# The Effects of Biofeedback-based Stimulated Recall on Self-Regulated Online Learning: A Gender and Cognitive Taxonomy Perspective

**Zhai, X., Fang, Q., Dong, Y., Wei, Z., Yuan, J., Cacciolatti, L. and Yang, Y.**

This is the peer reviewed version of the following article: Zhai, X., Fang, Q., Dong, Y., Wei, Z., Yuan, J., Cacciolatti, L. and Yang, Y. (2018), The Effects of Biofeedback-based Stimulated Recall on Self-Regulated Online Learning: A Gender and Cognitive Taxonomy Perspective, *Journal of Computer Assisted Learning*, which has been published in final form at:

https://dx.doi.org/10.1111/jcal.12284.

Journal of Computer Assisted Learning

# The Effects of Biofeedback-based Stimulated Recall on Self-Regulated Online learning: A Gender and Cognitive Taxonomy Perspective

| | |
|---|---|
| Journal: | *Journal of Computer Assisted Learning* |
| Manuscript ID | JCAL-17-157.R2 |
| Manuscript Type: | Article |
| Technology and Tools: | Handheld/Mobile devices, Intelligent systems, Information systems |
| Learning process / Pedagogy: | Individual learning, Tutorial |
| Paradigm: | Cognitivism, Constructivism, Student-Centred |
| Level of education : | Undergraduate |
| Place of learning: | Formal Learning |
| Type of research: | Mixed methods |
| Research technique: | Experimental, Survey / Questionnaire |
| Analysis/evaluation paradigm: | Participatory Evaluation |
| Issues: | Attitude, Skills, Self-direction / Self-regulation, Knowledge |
| | |

# The Effects of Biofeedback-based Stimulated Recall on Self-Regulated Online learning: A Gender and Cognitive Taxonomy Perspective

Xuesong Zhai[123], Qiansheng Fang[2], Yan Dong[13*], Zhihui Wei[4], Jing Yuan[5], Luca Cacciolatti[6], Yalong Yang[2]

[1] *Faculty of Education, School of Educational Technology, Beijing Normal University, Beijing PRC*
[2] *Anhui Provincial Key Laboratory of Intelligent Building and Building Energy Saving, Anhui Jianzhu University, Hefei, PRC*
[3] *Smart Learning Institute, Beijing Normal University, Beijing PRC*
[4] *Open Education Research Publisher, Shanghai Open University, Shanghai, PRC*
[5] *School of Foreign Studies, Anhui Sanlian University, Hefei, PRC*
[6] *Westminster Business School, University of Westminster, London, UK*
*corresponding author

## Abstract

Previous studies posited the effectiveness of Stimulated Recall (SR) by exposing learners to recorded videos enhancing their personal perceptions and authentic understanding of knowledge in an interactive classroom. However, few studies explored how SR is implemented in a relatively static context, e.g., online self-directed learning, or took human factors, e.g., cognitive style and gender, into consideration in such a context. To fill this gap, the current study, based on previous psychological research findings, aims to introduce biofeedback as a stimulus for learners to engage in retrospection regarding their learning behavior. A quasi-experimental design study was carried out over a 12-week set of EFL (English as a Foreign Language) self-regulated online reading activities. The participants consisted of an experimental group (54 undergraduate students) and a control group (52 undergraduate students) at one Chinese university. Pre-post tests on reading performance and their association with a specific cognitive taxonomy were assessed through a developed scale instrument, whereas physiological signals (e.g., gazing duration, verbal fixation and brain wave) were captured via eye-tracking and electroencephalograph (EEG) technology. The results emphasized that (a) students' reading ability and cognitive hierarchy significantly improved through biofeedback stimulation. Moreover, (b) learners in single level-one cognitive hierarchic groups had significant improvements in both cognitive abilities and reading comprehension, whereas learners in multi-level hierarchic groups had no significant enhancements. Finally, (c) the optical data results and EEG reports showed that males favor procedural feedback and females have a preference for a conclusive assessment.

**Keywords:** Biofeedback; Stimulated Recall; Self-Regulated; Taxonomy

## 1. Introduction

The technique of stimulated recall (SR) is considered a valuable tool for learners to capitalize on introspection and cognitive processes (King, 1980; Peterson & Clark, 1978). Although many researchers have applied SR to assess learners' thoughts in a traditional learning pattern, the challenge is how to implement this method in the context of self-regulated online learning (Meier & Vogt, 2015), the reason being that online self-regulated learning is generally carried out in a relatively static mode, which lacks the interactivity to generate a stimulus (Duo & Song, 2012). In addition, some researchers maintain that learners' response to the SR, mainly in the form of a self-reported verbal protocol, suffers from a lack of validity (Meade & McMeniman, 1992). Moreover, a plethora of studies suggest that learning performance is closely associated with human factors (Li & Kirkup, 2007; Lu & Chiou, 2010), which has led us to take learners' human factors into consideration in the current research.

The prior psychological findings suggest that the biofeedback training in SR could significantly improve children's attentive behavior (Linden, Habib, & Radojevic, 1996), and we propose that biofeedback techniques may be incorporated into SR as a stimulus in the context of online autonomous learning. Specifically, the inclusion of physiological information increases accuracy and provides an intelligent identification of users' individual emotional and learning status, allowing for more personalized pedagogical design. Additionally, the biological signals captured from learners minimize the superficiality of self-reported data. Considering the popularity and overall widespread usage of various physiology measurement devices, eye trackers and portable electroencephalograph (EEC) readers were chosen for the current study.

The participants were selected from a university in China and assigned to either an experimental group or to a control group and administered an EFL online reading task for a period of 12 weeks. Their reading abilities and cognitive taxonomy levels were tested before and after the experiment, and their physiological measurements were incorporated in the study. With this quasi-experimental setup, this study tries to answer the following questions:

(1) Can biofeedback as a stimulus significantly influence students' reading comprehension and cognitive hierarchy when in an online autonomous learning mode?

(2) In light of different personal cognitive hierarchic levels in learners, how does biofeedback affect students' cognitive taxonomies and reading abilities?

(3) In light of gender differences, how does biofeedback influence students' learning behavior?

The paper is structured in the following way: After this introduction, section two deals with the research background, section three describes the methodology, section four reports the results, section five provides a discussion in light of the current literature, and section six concludes the paper, highlighting some implications.

## 2. Literature Review

### *2.1 Self-Regulated Learning*

Self-regulated learning (SRL) ~~are~~is defined as an active, constructive process by which learners initiate monitor, regulate, and control their cognition, motivation, and behavior processes to achieve their learning goals (Pintrich, 2000). Zimmerman (1989) posited that self-regulated learning is the triadic interaction between self-observation, self-judgment and self-reaction for their thoughts, feelings and actions (Zimmerman, 1989). With the increasing development of information technology, SRL has been closely integrated with an online context, which provides flexible accessibility and additional resources for learners to perform asynchronous learning without the barriers of space and time.

Although many previous studies have examined the positive effect of an appropriate online SRL strategy on leaning outcomes and perceptions (Devolder, van Braak, & Tondeur, 2012; Panadero, Kirschner, Järvelä, Malmberg, & Järvenoja, 2015), some critical arguments requiring further exploration remain, and learners' disengaging from online SRL suggests that strategies are of special value for achieving SRL. Firstly, although the content and results of learning behavior could be observed in some online SRL systems, learners' feelings related to actual behaviors are difficult to be examined. Additionally, learners may fail to make a corrective self-judgment of their personal characteristics, which may lead to the impairment of individualized learning environments in online SRL (Kizilcec, Pérez-Sanagustín, & Maldonado, 2017). Moreover, the measurements of SRL are commonly limited to self-reported instruments and/or a think-aloud approach, which possibly distract learners from the target task and cause cognitive overload (Mey & Mruck, 2010). It is suggested that informative assessment and process mining techniques be employed in online SRL (Houben, 2016).

Stimulated recall is regarded as an applicable approach for recollecting and assessing learners' thoughts about their self-regulated learning, because the retrospection can be conducted without distracting students from their learning tasks and provide an additional description of a particular event. Furthermore, some open questions can be designed during the process (Meier & Vogt, 2015). Moreover, simulated recall with a biofeedback stimulus can offer learners' emotional situations, which helps us to investigate and explain learners' performances from the perspective of human feelings, which may be valuable for the construction of an individualized learning context in SRL. Additionally, stimulated recall with biofeedback stimulus is considered a useful formative measurement that provides reliable objective information about learners to assess their learning behavior in the context of online SRL.

### *2.2 Stimulated Recall*

Stimulated recall comprises introspective procedures through which participants' cognitive processes help learners engage in more effective learning by adopting a stimulus (normally a recorded video) to be delivered to the student at the time of learning (Iovane, Salerno, Giordano, Ingenito, & Mangione, 2012). Bloom (1953) observed that SR could be useful for examining humans' covert cognitive behavior. In addition, many constructivists, based on the theory of constructivism, have found that stimulated recall is a valid approach to aid students' learning strategies (Jensen, 2000).

The decades of research in the domain of SR-enhanced learning can be categorized into three main categories. The *first* is the effectiveness of this method with regards to both learning outcomes and interactive cognitions. Lindgren's (2002) research presented that learners' EFL writing skills were significantly improved when SR was adopted (Lindgren, 2002). Furthermore, a study using video clips and photographs to stimulate primary school children to recall science center exhibits resulted in higher engagement with the science center (Lindgren, 2002). Additionally, SR is regarded as a useful method for constructing individuals' relationship between cognition and behaviors within learners (Meade & McMeniman, 1992). *Second,* SR has been extensively implemented in various research contexts within a variety of academic subjects as diverse as second language learning (Gass, 2001; Selinker & Gass, 2008) and nursing education (Wang, Liang, Blazeck, & Greene, 2015), as well as a variety of learning setups (e.g., traditional face-to-face delivery versus online setups) and a variety of participants (e.g., primary school students and mature students). *Third,* the stimulus source may differ from one research context to another. For instance, some studies indicated that although SR generally includes audio-video replay, another variant of the stimulus could include participants' physiological data (Jennett & Affleck, 1998).

Although the majority of studies have shed light on the strengths of SR, the different applications of SR and the variety of stimuli that may potentially be used highlight some questions. For instance, current research does not explain an issue that is mostly concerned with SR as a method (Tjeerdsma, 1997): the supplement of information to incomplete memories or rather introspection. This question may be attributed to the observation that the stimulus sources, normally presented in audio-video narrative episodes, may not be able to produce cognition per se (Wilcox & Trudel, 1998). This assumption is consistent with Gass' (2001) research, which points out that recall may decay with delayed protocols since learners may treat a stimulus as a recollection instead of a reflection. It is therefore plausible that the type of stimulus used in SR may stimulate users' cognitive activity in different ways. In what follows, we discuss the effects of biofeedback on learning.

### *2.3 Biofeedback*

Biofeedback, which includes a series of physiological stimuli, has been widely employed to investigate users' emotional states when operating many smart devices (Huang, Hwang, & Chen, 2014; Picard & Picard, 1997). Biofeedback makes it possible to narrow the gap between the human and the machine (Sano & Picard, 2013) by computing humans' affection through the recognition and analysis of humans' physiological signals. We propose that the usage the biofeedback could benefit the enhancement of personalized human-computer interactions.

With the increase of research improvements in physiology, many smart devices and systems have been improved and are being used extensively in different fields of application (e.g., psychology, neuro-sciences, and education). Among these tools, eye trackers and portable EEGs are normally recommended as a method, especially in E-learning settings, due to their portability and economical costs. As far as eye tracking is concerned, researchers take into consideration seven research themes and three eye-tracks measurements, namely, the position of fixation (to test location of interest), the fixation duration (to examine the extent to which readers focus on a target), and the scan-path (to explore reading habits); all of these provide a promising channel to help connect learners' cognition to learning outcomes (Lai et al., 2013).

Current studies in technology and education often utilize eye-tracking techniques to test online reading activities. For instance, Kang (2014) used eye-trackers to compare online reading patterns and comprehension between readers whose reading language is their first language and those whose reading language is their second language. Another study analyzed readers' scan-path data collected by an eye tracker to explore the ways in which readers view the different features of different genres, or topics, in a text document (Clark, Ruthven, Holt, Song, & Watt, 2014).

In addition to eye tracking, EEGs have also been extensively used in many research fields ranging from psychology to education. Some psychologists found that EEG signals could be used in the biofeedback training mode and that the attentive behavior of children affected by Attention Deficit Disorder (ADD) or Attention Deficit Hyperactivity Disorder (ADHD) were significantly improved (Linden, Habib, & Radojevic, 1996).

Furthermore, the relationship between EEG features and corresponding emotional states have been tested and confirmed in many learning contexts (Wang, Nie, & Lu, 2014). For instance, some research based on EEG analysis found that personal local features can significantly enhance students' prediction performance in a self-paced learning environment (Yamauchi, Xiao, Bowman, & Mueen, 2015). In light of the current literature, there is strong evidence of the effectiveness of biofeedback in enhancing the human-machine interaction and, as a consequence of this interaction, the ability of the biofeedback to generate a change in behavior. Thus, we propose that eye movement tracking and EEG data can be utilized as personalized feedback to enhance online learners' reading outcomes and belonging to multi-level cognitive taxonomies. In what follows, we deal with human factors and their effects on learning behavior.

### *2.4 Human Factors and Learning Behavior*

A number of empirical studies have demonstrated that learning behavior has a close relationship with different human factors, such as cognitive style, knowledge level and gender. For this very same reason educational technology development has been increasingly highlighted in personalized learning systems and applications. Learners' cognition has been demonstrated to be a significant variable predicting students' learning performance (Hung, Lin, Fang, & Chen, 2014).

Some authors maintain that the cognitive taxonomy may be utilized as a significant educational instrument in teaching critical reading in EFL classes (Surjosuseno & Watts, 1999). Most of the various cognitive hierarchy instruments are essentially similar to Bloom' cognitive taxonomy. Bloom's cognitive taxonomy is a six-level classification system whose categories are knowledge, comprehension, application, analysis, synthesis and evaluation. These are ranked from "lower order" to "higher order" thinking and used to measure the level of cognitive achievement (Krathwohl, Bloom, & Masia, 1964). These six categories in the taxonomy are useful tools for planning and guiding various teaching activities to encourage students' critical reading in EFL (Athanassiou, McNett, & Harvey, 2003). Some research on EFL reading has indicated that thought-provoking exercises based on Bloom's taxonomy can guide learners to develop reading skills (Khorsand, 2009).

Furthermore, a copious amount of studies have demonstrated that learning tendencies and behavior are reflected in different manners by gender (Tsai & Tsai, 2010). For instance, Brantmeier (2001) found that gender was a key concern associated with reading comprehension in a group of readers whose reading language was their second language. Moreover, Pae (2004) investigated the effect of gender on EFL reading comprehension, and the results showed females were in favor of Mood/Impression/Tone items whereas males preferred Logical Inference items. From the perspective of tech-supported learning, Terzis and Economides (2011) found that males focused on the usefulness of computer-based assessment whereas females focused on how easy or difficult it was. Thus, in light of the human factors affecting behavior, we propose that biofeedback will affect students' cognitive taxonomies and reading abilities in light of different personal cognitive hierarchic levels and genders.

## 3. Methodology

In this section, we explain the detailed procedures of this experiment to investigate if learners' reading performance and cognitive levels are affected by SR though biofeedback, and what roles learners' cognitive taxonomy and gender play. The biofeedback data were captured by eye-tracking and EEG devices. In addition, the scores from Bloom's taxonomy survey and reading tests were collected by means of a questionnaire and standardized test materials, respectively, and then analyzed with IMB SPSS 19.

### *3.1 Quasi-Experimental Design and Participants Selection*

A quasi-experimental design was carried out with an experimental group and a control group. Participants were recruited from a university in China, which has offered since 2004 an EFL self-regulated online learning program. The experimental design criteria were as follows. *First*, participants should strictly adhere to the arrangements made by their instructors in a specific set time and location. This requirement was set to limit environmental biases considering the complexity of factors that may affect the results of experiments involving physiological measurements. *Second*, comparisons should be as accurate and objective as possible. To improve consistency within groups, we recruited participants based on their knowledge background and learning experience to try to improve the between-group relative homogeneity. *Third*, since gender is an important human factor affecting SR and biofeedback, the gender ratio should be balanced between the two groups.

By following the selection criteria above, 106 participants majoring in economics at undergraduate level grade one were selected randomly from lists of original university cohorts. After a random selection and screening to the set criteria, random student allocation generated an experimental group with 54 students and a control group with 52 students. Compared with other majors in this university, the gender ratio in economics is relatively balanced, and students' general proficiency of EFL reading ranged from band three to band four of the College English Test (CET 4), which is a standardized test adopted by the Chinese education system. All participants had little or no familiarity or background on physiology tests.

### *3.2 Procedures and Instruments*

The current teaching experiment was conducted from September 2015 to Jan 2016, and provided students with the opportunity to engage in multi-level learning activities. The experiment consisted of five EFL classes each week, of which two were 45 minute *reading and writing* (R&W) classes in a physical classroom, two were 45 minute *collaborative learning* (CL) classes in an interactive classroom, and one was 1 hour *autonomous learning* (AL) classes in a language lab with a computer for each individual participant. In addition, the teaching experiments were fairly designed for both the control group and experimental group, because the learning conditions were exactly the same for both groups; that is, the same material, contents and instructors were used. Biofeedback was administered only for the experimental group.

The above arrangement ensured that almost all the teaching and learning processes were given utmost control, so as to minimize sources of bias and to enable us to identify whether SR is a valid tool for enhancing learners' performance. The current experiment was performed for nearly a semester, and figure 1 shows that the learning experiments consisted of three stages with the different instruments.

### *3.2.1 Homogeneity Test*

In the first stage, both Bloom's taxonomy scale and standard EFL reading material were used to test the homogeneity of the subjects. To ensure the quality of the pre-post test, both a cognitive hierarchy questionnaire and a reading comprehension test were conducted during the R&W class under the teachers' supervision. Furthermore, to minimize psychological interference on the participants, this experiment employed a double-blind approach.

A rating-scale questionnaire was developed from Nicholas' study (Athanassiou et al., 2003) and consisted of 6 items (shown in Appendix 1), which correspond to Bloom's six cognitive hierarchies. Achievement was coded 1 to 6 respectively to represent knowledge (1), comprehension (2), application (3), analysis (4), synthesis (5), and evaluation (6). Two experienced EFL teachers were involved in translating the instrument into Chinese for its reliability and validity with a Chinese audience. Participants were required to select the items they supposed they had achieved according to their perception of their current cognitive level. Then, an average score was computed for each of their submissions. Learners' perceived cognitive taxonomy was tested using a developed self-reported questionnaire, which explained to both groups in detail their learning activities. A series of individual assignments and group discussion were conducted to ensure students' comprehension and the measurement of the levels with which they engaged in the EFL context.

·*Knowledge:* This cognitive category simply focuses on recalling learned concepts.

In the current EFL learning context, *knowledge* represents remembering concepts, such as words, collocations, and grammatical knowledge. These are the primary learning activities for EFL learning.

·*Comprehension:* This is the cognitive category that highlights the capability to seize

the inner meaning of a text through comparison and contrast. In the current study, *comprehension* was presented to EFL learners as the understanding of the text based on comparison and contrast of the material given as text and the extrapolation of meaning rather than simple information recall.

·*Application:* This concept in the current EFL learning context was explained to

students as their ability to recall what they learned before, such as a related linguistic approaches or logical features, and apply them to the current text analysis, which led them to develop a close relationship between current materials and previously learned principles and methods.

·*Analysis:* This was explained to learners as the ability to understand both the content and structural forms of the material given to them.

·*Synthesis:* This was presented as the cognitive level at which students can pull together different ideas and creative thinking as an output generated from original materials.

·*Evaluation:* This last element was explained to learners as the ability to appraise the value of some material. In addition to making sense of the concept, data and theory, learners can build conscious value judgments derived from their existing schemata to solve the problem with which they are confronted.

Finally, the reading materials that were adopted from CET 4 consisted of four short articles (250-280 words) with 5 single choice items per article, and 40 minutes in total was allotted for students to take the pretest. To estimate whether the two groups were homogeneous in terms of their cognitive level and language proficiency, the Wilcoxon matched-pairs test of analysis and an independent t-test were utilized.

### *3.2.2 Physiological Computing and SR*

In the second stage of the experiment, the instructor explained the harmlessness of the eye-trackers and EEG to the experimental group during their first meeting. Both the experimental and the control groups were given one hour of self-regulated online reading activities every week. Language labs were open to students from Monday to Friday, and the control group students could freely choose their study time at their convenience and perform their reading activities online. However, only three sets of biofeedback devices were utilized in the current research; the experimental group students were thus required to attend a pre-scheduled appointment in specialized labs (figure 2) under the guidance of a lab assistant.

The stimulated recall for learners lasted around 10 minutes after their 1-hour reading tasks. To improve the validity of the SR experiment, time delay between the reading task and the recall was minimized (Lyle, 2003; Schepens, Stapley, & Drew, 2008). The biofeedback data were presented to learners immediately after they finished with their reading tasks. The physiological data collected by eye trackers and portable EEG devices were utilized as feedback stimulating learners to recall their learning behaviors.

The optical data captured by the eye tracker involved fixation allocation, fixation duration and scanning path data, which are presented in the forms of both descriptive data and a heat map. These key clues could help learners recognize their reading behavior as follows.

·*Fixation allocation* is the point on which eyes focus. According to Rayner's (Rayner, 2009) suggestion based on prior research, the current fixation parameter should be set at 200 millisecond (ms). The fixation presented was offered for learners to examine their interested areas, which may help learner to engage in retrospection if they seized the key components during reading activities. Furthermore, the visualization of the fixation allocation reminds learners about the neglected reading areas, encouraging speculation on the text-related mental space. For example, although some logic connectors were emphasized in the R&W class to comprehend the structural style, learners may still neglect their important roles in their reading task, and the worse is they barely realize it. Fixation location may intuitively help them recall the key components for the analysis of the logic relationship amongst parts.

·*Fixation duration* is the total time spent on fixation. For the heat map, ranging gradually from red to blue, passing through yellow and green represents the fixation duration (from long to short time) on a specific location. Biofeedback helps learners to review rationally what proportion of time they assigned to a specific point on the text. Students are encouraged to make a comparison between their previous reading behavior and current thinking. Furthermore, the fixation duration may offer opportunities for students to remember teachers' instructions and switch on starting to follow them in subsequent reading. Although opportunities to encourage exist, comments from peers and instructors may challenge them in their behavior. For example, although readers should pay more attention to predictive verbs than non-finite verbs in general, within an EFL context, the concentration on non-finite verbs may stimulate learners to ratiocinate on authors' ideas beyond words, which is of great significance to enhance learners' creative thinking.

·*Scanning path,* which is a valid approach to identify the patterns of fixation (Just & Carpenter, 1980), show students' logical sequence during the reading activities. Learners' cognitive levels were inspired through the recalling of their previous mental logic and the processing of psychological conflicts. For example, in an EFL context, it is difficult for Chinese native speakers to deal with the transformation of intertextuality. Since English is a language characterized by hypotaxis, the achievement of textual coherence is dependent upon several contextual themes, and meaning can be identified through a logic of coordination and subordination of the words into sentences; in contrast, Chinese is a paratactic language, and therefore the meaning is built on a logic constructed by sequential and non-subordinated ordering of words. Thus, the scanning path can clearly show learners' processing, be it hypotaxic or paratactic, and may stimulate learners to understand the structural style and pragmatic characteristics of the language they read.

In addition, the portable EEG detector named Neurosky could collect four original wavebands on a real time basis: Alpha (α), Beta (β), Theta (θ) and Delta (δ). The supporting software Minxp was used to analyze the original brain wave data and generate the mind-wave report containing both real-time information and the cumulative state of the users throughout the process. Specifically, the four-page reports contained three sections: the first section in the first page present the demographic information of the participants inputted before the experiment; the second section in the second and third pages consists of line/pie/bar charts recording the detailed information about learners' instant EEG parameters, and learners' attention and relaxation via procedural evaluation of visual signals was reported on the basis of the centesimal system; the last section in the forth generated verbal report in a conclusive assessment, evaluating learners' conclusive learning status during the online autonomous learning period with suggestions in terms of attention and relaxation. Learners are therefore guided towards a comprehensive understanding of their attention and fixation features, and through questions, they are prompted to recall, for example, why they felt relaxed or concentrated on a particular section of the text.

To investigate learners' reactions to different EEG recorded events, we used eye trackers to collect users' optical data during their SR stage and when reading their EEG reports. Their optical data corresponded to their EEG reports during the reading phase. Due to the experimental settings and application of portable devices, all students reported that they did not feel distracted by the eye tracker or EEG devices after the SR.

#### *3.2.3 Post-test*

Both the experimental and the control groups were required to complete a reading comprehension test and Blooms' cognitive hierarchy questionnaire in the 12th week of the class. To increase the validity of the pre-post tests between the two groups, a reading comprehension quiz was implemented similarly to the pretest in terms of the format (four passages with 5 items each), test time (40 minutes) and exam level (CET band 4); additionally, participants were required to fill the same cognitive taxonomy questionnaire.

### *3.3 Data Collection and Analysis*

Both qualitative and quantitative research methods were employed in the current research. To compare participants' cognitive taxonomy, a Wilcoxon matched-pairs test analysis was used to examine the rating-scale questionnaire data. Then frequency analysis was applied to distinguish groups with relative lower cognition from the higher group in the experimental class. Regarding learning performance, paired sample *t*-test were conducted to measure the score changes between the pre-post tests scores. To find the statistical differences of the reading scores of the experimental group and control group, some independent *t*-tests were employed in both the pretest and post-test measurement phases.

Furthermore, to investigate the effect of the two human factors in terms of gender and cognitive levels on learning performance, independent *t*-tests were performed to compare statistical differences in scoring. To explore the effect of learners' gender differences on EEG reports, data on the fixation duration and fixation allocation collected by the eye tracker was compared with *t*-test results as well. Finally, to explain and analyze the statistical results, 12 participants (experimental group and control group each half with equal ratio of gender) were randomly selected to have a face-to-face interview with the instructors.

## 4. Results

### *4.1 Effects of Biofeedback*

To answer the first research question, this research uses the Wilcoxon test and *t*-test to compare the pre-post scores, which indicate the effects of biofeedback on learners' cognitive taxonomy and reading abilities. As shown in Table 2a, no significant difference existed between the experimental group and control group in terms of cognitive taxonomy ($z$ = -0.17, $p$ = 0.87) and reading abilities ($t$ =-0.36, $p$ = 0.72) in the pretest, which verified the homogeneity of the two groups. However, the results from the pre-post tests showed that the experimental group had significant improvements in cognitive taxonomy ($z$ = -4.35, $p$ < 0.001) and reading scores ($t$ =-2.47, $p$ = 0.017), whereas no significant distinction existed in the control group for cognitive taxonomy ($z$ = -0.44, $p$ = 0.66) and reading ($t$ = 1.38, $p$ = 0.17). However, all six of the interviewees in the experimental group reported a special interest in the biofeedback technique. Synthesizing those data leads to the result that students adopting biofeedback as a stimulus demonstrated significant improvement in the dimensions of the cognitive level and reading comprehension, compared to those who studied in the traditional self-regulated online settings without SR.

### *4.2 Cognitive Hierarchy and SR*

With regards to the results of the cognitive hierarchy in the experimental group and the effects of SR, Table 1 shows that in the experimental group, 52 (96.3%) students asserted their taxonomy in *knowledge*, followed by *comprehension* (28 students, 51.85%), *application* (7 students, 12.96%), *analysis* (7 students, 12.96%), *synthesis* (4 students, 7.41%), and *evaluation* (none). The proportions indicate that the tendency was towards lower-level cognitive taxonomies on the whole. According to Table 1, 25 participants who selected *knowledge* scored 0.17, 17 participants scored 0.5, 1 participant scored 0.67, 1 participant scored 0.83, 4 participants scored 4, 2 participants scored 1.17, 2 participants scored 2, and 2 participants scored 2.5. For the frequency analysis based on multiple responses, nearly half of the participating students (46.3%) recognized only *knowledge* as their cognitive status, and students with such status were defined as the "single level-one" group, whereas the rest of students with at least two taxonomies were labelled as the "multi-levels" group.

To understand the role of cognitive hierarchy and learning performance in the context of biofeedback, a series of Wilcoxon nonparametric tests and independent samples $t$-tests were conducted. As shown in Tables 2b and 2c, surprisingly, the single-level students in the experimental group had satisfactory significant enhancement of taxonomy ($z$ = -4.36***, $p$ < 0.001) and reading capacity ($t$ = -5.29***, $p$ < 0.001) according to the pre-post scores, whereas the reading scorings from single-level students in the control group significantly decreased ($t$ = 2.51*, $p$ = 0.02). However, beyond what we expected, the students in high-cog group of both groups had no significant improvement in cognitive taxonomy ($z$ = -1.42, $p$ = 0.16; $z$ = -0.24, $p$ = 0.81) or reading capacity ($t$ = -0.96, $p$ = 0.34; $t$ = -0.22, $p$ = 0.81).

### *4.3 Gender and SR*

To answer how the gender differences relate to the effect of biofeedback on students' self-regulated learning performance, a family of Wilcoxon matched-pair tests and paired sample t-tests were implemented to explore the differences in the pre-post scores of male/female groups. Table 2d shows that the ratio of gender in the experimental class is coincidentally 50:50 (27 males, 27 females), and in the pretest, there was no significant distinction between males and females in terms of taxonomy ($z$ = -0.43, $p$ = 0.66) and reading abilities ($t$ = 0.59, $p$ = 0.56). Unexpectedly, although many previous studies verified that gender difference was a significant variable predicting learners' reading behavior, the results of the current study show no significant relationship between gender and SR.

However, according to the comparison analysis of the optical data on EEG reports, we surprisingly found that there is a significant gender difference regarding learners' average fixation duration and average fixation count. As shown in Table 3, the males' average fixation duration on page 2 of the reports is longer than the females' ($t$ = 2.4, $p$ = 0.02). In addition, males had higher fixation counts than females ($t$ = 3.84, $p$ < 0.001). Similarly, these significant distinctions were also found on page 3 of the reports ($t$ = 3.1, $p$ = 0.03; $t$ = 4.1, $p$ < 0.001 respectively). In contrast, the eye movements on page 4 were totally reversed such that females had longer fixation durations ($t$ = -2.91, $p$ = 0.005) and higher fixation counts ($t$ = -4.72, $p$ < 0.001) than males. The average fixation duration and average fixation count in paper one indicate that insignificant gender difference was found in terms of attention of demographic information ($t$ = -0.59, $p$ = 0.56; $t$ = -1.91, $p$ = 0.06 respectively). The surprising findings prompted the researchers in this study to carefully analyze all report pages for gender differences, and two screenshots in Figure 3 exported from the eye-tracker system display the heat map for two interviewees (one male and one female). According to the heat map. The, forth pages 1 and 4, with a conclusive evaluation of readers' EEG, attracted much more attention from females, whereas pages 2 and 3, presenting various charts filled with detailed and procedural information, highlighted males' preferential style.

## 5. Discussion

The pre-post scores, in terms of cognitive taxonomies and reading abilities, were compared, and the results showed that there was no significant improvement in the post-test outcome for the control group. It is suggested that the current efforts on traditional online autonomous learning are not working. One possible explanation is that metacognition and critical thinking have a significant positive relationship with SRL achievement (Broadbent & Poon, 2015); however, the traditional SRL model sometimes is so flexible that learners may tend to select materials in which they are interested while ignoring their weaknesses and shop boards through reliable feedback. Additionally, from the perspective of the subject, foreign learning environments fail to provide learners with sufficient input, output, or interaction opportunities, and a high level of language achievement is difficult to obtain without the effective regulation of learning behavior and the context of learning outside the classroom (Kormos & Csizer, 2014).

However, learners' using physiology signals to recall their experience in retrospection did facilitate their reading performance and cognitive hierarchy. The findings were consistent with some psychologists' suggestion that biofeedback can be used as an effective method to treat some psychology issues, such as attention deficit disorder (Linden et al., 1996). One possible reason is the following aspects:

First, eye movement data leave students much mental space for re-examine their areas of interest and neglected areas, which may transform their rigid thinking mode into an open and speculative style. Furthermore, the visual scanning path may stimulate them to compare and contrast their prior learning behavior and current retrospection, which helps to enhance their cognitive level and learning performance. Second, students may come across many emotional issues, such as a sense of anxiety and helplessness, in the process of their autonomous learning activities. Biofeedback may act as a metacognitive method to help learners realize their personalized learning habits and cognitive modes and to provide retrospection on their personal learning strengths and weaknesses. This benefits their cognitive structure and learning habits when studying in a self-directed mode. Third, as mentioned in section 4.1, the students in the experimental group presented their special interests in their biofeedback information as evidence of learning references. That may be attributed to the fact that learners showed robust belief in their physiology signals, since the signals were very personalized and unique to their own learning status. They were thus willing to adjust their autonomous learning to their personal traits.

Furthermore, the pre-post results showed that the multi-level students had higher reading comprehension mean scores than that of the single level students' scores in both groups, which supports that high-order cognitive skills are usually associated with better performance (Goradia & Bugarcic, 2017) However, the results showed that the students in the "single level-one" group showed significant enhancement, whereas the "multiple-levels" group students showed no significant improvement. One possible explanation is that biofeedback, such as attention, relaxation, and fixation, are superficial and basic information. These have a close relationship with learning habits but rarely a relationship with deep cognitive behavior. Therefore, physiological information may be helpful for ameliorating some superficial and inappropriate learning behavior, whereas it is difficult to help "multiple-levels" learners improve to a higher taxonomy. In addition, compared with the multi-level students with top-tier language proficiency, single-level students who performed comparatively poor had a larger possibility of making significant improvements.

According to the pre-post test results in Table 2d, there is no significant distinction for reading and cognitive scores between males and females through biofeedback, which indicated that stimulated recall with physiological signals is a suitable learning instrument for both males and females in terms of learning outcomes and cognitive level. However, regarding the examination of eye movement data on EEG reports, researchers surprisingly found that females were more in favor of conclusive evaluation, while males tend to prefer procedure assessments. Pae's (2004) research about the effects of gender on EFL reading comprehension supported the finding that males were more likely to favor logical inference than females. Another study by Terzis and Economides (2011) demonstrated that females were more likely to emphasize the ease of use, whereas males focused on usefulness in the context of computing-based assessment. Therefore, it is suggested that males tend to care more about useful information through various data charts, whereas a conclusive assessment would be accepted by females to guide their learning strategy directly.

## 6. Conclusion and limitation

This research provides, through empirical evidence, a variety of insights into the domain of autonomous learning. With the explosive development of Artificial

Intelligence, constructivism needs to be highly considered in the context of information technology. Previous studies have addressed the significance of human-computer interaction to constructivism (Al-Huneidi & Schreurs 2011; Reidsma, Nijholt, Tschacher, & Ramseyer 2010), and this study empirically tested whether biofeedback could be used as a variable in the interaction between human and computers, through which constructists may be provided a special perspective when considering the construction of a student-centered learning context.

From the perspective of learners, with the rapid expansion of the online learning model, it is important to understand how to enhance effective learning performance in an autonomous learning setting. Considering that computers can read learners' mental mechanisms via biofeedback, machines are more desirable for learners who want to learn independently without a human teacher but who do not want to miss on the useful learning feedback traditionally communicated by teachers.

We suggest that SR through biofeedback be utilized to help learners not only engage in retrospection about their learning behavior to improve their learning performance by refining their study skills but also embark in the often daunting journey of independent learning. Furthermore, as far as instructors are concerned, learners' biofeedback is valuable data for adjusting their pedagogic design and improving teaching arrangements according to learners' emotional status. This could have great applications to enhancing the learning of students with mild learning and cognitive impairments. Thus, we see it possible for some firms operating in the education sector to develop tools based on Affective Computing and Physiology to be used within the traditional education system. Moreover, this study has implications for software developers, providing them with some ideas for the application of technology to integrate biofeedback-based pedagogy and information technology. These findings may enable, through purposefully built software, easily transferrable learner profiles based on the identification of learners' personal traits, such as cognitive levels, gender, and the need to pay attention, which can be used to offer an individually tailored education experience.

This study also bears some limitations. For instance, this study limited the biofeedback administration via eye trackers and EEG to participants recruited amongst university freshmen, making the study not generalizable to the overall student population but rather specific to that student cohort. Furthermore, the material used for the experiment was selected from EFL learning materials, and biofeedback and SR may differ with the exposure of learners to different subject areas, some of which are more suitable for artistic minds and some of which are more suitable for more scientific minds. Furthermore, the variety of physiological information of learners was limited because of the limited functionality of the technology used.

Finally, future research could try to replicate this study by differentiating some elements of our quasi-experimental design, for instance by increasing the sample size and by looking at different education stages and subjects, which may provide other insights into the effect of SR by biofeedback and a wider capture of physiological data.

## References:

Al-Huneidi, A., & Schreurs, J. (2011, September). Constructivism based blended learning in higher education. *In World Summit on Knowledge Society* (pp. 581-591). Springer Berlin Heidelberg.

Athanassiou, N., McNett, J. M., & Harvey, C. (2003). Critical thinking in the management classroom: Bloom's taxonomy as a learning tool. *Journal of Management Education, 27*(5), 533-555.

Bloom, B. S. (1953). Thought-processes in lectures and discussions. *The Journal of General Education, 7*(3), 160-169.

Brantmeier, C. (2001). Second Language Reading Research on Passage Content and Gender: Challenges for the Intermediate□Level Curriculum. *Foreign Language Annals, 34*(4), 325-333.

Broadbent, J., & Poon, W. (2015). Self-regulated learning strategies & academic achievement in online higher education learning environments: A systematic review. *The Internet and Higher Education, 27*, 1-13.

Clark, M., Ruthven, I., Holt, P. O. B., Song, D., & Watt, S. (2014). You have e-mail, what happens next? Tracking the eyes for genre. *Information Processing & Management, 50*(1), 175-198.

Devolder, A., van Braak, J., & Tondeur, J. (2012). Supporting self□regulated learning in computer□based learning environments: systematic review of effects of scaffolding in the domain of science education. *Journal of Computer Assisted Learning, 28*(6), 557-573.

Duo, S., & Song, L. X. (2012). An e-learning system based on affective computing. *Physics Procedia, 24*, 1893-1898.

Gass, S. M. (2001). Innovations in second language research methods. *Annual Review of Applied Linguistics, 21*, 221-232.

Goradia, T., & Bugarcic, A. (2017). A social cognitive view of self-regulated learning within online environment. *Advances in integrative medicine, 4*(1), 5-6.

Houben, G.-J. (2016). *Retrieval Practice and Study Planning in MOOCs: Exploring Classroom-Based Self-regulated Learning Strategies at Scale.* Paper presented at the Adaptive and Adaptable Learning: 11th European Conference on Technology Enhanced Learning, EC-TEL 2016, Lyon, France, September 13-16, 2016, Proceedings.

Huang, Y.-M., Hwang, J.-P., & Chen, S. Y. (2014). Matching/mismatching in web-based learning: a perspective based on cognitive styles and physiological factors. *Interactive Learning Environments*, 1-17.

Hung, I., Lin, L.-I., Fang, W.-C., & Chen, N.-S. (2014). Learning with the body: An embodiment-based learning strategy enhances performance of comprehending fundamental optics. *Interacting with Computers, 26*(4), 360-371.

Iovane, G., Salerno, S., Giordano, P., Ingenito, G., & Mangione, G. R. (2012). *A computational model for managing emotions and affections in emotional learning platforms and learning experience in emotional computing context.* Paper presented at the Complex, Intelligent and Software Intensive Systems (CISIS), 2012 Sixth International Conference on.

Jennett, P., & Affleck, L. (1998). Chart audit and chart stimulated recall as methods of needs assessment in continuing professional health education. *Journal of Continuing Education in the Health Professions, 18*(3), 163-171.

Jensen, J. W. (2000, April). *Application of Constructivism to Teacher Education.* Paper presented at the Annual meeting of American Educational Research Association, New Orleans, LA.

Just, M. A., & Carpenter, P. A. (1980). A theory of reading: from eye fixations to comprehension. *Psychological review, 87*(4), 329.

Kang, H. (2014). Understanding online reading through the eyes of first and second language readers: An exploratory study. *Computers & Education, 73*, 1-8.

Khorsand, N. (2009). Cognitive Levels of Questions Used by Iranian EFL Teachers in Advanced Reading Comprehension Tests. *Online Submission*.

King, L. H. (1980). *Student thought processes and the expectancy effect*: Centre for Research in Teaching, Faculty of Education, University of Alberta.

Kizilcec, R. F., Pérez-Sanagustín, M., & Maldonado, J. J. (2017). Self-regulated learning strategies predict learner behavior and goal attainment in Massive Open Online Courses. *Computers & Education, 104*, 18-33.

Kormos, J., & Csizer, K. (2014). The Interaction of Motivation, Self□Regulatory Strategies, and Autonomous Learning Behavior in Different Learner Groups. *Tesol Quarterly, 48*(2), 275-299.

Krathwohl, D. R., Bloom, B. S., & Masia, B. B. (1964). Taxonomy of educational objectives, handbook ii: affective domain. New York: David McKay Company. *Inc. ISBN 0-679-30210-7, 0-582-32385, 1*.

Lai, M.-L., Tsai, M.-J., Yang, F.-Y., Hsu, C.-Y., Liu, T.-C., Lee, S. W.-Y., Tsai, C.-C. (2013). A review of using eye-tracking technology in exploring learning from 2000 to 2012. *Educational Research Review, 10*, 90-115.

Li, N., & Kirkup, G. (2007). Gender and cultural differences in Internet use: A study of China and the UK. *Computers & Education, 48*(2), 301-317.

Linden, M., Habib, T., & Radojevic, V. (1996). A controlled study of the effects of EEG biofeedback on cognition and behavior of children with attention deficit disorder and learning disabilities. *Biofeedback and self-regulation, 21*(1), 35-49.

Lindgren, E. (2002). The Effect of Stimulate Recall on 14-Year-Olds' L1 Swedish and EFL Writing and Revision. *Language teaching research, 6*(3), 267-268.

Lu, H. P., & Chiou, M. J. (2010). The impact of individual differences on e□learning system satisfaction: A contingency approach. *British Journal of Educational Technology, 41*(2), 307-323.

Lyle, J. (2003). Stimulated recall: A report on its use in naturalistic research. *British educational research journal, 29*(6), 861-878.

Meade, P., & McMeniman, M. (1992). Stimulated recall—an effective methodology for examining successful teaching in science. *The Australian Educational Researcher, 19*(3), 1-18.

Meier, A. M., & Vogt, F. (2015). The potential of stimulated recall for investigating self-regulation processes in inquiry learning with primary school students. *Perspectives in Science, 5*, 45-53.

Mey, G., & Mruck, K. (2010). *Handbuch qualitative Forschung in der Psychologie*: Springer.

Pae, T.-I. (2004). Gender effect on reading comprehension with Korean EFL learners. *System, 32*(2), 265-281.

Panadero, E., Kirschner, P. A., Järvelä, S., Malmberg, J., & Järvenoja, H. (2015). How individual self-regulation affects group regulation and performance: A shared regulation intervention. *Small Group Research, 46*(4), 431-454.

Pintrich, P. R. (2000). The role of goal orientation in self-regulated learning *Handbook of self-regulation* (pp. 451-502): Elsevier.

Peterson, P. L., & Clark, C. M. (1978). Teachers' reports of their cognitive processes during teaching. *American educational research journal, 15*(4), 555-565.

Picard, R. W., & Picard, R. (1997). *Affective computing* (Vol. 252): MIT press Cambridge.

Rayner, K. (2009). Eye movements and attention in reading, scene perception, and visual search. *The quarterly journal of experimental psychology, 62*(8), 1457-1506.

Reidsma, D., Nijholt, A., Tschacher, W., & Ramseyer, F. (2010, October). Measuring multimodal synchrony for human-computer interaction. In *Cyberworlds (CW), 2010 International Conference on* (pp. 67-71). IEEE.

Sano, A., & Picard, R. W. (2013). *Stress recognition using wearable sensors and mobile phones.* Paper presented at the Affective Computing and Intelligent Interaction (ACII), 2013 Humaine

Association Conference on.

Schepens, B., Stapley, P., & Drew, T. (2008). Neurons in the pontomedullary reticular formation signal posture and movement both as an integrated behavior and independently. *Journal of neurophysiology, 100*(4), 2235-2253.

Selinker, L., & Gass, S. M. (2008). *Second language acquisition: An introductory course*: Taylor & Francis.

Surjosuseno, T. T., & Watts, V. (1999). Using Bloom's Taxonomy to teach critical reading in English as a foreign language classes. *Queensland Journal of Educational Research, 15*(2), 227-244.

Terzis, V., & Economides, A. A. (2011). The acceptance and use of computer based assessment. *Computers & Education, 56*(4), 1032-1044.

Tjeerdsma, B. L. (1997). Comparison of teacher and student perspectives of tasks and feedback. *Journal of teaching in physical education, 16*, 388-400.

Tsai, M.-J., & Tsai, C.-C. (2010). Junior high school students' Internet usage and self-efficacy: A re-examination of the gender gap. *Computers & Education, 54*(4), 1182-1192.

Wang, W., Liang, Z., Blazeck, A., & Greene, B. (2015). Improving Chinese nursing students' communication skills by utilizing video-stimulated recall and role-play case scenarios to introduce them to the SBAR technique. *Nurse education today, 35*(7), 881-887.

Wang, X.-W., Nie, D., & Lu, B.-L. (2014). Emotional state classification from EEG data using machine learning approach. *Neurocomputing, 129*, 94-106.

Wilcox, S., & Trudel, P. (1998). Constructing the coaching principles and beliefs of a youth ice hockey coach. *AVANTE-ONTARIO-, 4*, 39-66.

Yamauchi, T., Xiao, K., Bowman, C., & Mueen, A. (2015, September). Dynamic time warping: A single dry electrode EEG study in a self-paced learning task. In *Affective Computing and Intelligent Interaction (ACII), 2015 International Conference on* (pp. 56-62). IEEE.

Zimmerman, B. J. (1989). A social cognitive view of self-regulated academic learning. *Journal of educational psychology, 81*(3), 329.

# Practitioner Notes

**What is already known about this topic:**

· The stimulated recall (SR) technique, considered being a valuable tool for learners to capitalize on introspection, have positive effect on learning outcomes and cognitive processes in physical context.

· The recorded audio and video are generally used as the stimulus in physical learning context.

· The stimulus source may differ from a research context to another.

· Learning performance has close relationship with different human factors, such as cognitive taxonomy and gender differences.

**What this paper adds:**

· Students adopting the biofeedback as stimulus demonstrated significant improvement in the dimensions of cognitive level and reading comprehension.

· Biofeedback, such as EEG and eye-movement, may be applicable stimulus for stimulated recall in online self-directed learning context.

· The lower-cognitive students have more significant enhancement of taxonomy and reading capacity, when conducting biofeedback-based stimulated recall.

· Eye-track reports showed that males favor procedural feedback and females have a preference for the conclusive assessment.

**Implications for practice and/or policy:**

·Biofeedback may act as a meta-cognitive method to help learners realize their personalized learning habits and cognitive modes, and encourage them to embark in the often daunting journey of autonomous learning.

·Biofeedback data could be used as valuable measurements for instructors to adjust their pedagogic design and improve teaching arrangements according to learners' emotional status and human factors.

·Procedural feedback is adaptive and should be considered utilizing for male students in self-access learning context; while conclusive assessment approach is more applicable for female learners.

# Tables

**Table 1.** The Cognitive Taxonomy Data in The Experimental Group

| students | *Pre-test* *Bloom's Level* | | | | | | *Pre-test* Student Score | *Post-test* *Bloom's Level* | | | | | | *Post-test* Student Score |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | *1* | *2* | *3* | *4* | *5* | *6* | | *1* | *2* | *3* | *4* | *5* | *6* | |
| 1 | 1 | 2 | | | | | 0.50 | 1 | 2 | | | | | 0.50 |
| 2 | 1 | | | | | | 0.17 | 1 | 2 | | | | | 0.50 |
| 3 | 1 | 2 | 3 | | | | 1.00 | 1 | 2 | 3 | | 5 | | 1.83 |
| 4 | 1 | 2 | | | | | 0.50 | 1 | | | | | | 0.17 |
| 5 | 1 | | | | | | 0.17 | 1 | 2 | 3 | 4 | | | 1.67 |
| 6 | 1 | 2 | | 4 | | | 1.17 | 1 | 2 | 3 | 4 | | | 1.67 |
| 7 | 1 | 2 | | | | | 0.50 | 1 | 2 | | | | | 0.50 |
| 8 | 1 | | | | | | 0.17 | 1 | 2 | | | | | 0.50 |
| 9 | 1 | | | | | | 0.17 | 1 | 2 | 3 | | | | 1.00 |
| 10 | 1 | 2 | 3 | | | | 1.00 | 1 | 2 | 3 | 4 | | | 1.67 |
| 11 | 1 | 2 | | | | | 0.50 | 1 | | | | | | 0.17 |
| 12 | 1 | | | | | | 0.17 | 1 | 2 | | | | | 0.50 |
| 13 | 1 | | | | | | 0.17 | 1 | 2 | 3 | | | | 1.00 |
| 14 | 1 | | | | | | 0.17 | 1 | 2 | | | | | 0.50 |
| 15 | 1 | 2 | 3 | | | | 1.00 | 1 | 2 | 3 | | | | 1.00 |
| 16 | 1 | 2 | | 4 | 5 | | 2.00 | 1 | 2 | 3 | 4 | 5 | | 2.50 |
| 17 | 1 | | | | | | 0.17 | 1 | 2 | | 4 | 5 | 6 | 3.00 |
| 18 | 1 | 2 | 3 | 4 | 5 | | 2.50 | 1 | 2 | 3 | 4 | | | 1.67 |
| 19 | 1 | | | | | | 0.17 | 1 | 2 | 3 | | | | 1.00 |
| 20 | 1 | | | | | | 0.17 | 1 | 2 | | | | | 0.50 |
| 21 | 1 | 2 | | | | | 0.50 | 1 | 2 | | | | | 0.50 |
| 22 | | 2 | 3 | | | | 0.83 | 1 | 2 | 3 | | | | 1.00 |
| 23 | 1 | 2 | | | | | 0.50 | | 2 | 3 | | 5 | | 1.67 |
| 24 | 1 | 2 | | | | | 0.50 | 1 | 2 | 3 | | | | 1.00 |
| 25 | 1 | 2 | | 4 | 5 | | 2.00 | 1 | 2 | 3 | | | | 1.00 |
| 26 | 1 | 2 | | | | | 0.50 | 1 | | 3 | | | | 0.67 |
| 27 | 1 | | 3 | | | | 0.67 | 1 | | 3 | | | | 0.67 |
| 28 | 1 | 2 | | | | | 0.50 | 1 | 2 | | | | | 0.50 |
| 29 | 1 | | | | | | 0.17 | 1 | 2 | | 4 | | | 1.17 |
| 30 | 1 | | | | | | 0.17 | 1 | 2 | | | | | 0.50 |
| 31 | 1 | | | | | | 0.17 | 1 | 2 | 3 | 4 | | | 1.67 |
| 32 | 1 | | | | | | 0.17 | 1 | 2 | | | | | 0.50 |
| 33 | 1 | 2 | | | | | 0.50 | 1 | 2 | | | | | 0.50 |
| 34 | 1 | 2 | | | | | 0.50 | 1 | 2 | | | | | 0.50 |
| 35 | | 2 | | 4 | | | 1.00 | 1 | 2 | 3 | 4 | | | 1.67 |
| 36 | 1 | 2 | | | | | 0.50 | 1 | 2 | | | | | 0.50 |
| 37 | 1 | | | | | | 0.17 | 1 | 2 | 3 | | | | 1.00 |
| 38 | 1 | | | | | | 0.17 | 1 | | | | | | 0.17 |
| 39 | 1 | 2 | | | | | 0.50 | 1 | | | | | | 0.17 |
| 40 | 1 | | | | | | 0.17 | 1 | 2 | | | | | 0.50 |
| 41 | 1 | | | | | | 0.17 | 1 | 2 | 3 | | | | 1.00 |
| 42 | 1 | | | | | | 0.17 | 1 | 2 | | | | | 0.50 |
| 43 | 1 | | | | | | 0.17 | 1 | 2 | | | | | 0.50 |
| 44 | 1 | 2 | | | | | 0.50 | 1 | 2 | 3 | | | | 1.00 |
| 45 | 1 | 2 | | | | | 0.50 | 1 | 2 | | | | | 0.50 |
| 46 | 1 | | | | | | 0.17 | 1 | 2 | | 4 | 5 | | 2.00 |
| 47 | 1 | | | | | | 0.17 | 1 | 2 | 3 | | | | 1.00 |
| 48 | 1 | 2 | 3 | 4 | 5 | | 2.50 | 1 | | | 4 | 5 | 6 | 2.67 |
| 49 | 1 | | | | | | 0.17 | 1 | 2 | 3 | | | | 1.00 |
| 50 | 1 | 2 | | | | | 0.50 | 1 | 2 | | | | | 0.50 |
| 51 | 1 | 2 | | 4 | | | 1.17 | 1 | 2 | 3 | | | | 1.00 |
| 52 | 1 | | | | | | 0.17 | 1 | 2 | | | | | 0.50 |
| 53 | 1 | | | | | | 0.17 | 1 | 2 | | | | | 0.50 |
| 54 | 1 | 2 | | | | | 0.50 | 1 | 2 | 3 | | | | 1.00 |
| total | 52 | 28 | 7 | 7 | 4 | 0 | | 53 | 47 | 25 | 11 | 6 | 2 | |

**Table 2.** The Comparisons of Learners' Taxonomies and Reading Capacity

| stages | Sample types | Wilcoxon Test of Taxonomy | | | *t*-test of Reading Capacity | | | |
|---|---|---|---|---|---|---|---|---|
| | | *n* | *z*-value | *p* | Mean | SD | *t*-value | *p* |
| ***a: The comparison of pre-post tests between experimental group and control group*** | | | | | | | | |
| pretests | Experimental group | 54 | -0.17 | 0.87 | 24.15 | 4.63 | -0.36 | 0.72 |
| | Control group | 52 | | | 24.46 | 4.31 | | |
| pretests | Experimental group | 54 | -4.35*** | <0.001 | 24.44 | 4.23 | -2.47* | 0.017 |
| post-tests | | | | | 25.93 | 3.98 | | |
| pretests | Control group | 52 | -0.44 | 0.66 | 24.46 | 4.31 | 1.38 | 0.17 |
| post-tests | | | | | 23.81 | 4.56 | | |
| ***b: The comparison of pre-post tests between the "single level" and the "multiple levels" in experimental group*** | | | | | | | | |
| pretests | single level students | 25 | -4.36*** | <0.001 | 21.28 | 3.55 | -5.29*** | <0.001 |
| post-tests | | | | | 24.56 | 3.68 | | |
| pretests | multiple levels students | 29 | -1.42 | 0.16 | 26.62 | 4.00 | -0.96 | 0.34 |
| post-test | | | | | 27.10 | 3.91 | | |
| ***c: The comparison of pre-post tests between the "single level" and the "multiple levels" in control group*** | | | | | | | | |
| pretests | single level students | 22 | -1.34 | 0.18 | 21.91 | 3.29 | 2.51* | 0.02 |
| post-tests | | | | | 20.18 | 2.54 | | |
| pretests | multiple levels students | 30 | -0.24 | 0.81 | 26.33 | 4.04 | -0.22 | 0.81 |
| post-test | | | | | 26.47 | 3.81 | | |
| ***d : The comparison of pre-post tests between male and female in experimental group*** | | | | | | | | |
| pretests | male | 27 | -0.43 | 0.66 | 24.52 | 3.87 | 0.59 | 0.56 |
| | female | 27 | | | 23.78 | 5.33 | | |
| post-test | male | 27 | -0.05 | 0.96 | 26.59 | 2.93 | 1.24 | 0.22 |
| | female | 27 | | | 25.26 | 4.78 | | |

**Table 3.** The Comparison Analysis of Eye-movements Between Genders

| Materials | Gender | *n* | Average Fixation Duration | | | | Average Fixation Count | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | | | **Mean** | **SD** | ***t*-value** | ***p*** | **Mean** | **SD** | **t-value** | ***p*** |
| Page 1 | Male | 27 | 498.26 | 50.38 | -0.59 | 0.56 | 39.59 | 8.68 | -1.91 | 0.06 |
| | Female | 27 | 510 | 90.92 | | | 43.89 | 7.82 | | |
| Page 2 | Male | 27 | 474.19 | 44.72 | 2.4* | 0.02 | 43.26 | 5.82 | 3.84*** | <0.001 |
| | Female | 27 | 441.26 | 55.60 | | | 37.81 | 4.48 | | |
| Page 3 | Male | 27 | 458.63 | 38.85 | 3.1* | 0.03 | 49.22 | 7.59 | 4.1*** | <0.001 |
| | Female | 27 | 420.37 | 51.05 | | | 41.44 | 6.3 | | |
| Page 4 | Male | 27 | 358.63 | 53.41 | -2.91** | 0.005 | 37.04 | 6.87 | -4.72*** | <0.001 |
| | Female | 27 | 407.11 | 68.26 | | | 45.44 | 6.22 | | |

**Appendix 1** The Blooms' cognitive hierarchy instrument on the reading comprehension

1. When conducting English reading, I focus on recalling learned concepts, such as vocabulary , collocations, and grammatical knowledge. (Knowledge)
2. My understanding of the text based on comparison and contrast of other materials, current events, etc. to extrapolate the meaning. (Comprehension)
3. I connect the ideas from the current text to other readings, class discussions, such as a related linguistic approach or logical features, and even my work or other experiences. (Application)
4. I identified the author's theories, assumptions, fallacies, and reconstruct the components and structure of the texts (Analysis)
5. I explore the reading material and use this exploration to build a new understanding and challenging of the material, or to formulate new ideas or solutions? (Synthesis)
6. I make use of course concepts,data, and theories rather than personal opinion as a criterion for evaluation of my study and work ? (Evaluation)

# figures

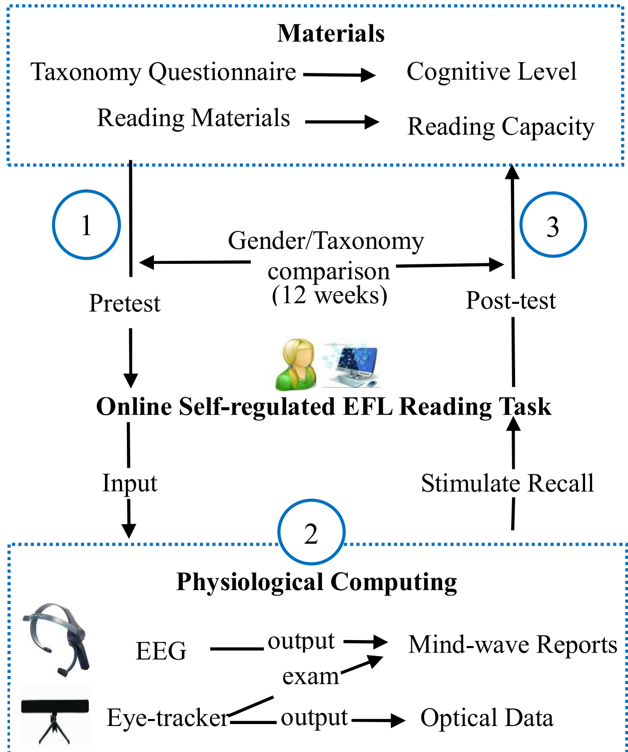


**Figure1.** The procedure of biofeedback in pre-post tests

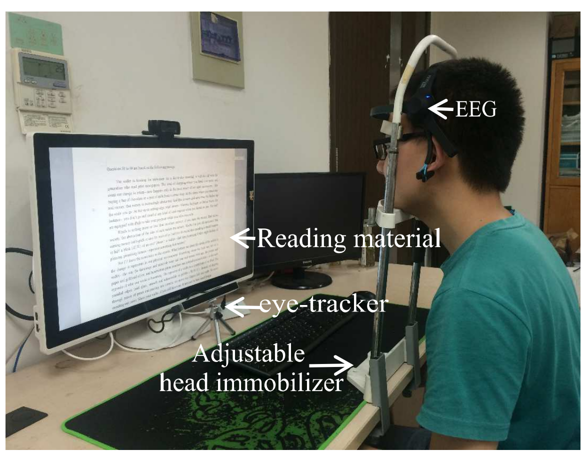


**Figure 2.** The sample test and physiological devices