Subtitles and Eye Tracking: Reading and Performance

Jan-Louis Kruger

North-West University (Vaal Triangle Campus), Vanderbijlpark, South Africa; and Macquarie University, Sydney, Australia

Faans Steyn

North-West University, Potchefstroom, South Africa

ABSTRACT

This article presents an experimental study to investigate whether subtitle reading has a positive impact on academic performance. In the absence of reliable indexes of reading behavior in dynamic texts, the article first formulates and validates an index to measure the reading of text, such as subtitles on film. Eye-tracking measures (fixations and saccades) are expressed as functions of the number of standard words and word length and provide a reliable index of reading behavior of subtitles over extended audiovisual texts. By providing a robust index of reading over dynamic texts, this article lays the foundation for future studies combining behavioral measures and performance measures in fields such as media psychology, educational psychology, multimedia design, and audiovisual translation. The article then utilizes this index to correlate the degree to which subtitles are read and the performance of students who were exposed to the subtitles in a comprehension test. It is found that a significant positive correlation is obtained between comprehension and subtitle reading for the sample, providing some evidence in favor of using subtitles in reading instruction and language learning. The study, which was conducted in the context of English subtitles on academic lectures delivered in English, further seems to indicate that the number of words and the number of lines do not play as big a role in the processing of subtitles as previously thought but that attention distribution across different redundant sources of information results in the partial processing of subtitles.

As digital contexts proliferate, the types of texts that children and adults encounter keep changing. Moreover, as the contexts in which reading occurs change, the reader has to adopt a variety of reading styles. These reading styles have to adapt to static text on a stable background on the one end of the spectrum (where the pace of reading is determined by the reader) and fleeting text on a dynamic background, as in subtitled or captioned audiovisual texts, on the other end of the spectrum (where the reader has to adjust the pace of reading to the pace of presentation). Furthermore, new media such as smartphones and tablets present information in increasingly hands-free modes that anticipate reading behavior by means of algorithms and scroll text (semi)automatically, presenting readers with a combination of dynamic and static texts.

It can be assumed that the exposure to and reliance on dynamic texts, such as video and interactive multimedia, in both formal and informal education will continue to increase. This could have far-reaching effects for reading instruction if subtitles and other text were to be introduced in a more conscious manner. However, before teachers and reading instructors can employ modes such as subtitles optimally, we require more insight into the way in which children and

Reading Research Quarterly, 49(1)
pp. 105–120 | doi:10.1002/rrq.59

adults read when interacting with dynamic texts. The complexity of measuring subtitle processing is (among other things) caused by issues related to cognitive load when reading competes with other forms of information processing, such as listening and visual search. The methodologies for studying reading in dynamic texts may therefore have to differ significantly from those for studying reading in static texts.

This article presents an investigation of the visual processing of subtitles as changing text (e.g., transcription or translation of dialogue) displayed at the bottom of the screen. In particular, the way in which students process subtitles when watching a recording of academic lectures is studied, as well as the impact the subtitles have on students' understanding of the content of the course. To investigate the processing of subtitles on dynamic text experimentally, the article also presents a way of measuring text processing by making use of eye-tracking data in the context of an absence of reliable techniques for measuring reading as opposed to looking at subtitles.

The study of eye movements as an index of attention allocation in the context of multimodal texts is well established. In fields such as cognitive science, educational science, psycholinguistics, and experimental psychology, eye tracking has been used to investigate matters such as cognitive load, processing of different channels, and attention distribution. However, although various studies have made use of eye tracking to investigate static reading, comparatively few have attempted to use eye tracking to investigate the processing of subtitles and other text on film when presented at the same time as dynamic images and sounds.

This is not to say that eye tracking has not been used productively to investigate subtitle processing to date. Attention allocation to subtitles as text added to film has been investigated by means of eye tracking since at least the 1980s, most notably by the Belgian scholar d'Ydewalle and colleagues between 1985 and 2007 (cf. d'Ydewalle & De Bruycker, 2007; d'Ydewalle, & Gielen, 1992; d'Ydewalle, Muylle, & van Rensbergen, 1985; d'Ydewalle, Praet, Verfaillie, & van Rensbergen, 1991; d'Ydewalle & van Rensbergen, 1989; d'Ydewalle, van Rensbergen, & Pollet, 1987). Between 1997 and 2000, Jensema and colleagues also conducted a (largely qualitative) research project on eye movement strategies of viewers of captioned television (cf. Jensema, El Sharkawy, Danturthi, Burch & Hsu, 2000). In the past five years, a number of other studies have appeared in which eye tracking was used to investigate attention to and processing of subtitles. This interest in the behavioral study of subtitle processing is largely due to the educational benefits found in various performance studies when information is presented in a multimodal manner.

In spite of the substantial number of studies on the processing of subtitles to date, it has not been investigated with the same rigor as, for example, the processing of static texts. In 1998, Rayner provided an overview of 20 years of research on eye movements in reading and information processing (in static texts), and research in this field has increased exponentially since. As a result, whereas the field of reading research has given us a good understanding of how the human mind processes static texts, studies on subtitle processing (or reading over dynamic texts) have not yielded a similar level of insight. In subtitled film, the viewer has to not only manage cognitive resources across different sources of information (verbal and nonverbal, visual and auditory) but also do so without having control over the speed of presentation, unlike in static reading of written or some multimedia texts.

The mere presence of various salient sources of information simultaneously already means that whatever we find in terms of cognitive processes is much harder to ascribe to any one source of information, particularly where there is redundancy between two or more sources of information. Furthermore, current eye-tracking technology does not allow for the analysis of eye movements in the dynamic text of subtitles with the aid of reading statistics, something that can be done with very little effort with most eye-tracking software programs on static texts. This is because subtitles appear (fleetingly) as part of the image and not as combinations of letters. As a result, most studies that have looked at the processing of subtitles by means of eye tracking have either been qualitative in nature (making use of, e.g., heat maps, focus maps, or scan paths) or merely investigated the amount of attention to subtitles.

Subtitle reading can of course be measured in more detail visually by inspecting scan paths of participants watching a subtitled video while their eye movements are recorded. In line with established findings from reading research, the following parameters could be used to measure processing of subtitles, although this list is far from exhaustive or without ambivalences and only serves to provide a rough indication of the mechanics of inspecting reading of subtitles:

- A word can be considered as having been processed if at least one fixation hit the word.
- Short words to the left of a fixation (within the perceptual span of three characters) can be considered processed.
- Full words within the perceptual span of eight characters to the right of a fixation can be considered processed even if they were not fixated.
- Short unfixated words (two or three characters) between two longer fixated words can be considered as processed on the basis of predictive

- reading, even if they are strictly speaking not inside the perceptual span.
- Only words fixated in succession from left to right can be considered to have been processed.
- In many subtitles, the landing site seems unrelated to the reading process because the first fixation may be toward the center of a subtitle, before a regression takes the eyes to the beginning of the subtitle. In such cases, a word on the landing site should not be considered as processed per se unless a subsequent fixation hits the word or includes the word in the perceptual span.
- Reading logic can also be applied in cases where doubt exists, favoring a reading pattern above random fixations.

Measuring reading in this manner could be used to arrive at a reading score or perceived visual processing (PVP) score by dividing the number of words in a subtitle that are considered to have been processed by a participant by the number of words in the subtitle and multiplying that by 100 to arrive at a percentage. Such a PVP score would potentially provide a more accurate index of subtitle reading than any automatic measure because it relies on an actual count of words read based on the scan path of each participant. However, because it relies on the judgment of a researcher or researchers for numerous successive subtitles for each participant, it is prone to errors of subjectivity and also takes a great amount of time.

Determining the PVP is therefore not feasible in any quantitative study of subtitle processing over extended texts. Determining the PVP for 20 participants over a 30-minute video containing around 300 subtitles means that the researcher will have to inspect and note down the visual processing of a total of 6,000 subtitle occurrences. Such inspection has to be done on a frame-by-frame basis, with the researcher going forward and backward to very specific frames in the video many times to verify whether and in what sequence words were looked at. This means that it could take as much as one to two minutes per subtitle per person, or between one and two full weeks of meticulous inspection for a 30-minute film for 20 viewers.

Research Aims

In this article, the impact of subtitle reading (processing) on academic performance is investigated. To achieve this aim, the article has the subsidiary aim of formulating and subsequently validating an index of reading behavior in the context of dynamic texts. This index is intended to make it possible to measure the visual processing of subtitles (in meaningful reading behavior) over long texts by making use of the vast amount of eye movement data obtained by means of eye tracking. Unlike in previous studies where the amount of attention rather than the nature of attention formed the main focus, this study attempts to establish a way of measuring reading behavior more meticulously—in other words, making it possible to determine largely automatically to what extent participants read subtitles. The value of such a validated index lies in its future application in dealing with theoretical issues such as cognitive load and attention allocation in the reception of subtitled film or any multimedia text that incorporates a dynamic audiovisual component with subtitles or other text from a behavioral rather than performance angle.

In addition to arriving at a reading index for dynamic texts, the article also investigates factors that impact reading behavior, such as interindividual differences, nature of the stimulus, and length of subtitles. Hopefully, the index will make it possible to gain more insight into the cognitive issues underlying reading in dynamic texts, which could also have a number of implications for reading instruction and for instructional design by making it possible to design multimedia audiovisual texts that optimize the learning experience by managing the cognitive load brought by the introduction of text.

Subtitling and Cognition: Performance Measures

There are different factors to bear in mind when investigating the reception and processing of an audiovisual text. In the first place, the multimodal nature of these texts results in multiple sources of information being present simultaneously. In the case of hearing, sighted audiences, there is always competition (and interaction) between information in the auditory channel (verbal auditory and nonverbal auditory) and visual channel (verbal visual and nonverbal visual) and, of course, within each channel.

This multimodality of audiovisual texts has been the source of much debate in studies on subtitling. Although subtitling as audiovisual practice has been around for decades and has increasingly received attention in academic research, the bulk of this research has been done on didactic or training issues as well as practical guidelines and translational issues, with comparatively little attention allocated to the processing and reception of subtitles (cf. McLoughlin, Biscio, & Mhainnín, 2011). The studies that have been done on the reception and processing of subtitles can broadly be divided into performance studies on the educational benefits of subtitling and behavioral studies on how, or how much, subtitles are attended to, using eye-tracking methodology.

Performance Studies: Benefits of Subtitling and Dual Coding

Performance studies focus on the use of subtitling in language learning, literacy training, and comprehension and take the form of empirical studies aimed at testing the benefits of delivering information through more than one channel (mainly in an educational context or for educational purposes) and on studies investigating issues of cognitive load in the context of multimedia design. There is extensive evidence in the literature that same-language subtitles, and other forms of subtitling, hold significant potential in education (cf. Bird & Williams, 2002; Danan, 2004; Garza, 1991, 1996; Linebarger, Piotrowski, & Greenwood, 2010; Markham, 1999; Vanderplank, 1988, 1990, 2010). These studies focus on aspects such as listening comprehension, vocabulary acquisition, and language learning.

The benefits of subtitling illustrated in these studies seem to support dual coding theory, which suggests that "a combination of imagery and verbal information improves information processing" (Sydorenko, 2010, p. 50; see also Paivio, 1986, 1991, 2007) as well as what Mayer, Heiser, and Lohn (2001) identify as the information delivery hypothesis, which states that delivering information by more paths results in improved learning. However, the positive findings reported by these scholars are questioned in a more fundamental manner by studies in the field of educational psychology and multimedia learning.

Performance Studies: Disadvantages of Subtitling and Cognitive Load

In the field of multimedia learning, studies that take the dual channel and limited capacity assumptions of cognitive theory (see Mayer, 2002) as a starting point have found a redundancy effect to impact negatively on learning when students are exposed to information in more than one channel (see, e.g., Diao, Chandler, & Sweller, 2007; Mayer, 2002; Mayer et al., 2001). The dual channel assumption is based on a view of the human cognitive system as consisting of two channels, namely the visual-pictorial channel (used for processing pictures that enter the cognitive system through the eyes as pictorial representations) and the auditory–verbal channel (used for processing words that enter the cognitive system through the ears as verbal representations; see Mayer, 2002). This is related to the limited capacity assumption, which holds that the limited capacity of each cognitive system could easily result in cognitive overload in the presence of, for example, too many spoken words, sounds, and images.

In cognitive load theory, the relationship between working and long-term memory means that instructional material should be designed in such a manner that unnecessary cognitive load is avoided. If this is not done, it could result in the redundancy effect when information in more than one channel has to be coordinated, impacting negatively on learning (cf. Diao et al., 2007; Paas, Renkl, & Sweller, 2004). This obviously applies directly to the design of material for reading instruction and language learning, as well as other educational fields.

Subtitling and Cognition: Behavioral Measures

Behavioral studies on subtitling have increased in frequency, and here the focus has been on, for example, determining how one- and two-line subtitles are read, what the difference in attention allocation is between subtitles and the rest of the screen, and the effort involved in reading subtitles. Specifically, subtitle reading was found to be largely automatic, requiring little additional cognitive effort, and viewers were found typically to have little difficulty in distributing visual attention (see d'Ydewalle & De Bruycker, 2007; d'Ydewalle et al., 1991; d'Ydewalle & Gielen, 1992; d'Ydewalle, Warlop, & Van Rensbergen, 1989). In other behavioral studies, the cognitive effectiveness of subtitling has been investigated in terms of film comprehension and dialogue and scene recognition (e.g., Lavaur & Bairstow, 2011; Marian, 2009; Perego, Del Missier, Porta, & Mosconi, 2010). In addition to largely using qualitative methodologies that tend to make use of heat maps and focus maps presented on one frame or screenshot, which accumulates eye movement data over a number of preceding frames of the dynamic text, the samples are generally extremely small—both in terms of participants and in terms of the number of subtitles studied. This often divorces a set of subtitles from a broader context that has the potential to skew the results and impact the behavior. Also, it often makes generalization impossible.

What has hampered more extensive studies to date has been the fact that data on subtitle reading (as opposed to data on visual attention to subtitles) has to be obtained in a largely manual manner, unlike eye tracking of static texts where statistics related to word frequencies, reading ease, and so forth can be extracted fairly effortlessly with most commercial eye-tracking systems. Furthermore, the use of moving window paradigms used in reading to establish perceptual span cannot be applied unaltered in the study of subtitle reading due to the fleeting nature of the text and because subtitles are not processed in isolation but have to contend with other visual and also auditory sources of information that often interrupt the reading process, potentially disturbing a linear reading pattern. The rest of this section provides an overview of studies done on subtitle reading, pointing out the need for a more meticulous mechanism to distinguish mere attention allocation from reading behavior.

The majority of behavioral eye-tracking studies on subtitle reading that do not merely deal with qualitative data, such as heat maps, focus maps, or scan paths, tend to divide the screen into subtitle area and image area and then report some eye-tracking measures for either the subtitles alone or the subtitles versus the image. The eye-tracking measures investigated in these studies are subsequently discussed. Although the findings of the different studies are interesting, this article is limited to a discussion of the measures used so as not to lose focus on the main purpose of arriving at a measure for the visual processing of subtitles.

Those studies that contrast subtitle viewing with image viewing tend to provide information either on the amount of attention to the subtitle area only or the distribution of attention between the two areas. In both of these types of studies, the most common measures of visual attention reported are fixation count, fixation time, fixation time as function of visible time, and/or average fixation duration. This is the case not only in a study by Pavakanun (1992) but also in the substantial body of work done by d'Ydewalle and colleagues (cf. d'Ydewalle et al., 1985, 1987, 1991; d'Ydewalle & Gielen, 1992; d'Ydewalle & van Rensbergen, 1989). These studies typically do not distinguish looking at subtitles from reading subtitles, and they tend to define crude areas of interest (AOIs), such as the entire subtitle area, which means that eye movement data are also collected for the subtitle area when there are no subtitles on screen, which further skews the data.

Specker (2008) investigated the reading of scrolling or upward-rotating, three-line captions produced by means of respeaking, mainly looking at fixation counts and average fixation duration. Her inspection of the fixation plots for the subtitles gives a good interpretation of how the subtitles were read. The main value of her study, which is limited in terms of the duration of the texts and the number of participants, lies in her investigation of consecutive fixations, thereby coming closer to describing reading behavior. Her study also provides a very meticulous multimodal analysis of subtitles together with the image. This qualitative analysis provides valuable insights into the processing of subtitles in the presence of competing sources of information in other channels. However, the qualitative nature and small sample of the study make it impossible to extrapolate to general reading behavior in subtitles.

Three further eye-tracking studies investigated the way in which subtitle reading takes place with different dependent variables. Perego et al. (2010) and Rajendran, Duchowski, Orero, Martínez, and Romero-Fresco (2013) both investigated the impact of text chunking or line divisions on subtitle processing, while Ghia (2012) investigated the impact of translation strategies on subtitle processing.

In their investigation of 28 subtitles viewed by 16 participants, Perego et al. (2010) made use of a threshold line between the subtitle region and the main film zone or upper area and then proceeded to analyze fixation counts, total fixation time, average fixation duration, path length (sum of saccade lengths in pixels), and number of shifts between subtitle region and upper area (revisits or glances).

Rajendran et al. (2013) also looked at the effect of text chunking on subtitling by analyzing the eye-tracking data of 24 participants viewing very short clips (under a minute) under different text chunking conditions. The eye-tracking measures investigated include mean fixation duration, proportion of gaze points and fixations in the subtitles, and number of times a viewer's gaze jumped from scene to subtitles and vice versa (what they call saccadic crossovers, also related to glance count or revisits).

Ghia (2012) looked at the eye movements of 13 Italian intermediate learners of English while watching a six-minute clip from a feature film subtitled into Italian from English as literal or nonliteral translations. She investigated the intensity of visual activity in subtitle reading by analyzing gaze paths to determine deflections between subtitles and image, and also fixations on specific words and regressions, providing a solid qualitative basis for her conclusions.

Two further studies in particular came closer to studying more than just the amount of attention in subtitles (apart from Specker's, 2008, brief investigations of consecutive fixations): d'Ydewalle and De Bruycker (2007) and Bisson, Van Heuven, Conklin, and Tunney (2012). The latter did a fairly detailed analysis of subtitle processing by looking at fixation counts, total fixation duration, average fixation duration, number of subtitles skipped, and proportion of successive fixations (number of successive fixations divided by total number of fixations).

Although Bisson et al. (2012) criticize d'Ydewalle and De Bruycker (2007) for not having a control group who saw the videos without subtitles, the latter is arguably the most thorough research on the reading or visual processing of subtitles to date. Earlier studies by d'Ydewalle and colleagues looked mainly at attention and processing issues, investigating global latency time in shifts from the visuals to the subtitles when they appear, and the percentage of time the observers spent on the subtitles (i.e., time on subtitle as a function of its presentation time). In some of these studies (see, e.g., d'Ydewalle & Gielen, 1992), the time spent on one-line versus two-line subtitles was also investigated. In a later study, d'Ydewalle and De Bruycker investigated reading patterns systematically to determine "to what extent there is regular word-by-word reading behavior in television subtitles among both children and adults" (p. 197). To determine reading behavior, they looked at various dependent variables, including measures of attention allocation (percentage of skipped subtitles, latency time, and percentage time spent in the subtitle area), characteristics of fixations (number of fixations, fixation duration, and word-fixation probability), and characteristics of saccades (saccade amplitude, percentage of regressive eye movements, and number of back-and-forth shifts between visual image and subtitle).

Nevertheless, even in the last two studies, measures of amount of attention dominate, and interpersonal differences are largely ignored. More importantly, none of these studies arrive at a combined index of eye-tracking measures that can be used in studies combining eye-tracking (behavioral) and performance measures. The rest of the article presents such an index, also elaborating on the experiment that was designed to test, verify, and validate the index, before using the index in an experimental study on subtitle reading and performance.

Toward a Reading Index for Dynamic Texts

Based on the previous literature survey and on visual inspection of reading behavior of participants when reading subtitles, a formula was developed with the potential to provide a reliable measure of the visual processing of subtitles, or reading behavior of viewers over extended text, arrived at by calculating the degree to which each individual subtitle was read. Before discussing the manual verification and statistical validation of this formula, its composition is described.

In very simple terms, the Reading Index for Dynamic Texts (RIDT) is taken to be a product of the number of unique fixations per standard word in any given subtitle by each individual viewer and the average forward saccade length of the viewer on this subtitle per length of the standard word in the text as a whole.

In a little more detail, the formula can be written as follows for video *v*, with participant *p* viewing subtitle *s*:

$$\mathrm{RIDT}_{vps} = \frac{\text{number of unique fixations for } p \text{ in } s}{\text{number of standard words in } s} \times \frac{\text{average forward saccade length for } p \text{ in } s}{\text{standard word length for } v}$$

This score then provides an indication of the degree to which a particular subtitle was read by a particular participant (video, participant, and subtitle specific).

If the average of the RIDT*vps* values is taken over all videos for one participant, it is said to be participant specific and is denoted by RIDT*p*, and if it is taken over all videos for all participants, it is simply RIDT without notation. In the same way, the average per video for all participants is the video-specific RIDT*v*, per video and per participant is the video- and participant-specific RIDT*vp*, and the average for all participants per subtitle is the subtitle-specific RIDT*s*.

The two main components of the formula are discussed next, and the rationale behind each measure is provided.

Unique Fixations per Standard Word

The assumption with which this first part of the formula is written is that a higher number of unique fixations per mean word can be interpreted as an indication of more complete processing. This is of course not a principle that can go unqualified, so the second part will compensate.

The number of unique fixations is obtained from the fixation count in the subtitle, excluding refixations, and penalized for regressions. This is intended to measure the unique information intake events in the subtitle by excluding refixations and regressions where no new information is likely to have been processed. As such, refixations and regressions are also taken as indexes of nonlinear reading. Although both refixations and regressions may also be an indication of the processing difficulty of a word, or of low reading proficiency, this should be less of a factor in dynamic texts than in static texts, where the viewer has control over the presentation speed of the text. Because subtitles are typically on screen only long enough to be read once, there is very little time for rereading a subtitle, which changes the role of regressions somewhat from reading static texts.

To identify refixations (or fixations during which no significantly new information is obtained when compared with the information obtained during the preceding fixation), a saccade between two successive fixations in the subtitle with a length of less than two characters qualifies the second fixation as a refixation.

In the case of regressions (or backward saccades when the eyes move from right to left in the same horizontal line), only regressions longer than two characters are taken into account. Also, to further avoid incorrect data, return sweeps are not counted when calculating the number of regressions. Return sweeps are defined as (long) saccades that have both horizontal movement from right to left (or against the direction of reading) longer than the length of a standard word, and vertical movement from top to bottom within the subtitle exceeding the distance between the two subtitle lines (typically around two characters).

The number of standard words is calculated as the number of characters in the subtitle divided by the number of characters in a standard word across the video. The standard word length is determined by dividing the total number of characters in all the subtitles across the video by the total number of words in all the subtitles. The standard word count is used instead of the actual word count to control for the effect of word length on the index.

Average Forward Saccade Length as Function of Standard Word Length

This part of the formula provides some nuance that sets the index further apart from mere mechanical calculations of amount of attention to subtitles found in some of the studies referred to previously. A clinical fixation count does not always provide a true index of reading behavior, and even when refixations and regressions are taken into account, the mechanics of the reading process are not necessarily reflected. However, because saccade length is dependent on more than two fixations separated by forward saccades and could therefore easily return a zero value even in the presence of single or unique fixations, the weighting of this measure is subsequently discussed in more detail. This part of the formula is also intended to factor in reading behavior in that shorter forward fixations should reflect less complete processing.

The forward saccade length is taken as the weighted average forward saccade length in the subtitle. The value for this element requires a number of adjustments involving the exclusion of perceptual jumps and refixations. First of all, the average saccade length is determined by dividing the sum of forward saccade lengths in pixels by the number of forward saccades in the subtitle. In order to not skew the formula, however, only forward saccades between two and 11 characters are included as forward saccades. Based on research into the reading of static text, saccades shorter than two characters are excluded as refixations, whereas saccades longer than 11 characters are excluded as perceptual jumps. This is based on the fact that the word identification span (or area from which words can be identified on a given fixation) is about eight character spaces to the right of a given fixation and about three characters to the left of the fixation (cf. Rayner & Liversedge, 2004; Vitu, 2011). Any saccade larger than 11 characters would therefore leave a perceptual gap between two successive fixations, which would affect the degree of processing. The same applies to very short saccades that could be taken to signal refixation and not a unique fixation.

Second, to avoid a false zero value, a standard value equal to the length of the standard word in the video is assigned as the average forward saccade length in cases where there is only one unique fixation or where there are more than one unique fixations but no valid forward saccades (if all forward saccades are either refixations or perceptual jumps).

The standard word length is taken as the length in pixels of the standard word calculated across the video. This length has to be determined manually based on fixation details in the eye-tracking software and the character length of the standard word.

Balancing Unique Fixations With Saccade Length

In reading, visual attention with visual span of around the same length as a word would result in an optimal factor of 1. Also, reading research suggests that only approximately two thirds of words are fixated, "some words being skipped in the first eye pass on the line of text; only a few skipped words are subsequently fixated following an inter-word regression" (Vitu, 2011, p. 732). However, as pointed out by Vitu, the probability of skipping five-letter words is about .65, and for seven or eight letters, it is only about .3. These factors would suggest that the second part of the formula acts as a normalizing factor. Although there are various other reading characteristics discussed in studies dealing with eye tracking of reading, only these basic principles were taken into account in the formula, and it was decided not to attempt an overview of the literature on static reading in this context.

In reading research, it has been found that individuals who are better readers should require fewer fixations per word because they are better able to predict the text and therefore tend to skip more words (see, e.g., Rayner, 1998, Rayner & Liversedge, 2004; Specker, 2008). In this formula, it may at first seem as though less proficient readers (who have more fixations) would receive a higher RIDT score. However, this is countered by the fact that the formula excludes refixations and also penalizes regressions. Furthermore, working on the assumption that more proficient readers will skip more words because of prediction factors, the second part of the formula rewards longer fixations in the calculation of the factor of average fixation length divided by the length of a standard word.

Gathering the Data

Population and Context

This study uses as context an educational environment where recorded lectures from a Psychology I course are screened to students in the particular course. The sample of participants was taken from first-year students in Psychology I who are second-language speakers of English studying via medium English and whose mother tongue is one of the indigenous languages of South Africa. This was done to control for the variable of home language and also means that the findings of the study reflect the situation for students with a South African indigenous language as mother tongue for whom English is a second language, looking at same-language subtitles in English (both the soundtrack and subtitles in English). Out of the approximately 400 students enrolled in Psychology I, a sample was obtained by inviting all students who fall into the language category mentioned and who failed their first semester in the course (for these students, motivation to participate would be higher, and in this way, a control is introduced for the variable of academic performance).

Eye-tracking and performance data were collected for 36 students in total who watched recordings of six individual classes from the course with or without subtitles (the videos constituted the core of the second semester of the first year of psychology). The six videos were presented during five individual sessions (with session 1 containing two shorter videos).

The 36 participants were randomly divided into a test group (n = 18) that saw the videos with subtitles and a control group (n = 18) that saw the videos without subtitles. The test group (mean = 53.06, SD = 6.54) and the control group (mean = 54.21, SD = 6.04) did not differ significantly on their average grade 12 or matriculation results, $t(34)$ = 0.55, p = n.s., and were therefore comparable.

After watching all six videos, the participants completed a comprehension test consisting of the multiple-choice section of the previous year's final examination paper in the module. Three participants were removed from the test group and two from the control group because they did not see all six videos. The test consisted of 40 items, with one correct option and three incorrect options each. The reliability of the test and the items was calculated based on the data of 399 students who wrote the paper the previous year, and four items that had a very low correlation with the total test score (<.05) were eliminated. The Cronbach's α for the remaining 36 items was .7, which is considered acceptable.

In terms of comprehension, the test group (mean = 44.90, SD = 10.22) and the control group (mean = 49.82, SD = 8.14) did not differ significantly: $t(29)$ = 1.49, p = n.s. This finding does not support cognitive overload, nor does it support the information delivery hypothesis, but once again it emphasized the need for an index to determine the degree to which the subtitles were actually read by the test group.

One participant had to be excluded because she did not watch all the videos and because her data on the remaining videos was incomplete. The final group consisted of three men and 14 women, with a mean age of 20.9 years. In terms of the language distribution, 47% of the participants were from the Nguni family of languages (seven spoke isiZulu as their home language and one isiXhosa) and 53% from the Sotho family of languages (three spoke Sesotho as their home language, five Setswana, and one Sepedi).

Experiment

Participants reported to the eye-tracking laboratory one by one, in five different sessions scheduled over a period of two weeks. When they arrived for the first session, the experiment was explained to them, and they had the opportunity to read and sign the informed consent form, which formed part of the ethics clearance for the project. In order not to influence their behavior, they were simply told that they would be watching a series of recorded lectures while their eye movements were recorded and that they would answer questions on the content afterward. The participants were also told that the aim of the experiment was to determine how students process educational material (without calling any attention to the subtitles in the instructions).

Each participant was asked to sit in a comfortable position in front of the stimulus monitor, at a distance of approximately 700 mm from the stimulus screen. Their eye movements were monitored and recorded using the iView X remote eye-tracking system (RED), which uses a built-in camera with a sampling rate of 50 Hz. The detection of events in the RED system uses fixations as the primary events, with the algorithm that is used to detect fixations being based on dispersion. Fixation detection parameters are set to detect a fixation when the eyes remain in the same area (a maximum dispersion of 100 pixels) for a period of more than 80 ms (SensoMotoric Instruments, 2010). The screen resolution was 1,280 pixels × 1,024 pixels, and the physical stimulus dimensions were 376 mm horizontally and 301 mm vertically.

Each participant's eyes were calibrated in iView X before each of the five sessions using a 9-point calibration and validated to ensure the accuracy of the data. Participants were also instructed to sit as still as possible to ensure accuracy, although the system does allow for some movement. The average calibration deviation for participants was 0.48 degrees on the x-axis and 0.61 degrees on the y-axis. The average tracking ratio or quality for participants across the trial was 90.1%. Tracking quality is the number of nonzero gaze positions divided by the sampling frequency, multiplied by the run duration, expressed as a percentage (see SensoMotoric Instruments, 2010).

Although the duration of the six videos presented in five sessions differed, the videos were set up in SMI's

Experiment Centre to begin immediately after calibration and to run without pause from beginning to end (i.e., the videos were not self-paced), after which the video stopped automatically. All videos were therefore displayed to all participants in exactly the same manner for exactly the same length of time.

To obtain the necessary data for the formula, the subtitles in the videos were first marked individually as AOIs. This was done by drawing a box around each individual subtitle covering the width and height of the subtitle, but also including a border around the subtitle of approximately two characters on each side. The AOI for each subtitle was toggled on exactly on the frame in which the subtitle appears and toggled off on the frame on which the subtitle disappears. This made it possible to collect all the data and only the data pertaining to the presentation of subtitles (in this case, the position and time data for each fixation made by each participant in all the subtitles).

From BeGaze 3.1, the data was then exported as AOI fixation details, including subject, AOI, fixation start, fixation duration, fixation end, horizontal position, and vertical position. From this data, sorted by subject, then AOI, and then fixation start, it was possible in Microsoft Excel to calculate the following per AOI: number of fixations, number of refixations, number of regressions, number of return sweeps (based on coordinates of successive fixations to identify longer regressions), saccade direction, and saccade length.

The data required for standard words were extracted from the subtitle files, whereas the length of characters in pixels was determined manually by obtaining the position on the *x*-axis of two fixations made by one participant, one near the beginning of a subtitle line and one near the end. The *x*-position of the first subtracted from the *x*-position of the second yielded the number of pixels between the two fixations, which was divided by the number of characters including spaces between the two fixations to arrive at the number of pixels per character. This was done for around 20 subtitles per video in BeGaze3.1, and the final pixel length for each video was calculated as the average of the average pixel length over the 20 subtitles.

Validation

Once the RIDT scores were calculated, the scores were validated. This was done on a random sample from the subtitles in each video. The randomizer function in Microsoft Excel was used, which generates random values between 0 and 1 for each subtitle. The subtitles with random values under .1 were taken for the random sample. Of the 1,579 subtitles in the first four videos, 145 subtitles were sampled, amounting to 9.2% of the total number of subtitles. Due to the sheer volume of data and the time-consuming nature of the visual inspection, it was decided to leave out the final video for the purposes of the validation. The details of the video duration, word length, and sample size for the four remaining videos are seen in Table 1.

To validate the scores for the 145 subtitles, each subtitle was inspected individually for each participant to determine to what extent the subtitle was read by each participant, or the PVP of subtitles. This meant that each of the 145 subtitles had to be inspected 17 times, first by a research assistant and then by the researcher, with significant discrepancies being inspected individually toward a consensus. This took a considerable amount of time but was essential for the validation of the index. The PVP was then calculated as a percentage of words in the subtitle that were processed or that fell within the perceptual range. In other words, the PVP score could be said to reflect the extent to which the subtitle was actually processed or "read" by each participant, something the final index is meant to determine automatically from the eye-tracking data.

Here, the notation PVP*vps* is used to denote the PVP value for video *v* for participant *p* who views subtitle *s*. As in the case of RIDT, PVP*v*, PVP*s*, PVP*p*,

TABLE 1
Statistics for Samples From Videos

Video	Approximate duration	Number of subtitles	Number of words	Number of characters	Standard word length in characters	Sampled number (and percentage) of subtitles
1	44 minutes	596	4,798	26,166	5.45	49 (8.2%)
2	30 minutes	427	3,284	18,197	5.54	42 (9.8%)
3	26 minutes	350	2,841	16,099	5.67	33 (9.4%)
4	15 minutes	206	1,597	9,103	5.70	21 (10.2%)
Total	115 minutes	1,579	12,520	69,565	5.56	145 (9.2%)

and PVP*pv* denote the video-specific, subtitle-specific, participant-specific, and participant- and video-specific values, respectively. PVP without notation is used for the average of all participants processing all subtitles in all videos.

In contrast to the laborious process of calculating reading of subtitles visually as explained in the introduction, the RIDT requires the researcher to mark each subtitle as an individual AOI and then export the eye movement data, calculate the pixel length of a character in the video, and enter the data in the spreadsheet containing the formula, all of which would take less than a day irrespective of the number of participants.

Factors Influencing the PVP

Because the number of words contained in a subtitle as well as the number of lines could play a role in the visual processing of the subtitle (cf., e.g., d'Ydewalle & Gielen, 1992), PVP*vps* scores were adjusted for these covariates in a two-way ANCOVA, with subjects (15 participants who watched all four videos) as a random factor and videos as a fixed factor on four levels because the videos were classes in the same course presented by the same lecturer. The subject × video interaction was found to be statistically highly significant ($p < .0001$) in the ANCOVA performed by the statistical computer package STATISTICA (StatSoft, 2011). Because this interaction made interpretations regarding the overall effects of subjects and videos difficult, it was therefore decided to repeat a one-way ANCOVA with a subjects factor for each video. For all videos, the between-subject effects were highly significant, pointing to a significant interindividual variance. However, although such interindividual variance could be expected due to the difference in the way individuals deal with cognitive load, this variance only accounted (i.e., the partial eta-squared values) for between 15% and 25% of the total variance of the PVP within videos.

The covariate of word count did not have a significant effect, meaning that shorter subtitles were not processed more fully than longer subtitles. In the case of the covariate of number of lines per subtitle, significance was only found in the case of video 1 ($p = .039$), as can be seen in Table 2. This would suggest that earlier findings on the effect of number of lines (and, by implication, word count) on subtitle reading by d'Ydewalle and Gielen (1992) are brought into question and may have been overstated or only applicable to a situation in which participants were used to reading a lot of subtitles.

Although the percentage of words in a subtitle that were processed cannot be expected to be normally distributed, it was found that the residuals resulting from the ANCOVA models were distributed near enough to a normal distribution to conclude the highly significant effects stated previously.

Validation of the RIDT

To establish the reliability of the RIDT, it was then correlated (using the Pearson correlation coefficient) with the PVP for all videos for participant and subtitles (i.e., RIDT*vps* and PVP*vps*) and at participant and subtitle level (RIDT*ps* vs. PVP*ps*), at participant-specific level (RIDT*p* vs. PVP*p*), and at subtitle- specific level (RIDT*s* vs. PVP*s*). A high correlation would mean that the RIDT is a reliable and robust measure of visual processing. The Pearson correlation coefficient was therefore calculated between the PVP scores and the RIDT values at all three levels. The correlations for all 15 participants for all 145 subtitles as well as per video are displayed in Table 3.

The distributions of PVP and RIDT scores are not necessarily normal, and this can influence the linear relationship and the correlation. Table 4 gives the descriptive statistics of both PVP and RIDT from which deviations from normality are apparent because the skewness value differs somewhat from 0 and because kurtosis value is smaller than the target of 3 (in the case of normality). After transforming both of these variables by means of the Box–Cox transformation to obtain normality, the scatterplot still indicated a strong linear relationship with a Pearson correlation of .798. The Spearman correlations were also calculated, which indicate a strong monotone positive relationship. Because the interpretation of both PVP and RIDT will be difficult when transformed into other scales, it was decided to stick to the original scales, which clearly did not influence their linear relation substantially.

The very strong linear relations that can be observed in Tables 3 and 5 between the PVP scores and the RIDT index values, which were all highly statistically significant, suggest that the RIDT is valid for measuring visual processing of text, or more specifically reading behavior in dynamic texts in a time-efficient and robust manner.

However, because the RIDT as such is not expressed as a percentage, it can be used to differentiate between subtitles based on the degree (higher vs. lower) of processing. Ideally, such an index should make it possible for the researcher to determine whether subtitles were processed fully, partially, or not at all. This would require the calculation of a baseline for full processing, something that will have to be investigated empirically in the future.

Interpretation of the RIDT

Figures 1 and 2 illustrate the index. For participant P16a, the PVP on subtitle 67 of the first video (AOI067)

TABLE 2
Results of One-Way ANCOVAs for the Different Videos

Video	Effect	Fixed or random	Sum of squares	Degree of freedom (df)	Mean square (MS)	Error df	Error MS	*F*	*p*	Partial eta-squared
1	Intercept	Fixed	239,529.572	1.000	239,529.572	37.613	1,898.994	126.135	.000	.770
	Words	Fixed	189.193	1.000	189.193	718.000	814.600	0.232	.630	.000
	Lines	Fixed	3,476.141	1.000	3,476.141	718.000	814.600	4.267	.039	.006
	Subject	Random	189,396.603	14.000	13,528.329	718.000	814.600	16.607	.000	.245
	Error		584,882.566	718.000	814.600					
2	Intercept	Fixed	70,221.467	1.000	70,221.467	73.583	1,382.276	50.801	.000	.408
	Words	Fixed	1,564.612	1.000	1,564.612	613.000	854.602	1.831	.177	.003
	Lines	Fixed	2,388.635	1.000	2,388.635	613.000	854.602	2.795	.095	.005
	Subject	Random	111,733.159	14.000	7,980.940	613.000	854.602	9.339	.000	.176
	Error		523,870.838	613.000	854.602					
3	Intercept	Fixed	129,449.284	1.000	129,449.284	90.657	1,124.182	115.150	.000	.560
	Words	Fixed	246.317	1.000	246.317	478.000	744.647	0.331	.565	.001
	Lines	Fixed	1,229.822	1.000	1,229.822	478.000	744.647	1.652	.199	.003
	Subject	Random	96,929.075	14.000	6,923.505	478.000	744.647	9.298	.000	.214
	Error		355,941.500	478.000	744.647					
4	Intercept	Fixed	68,015.990	1.000	68,015.990	96.545	1,150.824	59.102	.000	.380
	Words	Fixed	233.217	1.000	233.217	298.000	801.002	0.291	.590	.001
	Lines	Fixed	0.558	1.000	0.558	298.000	801.002	0.001	.979	.000
	Subject	Random	79,171.778	14.000	5,655.127	298.000	801.002	7.060	.000	.249
	Error		238,698.508	298.000	801.002					

TABLE 3
Pearson Correlations Between PVP_{pvs} and $RIDT_{pvs}$ (with Spearman correlations) for All Participants Over All Videos and Each Video Individually

PVP_{pvs}: RIDT*pvs*	PVP_{v1}: $RIDT_{v1}$	PVP_{v2}: $RIDT_{v2}$	PVP_{v3}: $RIDT_{v3}$	PVP_{v4}: $RIDT_{v4}$
.81 (.77)	.77 (.73)	.81 (.76)	.82 (.82)	.82 (.84)

Note. PVP = perceived visual processing. RIDT = Reading Index for Dynamic Texts.

TABLE 4
Descriptive Statistics of PVP_{pvs} and $RIDT_{pvs}$

Processing	*N*	Mean	Median	Min	Max	Lower quartile	Upper quartile	Standard deviation	Skewness	Kurtosis
$RIDT_{pvs}$	2,175	0.69	0.74	0.00	2.62	0.38	0.97	0.41	−.08	2.61
PVP_{pvs}	2,175	54.38	61.54	0.00	100.00	28.57	80.00	31.97	−.40	1.94

Note. PVP = perceived visual processing. RIDT = Reading Index for Dynamic Texts.

TABLE 5
Descriptive Statistics and Pearson Correlations Between RIDT and PVP

Processing	*N*	Mean	Standard deviation	Correlation
Person specific				
RIDT*p*	15	0.68	0.12	
PVP*p*	15	54.36	11.77	0.85
Subtitle specific				
RIDT*s*	145	0.69	0.20	
PVP*s*	145	54.38	14.39	0.87

Note. PVP = perceived visual processing. RIDT = Reading Index for Dynamic Texts.

was 53% (eight of the 15 words identified as falling in the perceptual range). For participant P07t, the PVP was 80% (12 of the 15 words falling in the perceptual range). The respective RIDT scores for the two participants were .730 and .938, as subsequently explained.

In the case of participant P16a, seven of the 16 fixations on this subtitle were preceded by regressions, resulting in nine fixations during which new information could be said to have been acquired. For participant P07t, three of the 12 fixations were preceded by regressions, likewise resulting in nine unique fixations. There were 11 standard words in subtitle AOI067. The average length of the eight forward saccades was 99.4 pixels for P16a and 127.75 pixels for P07t, while the standard word length in pixels for the video was 111.49. The RIDT score therefore clearly reflects the higher degree of reading by P07t, who attended to both lines of the subtitle, whereas P16a only attended to the first line, possibly triggered by the spelling mistake. Here are the calculations of the RIDT scores:

$$\text{RIDT P16}a = \frac{9}{11} \times \frac{99.4px}{111.49px} = .73$$

$$\text{RIDT P07}t = \frac{9}{11} \times \frac{127.75px}{111.49px} = .94$$

Table 6 provides the details of the fixations and saccades.

Determining the Effect of Subtitles on Performance

To achieve the first aim of the article—namely, to determine the impact of subtitle reading (processing) on academic performance—correlations were performed between the performance in the posttest and the reading of the participants. It is only when we consider the correlation between comprehension and the degree to which participants actually read the subtitles that we begin to approach an understanding of the role of subtitles in comprehension. In fact, any conclusions on the impact of subtitles on language learning, comprehension, reading skills, or academic literacy levels have to be qualified strongly if the research is unable to provide an indication

FIGURE 1
Scan Path of Participant P16a on AOI067

FIGURE 2
Scan Path of Participant P07t on AOI067

of whether participants actually read the subtitles. In this study, we therefore correlated the participant-specific RIDT*p* score for each participant across all six videos used in the experiment with the performance of each participant on the posttest, which was written within two weeks after each participant watched the final video.

Of the 18 participants in the initial test group, three were removed because they did not see all six videos. One

Table 6
Fixations and Saccades for Participants P16a and P07t

Subject	Saccade length (in pixels)	Forward saccade length	Fixations	Regressions
P16a	−56.8	0	1	−1
	−58.4	0	1	−1
	192.9	192.9	1	0
	95.7	95.7	1	0
	90.3	90.3	1	0
	45.6	45.6	1	0
	88.5	88.5	1	0
	−125.2	0	1	−1
	−55.5	0	1	−1
	90.3	90.3	1	0
	100.4	100.4	1	0
	−91.7	0	1	−1
	91.5	91.5	1	0
	−97.6	0	1	−1
	−208.3	0	1	−1
	0	0	1	0
P07t	−131	0	1	−1
	85.3	85.3	1	0
	96.2	96.2	1	0
	91.3	91.3	1	0
	148.3	148.3	1	0
	−13.6	0	1	0
	151.7	151.7	1	0
	−46.5	0	1	−1
	−267.3	0	1	−1
	−23.9	0	1	0
	193.7	193.7	1	0
	0	0	1	0

additional participant's data were removed for the purposes of calculating the RIDT*p* score because that participant's eyes tracked very poorly for three of the six videos. For the remaining 14 participants, a correlation was performed between the average RIDT*p* score and the posttest comprehension score of each participant. This yielded a statistically significant positive correlation across all six videos: $r = .56$, $p < .05$ (the Spearman rank order correlation was .48). This finding adds another dimension to the findings on the impact of (the presence of) subtitles (as opposed to subtitle reading) on comprehension already discussed in the methodology section. In other words, although no significant difference could be found between the performance of those participants who saw the videos with subtitles and those who saw the videos without subtitles (which may result in the interpretation that subtitles did not influence comprehension), those who saw the videos with subtitles and also read the subtitles performed better than those who saw the videos with the subtitles but did not read the subtitles as fully.

Conclusion

Although no evidence could therefore be found that the mere presence of subtitles improves performance, this study has shown that subtitled audiovisual material may still add significant value to education because of the high correlation between subtitle reading and performance. This study investigated the impact of subtitle processing on performance in the context of an English as a second language learning environment and also investigated this over a relatively extended period (two weeks and six audiovisual texts with a delayed posttest). This, as well as the addition of a reading index, enhances and clarifies the results of previous studies in the fields of educational design and language acquisition where claims have been made on the impact of subtitled material in an educational context without the studies concerned taking the actual processing of subtitles into account.

It can therefore be confirmed that reading instruction and language acquisition courses, among other courses, stand to benefit greatly from the use of reading over dynamic texts—not least because this form of text has proliferated so much over the past decades. To utilize such texts, instructional design has to interrogate exactly how children and adults process dynamic texts. This cannot be done without a reliable measure of reading in these texts. Past studies discussed in the literature overview would have benefited greatly from a distinction between attention allocation to and reading of the text (subtitles), something that cannot be deduced accurately without a formula such as the RIDT proposed here. Without a reliable measure of the extent to which children or adults read the text in subtitled video or in other multimedia texts, it is impossible to deduce that any benefits can be ascribed to the text (subtitles or other text on dynamic sources). Likewise, conclusions about cognitive load, and specifically cognitive overload, would benefit greatly from the nuance provided by this index precisely because it allows for the quantification of reading.

In conclusion, then, this article set out to investigate the impact of subtitle reading on performance and, to do so, to formulate and validate an index for reading behavior in dynamic texts or for determining the extent to which text such as subtitles are read by viewers of dynamic texts. With this validated RIDT, the article therefore presents a way in which to make sense of behavioral eye-tracking data in the processing of text in dynamic scenes that can now be used in studies that combine performance and behavioral data. In particular, this index could make a significant contribution to future research in media psychology, where the aim is to determine the contribution of written text in dynamic media to the cognitive processing of the text as a whole. For example, the quantification of reading in this context makes it possible to isolate this source of information more scientifically over extended texts, which would enable researchers to determine what the impact of a source such as subtitles is on cognitive load. This would in turn make it possible to arrive at more reliable foundations for multimedia learning and to test the redundancy effect as well as cognitive overload against the information delivery and dual coding hypotheses.

In the field of audiovisual translation, the index could also play a significant role in providing researchers with a way to test the impact of factors such as line segmentation, word frequency, and font size on the processing of subtitles.

The most important contribution of the index is situated in the potential it holds for researchers on the processing of written text in dynamic scenes to obtain quantitative data to supplement the largely qualitative data interpretation that has been used to date. What the study already indicates is that word count does not seem to impact significantly on subtitle reading and that the number of lines in a subtitle has a much smaller effect on subtitle reading than found in other studies (see, e.g., d'Ydewalle & De Bruycker, 2007; d'Ydewalle & Gielen, 1992). The statistically significant positive correlation between performance and RIDT also suggests that, in an educational context at least, students who read subtitles in a subtitled video stand to benefit from the mode.

NOTES

We would like to thank Bertus van Rooy, director of the UPSET research focus area at North-West University, for his input; Este Hefer and Gordon Matthew from the eye-tracking laboratory for their assistance in the collection of the data; Gordon Matthew for his assistance in the automatization of the formula in Excel, which can be obtained from the corresponding author; and Marqia van der Walt, who assisted with the visual inspection of subtitles for the PVP.

REFERENCES

Bird, S.A., & Williams, J.N. (2002). The effect of bimodal input on implicit and explicit memory: An investigation into the benefits of within-language subtitling. *Applied Psycholinguistics*, *23*(4), 509–533. doi:10.1017/S0142716402004022

Bisson, M.-J., Van Heuven, W.J.B., Conklin, K., & Tunney, R.J. (2012). Processing of native and foreign language subtitles in films: An eye tracking study. *Applied Psycholinguistics*. Advance online publication. doi:10.1017/S0142716412000434

Danan, M. (2004). Captioning and subtitling: Undervalued language learning strategies. *Meta: Translators' Journal*, *49*(1), 67–77. doi:10.7202/009021ar

Diao, Y., Chandler, P., & Sweller, J. (2007). The effect of written text on comprehension of spoken English as a foreign language. *The American Journal of Psychology*, *120*(2), 237–261.

d'Ydewalle, G., & De Bruycker, W. (2007). Eye movements of children and adults while reading television subtitles. *European Psychologist*, *12*, 196–205. doi:10.1027/1016-9040.12.3.196

d'Ydewalle, G., & Gielen, I. (1992). Attention allocation with overlapping sound, image, and text. In K. Rayner (Ed.), *Eye movements and visual cognition: Scene perception and reading* (pp. 415–427). New York, NY: Springer-Verlag. doi:10.1007/978-1-4612-2852-3_25

d'Ydewalle, G., Muylle, P., & van Rensbergen, J. (1985). Attention shifts in partially redundant information situations. In R. Groner, G.W. McConkie, & C. Menz (Eds.), *Eye movements and human information processing* (pp. 375–384). New York, NY: Elsevier.

d'Ydewalle, G., Praet, C., Verfaillie, K., & van Rensbergen, J. (1991). Watching subtitled television: Automatic reading behavior. *Communication Research*, *18*(5), 650–666. doi:10.1177/009365091018005005

d'Ydewalle, G., & van Rensbergen, J. (1989). Developmental studies of text–picture interactions in the perception of animated cartoons with text. In H. Mandl & J.R. Levin (Eds.), *Knowledge acquisition from text and pictures* (pp. 233–248). New York, NY: Elsevier. doi:10.1016/S0166-4115(08)62157-3

d'Ydewalle, G., van Rensbergen, J., & Pollet, J. (1987). Reading a message when the same message is available auditorily in another language: The case of subtitling. In J.K. O'Regan, & A. Levy-Schoen (Eds.), *Eye movements from physiology to cognition* (pp. 313–321). New York, NY: Elsevier.

d'Ydewalle, G., Warlop, L., & Van Rensbergen, J. (1989). Television and attention: Differences between young and older adults in the division of attention over different sources of TV information. *Medienpsychologie: Zeitschriftfür Individual- und Massenkommunikation*, *1*, 42–57.

Garza, T.J. (1991). Evaluating the use of captioned video materials in advanced foreign language learning. *Foreign Language Annals*, *24*(3), 239–258. doi:10.1111/j.1944-9720.1991.tb00469.x

Garza, T.J. (1996). The message is the medium: Using video materials to facilitate foreign language performance. *Texas Papers in Foreign Language Education*, *2*(2), 1–20.

Ghia, E. (2012). The impact of translation strategies on subtitle reading. In E. Perego (Ed.), *Eye tracking in audiovisual translation* (pp. 157–182). Rome, Italy: Aracne.

Jensema, C.J., El Sharkawy, S., Danturthi, R.S., Burch, R., & Hsu, D. (2000). Eye movement patterns of captioned television viewers. *American Annals of the Deaf*, *145*, 275–285. doi:10.1353/aad 2012 .0093

Lavaur, J.M., & Bairstow, D. (2011). Languages on the screen: Is film comprehension related to viewers' fluency level and to the language in the subtitles? *International Journal of Psychology*, *46*, 455–462. doi:10.1080/00207594.2011.565343

Linebarger, D., Piotrowski, J.T., & Greenwood, C.R. (2010). On-screen print: The role of captions as a supplemental literacy tool. *Journal of Research in Reading*, *33*(2), 148–167. doi:10.1111/j.1467-9817.2009.01407.x

Marian, V. (2009). Audio-visual integration during bilingual language processing. In A. Pavlenko (Ed.), *The bilingual mental* *lexicon: Interdisciplinary approaches* (pp. 52–78). Tonawanda, NY: Multilingual Matters.

Markham, P.L. (1999). Captioned video-tapes and second language listening word recognition. *Foreign Language Annals*, *32*(3), 321–328. doi:10.1111/j.1944-9720.1999.tb01344.x

Mayer, R.E. (2002). Cognitive theory and the design of multimedia instruction: An example of the two-way street between cognition and instruction. *New Directions for Teaching and Learning*, *2002*(89), 55–71. doi:10.1002/tl.47

Mayer, R.E., Heiser, J., & Lohn, S. (2001). Cognitive constraints on multimedia learning: When presenting more material results in less understanding. *Journal of Educational Psychology*, *93*(1), 187–198. doi:10.1037/0022-0663.93.1.187

McLoughlin, L.I., Biscio, M., & Mhainnín, M.A.N. (Eds.) (2011). *Audiovisual translation subtitles and subtitling: Theory and practice*. Oxford, UK: Peter Lang.

Paas, F., Renkl, A., & Sweller, J. (2004). Cognitive load theory: Instructional implications of the interaction between information structures and cognitive architecture. *Instructional Science*, *32*, 1–8. doi:10.1023/B:TRUC.0000021806.17516.d0

Paivio, A. (1986). *Mental representations: A dual coding approach*. Oxford, UK: Oxford University Press.

Paivio, A. (1991). Dual coding theory: Retrospect and current status. *Canadian Journal of Psychology*, *45*, 255–287. doi:10.1037/h0084295

Paivio, A. (2007). *Mind and its evolution: A dual coding theoretical approach*. Mahwah, NJ: Erlbaum.

Pavakanun, U. (1992). *Incidental acquisition of foreign language through subtitled television programs as a function of similarity with native language and as a function of presentation mode* (Unpublished doctoral dissertation). University of Leuven, Belgium.

Perego, E., Del Missier, F., Porta, M., & Mosconi, M. (2010). The cognitive effectiveness of subtitle processing. *Media Psychology*, *13*(3), 243–272. doi:10.1080/15213269.2010.502873

Rajendran, D.J., Duchowski, A.T., Orero, P., Martínez, J., & Romero-Fresco, P. (2013). Effects of text chunking on subtitling: A quantitative and qualitative examination. *Perspectives: Studies in Translatology*, *21*(1), 5–21. doi:10.1080/0907676X.2012.722651

Rayner, K. (1998). Eye movements in reading and information processing: 20 years of research. *Psychological Bulletin*, *124*(3), 372–422. doi:10.1037/0033-2909.124.3.372

Rayner, K., & Liversedge, S.P. (2004). Visual and linguistic processing during eye fixations in reading. In J.M. Henderson, & F. Ferreira (Eds.), *The interface of language, vision, and action: Eye movements and the visual world* (pp. 59–104). Hove, UK: Psychology.

SensoMotoric Instruments. (2010). *BeGaze™ manual: Version 2.4*. Teltow, Germany: Author.

Specker, E.A. (2008). *L1/L2 eye movement reading of closed captioning: A multimodal analysis of multimodal use* (Unpublished doctoral dissertation). University of Arizona, Tucson.

StatSoft. (2011). *STATISTICA* (V. 10) [Data analysis software system]. Available: www.statsoft.com

Sydorenko, T. (2010). Modality of input and vocabulary acquisition. *Language Learning & Technology*, *14*(2), 50–73.

Vanderplank, R. (1988). The value of teletext sub-titles in language learning. *ELT Journal*, *42*(4), 272–281. doi:10.1093/elt/42.4.272

Vanderplank, R. (1990). Paying attention to the words: Practical and theoretical problems in watching television programmes with uni-lingual (CEEFAX) sub-titles. *System*, *18*(2), 221–234. doi:10.1016/0346-251X(90)90056-B

Vanderplank, R. (2010). Déjà vu? A decade of research on language laboratories, television and video in language learning. *Language Teaching*, *43*(1), 1–37. doi:10.1017/S0261444809990267

Vitu, F. (2011). On the role of visual and oculomotor processes in reading. In S.P. Liversedge, I.D. Gilcrest, & S. Everling (Eds.), *The Oxford handbook of eye movements* (pp. 731–749). New York, NY: Oxford University Press.

JAN-LOUIS KRUGER (corresponding author) is the director of and an associate professor in the School of Languages on the Vaal Triangle Campus of North-West University, Vanderbijlpark, South Africa; e-mail janlouis.kruger@nwu.ac.za. Beginning December 2013, he will be an associate professor in translation and interpreting at Macquarie University in Sydney, Australia.

FAANS STEYN is a professor and consultant in the Statistical Consultation Services at North-West University in Potchefstroom, South Africa.