

Eye Tracking Analysis of EAP Student's Regions of Interest in Computer-based Feedback on Grammar, Usage, Mechanics, Style and Organization and Development

Khaled El Ebyary
Damanhour University
Egypt

Scott Windeatt
Newcastle University
UK

Abstract

Much research has been conducted on the nature, role and effect of feedback. While it is generally claimed that feedback on L2 writing can improve learning, the role which feedback plays in the development of a learner's writing remains unclear. An attempt can be made to infer the effect of feedback from a student's subsequent written products, however the way in which a student engages with feedback messages and acts on them is hard to investigate, principally because this is generally an individual, private process, and so hard to externalise. The current study is concerned with using a range of methods for examining such student reactions to computer-based feedback, including eye-tracking data which provides a record of which errors and feedback messages the participants appear to read, in which order and for how long. The results suggest a marked tendency to focus on feedback on grammar, and on organisation and development. It is hypothesised that the former is due to participants' perception of the role of grammar in writing in their EFL exam-oriented contexts. The latter appears to be the result of the nature of the computer-based feedback they received.

Keywords: Writing, Automatic Writing Evaluation, Feedback, Eye-tracking

1. Introduction

Research on writing generally describes it as a productive skill involving various processes that occur before, during and after the act of writing itself. It is usually an individual activity, although it can involve other parties (tutor, peer, self or computer) in providing inspiration, encouragement, advice and feedback. It is generally claimed that the provision of feedback (also known as response and expert input) can improve learning (Anderson, 1982; Vygotsky, 1978) and such a claim is widely supported in relation to L2 writing (Ferris, 2006; Gibbs & Simpson, 2004; Hyland, 1998). Nevertheless, the precise role which feedback plays in the development of a learner's writing remains unclear, partly because the act of interpreting and acting upon feedback is usually carried out individually, and the only evidence of the effect of such

feedback is a student's subsequent written product. Evidence of the processes involved when students access, interpret and make use of such feedback is, therefore, difficult to obtain because they are difficult to externalise. This study is therefore concerned with exploring students' visual perception of feedback on their written work, i.e. identify regions of interest (RoI) in the feedback on language errors by tracking the student's eye fixation time on each of these areas. A pre-study questionnaire was used to examine participants' attitudes towards writing and feedback, including any previous experience with receiving feedback from a computer. Using the Criterion Online Writing Evaluation Service, which was developed and validated by ETS to provide both annotated diagnostic feedback on 5 language areas (classified by ETS) and holistic scoring based on level-specific models built from essays pre-scored by ETS-trained readers (see <http://www.ets.org/criterion>), each week for 6 weeks participants logged in to Criterion and wrote two drafts of an essay on a topic assigned by the classroom teacher. They received feedback from Criterion on the first draft, and then again on a second draft. Eye-tracking technology (see <http://mirametrix.com/>) was then used to identify which errors and feedback messages appeared to be the object of a student's visual perception and whether or not the pattern of perception in a first draft was repeated in the second draft. Eye-tracking provides information about the parts of the screen which a student's gaze rests on, and for how long, thereby allowing the researchers to identify which of the errors identified by Criterion, and which of the feedback comments, the students provide themselves with the opportunity to engage with. On its own, however, the eye-tracking data provides no information about what a student was actually doing or thinking when his/her gaze rested on a particular part of the screen. Follow up video stimulated recall sessions were therefore carried out with individual participants. As with all such methods, stimulated recall can only provide an indirect insight into the thought processes of the participants, and the accuracy of the information will be further affected by the fact that the data is gathered retrospectively. Whilst bearing these caveats in mind, the stimulated recall data can nevertheless provide a useful supplement to the information provided by the eye-tracking data.

2. Literature Review

Research on the effectiveness of feedback has focussed on questions such as whether feedback should be provided on *all* student errors (e.g.Lalande, 1982), *selectively* (e.g.Ferris, 1995) or whether *delayed* or even *no* error correction should be provided (e.g.Truscott, 1996). Other areas for research have included the role of self and/or peer feedback (e.g.Caulk, 1994; Connor, 1994; Kim, 2008; Taras, 2001), the impact of different feedback strategies (e.g. corrective vs. model answer, clarifying vs. directive, direct vs. indirect) on students' performance (e.g.Bitchener, Young, & Cameron, 2005; Ferris & Roberts, 2001; Huxham, 2007; Robb, 1986), the impact of feedback on teaching (e.g.Brinko, 1993; Cook-Sather, 2008) and automated feedback (Attali, 2004a; Attali, 2004b; Deane, Quinlan, & Kostin, 2011). Student reaction to feedback has generally been investigated in terms of issues such as perception of quality, effectiveness, value, or fairness (e.g.Cohen & Cavalcanti, 1990; Hedgcock & Lefkowitz, 1996; Leki, 1991; Lizzio & Wilson, 2008; Peterson & Irving, 2008), students' strategies for using feedback (e.g.Burke, 2009; Huxham, 2007; Orsmond, Merry, & Reiling, 2005; Walker, 2008), the impact of feedback on learning (e.g.Haigh, 2007; Lee, 2007; Miller, 2008; Torrance, 2007), and on students' subsequent submission of assessable work (e.g.Covic & Jones, 2008; Crisp, 2007).

Hyland and Hyland (2006), suggest that feedback offers the assistance of an expert, guiding the learner through Vygotsky's (1978) 'zone of proximal development' and providing opportunities for students to see an example of how others might respond to their work and to learn from these responses. In most feedback studies however, the need for students to take an active role in accepting, modifying or rejecting feedback Kulhavy (1977) has been emphasized. Winne and Butler (1994,p.5740), for example, claim that: '*feedback is information with which a learner can confirm, add to, overwrite, tune, or restructure information in memory, whether that information is domain knowledge, meta-cognitive knowledge, beliefs about self and tasks, or cognitive tactics and strategies*'. To encourage the active involvement of the student, a variety of strategies have been used to introduce a greater degree of interaction into the writing process and into the provision of feedback (e.g. feedback conferencing, video/audio recorded feedback). Such strategies are, however, not feasible in many EFL contexts where students generally get little feedback on their written work, primarily because of large student numbers.

Despite this considerable body of work, it is nevertheless claimed that feedback remains under-conceptualized and under-researched (Walker, 2009; Weaver, 2006). For example, as an act of communication, feedback may convey the intended message, but may also be misinterpreted, and we have limited access to information that would help us understand how feedback is processed and acted upon by students because:

- a) most student writing is done individually, usually outside the classroom, and generally does not involve another party (e.g. teacher or peer);
- b) most teacher feedback on students' written work is created outside the classroom and generally not in the presence of the student writers, and
- c) most student reading of the feedback, and editing or production of a second draft, is done by the students individually, and not in the presence of the teacher.

In an attempt to uncover students' reactions to feedback, a range of interactive qualitative research methods (e.g. stimulated recall) have been used (Caldwell & Leslie, 2010). Computers have been used to investigate writing strategies by logging keystrokes (e.g. Leijten & Waes, 2013), and by using eye-tracking technology (e.g. Hacker, Keener, & Hirscher, 2009). Eye-tracking technology has also been used to investigate student reactions to errors in other students' writing (e.g. Anson & Schwegler, 2012) but not to errors in their own writing.

The current study has the aim of investigating EFL students' visual perception of feedback presented on-screen, primarily by identifying how they appear to fixate on feedback messages. Criterion Online Writing Evaluation Service, which was developed and validated by ETS (see <http://www.ets.org/criterion>) was used to provide an holistic score and analytic feedback on 18 student essays (9 students, with two drafts for each participant). Eye-tracking equipment was then used to track the students' visual perception (used interchangeably here with regions of interest-RoI) of the feedback on the screen by calculating fixation time on errors and on related feedback for each of 5 language traits assessed by the Criterion. The study also examined whether or not such reading pattern preferences changed in second drafts on the same topic.

3. Research Questions

The main aim of the study is to identify students' visual perception of feedback on different kinds of error in their writing, as identified by Criterion, and what that can tell us about their reactions to such feedback.

The specific research questions we posed were;

1. What does the feedback provided by the computer tell us about the participants' writing product?
2. What does the video screen recording data generated by the eye tracking software tell us about participants' reactions to feedback?
3. What does the eye tracking data tell us about participants' visual perception of feedback on different language areas?
4. What does video stimulated recall data tell us about the participants' eye-tracking data?

4. Methodology

4.1. Participants and their Background

The participants of the current study (henceforth participants or students) were 9 international students enrolled on an English language programme at a UK University, in preparation for starting an undergraduate or postgraduate level degree. The programme aims to develop the students' general English language skills as well as knowledge and skills in areas such as academic speaking and listening, academic writing, note-taking, research and project skills, report writing, and oral presentations. The nine participants comprised four males and five females. Four were from China, two from Japan, one was Kuwaiti, one Korean and one Iraqi, and most were intending to study arts or social science courses.

Data from the pre-study questionnaires examined participants' computer literacy skills, previous experiences of computer-based feedback and attitudes towards writing in English. All but one expressed considerable confidence in their computer skills, as well as in their typing speed, and around half of them claimed to look at the screen most of the time when typing. Asked about writing processes (planning, editing and feedback), few claimed to pre-plan their writing, and, although most claimed to edit their work, some claimed they would edit their writing infrequently or not at all. Around half of them claimed to have used computers to practise writing in English, and most expressed either a positive or neutral attitude towards using computers for this purpose. Most students claimed it was important for them to correct their own work, but that the teacher had an

important role to play in motivating, modelling, facilitating and giving feedback on their writing. However, most were also positive about the idea of getting computer-based feedback on their written work, with the opportunity for immediate individual feedback as the most quoted reasons, although they expressed doubts about the nature of the feedback that could be provided, citing concerns that the computer could not “reason”, “has no mind, no judgement” and so could not correct the internal coherence of the writing, but only grammar or vocabulary.

Participants were initially asked to describe the writing courses they had previously attended in their home countries. They were also asked to describe the role that they understood their writing skills would play in their future and, finally, to rate their attitudes towards writing in L2 and towards writing courses in their home countries. Their main comments, however, related to limitations in these courses due to large student numbers and poor student-teacher ratios, with the result that students received very little feedback on their written work. All students emphasized that writing was a highly valued skill, but this was partly due to the highly examination-oriented contexts they came from. In such contexts, writing (as well as grammar and vocabulary) is a major component of any high-stakes English exam and their responses suggested that they saw improvements in their grammar and vocabulary as the key to improving their writing.

4.2. The Study Design

The students attended a writing class held in a computer cluster and used the Criterion Online Writing Evaluation Service. A training session involved registering participants on Criterion and a brief introduction on how the system works. The classes lasted 6 weeks and were integrated into their normal EAP course. The essay topics were chosen by their class teacher. Each participant wrote an essay on the computer, in Criterion, receive a holistic score out of 6 (see Criterion scoring guide on <https://www.ets.org/Media/Products/Criterion/topics/co-1s.htm>) and analytical feedback from Criterion, and then revised the essay and submitted the revised version to Criterion and received feedback again on a second draft. The analytical feedback involved five primary traits and associated language areas, which Criterion defines as Grammar (e.g. subject-verb agreement), Usage (e.g. confused words), Mechanics (e.g. spelling, punctuation errors), Style (e.g. repetition), and Organization and Development (e.g. thesis statement, transitional words and phrases).

These are Criterion's own definitions, and further information can be found on the Criterion website (<http://www.ets.org/criterion/>).

Criterion saves the essays students submit, as well as the scores, feedback, and details of the time the students spend using the programme. In addition, eye-tracking technology was used to provide video-captures of the screen showing not only the process of making changes to their essays, but also the areas of the screen where the students' gaze fell (visual perception or regions of interest – RoI hereafter), how they moved between those areas (saccades), and how long they spent looking at those areas (fixations). The process of scanning the information on the screen involves alternating rapid jumps (saccades) between locations on the screen, followed by stops of varying lengths (fixations). The start- and end-points of these jumps are determined by the gaze resting for more than a specified amount of time at a particular location (usually a minimum of 150 milliseconds). The location of start- and end-points, and the amount of time spent at those points, are recorded by the eye-tracking technology. When this recorded data is replayed in the eye-tracking software the saccades appear as a series of lines between the fixations, and these fixations appear as circles, whose size is determined by the amount of time the gaze remains at those locations. For researchers there may be certain areas on the screen that are especially relevant to the matter under investigation, hence the term regions of interest. In the case of the current study, the main RoIs correspond to those parts of the text which have, or have not, been given higher fixations by the participant. Evidence of this type is recorded by the eye tracking software. Video screen capture, also generated by the eye tracking software, shows additional information such as students' access to the Writer's Handbook, explanations of the marking criteria, and sample essays, each of which might contain RoIs. Using the eye-tracking technology it was therefore possible to collect real-time data about the RoIs students spent the most time gazing at, i.e. there were more fixations in RoIs containing errors in some of the five language areas that Criterion provides feedback on than in others. However, the reasons why more time might be spent on one RoI rather than another could only be inferred from other data, such as that provided by video stimulated recall, or perhaps from an analysis of the nature of the feedback provided.

The nine students agreed to use the eye-tracking equipment to have their eye-movement and video screen data recorded for both submissions of one of their essays. The eye tracking and screen video-capture data was collected in week two. In the first session of week two they were asked to write and submit the first draft of an essay on the topic 'Changing your

Hometown' to Criterion, and to read feedback on errors in grammar, usage, mechanics, style and development and organization. In session two, the participants edited their essays in Criterion based on the feedback received, re-submitted their essay and received feedback on the second draft.

4.3. Instruments, Data Collection and Analysis

Four sources of data were used in this study: a) a pre-study questionnaires, b) the essays and feedback recorded by Criterion, c) eye-tracking data and d) video stimulated recall.

The questionnaire was administered to all nine students at the start of the course prior to any eye tracking data collection. The aim was to gather demographic information and to identify participants' computer skills and predispositions towards L2 writing and feedback, and the use of computers.

Criterion provides automated holistic scores, as well as feedback at word, sentence, paragraph and text level, and other sources of help (such as planning templates, sample essays, a Writer's Handbook) are available to participants in the student portal (see <https://criterion.ets.org/>). The planning templates, essays, scores and feedback are saved automatically. The third source of data was provided by the eye tracking software and hardware. The hardware consists of a slim rectangle-shaped box that fits under a computer monitor and which bounces an infra-red beam off the pupils' of the subject's eyes onto the computer monitor. Once calibrated, the system can identify where on the screen the user's gaze is directed, and so what the reader appears to be reading. The system claims to be capable of collecting real time eye gaze data with a 0.5-1 degrees of accuracy and can cope with a limited range of head movements up and down, side to side, and backwards and forwards (25 Width cm x 11 Height cm x 30 Depth cm), using a 9-point calibration system, so obviating the need for a chinrest or portable eye tracking headset. The software which accompanies the system creates a video recording of all participants' on-screen activity, i.e. which part of a text is being read and, any text editing that takes place (see <http://mirametrix.com/> for detailed technical description). The recorded eye-gaze data is overlaid on an image of the on-screen text, allowing offline viewing and analysis using the Mirametrix Viewer software, which displays a scan-path showing saccades and fixations (i.e. the direction in which the eyes move, the places where they stop, and for how long). The eye-tracking

system also generates an XML file containing numerical saccade and fixation data.

The fourth source of data was the video stimulated recall, which took place after the eye-tracking session. The video screen capture sessions were replayed for each participant, who was asked to comment on what they were doing or thinking when reading and acting on the feedback. The screen capture showed the path that the student's gaze followed when the feedback was displayed, with the errors identified by Criterion. When a circle appeared on screen showing an area of text that the student seemed to be gazing at, if the student did not spontaneously comment, the video was paused and the student was asked to try and recall why his or her gaze had paused at that point, and what he or she was thinking at that stage. This provided some indication of the nature of the attention that was being paid to that RoI.

There was no evidence that the eye-tracking equipment, which was placed between the monitor and the keyboard, caused any problems or was a distraction for the students. As part of the pre-study questionnaire, students were asked about their keyboarding skills and whether they prefer to look at the keyboard or the screen when typing. This, along with other information (e.g. wearing glasses), was taken into consideration when calibrating the eye-tracking equipment. In practice, glancing at the keyboard, as some participants did, did not result in sufficient head movement to disturb the calibration.

5. Results

5.1. What does the feedback provided by the computer tell us about the participants' writing product?

The first research question was intended to identify what computerized-feedback can tell us about the participants' writing product. For this aspect of the study it was decided to compare first and second drafts in terms of *a) word count*, *b) total number of errors (in grammar, usage, mechanics and style)*, *c) total number of comments (on organization and development)* and *d) holistic scores*. The preliminary analysis of the participants' written work (see table 1) indicates that:

- a) word counts increased slightly in second drafts,*
- b) holistic scores for some second drafts improved, though most remained the same as for first drafts,*
- c) second drafts included fewer errors (see table 2).*

Eye Tracking Analysis of EAP Student's Regions of Interest in Computer-based Feedback on Grammar, Usage, Mechanics, Style and Organization and Development

Table 1. Preliminary analysis of the participants' written work

Student	Draft/No of words*	Criterion Score	No of Grammar Errors	No of Usage Errors	No of Mechanics Errors	No of Style Errors	No of Comments
Participant 1	1 st (260)	(4/6)	7	3	14	22	7
	2 nd (289)	(5/6)	10	3	10	0	8
Participant 2	1 st (200)	(4/6)	2	6	7	35	8
	2 nd (234)	(4/6)	1	6	1	36	8
Participant 3	1 st (440)	(5/6)	11	12	13	7	8
	2 nd (461)	(6/6)	6	12	13	0	8
Participant 4	1 st (226)	(4/6)	3	2	2	12	8
	2 nd (232)	(5/6)	1	3	1	8	7
Participant 5	1 st (265)	(5/6)	3	2	5	23	7
	2 nd (266)	(5/6)	0	0	1	23	7
Participant 6	1 st (276)	(5/6)	2	6	1	13	7
	2 nd (275)	(5/6)	0	0	0	12	7
Participant 7	2 nd (319)	(5/6)	5	8	16	36	7
	1 st (320)	(5/6)	0	9	1	38	7
Participant 8	1 st (275)	(5/6)	3	10	2	0	7
	2 nd (320)	(5/6)	1	2	2	0	7
Participant 9	1 st (301)	(5/6)	6	5	6	10	7
	2 nd (298)	(5/6)	2	1	2	10	7
TOTAL			63	90	97	285	195

*No of draft + no of words produced.
 ** Data here is based on data from the assignment used for the eye-tracking session.

Language Area	Grammar	Usage	Mechanics	Style
Total	695	791	595	3107

Aggregated totals for errors in each language area in all essay writing attempts for all 9 participants on all topics (see Table 2) show that style errors were the most frequent by a considerable margin, occurring around 4 times more frequently than errors of grammar, usage or mechanics.

5.2. What does the video screen recording data tell us about participants' reactions to feedback?

The writing portal used in this study (Criterion) allows the instructor to tailor the tasks by allowing or disallowing options, including the following:

- a) *choosing whether to use a Criterion template to plan an essay;*
- b) *choosing between 8 planning templates;*
- c) *viewing sample essays;*
- d) *viewing the scoring guide;*
- e) *consulting the Writer's Handbook;*
- f) *posting comments to the instructor.*

It was decided to include as many options as possible in order to monitor the extent to which students would choose to use them at any given stage of the writing process or while reading feedback, so that the reasons for their choices could be explored later in the verbal protocols. As Criterion does not record student use of most of these options, in this study eye-tracking and screen video-recording allowed access to much of this information.

Analysis of eye-tracking and screen-recording data, and data recorded by Criterion (see table 3) shows that for 5 out of the 9 participants the first draft generally took longer to produce than editing, revising and submitting a second draft, lending some support to findings in other studies (e.g. Silva, 1993). In the training session in the first class the students' attention was drawn to the eight planning templates provided by Criterion. Data from screen video-capture showed that six out of nine participants clicked the link to *view* the different planning templates before writing the first draft, and five participants actually *used* a planning template (see example in Figure 1). However, there were no attempts either to view or edit an already used planning template, or to use a totally different one, when participants wrote their second drafts.

Table 3. Analysis of Participants' Writing Processes

	Drafts for P1		Drafts for P2		Drafts for P3		Drafts for P 4		Drafts for P5		Drafts for P6		Drafts for P7		Drafts for P8		Drafts for P9	
	1 st	2 nd																
Time spent on writing & editing	53m	26m	33m	22m	46m	60m	37m	18m	41m	46m	26m	6m	62m	67m	56m	63m	70m	9m
Viewing planning templates	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No						
Using planning templates	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No								

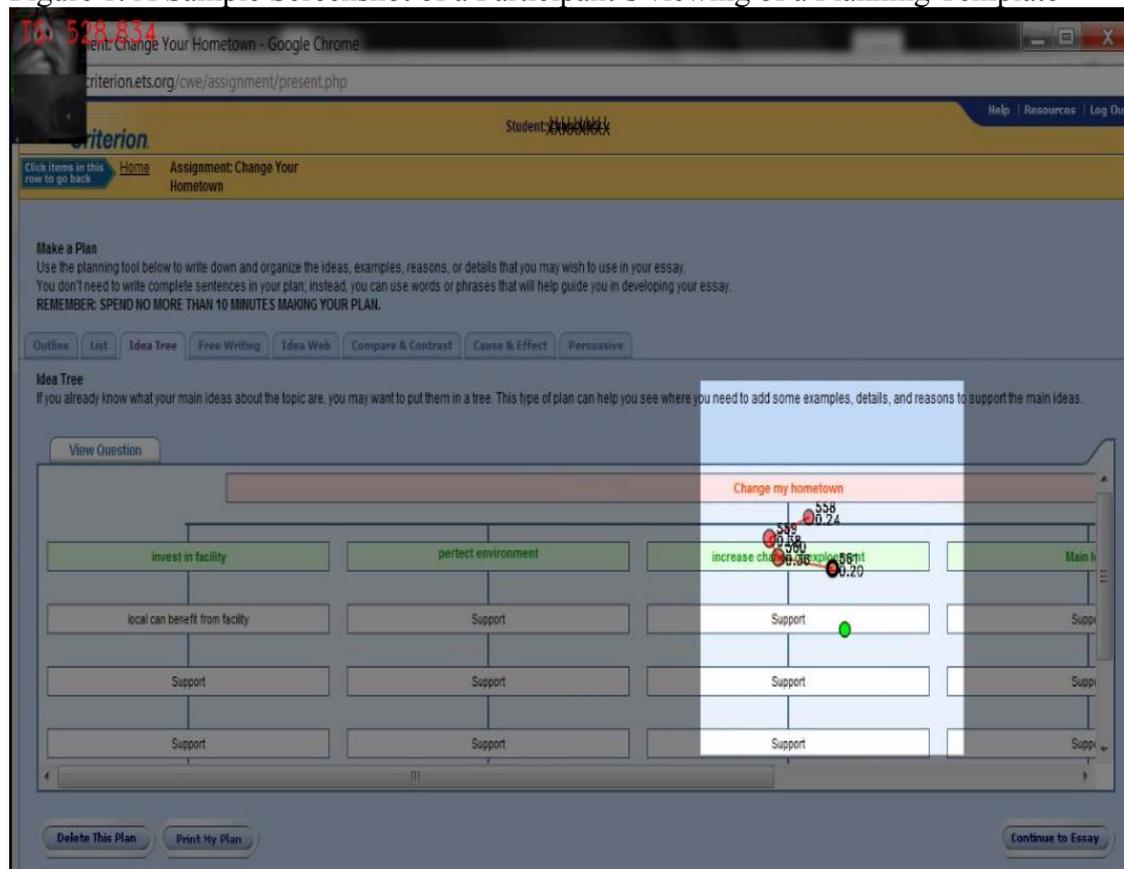
Eye Tracking Analysis of EAP Student's Regions of Interest in Computer-based Feedback on Grammar, Usage, Mechanics, Style and Organization and Development

Table 3. Analysis of Participants' Writing Processes

	Drafts for P1		Drafts for P2		Drafts for P3		Drafts for P 4		Drafts for P5		Drafts for P6		Drafts for P7		Drafts for P8		Drafts for P9	
	1 st	2 nd																
Viewing scoring guide	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes	No	Yes						
Viewing sample essays	No																	
Viewing Writer's Handbook	No	Yes	No	Yes	No	Yes												

*P= Participant

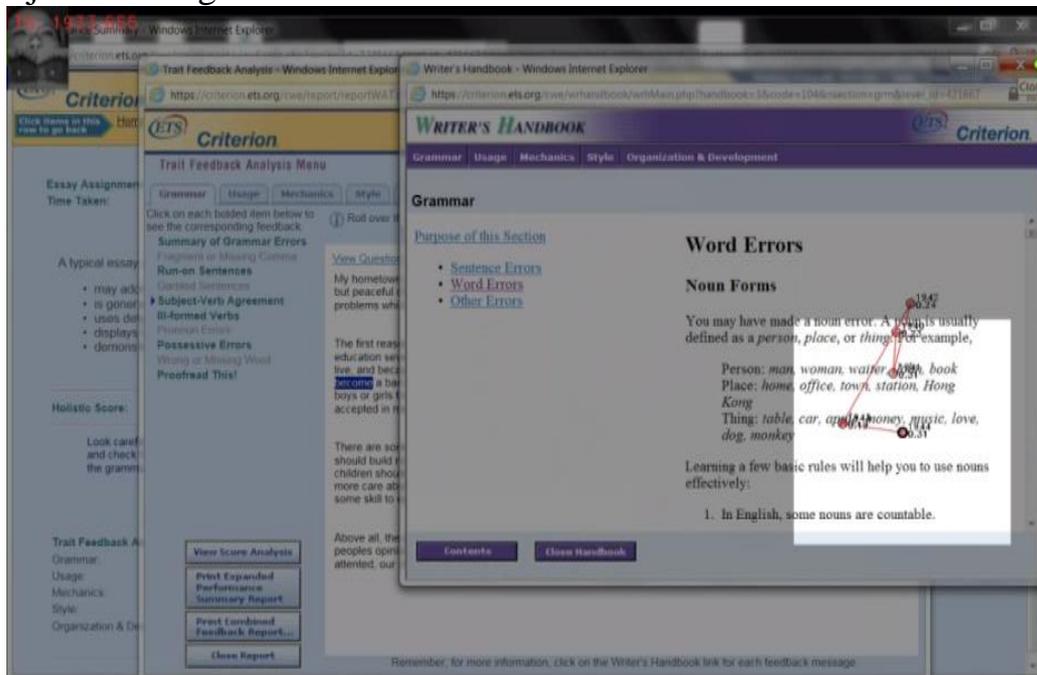
Figure 1. A Sample Screenshot of a Participant's viewing of a Planning Template



The guide setting out the criteria for the scores was available to the participants at all times, but none of them appear to have paid much attention to this information while writing, editing or reading feedback. Although screen video-capture data for some participants showed that they visited the scoring guide page, eye tracking data showed no fixations, i.e. their gaze did not rest on the text for long enough to indicate they were reading the guide. Six participants did spend time reading the criteria but only after receiving feedback on their second drafts. Participants whose scores showed no improvement in their second draft did not read the scoring guide.

In order to scaffold their writing, sample essays corresponding to each level of the marking scale were made available to participants. However, eye-tracking data confirmed that no participant read the sample essays before, during or after writing any of the drafts, or after receiving feedback. Similarly, only three participants consulted the online Writers' Handbook, which provided guidance on feedback (see example in figure 2 below). It was not clear why these participants chose to consult this resource while others did not.

Figure 2. Example of Viewing the Writer's Handbook in Relation to Subject-verb Agreement Errors on a 2nd draft



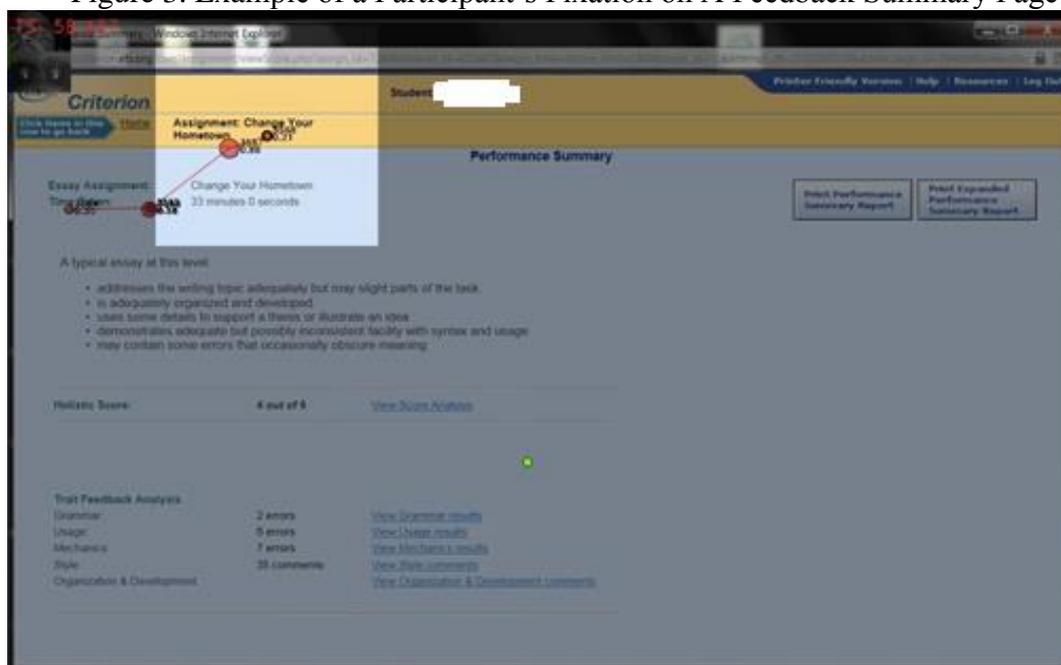
5.3. What does the eye-tracking data tell us about the participants' visual perception of feedback?

The eye-tracking data highlights the places the students' gaze fixes on and for how long, and so can show what feedback they may be reading. In answering this question, results are presented in the form of a) regions of interest (areas of immediate interest in the feedback) and b) comparisons between error type/number on the one hand and fixation time on the other.

5.3.1 Regions of Interest (ROI)

As students submit their written work, they receive instant feedback on a summary page (see example in figure 3 below), which provides a) a holistic score with a link to an explanation of the score and b) analytical feedback showing the number of errors in grammar, usage, mechanics, style and development and organization.

Figure 3. Example of a Participant's Fixation on A Feedback Summary Page



The first aim for the researchers at this stage was to examine students' RoI. This was examined by comparing fixations on the *holistic* and the *analytic feedback*, i.e. how long they gazed at each area on the screen, whether they looked first at the holistic score or at the analytic feedback, and, when they looked at the analytic feedback, which errors (i.e. *grammar, usage, mechanics, style and organization and development*) students fixated on first and for how long. As seen in table 4 below, the time spent gazing at the feedback summary page was generally higher in the first draft for all participants. The first fixations of almost all participants were on the holistic scores, and errors of grammar were viewed by all participants, before those of usage, style, mechanics or development and organization (though the latter might be due to the fact that grammar is first in the list of errors on the feedback summary page - see figure 3 above).

Table 4. Participants' RoI and Fixations on Feedback Summary Page

Participant	First Draft			Second Draft		
	Fix Time on Summary Page	1 st RoI on Summary Page	1 st area of interest	Fix Time on Summary Page	1 st RoI on Summary Page	1 st area of interest
P1	1:15	Hol Score	Grammar	Hol Score	00:42	Grammar
P2	1:22	Hol Score	Grammar	Hol Score	0:36	Grammar
P3	1:39	Missing	Grammar	Hol Score	1:22	Grammar

Eye Tracking Analysis of EAP Student's Regions of Interest in Computer-based Feedback on Grammar, Usage, Mechanics, Style and Organization and Development

P4	0:39	Hol Score	Grammar	Hol Score	0:30	Grammar
P5	0:33	Hol Score	Grammar	Hol Score	0:21	Grammar
P6	1:11	Hol Score	Grammar	Missing	0:29	Grammar
P7	0:39	Hol Score	Grammar	Hol Score	0:20	Grammar
P8	0:41	Hol Score	Grammar	Hol Score	0:22	Grammar
P9	1:12	Hol Score	Grammar	Hol Score	0:24	Grammar

P1 = Participant 1

Fix = Fixation

RoI = Regions of Interest

5.3.2 Error type/number and fixation time

Data showing fixation times for specific RoIs in the analytic feedback can show which types of error the students gazed at and so which types of error they may, for one reason or another, be most interested in. Fixation times in both first and second drafts were plotted against the number of errors in each category for 4 students (Participants 1, 2, 4 & 9). The selection was based on the four nationalities involved in the sample (i.e. Chinese, Kurdish, Japanese and Kuwaiti correspondingly), and the results reflect patterns which are common to all 9 participants. Table 5 presents a summary of the number of errors for grammar, usage, mechanics, style, organisation and development identified by Criterion for both drafts of one essay, for each of the four students.

Table 5. Participants' 1, 2, 4 & 9 Errors as Logged by Criterion

Participant and Draft	Grammar	Usage	Mechanics	Style	Organisation and Development
<i>Participant 1 - 1st</i>					
Number of errors	7	3	14	22	20
<i>Fixation Time (seconds)</i>	339	27	194	69	218
<i>Participant 1 - 2nd</i>					
Number of errors	10	3	10	0	30
<i>Fixation Time (seconds)</i>	156	16	87	0	140
<i>Participant 2 - 1st</i>					
Number of errors	2	6	7	35	16
<i>Fixation Time (seconds)</i>	78	9	57	15	72
<i>Participant 2 - 2nd</i>					
Number of errors	1	6	1	36	22
<i>Fixation Time (seconds)</i>	76	15	26	20	68
<i>Participant 4 - 1st</i>					
Number of errors	3	2	2	12	8
<i>Fixation Time (seconds)</i>	223	9	39	20	115
<i>Participant 4 - 2nd</i>					

Number of errors	1	3	1	8	8
<i>Fixation Time (seconds)</i>	<i>100</i>	<i>8</i>	<i>20</i>	<i>10</i>	<i>105</i>
Participant 9-1st					
Number of errors	6	5	6	10	33
<i>Fixation Time (seconds)</i>	<i>185</i>	<i>17</i>	<i>72</i>	<i>36</i>	<i>94</i>
Participant 9-2nd					
Number of errors	2	1	2	10	33
<i>Fixation Time (seconds)</i>	<i>142</i>	<i>11</i>	<i>42</i>	<i>29</i>	<i>118</i>

These results show that a) the number of errors identified by Criterion tends to be lower in most second drafts for most areas, suggesting that the analytic feedback has had some effect on the editing of individual items when preparing a second draft of their essay; b) all students spend more time focussing on feedback and areas of text identified as containing errors in the first draft than the second for almost all types of feedback; c) individual students spend quite different amounts of time on this task (Participant 1, for example, spends almost twice as much time as Participant 9 on almost the same number of grammar errors in the 1st draft); and d) participants generally spend more time on Grammar and Organisation and Development errors than on errors of usage, mechanics, and style.

Figure 4 presents an overview of the relationship between the number of errors in each of these 5 categories and the time each participant spends focussing on those errors. The graphs show clearly that a) the amount of time spent on different kinds of error was not linked to the number of errors in that category, b) this was particularly true of Grammar, and of Organisation and Development, where the amount of time was quite out of proportion to the number of errors, c) this pattern was repeated across all four participants and d) although the number of errors was generally lower in the second draft, and less time was spent in the relevant Regions of Interest, the same relationship between reading time and error-type was, with a few variations, repeated for the second draft. Data collected were cross-checked for clues as to the reason for these results.

5.4. What does the video stimulated recall data tell us about the participants' eye-tracking data?

Information provided by the students in their pre-treatment questionnaires and in their post-treatment video stimulated recall was analysed in order to help interpret the eye-tracking data.

5.4.1 Participant 1

In producing her first draft, Participant 1 spent 53 minutes writing 260 words and was given a score of 4 out of 6 by Criterion. Her score for the second submission increased to 5. Visual examination of screen video-capture while she was reading feedback on first draft errors showed that the student accessed the Writer's Handbook in both drafts to check guidance on Grammar, but not on the other types of error. Visual analysis also showed that she did not view the scoring guide, but did access sample essays written on the same topic while reading her feedback on first and second drafts. In the video stimulated recall this student claimed that language teachers in China did not focus much on Style and so she was not used to receiving much feedback on this language area. In contrast, Mechanics (mainly spelling) is seen as essential to good writing in China. However, despite practice in spotting misspelt words, checking her own electronic dictionary and then correcting the spelling, having corrected 14 misspelt words in her first draft, another 10 misspellings appeared in her second draft, raising questions as to the nature and efficacy of the uptake in this area. Furthermore, when asked about the reason she viewed sample essays and the Writer's Handbook, she explained that she was keen to obtain a better mark (holistic score), suggesting, perhaps, the influence of her experience in a highly examination-oriented educational system, and a focus on scores rather than learning.

Figure 5. A screen Shot of Participant 1's feedback on Organization and Development

Change Your Hometown Attempt 1

[Export](#)

Change Your Hometown

If you could change one important thing about your hometown, what would you change? Use reasons and specific examples to support your answer.

[Comments](#)

Organization & Development	Grammar (7)	Usage (3)	Mechanics (14)	Style (22)
Organization & Development				
<p style="font-size: small;">My hometown is a beautiful city,near sea and mountants.Individuals are very friendly and kind in my hometown ,it is the education. There are no pollution and less crimes happenS. Nonetheless, my hometown still has some problems which need to solve. This essay will discribe one issue in my hometown , it is the education. The first reason of education problem may is that in my hometown people did not pay any attention to childrens education several years ago, when child grew up, they went out of the city to find a job to afford themselves live, and becuse they thought go to school is not important, so they did not send their children to school ,either. It become a bad chain.On the other side, my hometown is always poor, most family have no money to let their children to study at school,sometimes they must study at home by themselves. There are some methods can be used to solve the education issue.Frist of all, the government of my hometown should build more school for child, and give them good teaching equipments and books. Then, children shoule to talk with the parents, to tell them the nessesary of education.The other thing should be done is take more care about childrens grown. Find out which subject is appropriate to children form poor family,who need some skill to earn money. Above all, the problem is not easy to be covered, but nothing is impossible when people really want to do,and only when education is attentioned, our children can have a wonderful future.</p>				

You have used only two main ideas to support your thesis statement. Your thesis statement will be clearer if you include at least one more main idea. Remember that you should introduce your main ideas with a topic sentence and give each main idea its own paragraph. Use examples, stories, and details to support and explain your main idea. Check the Writer's Handbook for ways to develop main ideas.

[Dialogue](#)

[Writer's Handbook](#)

5.4.2 Participant 2

Participant 2 spent 33 minutes in producing 200 words for his first draft, scoring 4 out of 6. While the absolute number of errors in the different categories was different from that of Participant 1, as was the fixation time for feedback on those errors, the fixation pattern was similar. Most time was spent reading feedback on Grammar and on Organisation and Development. As with Participant 1, the fixation time for feedback matched the number of errors most closely for usage, and least closely for Grammar and organisation and development, although the fixation pattern for mechanics and style was rather different from that for Participant 1 (and for Participants 4 and 9). Video stimulated recall data suggested the time spent on Organization and Development was mainly because that feedback was difficult to understand. Screen video-capture data showed that, unlike Participant 1, this student consulted the scoring guide when reading first draft errors, although not in the second draft. In the video stimulated recall, he explained his teachers had never shared marking criteria with him and that he was keen to understand them in order to improve his scores. Again unlike Participant 1, this student did not check any sample essays while reading feedback on first draft errors, but did check sample essays that scored 4, 5 and 6 while reading feedback on the second draft. Unlike Participant 1, he made little use of the writers' Handbook. During the stimulated recall, the student explained that he wanted to 'get 6 out of 6' and was eager to improve scores by editing the sentences that contained errors in the second draft in order to 'avoid'

being marked down for such errors as feedback in his hometown schooling was oral and generally geared towards avoidance strategies. What feedback was provided was minimal and focused on test taking strategies in which simple sentences meant fewer grammar mistakes and higher marks.

5.4.3 Participant 4

This participant spent 37 minutes producing a 226-word essay and as with Participants 1 and 2, his holistic score for his first draft was 4 out of 6. Fixation times for the different types of error showed a similar pattern to those for Participants 1, and for Participant 2's fixation times for Grammar and Organisation and Development. He scored 5 out of 6 for his second draft. Fixation patterns in reading feedback for the second draft showed a similar pattern to those for the first draft.

Screen video-capture data showed that, like Participant 2, Participant 4 viewed the scoring guide while reading feedback on the first draft, and one of the sample essays written on the same topic. While writing the second draft he chose to read sample essays with the same mark and another with a score of 6. Although he did not access the Writer's Handbook while reading the first draft, he did while reading the second draft. Stimulated recall data indicated that he read both sample essays and the handbook in the hope that would help him improve his mark.

5.4.4 Participant 9

Participant 9 spent 1 hour and 10 minutes writing his 301 word first draft which scored 5 out of 6. The pattern of fixations was similar to those for Participants 1 and 4, and to Participant 2's for Grammar and Organisation and Development, in both drafts, with Grammar and Organization and Development as his major RoI, followed by Mechanics. For his second draft, Participant 9 spent just 9 minutes editing or rewriting, and his holistic score remained the same. The number of errors for Organisation and Development was large - similar to the number for Participant 1 - and eye-tracking data showed long fixation times for the feedback. Nevertheless, he explained that he was not sure he had understood the feedback and that he spent time trying to interpret it as he was keen to improve his score.

6. Discussion

The clear pattern revealed by the eye-tracking data for long fixation times on Grammar errors and feedback, irrespective of the actual number of errors in that category, appears to be a reflection of the importance the students place on that aspect of writing. Data suggests that their priorities in dealing with errors are likely to be, at least in part, the result of

previous writing tuition. Participant 1, for example, explained during the stimulated recall that her short fixation times on Style errors were because teachers in her home country did not provide much feedback on Style.

Although most of the participants claimed that they generally edit their written work, their answers in the pre-study questionnaire did not provide enough detail to confirm whether they were referring to editing, or to proof-reading, i.e. dealing with surface level errors such as grammar, punctuation and spelling, rather than structure and organisation. The data on their fixation times in different Regions of Interest suggests that they appeared to pay similar amounts of attention to feedback on Grammar, a surface-level, proof-reading, issue, and to feedback on Organisation and Structure, an editing issue. However, they made few changes in response to feedback on Organisation and Development, and their long fixation times on that aspect of their writing seemed to be the result of difficulty in understanding or interpreting the feedback.

Most participants had been positive in their questionnaire responses about the idea of getting computer-based feedback on their written work. However, the extent to which they took advantage of the various forms of scaffolding provided by Criterion was limited. They claimed that they did not generally pre-plan their essay, and this was the case with the four participants for whom detailed eye-tracking data is presented here, with the exception of Participant 1, who reported a preference for writing with a planned outline or spidergram. This apparent lack of pre-planning was confirmed by the failure of any of the four participants to use the Criterion planning templates. Two of them consulted the scoring guide, which provides a description of the criteria for awarding each score, two consulted the Writer's Handbook for help on grammar, and two of them read sample essays. Only one of the four made use of three of these resources, however, and none made use of all four.

From the way in which they reacted to the feedback it is unclear to what extent they were using it in order to learn, rather than simply to correct mistakes in order to gain a better score. Most of the participants claimed – understandably – that they wanted a better score, but strategies for improving a score do not necessarily result in learning. While there is not space to explore this here, students appear to use a range of strategies in dealing with feedback from Criterion. There is evidence, for example, that participants sometimes simply delete an error rather than correct it (Participant 2 referred to his wish to “avoid” errors in his second submission), while some make a clear attempt to correct their errors. In

most cases the corrections are an improvement over the original, but not in all cases (and often that depends on the nature of the error and the clarity or specificity of the feedback that Criterion provides). In addition, sometimes, in the course of revising a first draft, students correct some errors, but introduce new ones in their second draft. A reduction, or even an increase, in the number of errors in a second draft is therefore not necessarily indicative of how much has been learned from the Criterion feedback.

7. Conclusion

The combination of research methods used in this study provide a useful insight into student visual perception of feedback on their writing and their priorities in dealing with feedback on different types of error. Grammar was clearly a major focus of interest. They read the feedback on Grammar first and they also spent more time focussing on grammar errors and feedback than on any of the other error-types. There is evidence from stimulated recall that this is linked to past experience of writing and writing tuition, and especially to their experience of what aspects of writing are tested in their own countries. The likelihood that they are relying on previous experience of testing is supported by the fact that only one of 4 participants read the Criterion scoring guide to check on how their writing was being assessed.

After grammar, all participants spent most time on feedback on Organisation and Development. In this case the reason seemed to be that such feedback is indirect and can be or is difficult to interpret. Interpretation of such feedback would probably be facilitated by accessing advice in the Writers' Handbook, and by reading model essays, though most of these participants made little or no use of these resources. These features of the writing process and, especially of participants' priorities and strategies in dealing with feedback on errors, would have been impossible or very difficult to observe without the combination of methods used in this study. Eye-tracking data on its own does not offer an explanation for student behaviour when errors are identified in their written work and feedback provided. It does, however, provide evidence of the extent to which they appear to direct their attention to those errors and feedback. Further investigation of the extent to which the feedback helps students to learn from their mistakes, but the results of this study already provide some clues as to how software such as Criterion might be used more effectively, and to ways in which teachers might help students reflect on their priorities when writing, and reacting to feedback whoever or whatever provides it.

References

- Anderson, J. R. (1982). Acquisition of cognitive skill. *Psychological Review*, 89(4), 369-406.
- Anson, C. M., & Schwegler, R. A. (2012). Tracking the Mind's Eye: A New Technology for Researching Twenty-First-Century Writing and Reading Processes. 151-171. Retrieved from <http://www.ncte.org/library/NCTEFiles/Resources/Journals/CCC/0641-sep2012/CCC0641Tracking.pdf> website:
- Attali, Y. (2004a). *Exploring the Feedback and Revision Features of Criterion*. Paper presented at the National Council on Measurement in Education (NCME), San Diego, CA.
- Attali, Y., and Burstein, J.C. (2004b). *Automated essay scoring with E-rater 2.0*. Paper presented at the the Conference of the International Association for Educational Assessment, Philadelphia, PA.
- Bitchener, J., Young, S., & Cameron, D. (2005). The effect of different types of corrective feedback on ESL student writing. *Journal of Second Language Writing*, 14(3), 191-205.
- Brinko, K. T. (1993). The Practice of Giving Feedback to Improve Teaching: What Is Effective? *Journal of Higher Education*, 64(5), 574-593.
- Burke, D. (2009). Strategies for using feedback students bring to higher education. *Assessment & Evaluation in Higher Education*, 34(1), 41-50.
- Caldwell, J., & Leslie, L. (2010). Thinking aloud in expository text: Processes and outcome. *Journal of Literacy Research*, 42, 308-410.
- Caulk, N. (1994). Comparing teacher and student responses to written work. *TESOL QUARTERLY*, 28, 181-187.
- Cohen, A., & Cavalcanti, M. (1990). Feedback on compositions: Teacher and student verbal reports. In B. Kroll (Ed.), *Second language writing: Research insights for the classroom* (pp. 155- 177). Cambridge: Cambridge University Press.
- Connor, U., & Asenavage, K. (1994). Peer response groups in ESL writing classes: How much impact on revision? *Journal of Second Language Writing*, 3, 257-276.
- Cook-Sather, A. (2008). From traditional accountability to shared responsibility: the benefits and challenges of student consultants gathering midcourse feedback in college classrooms. *Assessment & Evaluation in Higher Education*, 99999(1), 1 - 11.
- Covic, T., & Jones, M. K. (2008). Is the essay resubmission option a formative or a summative assessment and does it matter as long as the grades improve? *Assessment & Evaluation in Higher Education*, 33(1), 75 - 85.
- Crisp, B. (2007). Is it worth the effort? How feedback influences students' subsequent submission of assessable work *Assessment & Evaluation in Higher Education*, 32(5), 571-581.

- Deane, P., Quinlan, T., & Kostin, I. (2011). Automated Scoring Within a Developmental, Cognitive Model of Writing Proficiency. *ETS Research Report Series, 2011*(1), i-93. doi: 10.1002/j.2333-8504.2011.tb02252.x
- Ferris, D. (1995). Teaching ESL composition students to become independent self-editors. *TESOL Journal, 4*(4), 18-22.
- Ferris, D. (2006). Responding to writing. In B. Kroll (Ed.), *Exploring the dynamics of second language writing* (pp. 119-140). Cambridge: Cambridge University Press.
- Ferris, D., & Roberts, B. (2001). Error feedback in L2 writing classes: How explicit does it need to be? *Journal of Second Language Writing, 10*(3), 161-184.
- Gibbs, G., & Simpson, C. (2004). Conditions under which assessment supports student learning. *Learning and Teaching in Higher Education, 1*(1), 3-31.
- Hacker, D., Keener, M., & Hirscher, J. (2009). Writing in Applied Metacognition. In D. J. Hacker, John Dunlosky & A. C. Graesser (Eds.), *Handbook of Metacognition in Education* (pp. 154–172). New York: : Taylor and Francis.
- Haigh, M. (2007). Sustaining learning through assessment: an evaluation of the value of a weekly class quiz. *Assessment & Evaluation in Higher Education, 32*(4), 457 - 474.
- Hedgcock, J., & Lefkowitz, N. (1996). Some input on input: Two analyses of student response to expert feedback in L2 writing. *Modern Language Journal, 287*-308.
- Huxham, M. (2007). Fast and effective feedback: are model answers the answer? *Assessment & Evaluation in Higher Education, 32*(6), 601 - 611.
- Hyland, F. (1998). The impact of teacher written feedback on individual writers. *Journal of Second Language Writing, 7*(3), 255-286.
- Hyland, K., & Hyland, F. (2006). Interpersonal aspects of response: Constructing and interpreting teacher written feedback. In K. H. F. Hyland (Ed.), *Feedback in ESL writing: Contexts and issues* (pp. 206-224). Cambridge: Cambridge University Press.
- Kim, M. (2008). The impact of an elaborated assessee's role in peer assessment. *Assessment & Evaluation in Higher Education, 99999*(1), 1 - 10.
- Kulhavy, R. W. (1977). Feedback in written instruction. *Review of Educational Research, 47*(1), 211-232.
- Lalande, J. (1982). Reducing Composition Errors: An Experiment. *The Modern Language Journal, 66*(2), 140–149.
- Lee, I. (2007). Feedback in Hong Kong secondary writing classrooms: Assessment for learning or assessment of learning? *Assessing Writing, 12*(3), 180-198.
- Leijten, M., & Waes, L. V. (2013). Keystroke Logging in Writing Research: Using Inputlog to Analyze and Visualize Writing Processes. *Written Communication, 30*, 358-392.

- Leki, I. (1991). The preferences of ESL students for error correction in college-level writing classes. *Foreign Language Annals*, 24, 203-218.
- Lizzio, A., & Wilson, K. (2008). Feedback on assessment: students' perceptions of quality and effectiveness. *Assessment & Evaluation in Higher Education*, 33(3), 263 - 275.
- Miller, T. (2008). Formative computer-based assessment in higher education: the effectiveness of feedback in supporting student learning. *Assessment & Evaluation in Higher Education*, 99999(1), 1 - 11.
- Orsmond, P., Merry, S., & Reiling, K. (2005). Biology students' utilization of tutors' formative feedback: a qualitative interview study. *Assessment & Evaluation in Higher Education*, 30(4), 369 - 386.
- Peterson, E. R., & Irving, S. E. (2008). Secondary school students' conceptions of assessment and feedback. *Learning and Instruction*, 18(3), 238-250.
- Robb, T., Ross, S., & Shortreed, I. (1986). Salience of feedback on error and its effect on EFL writing quality. *TESOL QUARTERLY*, 20, 83-93.
- Taras, M. (2001). The Use of Tutor Feedback and Student Self-assessment in Summative Assessment Tasks: towards transparency for students and for tutors. *Assessment & Evaluation in Higher Education*, 26(6), 605 - 614.
- Torrance, H. (2007). Assessment "as" Learning? How the Use of Explicit Learning Objectives, Assessment Criteria and Feedback in Post-Secondary Education and Training Can Come to Dominate Learning. *Assessment in Education: Principles, Policy & Practice*, 14(3), 281-294.
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46, 327-369.
- Vygotsky, L. S. (1978). *Mind in Society: Development of Higher Psychological Processes*. USA: Harvard University Press.
- Walker, M. (2008). An investigation into written comments on assignments: do students find them usable? *Assessment & Evaluation in Higher Education*, 99999(1), 1 - 11.
- Walker, M. (2009). An investigation into written comments on assignments: do students find them usable? *Assessment & Evaluation in Higher Education*, 34(1), 67-78
- Weaver, M. (2006). Do students value feedback? Student perceptions of tutors' written responses. *Assessment & Evaluation in Higher Education*, 31(3), 379 - 394.
- Winne, P., & Butler, D. (1994). Student cognition in learning from teaching. In T. Husen & T. Postlewaite (Eds.), *International encyclopaedia of education* (2nd ed., pp. 5738-5745). Oxford, UK: Pergamon.