

# Writing ePlatform: A Corpus-based Resource to Support Learning and Assessment in Writing

Sean McMinn

The Hong Kong University of Science and Technology

Flora Fung Yin Leung

Education Bureau, Hong Kong

## Abstract

Based on the findings of a recent study conducted by Education Bureau (EDB) and Hong Kong Examinations and Assessment Authority (HKEAA) on using assessment data to enhance the learning and teaching of Speaking and Writing, an online corpus-based resource tool that provides instant informing feedback has been developed to assist Key Stage Three students writing. This paper discusses the rationale and development of the tool, the Writing ePlatform.

## Introduction

In a study conducted by Education Bureau and Hong Kong Examinations and Assessment Authority aiming to identify possible problems in the learning and teaching of Speaking and Writing in the English Language at Secondary Three, Cheung and Leung (2012) observed that students' performances in speaking and writing were highly correlated. In the sub-construct of "vocabulary and language patterns", they also saw evidence of transfer between spoken and written. However, transfer from writing to speaking seems more likely. These imply that it is worthwhile to address students' difficulties in writing, which would then benefit students' speaking. A qualitative analysis of the speaking and writing performance of the same students shows that different error types were observed among students of varied ability. It is also found that "grammar accuracy index" is one of the strongest predictors of students' writing

performance. As Harmer (1983: p.35) pointed out, “An error is the result of incorrect rule learning; language has been stored in the brain incorrectly”. Hence, students need to learn about common error types, how they arise, and how to avoid them. The findings moreover indicate that students’ self-ratings bear little relationship to the ratings which their performances received, suggesting that students do not clearly know where they are in their learning process. Therefore, it would also be useful to step up e-learning resources that can give students instant feedback, and provide them with advice on different learning strategies and metacognitive skills, with the intent to help them become more autonomous learners.

When findings from the study and recommendations were presented to the participating schools, most teachers shared the vision of developing a computer corpus-based identification and classification system for students’ errors. It was decided that developing an online system to enhance writing skills was desirable given the ready availability of information technology (IT) facilities in schools. The Center for Language Education (CLE) at the Hong Kong University of Science and Technology (HKUST) was commissioned to design, produce and test an online system as the first stage of developing online tools that could assist or enhance the teaching and learning of spoken and written English. This paper discusses the background, development and piloting of the system.

## The Aim and Objectives

It should be noted that the creation of this interactive resource is part of a larger research project, which has its own aim and objectives. Because of space, this paper focuses only on the construction of the ePlatform, which provides students with interactive feedback that is geared towards their ability level and given immediately after they have submitted their writing to the online system. The feedback focuses on common writing problems, especially for low achievers. One distinctive feature of this system is that it could be incorporated into process writing.

The platform aims to: (1) assist with constructing the lexico-grammatical and discoursal/rhetorical knowledge of the target language and the skills required to access and apply that language; (2) encourage reflection and metacognition, where students are encouraged in independent learning and self-confidence; and (3) develop ‘cognitive apprenticeship’, where coaching and modeling occur, and where scaffolding is provided to support language learning. Two key features of the system are given as follows:

- (1) **the eLab**: an interactive web-based platform where students can submit writing and receive instant feedback and suggestions for how to improve their written English. The eLab is designed to be flexible according to student proficiency levels and needs. Additionally, the eTutor, student-oriented concordancers, vocabulary tools, and a vocabulary profile that refers to Key Stage 1, 2 & 3 vocabulary lists are accessible features of the eLab that assist students with developing their writing and provide teachers with diagnostic information for face-to-face lessons; and
- (2) **the eTutor**: a web-based portal that contains English learning materials and interactive learning objects based on common errors made by Hong Kong students. The eTutor provides guidance on how to address language learning issues relating to vocabulary, grammar and mechanics in response to common errors made by students in their writing.

To the best of our knowledge, no similar corpus-based system has been developed in the market for the Hong Kong, or Chinese, context and student age range.

## Theoretical Underpinnings of the ePlatform

The Writing ePlatform is a corpus-based resource. Corpora have become research tools for exploring the lexical features of learners’ language production. Meyer (2002) and Nesselhauf (2005) create corpora from student-written essays collected in classes and Kuo (2005) explores employing test takers’ writing passages to create a learner corpus. Coniam (1997) investigates the extent to

which it is possible to produce tests using corpus word frequency data with currently available computer technology. Coniam (1999) further identifies word frequency as an indicator of language proficiency in the written English of Grade 13 learners of English in Hong Kong. The study extends Laufer and Nation's (1995) work involving the Lexical Frequency Profile (LFP), in which vocabulary profiles were extracted from student writing on the basis of the frequency of the words. With regard to vocabulary, corpus data analysis and corpus tools have been causing a significant refocusing of views concerning the nature of English language assessment and the making of important pedagogical decisions.

The system is also grounded in cognitivist theories of Second Language Acquisition (SLA), from the interlanguage concepts of Selinker (1972), to the emphasis on lexical forms by Felix (1981) and Hoey (2005). This approach also meets the requirements laid out for corpus-based language learning technology in Ghadessy et al. (2001) and Römer (2006); it takes advantage of advances in computational linguistics and has been implemented according to the latest developments of human language technology. The system incorporates techniques that can help KS3 students acquire accuracy and fluency in written English and develop life-long writing habits in learning English. As suggested by Milton (2006; 2011), we are taking advantage of online resources to help KS3 students and teachers shift from a machine- or teacher-centered pedagogy to one that puts the KS3 students at the center of the writing process by making the learner accountable, and ultimately more confident and independent.

Furthermore, based on and adapted from Bates' (2007) e-learning rationale, the system increases access to learners' resources, enhances teaching and learning, better prepares students for communication skills required in the international setting of Hong Kong, develops independent learning skills through online programming and mobile learning, and better accommodates the differing styles and background of students.

This rationale is also related to the term 'human-assisted' in that we still require expert analysis of potential 'errors' and/or problems Hong Kong students

make in their writing to assist us with writing instant prompts. The ‘human-assisted’ is also in relation to the expertise and experiences from Hong Kong teachers – by their use of data collected through the ePlatform of students' writing to diagnose individual problems. Sample writings from KS3 students were analysed to develop a better understanding of what type (and scope) of feedback is required for students when using the program; this allows for accurate and meaningful feedback prompts (and vocabulary profiles) for both students and teachers to use. *Figures 1.1 and 1.2* outline how we envisioned the Writing ePlatform to be used within a process writing approach to teaching and learning. The affordances of the Writing ePlatform vary and are not restricted to classroom teaching, nor encouraged, to be used only by teachers as a teaching tool in the classroom. In fact, the design of the ePlatform takes into consideration methodologies that include: independent learning, blended learning, and assessment as/for learning. For example, *Table 1* outlines an adapted version of Earl’s (2003) assessment roles and goals, where both the teacher and the Writing ePlatform have shared roles and goals in a blended learning approach.

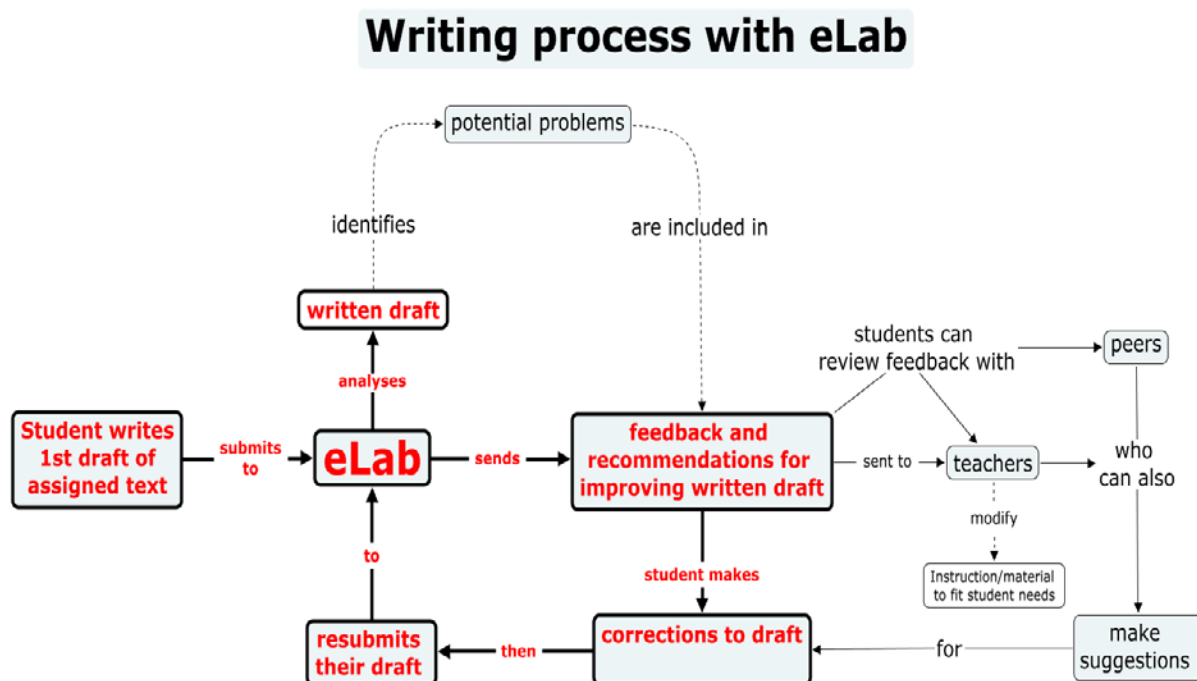


Figure 1.1 Suggested process writing approach with the Writing ePlatform

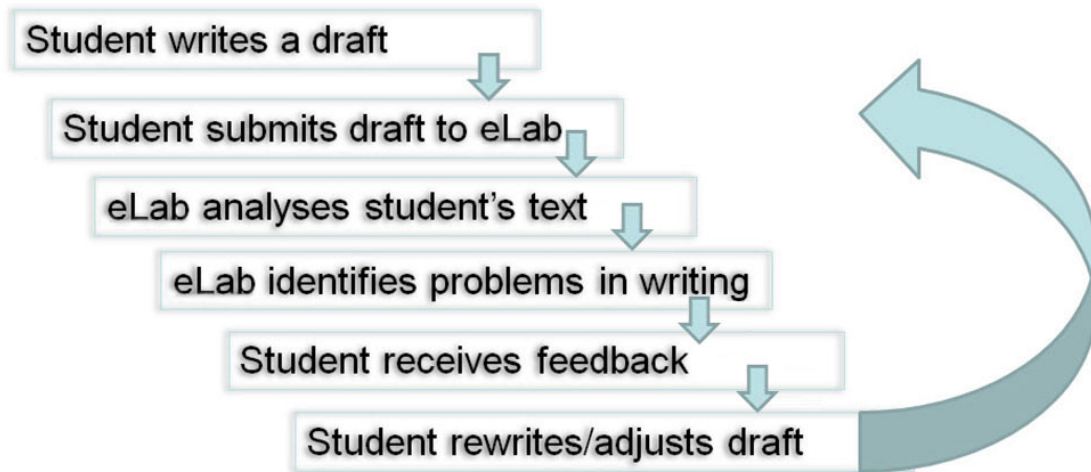


Figure 1.2 How Writing ePlatform assists in process writing

Table 1. Earl’s Assessment Roles and Goals

<b>Assessment Roles and Goals</b>	
<b>Role</b>	<b>Goal</b>
Teacher + <b>ePlatform</b> as mentor	Provide feedback and support to each student in <b>both formal (classroom) and informal (home) environments.</b>
Teacher + <b>ePlatform</b> as guide	Gather diagnostic information to lead the group <b>and individual students</b> through work at hand <b>and/or process.</b>
Teacher + <b>ePlatform</b> as accountant	Maintain records of student progress and achievement, <b>allowing review of entire processes and accumulation of learner corpora.</b>
Teacher + <b>ePlatform</b> as reporter	Report to parents, students, and the school administration about the student progress and achievement, <b>showing entire learning process.</b>
Teacher + <b>ePlatform</b> program director	Make adjustments and revisions to instruction practices <b>that focus on individual student needs based on empirical data.</b>

Source: Adapted from *Assessment As Learning* (Earl, 2003).

## Developing and Improving Error Rules for the System

The establishment of error rules and the improvement of the reliability and validity of these rules are two key tasks in establishing this online system. These rules were compiled based on previous research at HKUST and the research data provided by EDB. Initially, a total of 1,800 written essays on 10 different topics (180 essays per topic) were collected from either Secondary 3 (S3) TSA or S3 students from six secondary schools. These were analysed and a list of common errors produced by local Hong Kong students was drawn up based on specialists' expertise. These common language problems were interpreted into error patterns that could be used to analyse text heuristically. Additional rules were added based on the work of Milton (2006, 2010, 2011) and through analysis of the corpus. To enrich the feedback given by the system, a repository of information about errors and reference tools was created. Students submitted written assignments set in class to the system for analysis. Their scripts were saved by the system and this allowed the organic growth of the corpus of Hong Kong S3 student essays. In this way, the accuracy of testing rules was continually improved and this feature will enable further modification and improvement of the system in the future. The following is an example of the process of identifying an error, writing rules, and designing the instant prompts for that error.

Common Error: Although + but in a sentence

Although the food was not great but I felt very happy that I was able to cook the meal by myself.

Although the food was not great, I felt very happy that I was able to cook the meal by myself.

These rules were then converted to a formula for the ePlatform to identify patterns in students' writing.

Error formula:

!{well|such|known}{as|since|although}<{(CC)|(IN)}> \*<0-6> (CC)<{but}>

By using the Stanford Natural Language Processing Group's parts of speech tagger, the rules were tested using a learner corpus.

Example:

View rules Update database Test rule **POS Tagger** Writing eLab POS Tag Set Logout

Although the food was not great but I felt very happy that I was able to cook the meal by myself.

Submit

The part-of-speeches of the words generated by the Stanford POS Tagger are shown below:

Although the food was not great but I felt very happy that I was able to cook the meal by myself .  
 IN DT NN VBD RB JJ CC PRP VBD RB JJ IN PRP VBD JJ TO VB DT NN IN PRP .

Then instant prompts were written and allocated to relevant rules.

Example:

E1

**Error description**  
 Although + but in a sentence

**Error pattern**  
 ![well | such | known] {as | since | although}<{(CC) | (IN)}> \*<0-6> (CC)<{but}>

**Error explanation**  
 Either "although" or "but" may not be needed. Please check. They are rarely used in the same sentence. Use Word Neighbors to look for examples of how "although" and "but" are used by native writers of English.

- ✗ Although it looked like a high-class hotel but I only paid fifty-five dollars to stay there.
- ✓ It looked like a high-class hotel but I only paid fifty-five dollars to stay there.
- ✗ Although the food was not great but I felt very happy that I was able to cook the meal by myself.
- ✓ Although the food was not great, I felt very happy that I was able to cook the meal by myself.

Use [Word Neighbors](#) to find out more about this!



We launched and completed two pilots with the same schools and students: the first in October 2012; the second in January 2013. Four topics were assigned to seven schools in October, and the remaining six topics were assigned to the schools in January. Each school was given two topics. Additionally, we collected feedback from schools on the use of the ePlatform at the end of each pilot. This included an online survey completed by 336 students; observation reports from teachers, providing feedback on any user or technical problems that have occurred, and suggestions on how to improve the program.

To inform the team on what rules needed adjustment or deletion, numerous batch tests to analyse, review and evaluate rules used in the Writing ePlatform were conducted during the project pilot. This included measuring the effectiveness and hit rate of rules. The batch tests include: (a) an analysis of the Learner Corpora on all 10 topics provided by EDB of students from previous years, and (b) an analysis of the Learner Corpora created from the October-December and January-May pilots. Results and analysis also included comparing the two learner corpora. Below is an example of findings from the batch test analysis:

**Issues:** Developmental challenges.

**Example:** Error identifies too many false positives.

**Other error type:** Wrong error identified.

**False positives:** Rule flags correct use of English as an error.

**Correct identification:** Rule identifies correct error.

---

**Error Rule 1:** Although + but in a sentence (start with a capital, end in a full stop)

**Hits:** 22

**Issues:** None.

**Other error type:** None.

**False positives:** None.

**Correct Identification (samples):**

{Although they were so old but }they were still very strong.

{Although it looked like a high-class hotel but }I paid fifthty-five dollars only.  
 {Although it is quite expensive, but }it is a fun way to learn English as we can know more about the culture of the place at the same time.  
 {Although we can't said Japan's language, but }we can said english with they.  
 {Although, this trip was very short, but }we were enjoyed for this tirp.  
 {Although, I know I will be fat, but }I have bought it!  
 {Although the food was not yummy but }I felt so happy that I can cook the meal by myself!  
 {although I had hurt, but }I really enjoy this trip.

**Error Rule 2:** A sentence / clause has more than 2 verbs

**Hits:** 23874

**Issues:**

This rule is very problematic – too many false positives. Detection issues.

**Other error type:** None.

**False positives:**

During the past summer holiday, my family and I {went to a village in the Mainland China to visit }my grandparents.

When we arrived, grandpa and grandma {were in excitement since we had }n't visit them for a long period of time.

They {took out a watermelon from the fridge to greet }us in ecstasy.

Our sweats from the travel {had gone }away immediately.

Because of summer holiday so there {were lots of children who living }near to me playing with me I was very joyful when I was playing with them.

Because of summer holiday so there were lots of children who living near to me {playing with me I was }very joyful when I was playing with them.

Because of summer holiday so there were lots of children who living near to me playing with me I was very joyful when I {was playing }with them.

On the second day, I {went to the farm to help }the farmer with my family.

I {tried to be }a farmer and started farming.

It can be seen in this example that the system was able to flag when Error Rule 2 required amendment. The use of the rule, “A sentence / clause has more than 2 verbs”, was problematic since it resulted in numerous false positives. (A “false positive” refers to a correct sentence being mistakenly judged as erroneous by the system. An online error correction system must seek to avoid false positives.) The many occurrences of false positives also correlate with other findings from student survey’s and teacher interviews, where frustration was expressed due to correct sentence patterns being identified as incorrect causing confusion among students. The second of the two examples suggests that the Although-but rule was applied correctly by the system.

Findings from the batch tests were then categorized to determine whether there was a relationship between error type and topic. Establishing the link between errors and writing topics led to us indexing the online material by both error type and topic in the eTutor, which teachers can use to inform teaching. For example, when teachers teach a new topic they can take into consideration errors identified as common in that topic or metalinguistic explanations available there for their students. The following table shows an example of rules and the hit rate in students writing according to topic. Topics were categorized into three main text types: describing, giving advice, and giving information.

**Table 2. Occurrences of Error Rules Applied by Topic and by Text Type**

Error Rule	Topic					Total	Text Type			Total
	1. An Accident	2. Hong Kong Sightseeing	3. Information about Schools in Hong Kong	4. Information and Advice about Going to S Secondary Schools	5. Sports Day		Describing (Topic 1 + Topic 2 + Topic 5)	Giving Advice (Topic 4)	Giving Information (Topic 3 + Topic 4)	
Go/ going/ went + shopping	5	5	0	0	0	10	10	0	0	10
Many + Singular	11	59	25	30	29	154	99	30	55	184
Modal verbs (will / can/ would) + adjective	4	6	3	12	5	80	15	12	15	42
Modal verbs (will/ can/ would) + noun	0	55	7	8	0	70	55	8	15	78

Further analyses on tallies such as those given in Table 2 together with other information such as the differing writing abilities of students, can provide useful information to the ePlatform team on more efficient ways to individualise feedback. The tallies would also have pedagogical implications for teachers. For example, teachers may know what grammatical points to emphasise in their pre-teaching of the writing tasks (i.e. the emphasis on teaching modal verbs before students are being assigned to write Sightseeing in Hong Kong).

## Writing ePlatform: Features and Functions

Based on our pilot and batch test findings, our team finalized and developed the following key eLab and eTutor functions:

- (1) A basic student user interface (*Figure 2.1*): This is main user interface for the eLab where students will submit their text for feedback. Students are given the option to choose the topic of their text, proficiency level to determine the type of instant feedback they will receive, and links to additional tools to assist them.
- (2) Instant prompts (*Figure 2.2*): After a student has submitted their written work for analysis, any problematic text will be identified and highlighted in the student user interface. Clicking on the highlight text will show an instant prompt providing feedback on how a student can improve their text. Feedback in the instant prompts will also direct students to learning materials in the eTutor.
- (3) Word Tag (*Figure 2.3*): This will give teachers and students a visual analysis of vocabulary frequency and type.
- (4) Vocab-Profile (*Figure 2.4*): This will provide students and teachers with an analysis outlining vocabulary frequency based on Key Stage 1, 2, & 3 word lists (available on the website of Education Bureau).
- (5) Web-based tools (*Figure 2.5 & 2.6*): These are web-based tools that operate within the eLab. These tools are available to users for discovering common vocabulary usage, collocations, and frequency.
- (6) eTutor (*Figure 2.7, 2.8 & 2.9*): This web-based portal containing interactive learning objects based on common errors is organized by topic, error type, and additional input (videos). The eTutor provides guidance on how to reflect on common errors made by students in their writing.

These features allow the ePlatform to function as a tool within assessment as/for learning methodologies. For example, the ePlatform: (1) provides feedback and support to each student in both formal (classroom) and informal (home) environments; and (2) gathers diagnostic information to lead the group and individual students through work at hand and/or process.

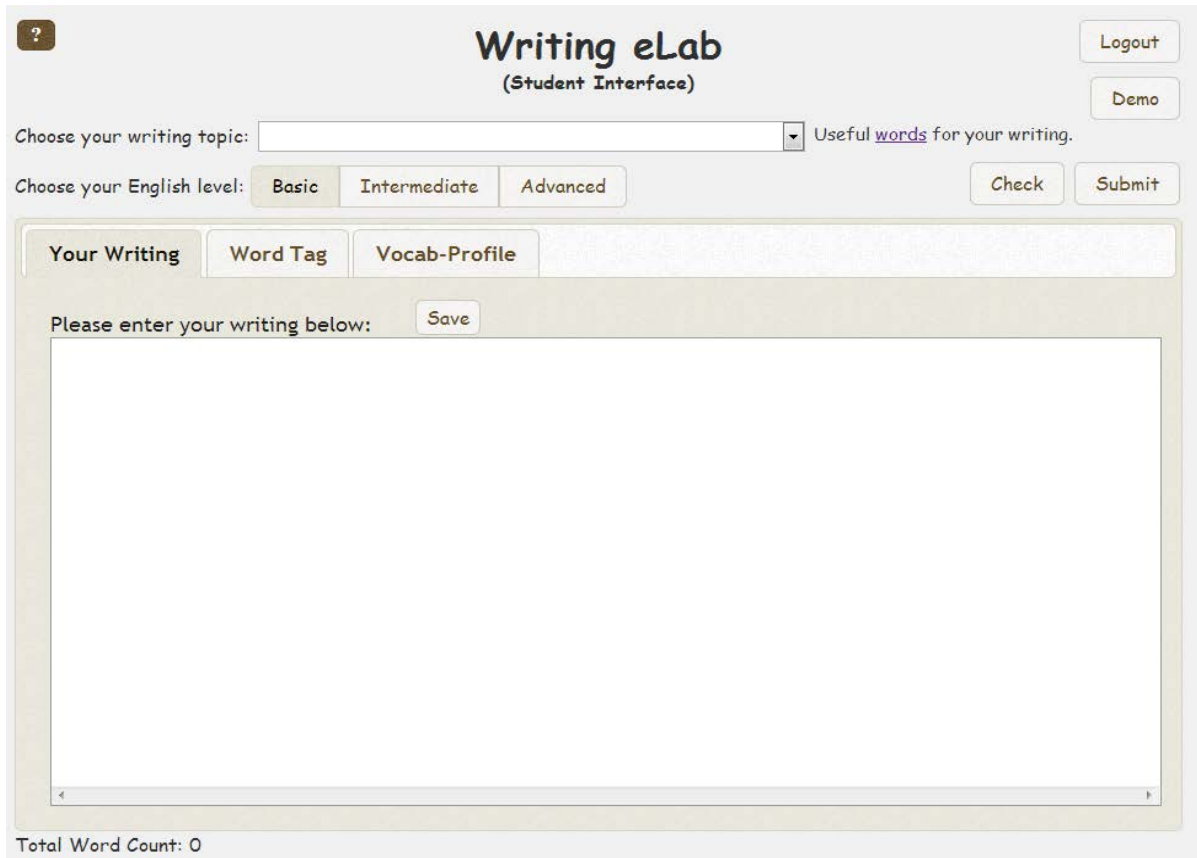


Figure 2.1 Writing ePlatform student user interface

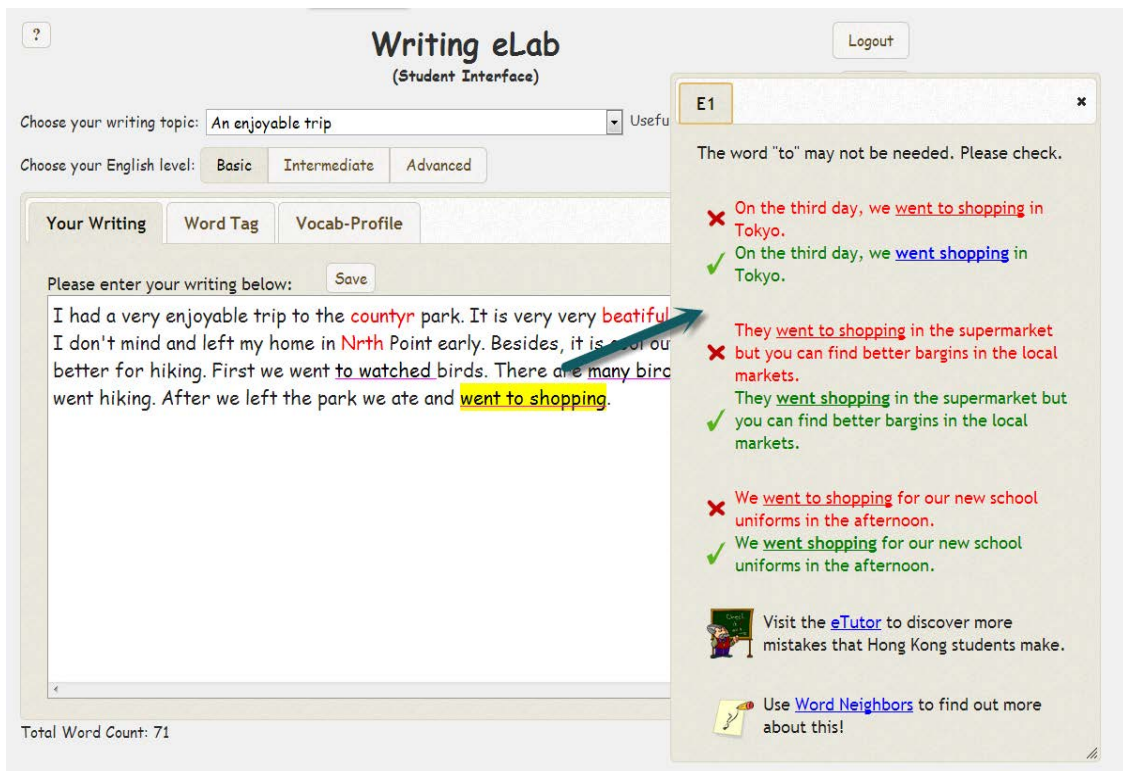


Figure 2.2 Instant prompt feedback

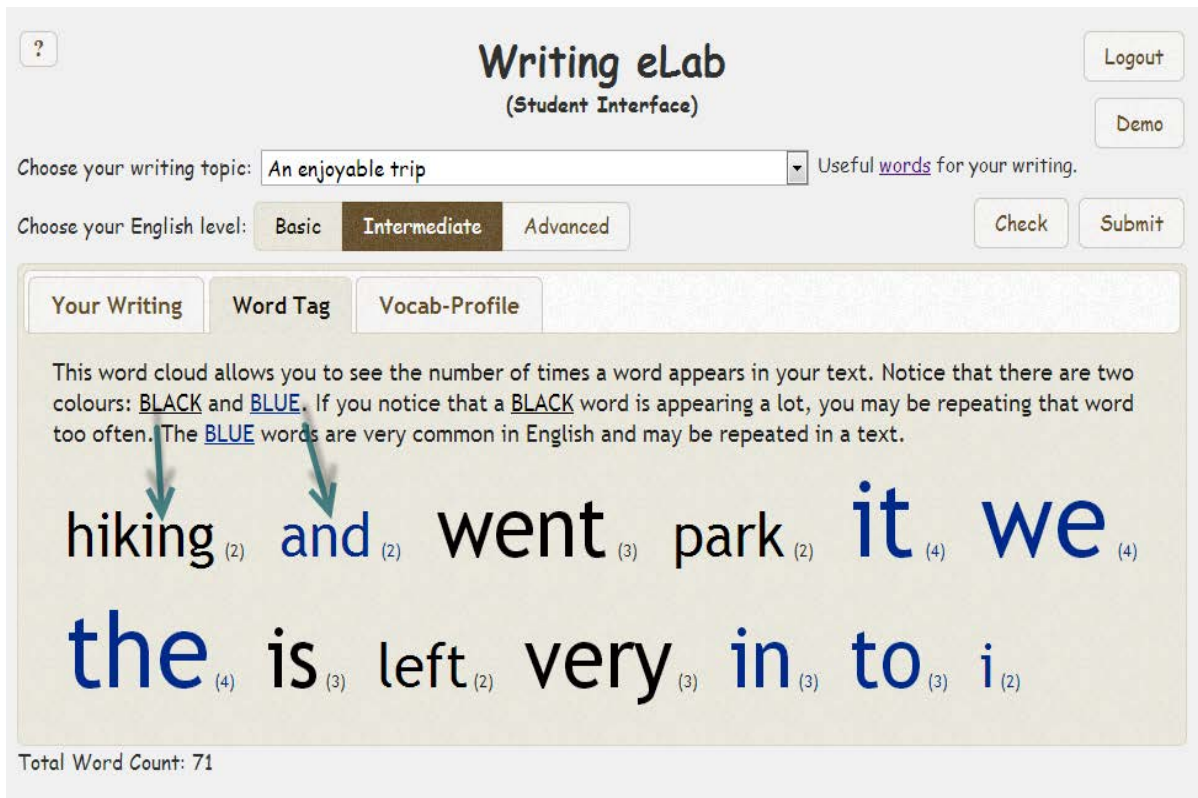


Figure 2.3 Word Tag

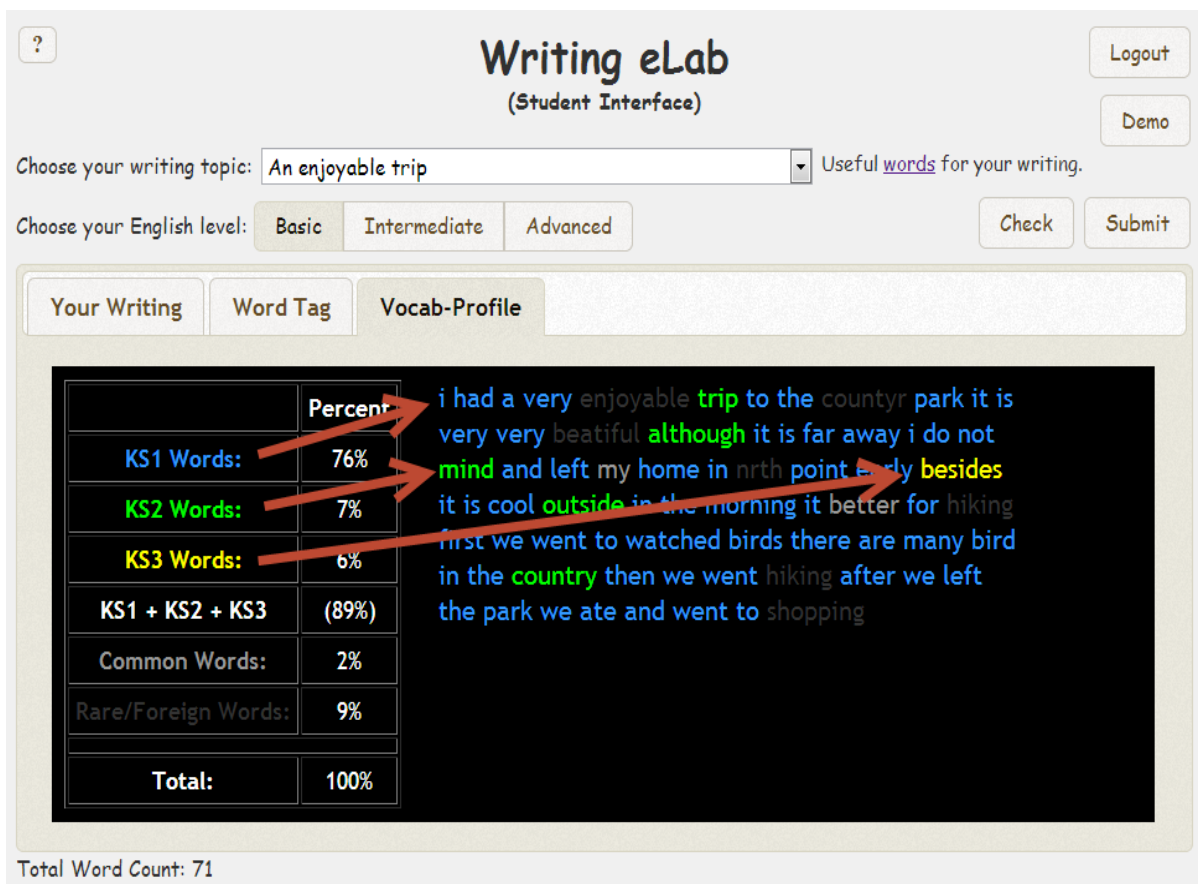


Figure 2.4 Vocab-Profile

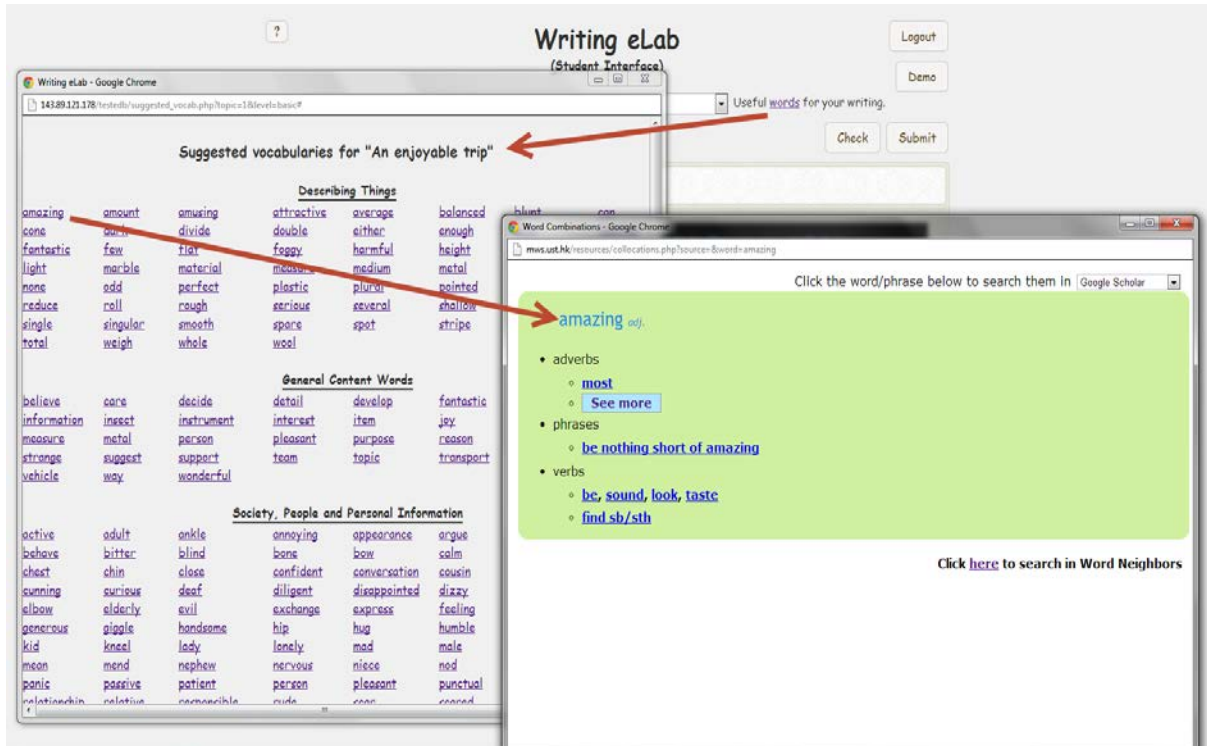


Figure 2.5 Useful Words for Your Writing



Figure 2.6 Word Neighbors (concordancer)



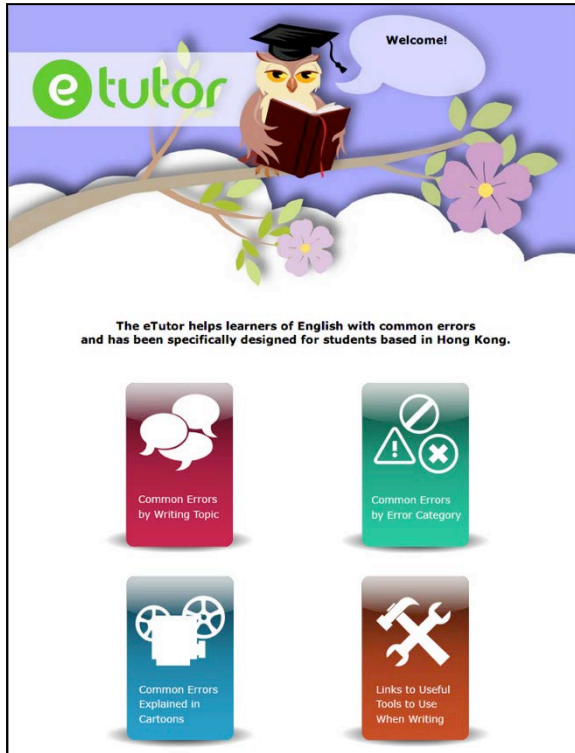


Figure 2.7 eTutor landing page



Figure 2.8 Category menu



Figure 2.9 eTutor error categorization

## Conclusion

Great efforts have been made by the team to improve the validity and reliability of the online feedback system. Pilots with participating secondary schools were conducted, and the rich data which we have collected would assist us in improving continuously the Writing ePlatform. The learner corpus generated from the project also enables our team to analyse students' writing, all of which are useful for designing feedback and support material and also future projects and research. We believe that the final product will be a positive addition for the Hong Kong secondary school community, and with proper training and planning, the Writing ePlatform can enrich and enhance the writing process for English language teaching and learning. By engaging individual students with feedback and feedforward, the ePlatform can potentially benefit students and teachers in the writing classroom and beyond.

## References

- Anderson, T. (2004). Introduction: Teaching in an online learning context. *Theory and practice of online learning*. Alberta: Athabasca University.
- Bates, A. (2007). Strategic planning for e-Learning in a polytechnic. In M. Bullen and D. Janes (Eds.), *Making the transition into e-learning strategies and issues*. Hershey, PA: Information Science Publishing.
- Cheung, A. K. M. & Leung, F. F. Y. (2012). Using assessment data to investigate the relationship between speaking and writing. *Assessment & Learning, 1*, pp. 140-156.
- Chickering, A. and Ehrmann, S. (1996). Implementing the seven principles: Technology as lever. *AAHE Bulletin*, October, pp. 3-6.
- Coniam, D. (1997). A preliminary inquiry into using corpus word frequency data in the automatic generation of English cloze tests. *CALICO Journal, Vol. 16*, Nos. 2-4, pp. 15-33.
- Coniam, D. (1999). Subjects' reactions to computer-based tests. *Journal of Educational Technology Systems, Vol. 23*, No. 3, pp. 195-206.
- Earl, L. (2003). *Assessment as learning*. Thousand Oaks, CA: Corwin Press.
- Felix, S. (1981). On the (in)applicability of Piagetian thought to language learning. *Studies in Second Language Acquisition, 3*(2), pp. 201-220.
- Ghadessy, M., Henry, A. & Roseberry, R. (Eds.) (2001). *Small corpus studies and ELT: Theory and practice*. Amsterdam: John Benjamins.

- Harmer, J. (1983). *The practice of English language teaching*. London: Longman.
- Hoey, M. (2005). *Lexical priming: A new theory of words and language*. London: Routledge.
- Kuo, Grace (2005). A preliminary corpus study on EFL test takers' writing proficiency. Proceedings of the 8th Academic Forum on English Language Testing in Asia: Assessment for Learning, Hong Kong.
- Meyer, C. F. (2002). *English corpus linguistics*. Cambridge: Cambridge University Press.
- Milton, J. (2006). Resource-rich web-based feedback: Helping learners become independent writers. In K. Hyland and F. Hyland (Eds.), *Feedback in second language writing: Contexts and issues* (pp. 123-138). Cambridge: Cambridge University Press.
- Milton, J. (2010). A toolkit to assist L2 learners become independent writers. In Proceedings of the NAACL HLT 2010 Workshop on Computational Linguistics and Writing (pp. 33-41), Los Angeles, California.
- Milton, J. (2011). A case study of a corpus-based, IT-enabled approach to improving English proficiency. Paper presented at ICTATLL2010 Kyoto Conference, Japan.
- Nesselhauf, N. (2005). *Collocations in a learner corpus*. Amsterdam: John Benjamins Publishing Co.
- Römer, Ute. (2006). Where the computer meets language, literature, and pedagogy: Corpus analysis in English Studies. In A. Gerbig & A. Müller-Wood (Eds.), *How globalization affects the teaching of English: Studying culture through texts* (pp. 81-109). Lampeter: E. Mellen Press.
- Selinker, L. (1972). Interlanguage. *IRAL*, 10, pp. 209-231.

Authors' email: Sean McMinn   lcmcminn@ust.hk  
Flora Leung   scdoase@edb.gov.hk