Seeing Is Not Perceiving:
A Preliminary EMMA Study of EFL Reading

Yueh-Nu Hung
National Taichung Teachers College

Abstract
EMMA (eye movement and miscue analysis) is a fairly new term coined by Ken Goodman and his colleagues at the University of Arizona. It is both a research and an evaluation tool for reading research. While eye movement research has a history of more than one hundred years, and miscue analysis research can be tracked back to the 1960s, the combination and relation of visual data from eye movements and the oral data from reading miscues are powerful tools to probe, unobtrusively, into the essence of reading. The study first introduces what EMMA methodology is, and then reports a preliminary EMMA research that involved two adult Mandarin Chinese speakers’ reading of an English story. Their eye movements were collected with ASL 504 eye tracker, and oral reading miscues were analyzed based on the guidelines in Reading Miscue Inventory written by Goodman, Watson, and Burke. The EMMA data shows that reading was more than a perceptual process of word or letter recognition, and the readers were consciously monitoring their meaning construction process. The results also show that reading is a perceiving, not seeing, process. Based on the EMMA data, the two readers’ English readings are also compared and evaluated.

Key words: eye movement, miscue analysis, reading process, EFL reading

INTRODUCTION
Reading researchers have long been studying the reading process based on overt and observable oral reading performance or comprehension tests. However, these methods do not reveal the thinking process that goes on in the reader’s head at the time of reading. How wonderful it would be to be able to see inside the reader’s head and watch what is happening in the process of meaning construction. Among the various reading research methods and techniques, two stand out because they are both naturalistic (that is, the reader does no more than read aloud and retell) and authentic (that is, real and whole texts are used instead of out-of-context pieces and bits of words or phrases): eye movement research and miscue analysis (Paulson & Goodman, 2001; Paulson, 2002). The acronym EMMA stands for eye movement and miscue analysis, a fairly new research methodology developed by Ken
Goodman and his colleagues at the University of Arizona in the United States of America. This study first briefly introduces this new reading research methodology. Then an EMMA study that involved the English reading of two Mandarin Chinese speakers is presented to show what EMMA enables us to do in the exploration of reading processes. Attention is also paid to the comparison and evaluation of the two subjects’ English reading based on the EMMA data. Finally, further EMMA studies are suggested.

**What Is EMMA?**

**Eye Movements**

Eye movement research has a long history that dates back more than a hundred years. Javal first reported in 1879 that the human eyes do not glide in a smooth manner but instead make a series of stops when they look at a scene or read lines of print (reported in Huey, 1908/1968). These stops are called fixations, from which the eyes gather information for the brain to process. Movements from one fixation to another are called saccades, which are so short in time length that nearly no information is gathered from them. Dodge’s research results (1900) indicate that since no useful information is received during the movements of the eyes, research should concentrate on the pauses, or fixations, the eyes make.

Since Javal and Dodge, eye movement research has been continually used in a large number of studies to answer questions from various disciplines, including reading. Researchers have studied fixations and saccades in eye movements to try to understand the nature and process of reading. O’Regan (1990) observes that earlier eye movement work in reading had been conducted with a view to understand the underlying perceptual process. Later work in the 1950s was done in order to find the most favorable conditions for reading. Then starting from around the 1980s, because of interest in linguistics and psycholinguistics, researchers used “eye movements as an indicator of the reader’s cognitive processes” (O’Regan, 1990, p. 395).

Since the inception of eye movement research at the turn of the previous century, many researchers have explored various aspects of the reading process with the help of eye movement information. From the physiological studies of the eyes, we know several basic facts about how the eyes behave when reading. For example, Huey (1908/1968) showed that the first fixation on a line is frequently not at the first word but at the second or even third. Likewise, the final fixation is usually not at the last word. Huey’s data also demonstrates that readers fixated on anywhere from 20 to 70 percent of the words on a line. Rayner and Sereno (1994) reported that during a fixation, the eye has access to three regions for viewing information: the foveal, parafoveal, and peripheral. The foveal region provides sharp and clear visual information, and it includes 2 degrees of visual angle (about 6 to 8 letters in focus). The parafoveal region extends to about 15 to 20 letters, and the peripheral region includes everything in the visual field beyond the parafoveal region. Research also
shows that content words are fixated far more than function words (Just & Carpenter, 1987).

Eye movement studies also reveal important information about the meaning construction process of reading. A myriad of researchers have investigated how different linguistic aspects (i.e., phonology, orthography, syntax, and semantics) and the higher level psycholinguistic and cognitive processes (like predicting, sampling, and the use of contextual information) shape word recognition and reading processes (e.g., Just & Carpenter, 1983, 1984, 1987; McClelland & O’Regan, 1981; Fisher & Shebilske, 1985; Balota, Pollatsek, & Rayner, 1985; Pollatsek, Lesch, Morris, & Rayner, 1992; Underwood, Clews, & Everatt, 1990; Rayner & Morris, 1992; Rayner & Well, 1996). The Eye movement research method has also been used to study Chinese reading. For example, Yang (1994) studied how properties of Chinese characters and words affect word perception and eye movement control in reading Chinese.

**Miscue Analysis**

Miscue analysis also has a long history that dates back to the 1960s. Any deviations between a reader’s oral reading and print could be considered mistakes or errors due to carelessness or laziness on the part of the reader, but miscue analysis takes a different view. Goodman has argued that the same linguistic processing explains both fluent reading and miscues (Goodman, 1996). Miscues are results of mis-use of language cues, and hence the name “miscues.”

In miscue analysis studies, whole real texts are used instead of out-of-context pieces and bits of words or phrases, and therefore the reading performance in miscue analysis studies resembles real life reading. Besides, miscue analysis provides a way for researchers to investigate the reading process at the time of reading, not after. From miscues, we are able to see the inner working of the socio-cultural psycholinguistic transactional meaning construction process of reading. Miscues provide a “windows on the reading process” (Goodman, 1973, p. 5). Several hundred miscue analysis studies have been conducted in a dozen or more different languages (Brown, Goodman & Marek, 1996), and miscue analysis remains a powerful tool in reading research.

**Eye Movements and Miscue Analysis**

Since 1999, Ken Goodman and his colleagues at the University of Arizona have been developing research methods by combining eye movement and miscue analysis, which they labeled EMMA (eye movement, miscue analysis). This new methodology relates eye movement data to miscue analysis data, and expands on what we learn from each methodology individually. With the new EMMA methodology, greater insights into the processes of perception during reading that have never been documented before have been developed. Several studies have been completed and some others are currently in progress.
using EMMA methodology. In Paulson’s study (2000), he carefully examined the eye movement data around where miscues occurred, and demonstrated that reading is a perceiving, and seeing, process. With EMMA methodology, Duckett (2001) was able to track where his first graders were looking when they read picture books and explored how much picture information compared to print was utilized in the reading process. EMMA data from Russian, Japanese, Arabic, and Korean reading is now being analyzed, and my own EMMA study (Hung, 2002) of Chinese reading adds a piece to the exploration of reading process in general with EMMA methodology.

A PRELIMINARY EMMA STUDY

A preliminary EMMA study was conducted to show what this new reading methodology can do and to understand native Mandarin Chinese speaker’s reading of English, an area of English reading that hadn’t been pursued with EMMA methodology.

Participants

Both participants, John and Jane (pseudo-names), in the study are adult native Mandarin speakers who speak English as a foreign language. They were born and educated in Taiwan, and at the time of the study they were both doing their doctoral studies. John was a 24-year-old male with a natural science background, and Jane a 36-year-old female with an educational psychology background. Both agreed to participate in the study voluntarily.

Reading Material

The well-known text “Boat in the Basement” was selected because of its long history of being used as experimental material to show the psycholinguistic reading process. It is a short story of 51 words (including the title) with 6 implanted errors (a repetition of the on the third line, boot for boat, though for through, he for she, a part for apart, and should of for should’ve). The full text is shown below.

Short as it might it, the text is nevertheless a good choice for this preliminary EMMA study of EFL reading. It was written by Gollasch (1980). He asked elementary and college students to read “Boat in the Basement,” and the reading errors (or miscues) of these students reveal the meaning construction process of reading. Goodman (1996) uses it in his book for readers to see their own reading strategies of predicting, sampling, confirming, and so forth. The selection of this material, then, is particularly crucial since it allows the researcher to see the oral reading miscue data, the visual eye movement data, and the readers’ reading performance around the implanted errors.
Title

The Boat in the Basement

Line 1  A woman was building a boat in her
Line 2  basement. When she had finished the
Line 3  the boot, she discovered that it was
Line 4  too big to go though the door. So he
Line 5  had to take the boat a part to get
Line 6  it out. She should of planned ahead.

Equipment

Eye tracking system Model 504 manufactured by Applied Science Laboratories of the United Stated was used for collecting eye movement data in the study. Unlike earlier eye tracking models that required the reader to put on a helmet, have the head stationed, or keep a bite bar in the mouth, this model involves non-obtrusive equipment that does not require the reader to wear or put on anything. The reader simply sits in front of the computer screen from which he or she will read, and a camera will follow the eye automatically, allowing approximately one square foot of head movement.

Procedure

When the data collection session began, I explained to the reader everything we were going to do in the lab to make her/him feel comfortable. First, the left eye was calibrated, and then the reader started reading “Boat in the Basement.” The oral reading was tape-recorded, and the eye movement performance was digitally video-recorded. After the reader finished reading the story, she/he was asked to retell the story as she/he remembered. At this point, the eye tracking had been stopped but audio recording continued. After the reader finished the retelling, I asked her/him questions to clarify her/his comprehension of the reading. The whole session took about 30 minutes.

Data Analysis

For eye movement data, raw data was first transferred into a fixation file, and fixation points were plotted on the page. Then the total fixation number of each reading was calculated, along with the mean fixation duration, regression number, regression rate, total reading time, and the percentage of words that were fixated. For miscue analysis data, oral reading miscues were first marked and then analyzed following Goodman, Watson and Burke’s oral miscue analysis guidelines (1987). Special attention was paid to places where substitutions, omissions, insertions, and meaning loss occurred.

Results and Discussion

In what follows, I present eye movement data and the miscue analysis data. Based on
the two data sources combined (i.e., the EMMA data), discussions are made about the reading process generally and the two subjects’ reading specifically.

**Eye movement data.** The following table shows the two subjects’ basic eye movement data regarding total fixation number, mean fixation duration, regression number and rate, total reading time, and percentage of words that were fixated.

<table>
<thead>
<tr>
<th>Subject</th>
<th>Fixation Number</th>
<th>Mean Fixation Duration (mini seconds)</th>
<th>Regression Number</th>
<th>Regression Rate</th>
<th>Reading Time (seconds)</th>
<th>Fixated Words</th>
</tr>
</thead>
<tbody>
<tr>
<td>John</td>
<td>47</td>
<td>478</td>
<td>2</td>
<td>4.2%</td>
<td>22.481</td>
<td>69%</td>
</tr>
<tr>
<td>Jane</td>
<td>42</td>
<td>441</td>
<td>2</td>
<td>4.8%</td>
<td>18.516</td>
<td>63%</td>
</tr>
</tbody>
</table>

The two mean fixation durations (478 and 441 mini seconds) are longer than those reported in other eye movement studies, which usually range from 200 to 300 mini seconds. This could be explained by the facts that (1) English is a foreign language for the two readers, (2) the reading text is short, and (3) there are implanted errors in the text. These facts probably could also explain the observation that the percentages of fixated word (69% and 63%) are higher than average. John spent a longer time reading the story (1.21 times longer than Jane), and had more fixations (47 vs. 42) and longer average fixation duration (478 vs. 441). Overall, based on the eye movement data, John seemed to be less proficient in English reading than Jane.

**Miscue analysis data.** Based on the oral miscues, the same distinction between the two readers is revealed. While Jane was more careful in her oral reading, John made a lot more miscues. Jane made miscues only on the second sentence of the story, while John made miscues on all sentences except the first. In the following transcripts of their reading, the miscues are underlined.

Jane’s oral reading of the second sentence:

When she had finished the **ba**-. the...boot, she discovered that it was too big to go **through** the door.

John’s oral reading of the whole text:

The boat in **a** basement. A woman was building a boat in her basement. When
she had finished the boat…boot, he discovered that it was too big to go through to the door. So he…he had to take the boat apart to get it out. So she should of planned ahead.

The second sentence, as orally read by John, was syntactically unacceptable with the insertion of a. It is semantically acceptable, but there was a meaning change because of the substitution of he for she. Looking at the numbers of their miscues, one is tempted to think that John is a sloppy and careless reader. However, their retelling says otherwise. Jane retold, in Chinese, that “A lady wants to build a boat, but she makes a boot instead. The boot is too big and she can’t get it out of the door. She should have planned ahead and should not make the boot too big to get out of the door.” When asked whether she noticed anything strange about the story, she reported that “I don’t know what boot has to do with the story. The English is strange, and the grammar is strange too. Also, the story plot is strange.” On the contrary, John’s retelling, also in Chinese, shows that he understood the story well. “This is a story about building a boat. A woman is building a boat, but the boat she makes is too big to get out of the door. She should have planned ahead.” When asked whether he noticed anything strange about the story, this is what John said: “I didn’t pay any special attention to the words. I was a little nervous but eager to finish reading the story.” John made a lot more miscues than Jane did, but showed better comprehension of the text, probably because he was willing to take risks. Jane, on the other hand, was conservative and remained literal in her understanding of the text.

**EMMA data.** When the visual eye movement data and the oral miscue analysis data are compared and related, we gain deeper insights into reading processes and behaviors. First of all, we see that reading is not a serial word recognition process. The eye movement data clearly shows that readers do not fixate every word on the reading material. John fixated 69% of the words in the text while Jane fixated 63%. Figure 1 shows Jane’s eye movements on the fourth line. There are only five fixations on this 9-word line.

![Figure 1. Jane’s eye movements on the fourth line.](image)

The data also shows that readers make regression fixations. Regressions are made out of the necessity to make sense of the print. Jane’s two regressions appear on the implanted error *should of* (Figure 2), and John’s two regressions appear on the implanted error *he* (Figure 3). It is obvious that when facing information from the implanted errors, both readers regressed to search for more information in order to construct a coherent meaning of
the text. In John’s case, he regressed first to the first letter of *had*, thus making two fixations on this word, and then back to the implanted error *he* which was the last word of the previous line, thus making a long swipe backward and up. These regressions show the comprehension monitoring of the readers.

![Figure 2. Jane’s regression fixations on the sixth line.](image)

![Figure 3. John’s regression fixations on the fourth and fifth lines](image)

EMMA data also shows that what you see physiologically is not what you understand. In other words, seeing is not perceiving. Both readers fixated on *though*, yet *through* was produced (Figure 4 and 5). The two words look and sound alike, but differ syntactically and semantically. The readers highly expected to see *through* based on the graphophonic, syntactic and semantic cues in the text. In other words, in the process of reading, the brain is in control, not the eyes. Smith goes further and says that reading is only incidentally visual (1994). The EMMA data shows that reading is a psycholinguistic meaning construction process, not merely a perceptual process of letter or word recognition.

![Figure 4. Jane’s eye movements and miscues on the fourth line.](image)

![Figure 5. John’s eye movements and miscues on the fourth line.](image)

**CONCLUSION**

This study is a preliminary investigation of native Mandarin Chinese speakers’ reading of English with EMMA methodology. It first introduces what this methodology is and what it allows us to learn about reading, and then presents an EMMA research model based on two
readings. The results show that readers did not fixate every word in the text sequentially. They regressed to monitor their meaning comprehension, and they highly predicted what would come next based on the graphophonetic, syntactic, and semantic cues of the text. The results also show that Jane, seemingly more proficient in her reading, was too reliant on the text and less confident in her meaning construction. John, on the other hand, was more risk-taking.

Due to the small number of subjects and readings, the findings cannot be generalized to other native Mandarin speakers’ reading of English. Further EMMA studies are called for to investigate various aspects of EFL reading of Mandarin Chinese speakers. For example, how is the EMMA data different from those of native English speakers? What do readers do when they encounter new or unfamiliar words? What do they do when the text is syntactically challenging? From the lens of EMMA data, more understanding and insights into these questions can hopefully be achieved.

REFERENCES


