modern Language Journal*, *93*, 741–753. https://doi.org/10.1111/j.1540-4781.2009.00970.x

Conklin, K., Pellicer-Sánchez, A., & Carrol, G. (2018). *Eye-tracking: A guide for applied linguistics research* (1st ed.). Cambridge University Press.

Danjou, C. (2020, January 29). Les applis linguistiques se multiplient, mais que valent-elles vraiment ? *Le Soir*. https://www.lesoir.be/280960/article/2020-01-29/les-applis-linguistiques-se-multiplient-mais-que-valent-elles-vraiment

Fievez, I., Montero Perez, M., Cornillie, F., & Desmet, P. (2020). Vocabulary learning through viewing captioned or subtitled videos and the role of learner- and word-related factors. *CALICO Journal*, *37*(3), 233–253. https://doi.org/10.1558/cj.39370

Fischer, R. (2007). How do we know what students are actually doing? Monitoring students' behavior in CALL. *Computer Assisted Language Learning*, *20*(5), 409–442. https://doi.org/10.1080/09588220701746013

Fischer, R. (2012). Diversity in learner usage patterns. In G. Stockwell (Ed.), *Computer-assisted language learning: Diversity in research and practice* (pp. 14–32). Cambridge University Press. https://doi.org/10.1017/CBO9781139060981.002

Godfroid, A. (2019). *Eye tracking in second language acquisition and bilingualism: A research synthesis and methodological guide*. Routledge. https://doi.org/10.4324/9781315775616

Godwin-Jones, R. (2019). Riding the digital wilds: Learner autonomy and informal language learning. *Language Learning and Technology*, *23*(1), 8–25. https://doi.org/10125/44667

Hegelheimer, V., & Tower, D. (2004). Using CALL in the classroom: Analyzing student interactions in an authentic classroom. *System*, *32*(2), 185–205. https://doi.org/10.1016/j.system.2003.11.007

Heift, T. (2006). Context-sensitive help in CALL. *Computer Assisted Language Learning*, *19*(2–3), 243–259. https://doi.org/10.1080/09588220600821552

Isbell, D. R. (2018). Online informal language learning: Insights from a Korean learning community. *Language Learning & Technology*, *22*(3), 82–102. https://doi.org/10125/44658

Jacob, R. J. K., & Karn, K. S. (2003). Eye tracking in human-computer interaction and usability research: Ready to deliver the promises. In J. Hyönä, R. Radach, & H. Deubel (Eds.), *The mind's eye: Cognitive and applied aspects of eye movement research* (pp. 573–605). Elsevier Science. https://doi.org/10.1016/B978-044451020-4/50031-1

Keuleers, E., Brysbaert, M., & New, B. (2010). SUBTLEX-NL: A new measure for Dutch word frequency based on film subtitles. *Behavior Research Methods*, *42*, 643–650. https://doi.org/10.3758/BRM.42.3.643

Latif, A. M. M. M. (2019). Eye-tracking in recent L2 learner process research: A review of areas, issues, and methodological approaches. *System*, *83*, 25–35. https://doi.org/10.1016/j.system.2019.02.008

Laufer, B., Elder, C., Hill, K., & Congdon, P. (2004). Size and strength: Do we need both to measure vocabulary knowledge? *Language Testing*, *21*(2), 202–226. https://doi.org/10.1191/0265532204lt277oa

Michel, M. C., & Smith, B. (2017). Eye-tracking research in computer-mediated language learning. In S. Thorne & S. May (Eds.), *Language, education and technology*. *Encyclopedia of language and education* (pp. 1–12). Springer International Publishing. https://doi.org/10.1007/978-3-319-02328-1_34-1

Montero Perez, M. (2019). Pre-learning vocabulary before viewing captioned video: An eye-tracking study. *Language Learning Journal*, *47*(4), 460–478. https://doi.org/10.1080/09571736.2019.1638623

Montero Perez, M., Peters, E., & Desmet, P. (2015). Enhancing vocabulary learning through captioned video: An eye-tracking study. *The Modern Language Journal*, *99*(2), 308–328. https://doi.org/10.1111/modl.12215

Muñoz, C. (2017). The role of age and proficiency in subtitle reading: An eye-tracking study. *System*, *67*, 1–10. https://doi.org/10.1016/j.system.2017.04.015

O'Rourke, B., Prendergast, C., Shi, L., Smith, B., & Stickler, U. (2017). Eyetracking in CALL - Present and future. In A. Gimeno, M. Levy, F. Blin, & D. Barr (Eds.), *WORLDCALL - Sustainability and computer-assisted language learning*. Bloomsbury.

Pellicer-Sánchez, A., Tragant, E., Conklin, K., Rodgers, M., Serrano, R., & Llanes, Á. (2020). Young learners' processing of multimodal input and its impact on reading comprehension. *Studies in Second Language Acquisition*, *42*(3), 577–598. https://doi.org/10.1017/S0272263120000091

Peters, E., Velghe, T., & Van Rompaey, T. (2019). The VocabLab tests: The development of an English and French vocabulary test. *International Journal of Applied Linguistics*, *170*(1), 53–78. https://doi.org/10.1075/itl.17029.pet

Poole, A., & Ball, L. J. (2006). Eye tracking in HCI and usability research. In C. Ghaoui (Ed.), *Encyclopedia of human computer interaction* (pp. 211–220). IGI Global. https://doi.org/10.4018/978-1-59140-562-7.ch034

Pujolà, J.-T. (2002). CALLing for help: Researching language learning strategies using help facilities in a web-based multimedia program. *ReCALL*, *14*(2), 235–262. https://doi.org/10.1017/S0958344002000423

Rayner, K. (2009). The 35th Sir Frederick Bartlett lecture: Eye movements and attention in reading, scene perception, and visual search. *The Quarterly Journal of Experimental Psychology*, *62*(8), 1457–1506. https://doi.org/10.1080/17470210902816461

Rayner, K., Reichle, E. D., Stroud, M. J., Williams, C. C., & Pollatsek, A. (2006). The effect of word frequency, word predictability, and font difficulty on the eye movements of young and older readers. *Psychology and Aging*, *21*(3), 448–465. https://doi.org/10.1037/0882-7974.21.3.448

Reinders, H., & White, C. (2016). 20 years of autonomy and technology: How far have we come and where to next? *Language Learning & Technology*, *20*(2), 143–154. http://dx.doi.org/10125/44466

Schiepers, M., Houben, A., Nordin, A., van Loo, H., van Nuffel, H., Verrote, L., & van den Branden, K. (2016). *How to create online practice opportunities for adult second language learners? Nedbox, an online platform for practising Dutch language skills* [Paper presentation]. INTED2016: 10th International Technology, Education and Development Conference, Valencia, Spain. https://doi.org/10.21125/inted.2016.1494

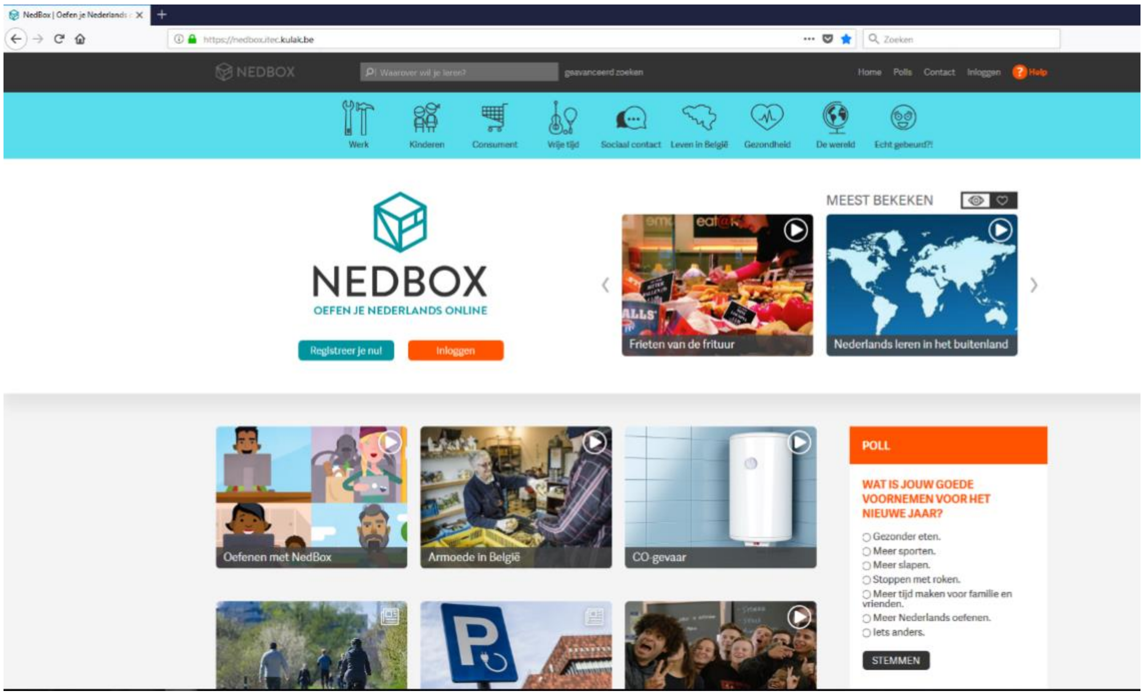
Tragant, E., & Pellicer-Sánchez, A. (2019). Young EFL learners' processing of multimodal input: Examining learners' eye movements. *System*, *80*, 212–223. https://doi.org/10.1016/j.system.2018.12.002

Weir, C., Hawkey, R., Green, A., Unaldy, A., & Devi, S. (2009). The cognitive processes underlying the academic reading construct as measured by IELTS. *IELTS Research Reports*, *9*, 157–189.

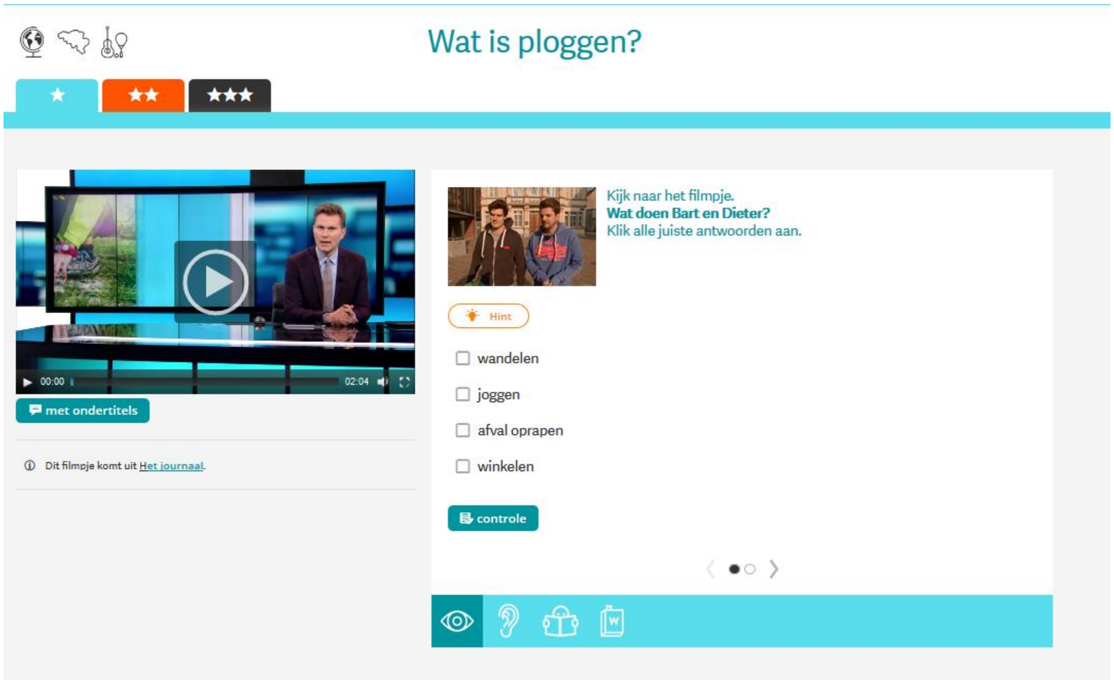
Winke, P., Sydorenko, T., & Gass, S. (2013). Factors influencing the use of captions by foreign language learners: An eye-tracking study. *The Modern Language Journal*, *97*(1), 254–275. https://doi.org/10.1111/j.1540-4781.2013.01432.x

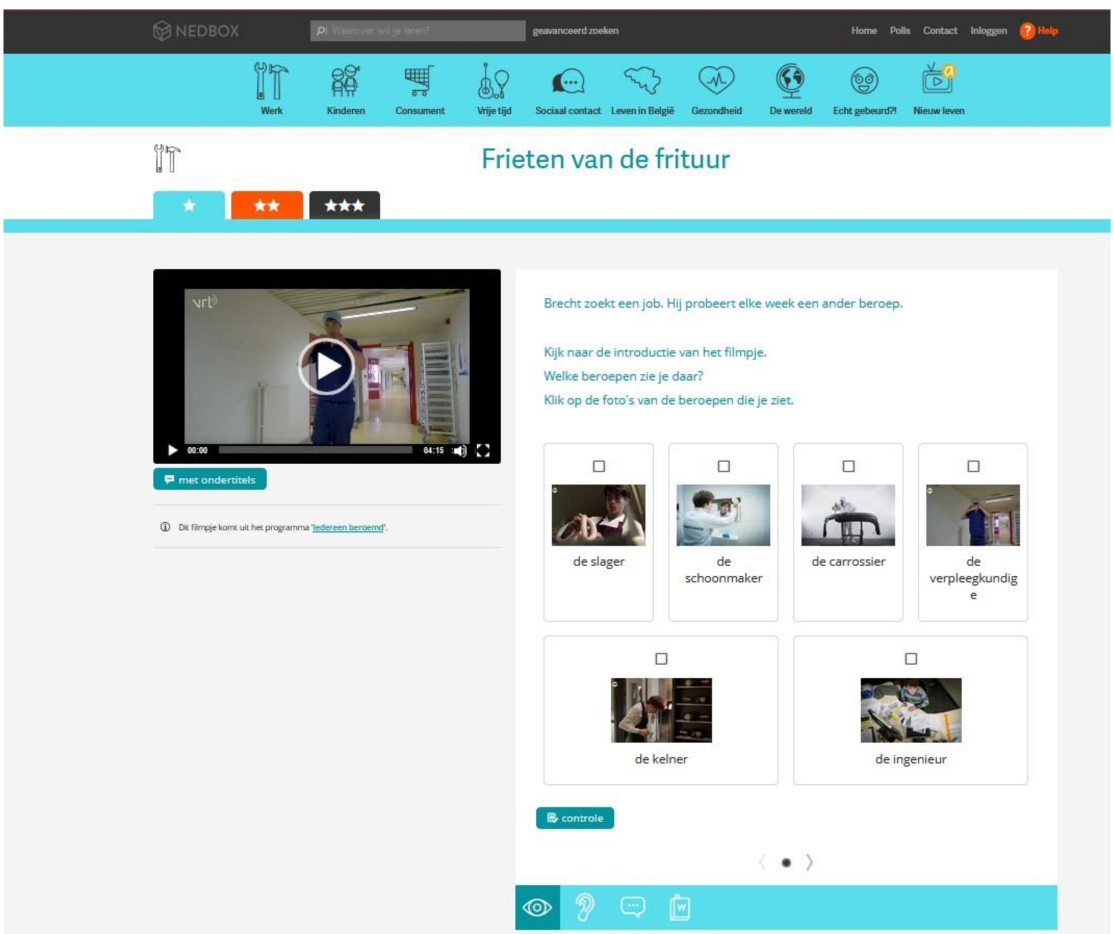
Appendix A. Screenshots from www.nedbox.be

NedBox Homepage

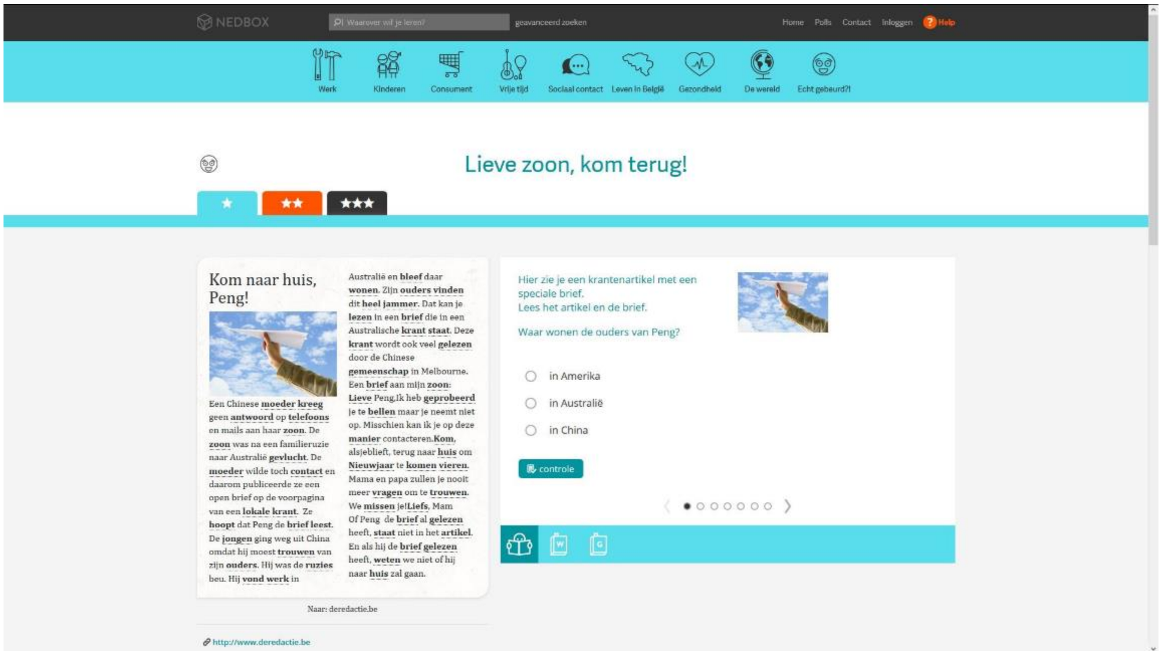


News Items with a Video

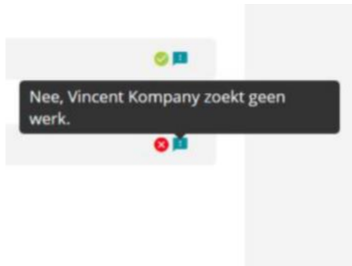
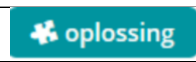




News Items with a Text



Appendix B. Complete List of Help Options Available on Learning Platform

Help Option	Representation on the Website	Description
CHECK		To check one's answer
FEEDBACK		To get feedback on one's answer
SOLUTION		To get the solution of the exercise
HINT		To get a hint on the solution of the exercise
GLOSSARY		To get a definition of a word in a newspaper article
SUBTITLES		To activate or deactivate the subtitles in the video

Appendix C. Complete List of Trials and Characteristics

Trial Name	Input	Length of Video / Words in Text	Number of Participants Who Completed It
Wat is ploggen?	Video	2:04	18/21
Wie betaalt de rekening?	Video	4:09	16/21
Lieve zoon, kom terug!	Text	185 words	14/21
Leer Nederlands met Vincent	Text	139 words*/195 words**	10/21
Werken in de thuiszorg	Video	2:36	5/21
Treinstaking	Video	2:25	2/21
Le Fidèle	Text	105 words	1/21
Chocopasta: gezond?	Text	130 words	1/21
Naar het eerste leerjaar	Video	4:53	1/21

About the Authors

Isabeau Fievez's PhD was part of an INTERREG V project. She obtained a BA in Germanic Languages (Université de Namur), followed by a MA in Linguistics and Literary Studies (Vrije Universiteit Brussel). She is currently working as a project manager for a non-profit organisation. In her doctoral research, she investigated second language vocabulary learning through watching audio-visual input and how the input can be enhanced in order to stimulate learning.

E-mail: i.fievez@hotmail.com

Maribel Montero Perez is Assistant Professor in the Department of Linguistics at Ghent University, Belgium. Her research interests include, amongst others, multimodal input for L2 learning, support tools for L2 learners in online learning environments, technology-enhanced listening, and vocabulary learning.

E-mail: maribel.monteroperez@ugent.be

Frederik Cornillie (PhD) is Research and Valorization Manager at ITEC. His main research interest concerns the effectiveness and use of learning support in technology-mediated environments, in particular in relation to individual differences.

E-mail: frederik.cornillie@kuleuven.be

Piet Desmet is Professor of French, Applied Linguistics and CALL at KU Leuven, Kulak Kortrijk campus, and coordinator of ITEC, imec research group at KU Leuven.

E-mail: piet.desmet@kuleuven.be