Arizona Working Papers in SLA & Teaching, 84-109 (2014)

AUTOMATED PREDICTION OF THE WRITTEN ERRORS OF TERTIARY LEVEL ESL AND EFL LEARNERS

Kelvin Ka Yu Chong, University of Arizona
Allen Ho, the Chinese University of Hong Kong
Olive Cheung, the Chinese University of Hong Kong
Ella Leung, the Chinese University of Hong Kong
Peter Clarke, exida Asia Pacific Pte Ltd

ABSTRACT

This paper summarizes a two-year project at the Chinese University of Hong Kong (CUHK) from 2006 to 2008. IELTS Online Writing Assistant (IOWA) is a computer-based teaching system designed for students in support of their preparation for the writing module of the International English Language Testing System (IELTS). Situated in Error Analysis (EA) (Corder, 1981; Mitchell, 2004; Lightbown, 1998) and Computer-Assisted Instruction (CAI), the system is designed to address both students' local (i.e., word and sentence) and global (i.e., discourse) errors. It attempts to predict which types of error tertiary level ESL and EFL learners in Hong Kong are most likely to make by assessing their ability to find errors in a prepared script. There are two major research questions: How effective is the system in predicting students' predispositions to committing writing errors and to creating discourse problems? How can teachers better use such a diagnostic tool to complement their teaching practice to address different issues of error categories? The study concludes that: (1) its predictive performance varies greatly depending on the types of error; (2) testing alone—with limited feedback—is effective at reducing the incidence of certain types of error, especially low-frequency and structural errors, in students' productive writing; (3) certain types of error, which are most difficult for IOWA to predict, could be better instructed by teachers in an ordinary classroom context. Overall, the study has laid important foundations for enhancing our students' IELTS writing test preparation.

Keywords: *Computer-Assisted Instruction (CAI), International English Language Testing System (IELTS), Error Analysis (EA)*

1. INTRODUCTION

The International English Language Testing System (IELTS), an international high stake English proficiency test, has been seen by many local universities and companies in Hong Kong as one of the most reliable tests for the English ability of incoming students, graduates, new job seekers, and employees. With a satisfactory score of the IELTS, English users display an important credential on their CV and resume, which in return opens doors of opportunities for their future career and academic development. For example, an overall band score of 6.5 or above with no individual score below band 6 obtained in one sitting in the Academic Module of IELTS within the two-year validity period is accepted as equivalent to the highest grade, "Level 2", in a test called *Use of English* in the government's *Common Recruitment Examination*, which is a compulsory test for those who are seeking a contracted civil servant position in Hong Kong.

IELTS tests a candidate's four areas of language abilities. Under the current exam format, IELTS can be taken as an academic or general training mode, both assessed on a 9-band scale. Each module (i.e., Listening, Academic Reading, General Training Reading, Academic Writing, General Training Writing, and Speaking), reported in either a half or whole band, carries equal weighting towards a total band score. According to the scoring descriptors, candidates scoring Band 6 are considered competent users who have a generally effective command of the English language, despite systematic errors that could interfere with meaning; those who score Band 7 are considered good users who generally handle complex language well with some systematic errors that do not interfere the meaning; Band 8 indicates very good users who have attained a full command of using complex language with occasional unsystematic errors; and Band 9 indicates expert users.

The academic version of the IELTS writing test consists of two tasks to be completed within one hour. Task 1 requires a 150-word description of a diagram or some data; Task 2 requires a 250-word argumentative essay. We analyzed error types separately for the two tasks, although the final selection of error types was based on the combined error frequency (see Methodology for more details).

In Hong Kong, under the 2010/11 Common English Proficiency Assessment Scheme (CEPAS), all graduating university students are currently encouraged and subsidized by the *University Grants Committee* (UGC) of Hong Kong to take the academic version of the IELTS. The numbers of candidates taking the exam have been constantly high: in 2009/10, for example, 68% of Hong Kong's graduands[i] accounting for 12,063 final-year students participatedii (UGC, 2010); in 2012/13, 60% of Hong Kong's graduands of UGC-funded undergraduate degree programs participated in the IELTS (UGC, 2013).

The Chinese University of Hong Kong (CUHK) offers workshops to prepare students for the test. These are generally popular, reaching around 22% percent of those intending to take the test. However, due to time constraints, the students are able to spend only one three-hour session on each of the four skills (reading, writing, speaking, and listening). It would clearly be beneficial for both students and teachers to optimize the training provided in such a limited time. In line with this belief, by focusing on writing skills, we have devised and tested an online system that attempts to predict the types of error an individual student would make in their writing. Based on these predictions, the student can then be directed to appropriate remedial teaching and exercises. For example, a 'task bank' of online remedial exercises has been constructed by our team.

Based on 295 authentic writing scripts (147 scripts of IELTS Task 1 plus 148 scripts of Task 2) produced by ESL students at the tertiary level in Hong Kong, we present an analysis of their typical error types. We found discernible patterns in the ability of the online test to predict and/or reduce the occurrence of errors in the students' IELTS scripts, depending on the types of error. In this paper, we describe our methodology for selecting the types of error to target as well as the design, construction, and testing of the predictive tool. Related work in the literature is discussed in the following section.

2. LITERATURE REVIEW

Over several decades, local tertiary institutes have faced the challenge of improving students' English proficiency, especially after the adoption of the IELTS by Hong Kong's University Grants Committee (UGC) as the English language exit test for university graduates in 2002. Helping tertiary students enhance their writing skills has become one of English instructors' toughest tasks, as they are confronted by timetabling constraints with limited contact hours with students. As Warschauer and Ware (2006) describe:

> At the same time we are cognizant of the high stakes for both ESL and EFL students in attaining English language writing proficiency, we are painfully aware of the steep odds that learners face in reaching this goal. The reality is that the need for expanded and improved instruction in English language writing simply cannot be matched by the capacity of educational institutions to offer corresponding instruction. (p.176)

To address the problem, Dodigovic (2002) maintains that developing and incorporating computer-based learning tools is a possible solution, as the flexible learning mode of such tools could enable writing practice and feedback to be ideally organized around the learners' individual study and casual work schedule.

2.1. History and Background: Error Analysis (EA) and Computer-Assisted Instruction (CAI)

In the 1950s, behaviorists viewed language learning as a process of habit formation via repeated responses to stimuli. This implies that foreign language (FL) or second language (L2) learning would be heavily influenced by a learner's mother tongue (L1). This tradition consequently held that effective L2 teaching and learning should focus on the differences between L1 and L2, which brought about Contrastive Analysis (CA). CA focuses on the scientific description of the language to be learned, carefully compared with a parallel description of the native language of the learner. Later research findings showed that CA did not perform as satisfactorily as expected in predicting learner difficulties by looking at the parallel linguistic features between languages. Meanwhile, increased attention was focused on learner language, i.e., language actually produced by learners themselves, which then brought about Error Analysis (EA), the systematic investigation of L2 learners' errors. Corder (1981) focused on this area, claiming that errors reflected learners' current understanding of the rules and patterns of the L2 (also, see Mitchell, 2004). EA research during the 1960s showed that most L2 learners' errors in fact did not originate from their L1 (Lightbown, 1998). In other words, to a certain extent, language produced by L2 learners neither resembles their L1 (as shown by EA) nor resembles their L2 (due to all the unexplainable mistakes made). The term interlanguage (IL) was then coined by Selinker to refer to the condition where the learner language lies somewhere in-between their L1 and the target language (TL) that they aim at mastering. Under a framework which investigates data on utterances observable from learners' NL, IL and TL, Selinker (1972) established various phenomena and processes underlying IL behavior, including fossilization, language transfer, and overgeneralization.

These breakthroughs in the areas of EA and IL had a significant impact upon research and pedagogy in L2 instruction throughout the world. As a result, research efforts and resources in Hong Kong were drawn to areas that were not given sufficient attention before. A wide range of rising issues at that point include language transfer interference (Chan, 1991; Chan, 2004; Sung, 1991; Green 1991b); error gravity (Green, 1991a); relation between errors and teaching contexts (Hepburn, 1991); and error avoidance (Lee, 1990), with particular consideration given to the local context (Yip and Matthews, 1991; Li and Chan, 1999) and to both the local and international contexts (Bunton, 1991). Data regarding IL produced by English as a Second Language learners in Hong Kong's classrooms have been collected at different levels of instruction and further processed and analyzed by researchers in various ways. For example, educational professionals have often supported the idea of categorizing errors into sub-types. Apart from ranking error types according to their gravity (degree of

error seriousness), Lee (1990) suggested that errors be distinguished according to their linguistic levels, i.e., morpho-syntactic errors, discourse errors, phonologically-induced errors, and lexical errors; whereas Li and Chan (1999) advocated the establishment of error taxonomies that collect lexical and structural errors with attention paid also to error teachability. Collection and analysis of this nature may very well contribute to the development of computer-assisted tools in correcting writing errors as well as teaching correct language forms.

In the 1960s, work in computer assisted instruction (CAI), or more precisely in question here, Automated Writing Evaluation (AWE), began; the first two products being *Project Essay Grade* developed by the College Board, a national network of universities in the United States, which aims to help score high school student essays, and the *Writer's Workbench*, which was a set of writing tools running in Unix systems providing feedback on writing quality. At that time, due to technological constraints, CAI/ AWE operated on a very narrow definition of such quality. For example, Writer's Workbench only allowed for limited functions, such as performing readability tests on the text to report a statistical count of sentence length, and for flagging wordy, clichéd, misspelled or misused phrases. However, the development of such pilot CAI/ AWE software essentially pointed the field in an important direction: evaluation and feedback (Burstein, 2004; Warschauer, 2006).

2.2. Functions and Design of Computer Assisted Instruction (CAI) Tools for Automated Writing Evaluation (AWE)

A wide range of CAI tools has gradually been developed either by educational practitioners or by researchers in commercial companies offering systems that perform AWE. The two most currently and widely used AWE systems in the field are *Criterion* and *MY Access!*, developed by Educational Testing Service and Vantage Learning respectively, both of which are commercial enterprises (Warschauer, 2006; Chen, 2006). Three recent examples of AWE systems developed by researchers/ practitioners within academia are the *UNED Grammar Checker* developed by the Universidad Nacional de Educación a Distancia in Spain (Lawley, 2004), the *Cyber Coach* piloted with a grant at Macquarie University in Australia (Dodigovic, 2002) and *HARRY*, developed by Loughborough University in the United Kingdom (Holdich, Chung, & Holdich, 2004).

In terms of function, most of the products run in two major steps: to evaluate and then to give feedback, though each of them may approach the two steps in different ways or with a different focus. In our paper, we use 'evaluation' and 'feedback' in a broad and loose sense to refer to the general design aims of AWE products instead of directly translating them into the more well-defined and narrowed definitions in the area of language testing. Generally speaking, although all forms of writing assessment include students' written output, followed by evaluation and feedback on students' work, these terms could differ in different contexts. In traditional language testing on writing, these terms of assessment, evaluation, and feedback usually include the process of learners' writing part of an essay, followed by the correction of errors, and showing rankings, scores or percentiles that inform students of their performance in a testing group or a learning community. On the other hand, in AWE, looser definitions of evaluation and feedback include any form of written or textual organizing output, which becomes the input to the system, followed by the diagnosis and comments on room for improvement or comments on categories that need suggestions, the direction for attention, and provision of remedial exercises.

The function types of the various AWE systems reviewed are summarized in Table 1.

Steps	Types of Functions	Details
Input	• Receive input from students	• Students copy and paste entire written texts (writing assignments/ exam compositions) into computer program • Students produce short answers to diagnostic prompts
System Evaluation	• Assessment/ Scoring of text quality • Diagnosis of students' weakness/ areas requiring attention	• Holistic scores for texts • System identifies the errors • System evaluates areas including: -spelling/punctuation -grammar -writing style - organization/ discourse structure
System Feedback	• Give comments to learners • Suggest ways to improve writing quality / to correct errors • Direct students to required teaching points	

Table 1. Steps in a typical Automated Writing Evaluation system

As summarized in Table 1, all of the AWE systems require learners to produce writing input for analysis, the input being either full word processed texts copied and pasted into the system or short answers in sentence form given by students in response to prompt questions. The systems will then process the input to give scores for the writing and/or to identify weaknesses and errors in terms of the writing style and textual organization (macro, top-down approach) and/or of the grammar and mechanics on the word or sentence levels (micro, bottom-up approach). (For more details, refer to Dodigovic, 2002, pp.11-15.) Lastly, the systems will give feedback in the form of comments, suggestions for possible changes, and/or explanation of error formation and correction.

Previous AWE approaches have focused mainly on addressing students' errors in retrospect, that is, after a writing task has been completed. Complementary to this, our project aims to adopt a proactive approach: predicting the errors that individual students are likely to make, and coaching them ahead of the writing task. A proactive approach is useful in the sense that it is preventive rather than curative. Since writing is a complicated process that requires high cognitive thinking, having an agenda of what to write before the actual writing can help students organize their ideas. For this purpose, a proactive approach can act as a reminder for learners' future writing.

3. METHODOLOGY

First, our IOWA team, consisting of six English language instructors, prepared a list of 75 types of error commonly committed in IELTS writing scripts. The list was narrowed down to 26 error types appearing in 174 mock IELTS writing scripts by selecting the most frequent error types. The 26 error types were further divided into 20 local (word and sentence level) errors and 6 global (structural) errors, e.g., weak or absent introduction, absent conclusion, and coherence.

We then carried out a computer-based test with ten students, which was a manageable size, considering the limited human power of the team. The test consisted of a teacher-written

IELTS script spiked with examples of all the local errors and some of the global errors. Students were asked to identify and correct all the local errors and evaluate the severity of the global errors on a 4-point scale. For instance, the script reads: "According to the bar chart, the number of cars in British was on an increase." The student should change the word "British" to "Britain." As for the global errors, they saw a list of potential global problems/errors (e.g., weak introduction, weak conclusion). They would then need to assign a number from 1 (the most severe) to 4 (the lease severe) to each of those listed global writing errors. The performance of the students in the test was compared with their tendency to commit the same errors in a mock IELTS writing test carried out immediately after the online test, as shown in Table 2.We also compared their performance with that of the other 148 students who attended one session of a two-hour long preparation workshop consisting of lectures and paper-and-pencil exercises in small groups rather than the online test.

We explored the effectiveness and challenges of the automated prediction system of the writing errors of tertiary level ESL learners. There were five research questions:

1. What are the most frequent written errors that appear in IELTS writing scripts?
2. Are there any notable patterns in the types of error and their frequencies of occurrence across the two writing questions of distinctive genres [iii]?
3. Do the coders code consistently? Is there high inter-rater reliability?
4. How effective is the system in predicting students' predisposition in committing writing errors and discourse problems?
5. How can teachers better use IOWA as a diagnostic tool to complement their teaching practice to address different issues of error categories?

3.1. Selection of Target Error Types

Based on our experience of preparing students for the academic IELTS test from 2002 until 2006 when the study started, we prepared a list of 75 error types that we considered both common and significant (in terms of adversely affecting the writer's score for the test). Examples of error types are: Inappropriate choice of verb, Faulty use of supporting data and Weak conclusion. The complete list is shown in Appendix 1.

Next, in order to make our task manageable, we narrowed this to a list of 26 target error types. To do this, we asked 147 CUHK students preparing for the test to complete a mock IELTS writing test, which was done after attending the conventional IELTS preparation workshop on writing skills. In the resulting scripts, every instance of the 75 error types was identified and coded. The error frequency was analyzed, and 26 of the most frequent error types chosen for the target list on the basis that these 26 error types comprised around 70% of all errors found. The error analysis results are also shown in Appendix 1. In order to validate the choices of our original 75-item list, we tried to see if the original list of over 9,000 errors (raw data) that we coded at the beginning of the study would fall into these 75 categories. Results showed that only five out of a total of 9,825 instances of error did not fit any of these categories; thus, the errors accounted for more than 99.949% of coverage. We divided the error types into two categories:

- 'local' errors: word- or sentence-level errors that could recur within a single script;
- 'global' errors: structural features (such as poor paragraphing) that would refer to a script as a whole.

Our initial list of 75 error types included 63 local errors and 12 global errors, in addition to an 'any other error' code. We analyzed the local and global errors separately because local errors can occur more than once in a script and, therefore, carry greater weight in the error

ranking.

Our final selection of error types on which to focus consisted of 20 local errors and 6 global errors. The 20 local errors are drawn from the 22 most frequently occurring errors; the two omitted from the list are *misspelling* (as this cannot be readily addressed with a short remedial task) and *faulty use of supporting data* (a logical or rhetorical error rather than a fault of language use). The 6 global errors are drawn from the 8 most frequently occurring errors; the two omitted are *over-complexity* (not readily addressed with a short remedial task) and *irrelevant content*. Full rankings are shown in Appendix 1. With regard to local errors, three types of error ranking were prepared for comparison:

- the total number of instances of each error type (as shown in Appendix 1);
- the number of scripts containing each error type; and
- the number of scripts containing each error type more than once.

The last of these rankings was designed to check how many single-instance errors were made, perhaps indicating lapses or slips rather than misunderstandings. In practice, we found that the rankings from all three approaches were very similar; 19 of our selected 20 local error types appeared in the top 22 of all three rankings.

The top four local error types were *singular-plural, verb tense, misspelling* and *missing article*, reflecting the typical interlanguage issues of the participating Chinese-L1 students. The top three global error types were perhaps more surprising: *poor conclusion, poor introduction* and *unsatisfactory answer to task question*; these are compositional skill issues rather than language mastery ones.

3.2. Comparing Error Types Selection with the Literature

In order to check if the error types under study are of importance to IELTS takers, we made a comparison between our list and three textbooks written by Berry (1961), Moore (2007) and Cullen (2007). Comparison of our target errors with the coverage in the books showed that all kinds of error covered in the books are within our scope of study. In addition, we included three kinds of error that we believe are crucial to IELTS writing, i.e. word order; missing connectives; and inappropriate connectives. These errors are not discussed in the three books. Also lacking in these textbooks are the structural aspects of IELTS writing as none of these books discuss global errors. In view of these problems, we include these additional error types in an attempt to conduct a more comprehensive study.

4. RESULTS & DISCUSSION

4.1. Patterns of error occurrence in relation to the genre differences

Based on the 147 scripts, we can see some types of error occurred more frequently in one task than in the other. Some examples are shown in Figure 1, in which error types near the top are more likely to occur in Task 1 (descriptive essay) while those near the bottom are more likely in Task 2 (argumentative essay). The distribution displays a pattern; for example, “sentence structure too simple” is more likely in descriptive Task 1 while pronoun errors occur more often in the argumentative Task 2. These findings could give insights into the relative weight that should be placed on various types of error when teaching writing of different genres.

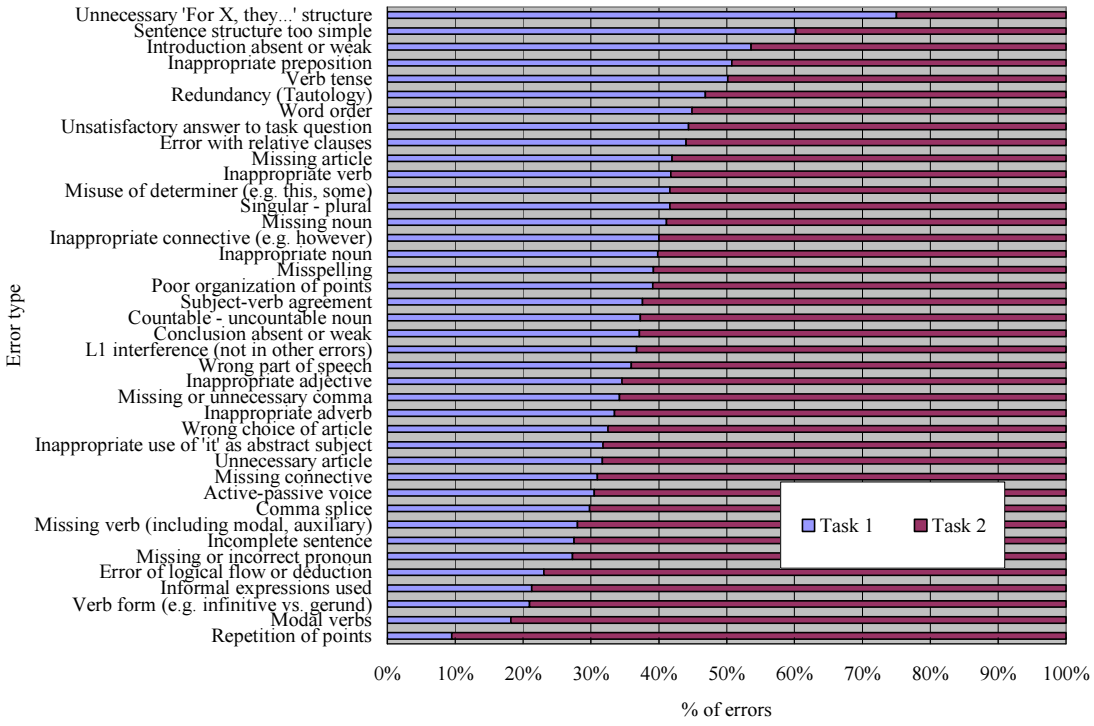

Figure 1. Distribution of selected common errors between Task 1 and Task 2 scripts

4.2. Inter-rater reliability- Checking for Error-coding Consistency

A detailed guideline for error coding was given to a team of ten teachers, who were experienced in preparing students for IELTS. Before coding, each coder was given the draft coding scheme and guideline, along with two 'standardization' scripts to code (the same scripts were issued to each coder). The resulting codings were analyzed for consistency.

We checked the coding in the first ten lines of the Task 1 and Task 2 scripts separately. The error codes used by each coder in those 20 lines were tabulated. For each error type, we counted how many coders used each error type. We also counted how many of the actual errors were found by each coder, and how many incorrect codes were used (e.g. against non-existent errors, or where the wrong code was selected to mark an error), terming these 'false positives'.

In the first ten lines of task 1, there were about seven error types, some occurring several times. Five of these types were found by more than half of the coders. Overall, 58% of the "true errors" were actually found. However, 28% of the codes used were 'false positives'. About two-thirds of the 'false positives' were error types that did not appear in the passage at all. Thus, the overall effect of the faults in coding would be to decrease the apparent occurrence of common errors, and increase that of uncommon errors—in other words, a 'levelling' effect.

The Task 2 passage studied contained about eleven actual error types. Of these, five were found by more than half of the coders. Overall, 49% of the "true" errors were found, and there were 10% false positives, nearly all of which were error types that did not appear in the passage. Thus, Task 2 seemed more difficult to code correctly, and many less common errors were under-represented in the coding (e.g. word order and transitivity errors).

Based on these findings, we revised the coding system and guidelines to coders,

removed ambiguities, and encouraged coders to find all errors in the scripts.

In order to conduct ongoing checks, each coder received 'calibration' scripts with their bundle of scripts for coding, at a ratio of one calibration script to 9 regular scripts. The coders were not aware of which scripts were for calibration. We compared the coding of the calibration scripts between coders. These showed a slight improvement in consistency compared with the earlier study; however, this may have been partly due to increased familiarity with our coding system.

Given that precisely accurate coding was not critical to the overall purpose of our study, the coding reliability achieved was reasonably adequate.

4.3. Prediction of Errors

The key objective of our project was to devise an online test that would predict which errors a particular student would be most likely to make. The criteria for the design of our predictive test were as follows:

- It should give accurate predictions, i.e. the actual errors made by a student in a mock IELTS writing test would match those predicted by the system.
- A short time (less than, say, 30 minutes) should be required for completion of the test, given our overall time constraints.
- It should be self-evident as to how to complete the test without detailed explanation given by the teacher.
- It should run as an online application in a normal web browser for maximum portability.
- It should look professional and be simple to navigate.
- It should work smoothly without ambiguous questions.
- It should give meaningful feedback.

4.4. Design of the Predictive test

In attempt to predict a student's predisposition to commit a certain error type, it is important to remember the cognitive distinction between 'production'—actual writing—and 'reaction'—responding to a test question about some aspects of writing. It is common and possible that students can be fully aware of the rules concerning, for example, subject-verb agreement, and are able to answer a related test question correctly. Nevertheless, they still commit this kind of error frequently in their own productive writing. Thus, the predictive power of a test question may well be weak, and we, therefore, elected to avoid this approach.

Logically, then, the best predictions could be expected to arise from analysis of a student's actual writing sample This, however, would require the software to have capabilities not yet developed even in state-of-the-art writing analysis systems. It may soon be possible for software to analyze grammatical errors in terms of their underlying cause—indeed, some parsers already have capabilities that take some steps towards this. However, higher-order errors relating to, for example, word choice, semantics, coherence, logical flow and structure, are likely to prove much more intractable to automated analysis.

Our chosen approach was to generate an extended text containing each of the target local errors and to have the user find and correct the errors—a 'proofreading' approach. Some global errors were also present in the text, and the user was required to assess their severity on a 4-point scale. When devising the text, we kept the following criteria in mind:

- The errors must be unambiguous—i.e., definite errors of a certain type, not open to question.
- The errors must not be made obvious from the way the text was presented. For

example, if *The* was missing from the start of a sentence, the following word would be capitalized, and the space closed up, so as to 'conceal' the error. Nor can an error be readily deducible by comparison with another point in the text; for example, in "Both man and women enjoy sport", the incorrect use of singular *man* is made obvious by the plural *women*.

- Errors must not be extended over more than 2-3 words, in order to keep them clearly defined. (An exception to this may be word order errors.)
- Errors must be discrete, i.e. each phrase must contain no more than one error, and there should be no overlap between erroneous phrases.
- Errors must have one or more clearly identifiable corrections, one of which could be to delete the erroneous word or phrase.

4.5. Implementation of the Predictive Test

Our system was written in C++ with .net framework 2.0 including JavaScript and AJAX and hosted on the CUHK e-learning server. We prepared two error-spiked texts—one for Task 1, one for Task 2—although only the Task 1 text was used for our studies. The texts are shown in Appendix 2.

The test was conducted in two parts. In the first part, our 20 targeted local errors are addressed. First, users of the IOWA were presented with the text and told that it contains 35 underlined errors. The users should correct each error (by replacing or deleting a word or phrase); their amendments appear in the text, whether correct or not.

The second part focuses on the 6 target global errors. The original text is shown again (without underlining). The 6 global errors are listed below the text, and users were asked to assess the text in terms of each global error, on a 3-point scale from 'satisfactory', 'weak' to 'very weak'; plus a 'don't know' option also available, since the user may not know what constitutes, for example, bad paragraphing.

After completing the two tests, users are offered detailed feedback. For the first part, the corrected text is shown, with the errors highlighted. Hovering the mouse over each error reveals the user's amendment, whether the amendment was correct, and all possible acceptable corrections. For the second part, the user's and 'correct' evaluations of the global errors are shown; explanations are offered via 'Help' buttons.

Finally, the system evaluates the user's performance and lists the top 10 error types in terms of significance to that particular user. The user is intended to use this list as a recommendation for further study. To carry out the evaluation, a "priority-score" is calculated for each error type: priority-score = mistake-score × priority-value.

For local errors, the mistake-score is 0 if the user amended an error correctly, 1 if the user made an incorrect amendment to the error, or 2 if the user did not attempt a correction. For global errors, the mistake score is 0 if the user selected the correct option on the scale, 1 if the user's selection was adjacent to the correct option, or 2 if the user's selection was far from the correct option or the user selected 'Don't know.' The priority-value is a list of fixed values per error (on a 1-9 scale), reflecting the significance or importance of each error type, in the view of a panel of teachers. The teachers were asked to assign a score, and an average was taken to be the priority-value. For example, 'weak introduction' was assigned a value of 7, whereas 'number incorrectly expressed' was assigned 3, reflecting the teachers' perception of how heavily these factors influence performance in the IELTS test.

When the priority scores have been calculated, they are ranked in descending order, and the top 10 are reported to the user for further study.

4.6. Evaluation of the Predictive Test

A group of 10 students were asked to complete the predictive test, and then to write a mock IELTS writing test script immediately afterwards. The students consisted of those who self-registered for our normal IELTS preparation workshop, unaware that they would be invited to attempt the predictive test. Thus, the students can be regarded as a representative sample of those who normally take our workshops. When they arrived at the workshop, all the students consented to participate in the study.

The students' performances in the predictive test were recorded by capturing screenshots at each stage of the test and analyzing their responses on an error-by-error basis. Their performance in the mock writing test was error-coded as described earlier.

To evaluate the effectiveness of the test, we prepared a grid showing each type of error (in columns) and each student (in rows). We inserted codes into the cells of the grid as follows:

Code	Meaning
W	The student made this type of error in their mock IELTS writing test.
P	The student made this type of error in the predictive test.

We then examined the correlation between occurrences of the codes W and P in the grid. Ideally, the codes would appear together in the same cells, indicating that predicted errors (P) and actual errors (W) were in agreement. In practice, we found varying degrees of correlation depending on error type. To assess this quantitatively, we calculated *correlation scores* (*r*) for each error type, where *c* ranges from +1 (perfect match between W and P) to –1 (no match between W and P). The detailed method of calculation is shown in Appendix 3. The values of *c* for each error are shown in Table 2.

4.7. Analysis of Global Errors

Naturally, it was not possible for the predictive test to contain examples of all the global errors. If it had, the test would have been ineffective because the students were required simply to evaluate the global error on a crude scale. Hence, it was not possible to analyze global error types in the same way as local errors and, thus, global errors are excluded from Table 2. However, the effectiveness of the test (and normal workshop) in reducing the occurrence of global errors could still be analyzed, and the results are presented in Figure 2.

Global errors should not be neglected, as they are likely to contribute significantly to the student's IELTS score, especially at higher band levels. Also, students tend to focus more on word- and sentence-level problems due to their relatively limited ability level.

Error code	Error description	*c*	Error category (see Table 3)
14	Inappropriate noun	0.33	A
61	Inappropriate connective	0.25	D
25	Missing word or phrase	0.20	B
13	Inappropriate verb	0.11	A
39	Redundancy	See note[IV]	E
35	Missing article	–0.10	A/C
63	Missing connective	–0.20	D
51	Comma	–0.43	D
20	Wrong part of speech	–0.50	B

17	Inappropriate preposition	–0.60	B
26	Singular-plural	–0.60	C
36	Unnecessary article	–0.60	E
29	Number incorrectly expressed	–0.67[v]	B/E
30	Verb tense	–0.67	C
56	Word order	–0.71	B
15	Inappropriate adjective	–0.78	B
27	Countable-uncountable	–1.00	C
32	Active-passive	–1.00	B
41	Subject-verb agreement	–1.00	C
45	Relative clause	–1.00	E

Table 2. Correlation scores *c* between errors in the predictive test and in students' actual writing; listed in descending order of *c*

4.8. Discussion of Evaluation Results

Examination of the results in Table 2 above shows that word choice and structural errors (e.g., punctuation) generally occur near the top of the table, showing better correlation than classic grammar errors such as subject-verb agreement and verb tenses. This suggests a general distinction between these two types of error; for the word choice errors, students are unaware of their weakness—and so commit the error both in the test and in their writing—whereas for the grammar errors, students are often well aware of the rules and so do well in the predictive test but still commit the errors in productive writing.

Based on close examination of the data, we divide the errors into five categories, as shown in Table 3. For each, we list our observations on the patterns of manifestation of this error, and a possible way the error could be more effectively predicted.

Error category (from Table 2)	**Observations from results of predictive test**	**Possible strategy for improved prediction**
A	These errors were made by most students, both in the test and in their writing.	Since these errors are so common, it can be assumed that students will commit them: no need to predict.
B	These errors were moderately common, but poorly predicted.	A different testing approach is required. These errors may tend to follow characteristic patterns (we call these recurrent errors) that can be identified; see Table 4.
C	Students generally corrected these errors appropriately in the test, but frequently committed them in their writing. This suggests high awareness of the principles involved, but failure to follow them in productive writing.	A productive writing task is required to test for these errors. Since the errors are, for the most part, well-defined issues of grammar, it may be possible to detect these errors automatically using a parser.

D	These errors were commonly made in the test, but not in students' writing. This suggests the students were either not attentive enough in looking for these types of error in the test, or not able to correct them despite the identification of the errors. However, these errors are unimportant as they do not appear in the actual writing.	A shorter proofreading task can be used, specifically focusing on these error types. The students can be informed which types of error to expect.
E	Similar to category C, except less common in students' writing.	Use a productive writing test, or make the proofreading test more difficult.

Table 3. Error categories

4.9. Effect of Predictive Test on Writing Performance

We were interested in discovering whether the process of conducting the predictive test—even without any follow-up remedial teaching—had an impact on students' performance in a mock writing test. To this end, we examined the numbers of scripts containing each of our 26 target errors, as written by the students who had just taken the predictive test described above. These were compared with the corresponding numbers from the original cohort of 147 students, all of whom had taken a 'traditional' workshop (two hour lecture style, with some individual or small group interactive activities, no computer usage) just before writing their scripts. The results are summarized in Figure 2.

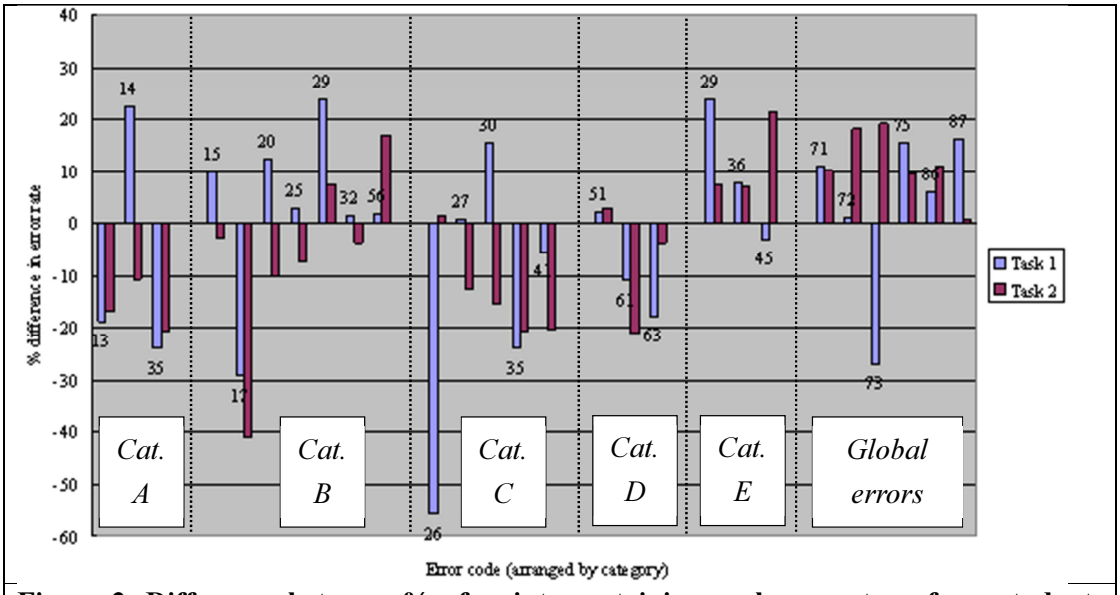

Figure 2. Difference between % of scripts containing each error type from students completing the 'traditional' workshop and those completing the predictive test. Positive values denote that scripts written after the predictive test having fewer occurrences of the error. The errors are arranged according to the categories shown in Tables 2 and 3; errors assigned to two categories appear twice in this Figure.

These results show a discernable pattern for all categories of error, except A and B. Categories C and D are errors which are poorly predicted by the test; Figure 2 shows that students who did the test instead of the normal workshop made significantly more of these types of error (as indicated by the negative difference scores in the Figure). This suggests that teacher input is important for reducing the incidence of these types of error; therefore, teachers should focus on these categories in the workshops.

On the other hand, the incidence of category E errors was lower among students who took the test. Category E errors are low-frequency errors that are poorly predicted by the test. This intriguing result suggests that the test itself is capable of increasing students' awareness of these less common errors, suggesting this is a useful teaching approach for this class of error. However, we do not know if these effects would last and result in eventual acquisition of those language structures.

Most interesting of all is the finding, shown in Figure 2, that the incidence of global errors was significantly reduced as a result of taking the test. This suggests that the process of working with a model text—albeit one with intentional weaknesses—helps the students focus on structural aspects of writing. This result has clear implications for classroom teaching of writing.

5. LIMITATIONS

5.1. Recurrent Patterns of Error

Our collection and error-coding of 147 mock IELTS writing scripts has generated a substantial corpus of authentic error examples. Many error types have hundreds of instances in this 'error bank'. In order to better understand how to predict and prevent such errors, we examined some of them in detail, aiming to identify recurrent patterns. One limitation that should be mentioned is that all the scripts were written in response to the same pair of IELTS questions, which may cause certain patterns of error to occur frequently in response to the specific subject matter of those questions. In IELTS, Question 1 is about describing some figures and patterns in one of the many different formats (e.g., pie charts, bar charts, flow charts, etc.), whereas Question 2 is pertinent to argumentative writing genre.

Error code	**Error description**	**Patterns identified (with approximate percentage of errors of that code)**
15	Inappropriate adjective	Unnatural collocation (25%), e.g., *impossible laws, reinforcing way, sustainable fossil fuels* Confused items (20%), e.g., *further (→future) generations, raising (→rising) numbers* Overly common word chosen (15%), e.g., *enough chance*
20	Part of speech	Adverbs (32%), especially writing adjective for adverb or *vice versa* Noun/adjective confusion (20%), e.g., *presence/present, Britain/British*
25	Missing word or phrase	Many different patterns observed; no single pattern particularly prominent
32	Active-passive	Misuse of passive (30%), e.g., *The number has been kept increasing; habits are difficult to be changed.*

		Misuse of active (30%), e.g., *These problems can be solving.* Malformed construction (25%), e.g., *The problem will be worsen in the future.*
36	Unnecessary article	Use of *the* for abstract or concept noun (40%), e.g.. *The public transport is greener.* Use of *the* for abstract plural nouns (30%), e.g., *Some ways are better than the others.* Use of *the* + singular noun when abstract plural noun would be better (25%), e.g., *encourage the citizen to...*
45	Relative clause	Use of unnecessary relative pronoun with contact clause (containing *-ing* or *-ed* verb) (25%), e.g., *problems that created by* Use of *that* before noun phrase (25%), e.g., *The graph shows that the number of cars in Britain.* Omission of relative pronoun (25%), e.g., *The number ^ took the test increased.*
51	Comma	The most common error is to omit commas, especially before *-ing* adverbial clauses, around non-defining subordinate clauses, and in lists.
56	Word order	Misplaced adverbials (25%), e.g., *it increased to 200 slowly* Misplaced modifiers (25%), e.g., *the number of choosing part-time education males; the percentage of CO_2 decrease* Misplaced auxiliary verb (15%), e.g., *There may be not enough space.*
72	Introduction	Failing to 'set the scene'—identify the subject area—is the most common error. Another common weakness is simply to repeat phrases from the question verbatim.

Table 4. Recurrent patterns identified in the error bank

One caveat to the results shown is that the sample of students taking the predictive test was small—only ten students completed it, compared with 147 taking the 'traditional' workshops before completing the writing test.

Another possible distortion arises from the fact that the scripts of the students taking the predictive test were coded by the project team, not by the original coding team, due to a different combination of the project members over the two-year project. The project team may have been more meticulous in finding every instance of the errors present; this would tend to increase the apparent incidence of errors among test-takers, making the difference scores in Figure 2 more negative.

It should also be emphasized that the test-takers in this study received only the limited feedback offered by the software. They did not work with the follow-up tasks in the 'task bank.' It is to be expected that working with the task bank should improve the students' performance further, especially if tasks are focused in the light of the findings reported above.

6. CONCLUSION

In this investigation, we have developed a comprehensive list of error types made by EFL students at the tertiary level in Hong Kong and rank-ordered the errors based on frequency by using a substantial sample of authentic scripts. We have also devised an online test system that attempts to predict which error types will be committed by a particular student, and we have demonstrated that:

1. Its predictive performance varies greatly depending on the type of error;
2. Testing alone—with limited feedback—is effective at reducing the incidence of certain types of error, especially low-frequency and structural errors, in students' productive writing;
3. Certain types of error benefit particularly from teacher input. Interestingly, these types of error are the ones that proved most difficult to predict under our testing system.

For teachers preparing students for the IELTS, implications are as follows. First, there should be different foci when teaching task 1 and 2 of the IELTS. Referring to Figure 1, error types near the top of the figure (e.g., *Sentence structure too simple*) should receive more attention for task 1, and error types near the bottom (e.g., *Informal expressions used*) for task 2.

Second, from a post-hoc analysis of the results of the improvement of each error type by students conducting the online test (not a control group), surprisingly, students' propensity for committing certain types of error is reduced more by teacher input than by conducting the online test. Therefore, teachers should focus on these types of error, shown as negative (downward) bars in Figure 2, e.g., *Inappropriate preposition, Inappropriate verb,* and *Missing/inappropriate connective*.

Third, students can effectively be reminded of their tendency to commit infrequent errors by performing a test. Therefore, a short test focusing on errors in Category E (as defined in Tables 2 and 3), such as *Redundancy* and *Number incorrectly expressed*, may be an efficient teaching aid for IELTS writing preparation.

Fourth, based on our data, working with error-spiked model texts helps students focus on structural aspects of writing, such as *Introduction absent or weak* and *Sentence structure too simple*. Therefore, students who are weak in these aspects may benefit from this kind of learning activity.

Fifth, analysis of recurrent instances of some errors (e.g., *Inappropriate adjective*) revealed specific topics that may benefit from teaching input (e.g., collocations), in order to reduce the incidence of these errors most effectively. Refer to Table 4 for details.

Overall, the study has laid important foundations for enhancing our students' IELTS writing test preparation. Further developments should focus on varying the approach to predictive testing according to the type of error, as detailed in Table 3. We also need to test the effectiveness of our bank of learning objects to determine their value compared with a traditional one-size-fits-all teacher-led classroom session.

Acknowledgements

The research was funded by the English Language Teaching Unit, The Chinese University of Hong Kong. We also would like to extend our gratitude to Dr. George Jor for his contribution to this project. *Correspondence concerning this article should be addressed to Kelvin, Ka Yu Chong, Department of Linguistics, University of Arizona, Tucson, AZ, 85722, or email at chongk@email.arizona.edu*

REFERENCES

Berry, T.E. (1961). The most common mistakes in English usage. Boston, MA: McGraw-Hill.

Bunton, D. (1991). A comparison of English errors made by Hong Kong students and those made by non-native learners of English internationally. ILEJ. (Special Issue No. 2), 9-22.

Burstein, J., Chodorow, M., & Leacock, C. (2004). Automated essay evaluation: The criterion online writing service. AI Magazine, 25(3), 27-36.

Chan, A. Y. W. (2004). Syntactic Transfer: Evidence from the Interlanguage of Hong Kong Chinese ESL Learners. The Modern Language Journal. 88(1), 56-73.

Chan, B. (1991). A study of errors made by F6 students in their written English with special reference to structures involving the transitive verb and the passive construction. ILEJ. (Special Issue No. 2), 43-51.

Cullen, P. (2007). Common mistakes at IELTS intermediate… and how to avoid them. Cambridge, MA: Cambridge University Press.

Corder, S. (1981). Error analysis and interlanguage. New York, NY: Oxford University Press.

Dodigovic, M. (2002). Developing writing skills with a cyber-coach. Computer Assisted Language Learning, 15(1), 9-25.

Green, C. F. (1991a). Teacher perceptions of the relative gravity of errors in written English. ILEJ. (Special Issue No.2), 69-79.

Green, C. F. (1991b). Typological transfer, discourse accent and the Chinese writer of English. Hong Kong Papers in Linguistics and Language teaching, 14, 51-63.

Hepburn, H. (1991). Teachers, textbooks and errors. ILEJ. (Special Issue No.2), 62-68.

Holdich, C. E., Chung, P. W. H., & Holdich. R. G. (2004). Improving children's written grammar and style: Revising and editing with HARRY. Computers & Education 42, 1–23.

Lawley, J. (2004). A preliminary report on a new grammar checker to help students of English as a foreign language. Arts & Humanities in Higher Education. 3(3), 331-342.

Lee, N. (1990). Notions of "error" and appropriate corrective treatment. Hong Kong Papers in Linguistics and Language Teaching, 13, 33-69.

Li, D. C. S., & Chan, A. Y. W. (1999). Helping teachers correct structural and lexical English errors. Hong Kong Journal of Applied Linguistics, 4, 79-101.

Lightbown, P., & Spada, N. (1998). How languages are learned. Oxford: Oxford University Press.

Mitchell, R., & Myles, F. (2004). Second Language Learning Theories. London: Arnold.

Moore, J. (2007). Common mistakes at IELTS advanced… and how to avoid them. Cambridge., MA: Cambridge University Press.

Selinker, L. (1972). Interlanguage. IRAL, 10(3), 219-231.

Sung, W. M. (1991). Typological transfer: a factor in the learner language of Hong Kong students? ILEJ. (Special Issue No.2), 52-61.

UGC, (2010). Results of 2009/10 Common English Proficiency Assessment Scheme Released PressRelease]. Retrieved from http://www.ugc.edu.hk/eng/ugc/publication/press/2010/pr10092010.htm

UGC, (2011). Results of 2010/11 Common English Proficiency Assessment Scheme Released [Press Release]. Retrieved from http://www.ugc.edu.hk/eng/ugc/publication/press/2011/pr03102011.htm

UGC, (2012). Results of 2011/12 Common English Proficiency Assessment Scheme Released [Press Release]. Retrieved from http://www.ugc.edu.hk/eng/ugc/publication/press/2012/pr16102012.htm

UGC, (2013). Results of 2012/13 Common English Proficiency Assessment Scheme

Released [Press Release]. Retrieved from http://www.ugc.edu.hk/eng/ugc/publication/press/2013/pr15112013.htm

Warschauer, M., & Ware, P. (2006). Automated writing evaluation: defining the classroom research agenda. Language Teaching Research, 10(2), 157-180.

Yip, V., & Matthews, S. (1991). Relative Complexity: Beyond Avoidance. CUHK Papers in Linguistics, 3, 112-124.

APPENDIX A
FREQUENCY OF ERRORS IN MOCK SCRIPTS

Tables 5 and 6 below show the total number of occurrences of all errors in descending order. The errors are divided into 'local' (word/sentence level) and 'global' (document level) errors. The 'error code' column shows our internal two-digit coding system.

Table 5. Number of occurrences of all local error types

Error code	Description	Total number of occurrences		
		Task 1	Task 2	Overall
26	Singular - plural	401	632	1033
30	Verb tense	287	240	527
11	Misspelling	178	346	524
35	Missing article	184	288	472
25	Other missing word or phrase (apart from those covered in other error types)	251	211	462
14	Inappropriate noun, i.e. a noun is needed, but the wrong one was chosen	161	281	442
17	Inappropriate preposition	218	176	394
29	Number incorrectly expressed (e.g. 1100 thousand)	360	31	391
13	Inappropriate verb (not modal verb, not phrasal verb)	129	247	376
39	Redundancy (Tautology)	159	148	307
36	Unnecessary article	66	218	284
20	Wrong part of speech (e.g. adjective vs. adverb, misuse of possessive)	105	170	275
41	Subject-verb agreement	78	167	245
79	Faulty use of supporting examples or data (e.g. too much, not enough, incorrect)	164	45	209
15	Inappropriate adjective	59	147	206
56	Word order	93	89	182
61	Inappropriate connective (e.g. *however*)	62	112	174
63	Missing connective	32	120	152
27	Countable - uncountable noun	54	94	148
45	Error with relative clauses (e.g. unnecessary, not used when needed, wrong relative pronoun)	68	73	141
32	Active-passive voice	39	94	133
51	Missing or unnecessary comma	43	88	131
54	Other punctuation error	80	38	118
42	Verb complementation: choice of bare verb, to+infinitive, gerund, past participle	14	101	115

24	Missing or incorrect pronoun	27	86	113
22	Missing noun	53	51	104
43	Inappropriate use of *it* as abstract subject of sentence	29	75	104
66	Error of logical flow or deduction	18	84	102
91	Use of superordinates when a better specific term exists (e.g. *people* instead of *students*)	38	60	98
21	Collocation error	43	50	93
81	Informal expressions used	16	76	92
38	Misuse of determiner (e.g. this/that, some)	32	52	84
92	L1 interference (where not covered by other error categories)	32	52	84
34	Mismatch between verb tense and time expression	57	25	82
23	Missing verb (including modal, auxiliary)	20	61	81
33	Modal verbs - omission, inappropriate usage, wrong choice	8	68	76
52	Comma splice	18	51	69
16	Inappropriate adverb	19	42	61
31	Verb transitivity (e.g. unnecessary preposition)	30	29	59
37	Wrong choice of article	15	40	55
49	Incomplete sentence	12	43	55
47	Faulty or missing parallelism	26	27	53
62	Unnecessary connective	19	28	47
12	Capitalization error	16	25	41
90	Unwarranted generalization	4	34	38
44	Inappropriate use of *there* as abstract subject of sentence	14	20	34
46	SVOVO structure (e.g. *The community objected to this policy was reasonable.*)	16	13	29
55	Unnecessary *For X, they...* structure	23	5	28
18	Inappropriate phrasal verb	8	18	26
68	Repetition of points	2	22	24
88	Use of clichés	8	15	23
85	Drawing attention to self	1	21	22
89	Expressing opinions too strongly	0	22	22
48	Run-on sentence	8	9	17
19	Morphology error (e.g. everyday / every day)	4	11	15
83	Addressing the reader directly ('you')	0	14	14
50	Dangling modifier	5	8	13
65	Topic sentence error (e.g. no topic sentence, multiple topics)	1	11	12
64	Awkward switch between general	4	2	6

	and specific			
82	Idiom or proverb used that is appropriate in style but inappropriate in meaning	4	2	6
84	Inappropriate use of rhetorical question	0	5	5
99	Any other error	2	3	5
28	Number - cardinal vs. ordinal	1	0	1
53	Inappropriate use of !	1	0	1

Table 6. Number of occurrences of all global error types (number of scripts shown, since these error types occur only once per script)

Error code	Description	Total number of occurrences		
		Task 1	Task 2	Overall
73	Conclusion absent or weak	34	58	92
72	Introduction absent or weak	31	27	58
75	Unsatisfactory answer to task question (e.g. not all important aspects covered)	23	29	52
87	Sentence structure too simple	24	16	40
71	Paragraphs: too long, too short or unclear	16	15	31
93	Over-complexity	15	16	31
77	Irrelevant content	9	20	29
86	Vocabulary too simple	9	16	25
74	Script too long or too short	10	14	24
67	Poor organization of points	9	14	23
78	Missing or unnecessary deduction or analysis	8	9	17
76	Opinion given inappropriately	8	5	13

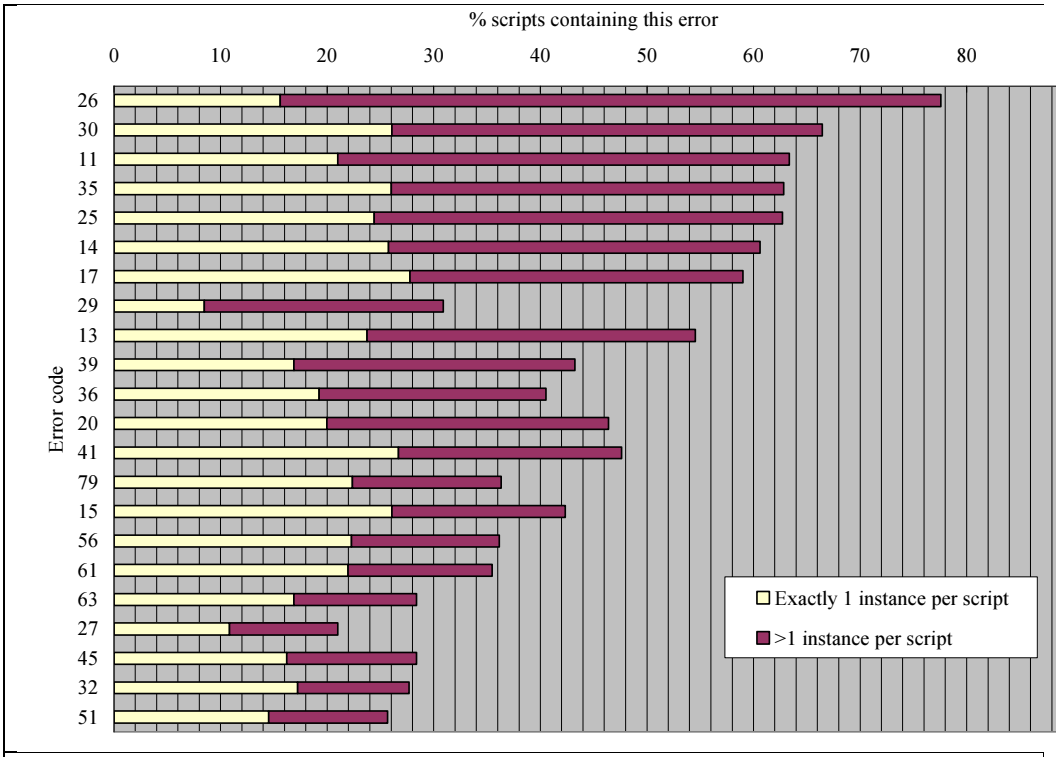

Figure 3. Local errors: Percentage of scripts in which the most common errors occur (counting Task 1 and Task 2 separately)

Figure 3 shows the top 22 error types from Table 5. For each error type, the percentage of scripts manifesting this error is shown. The percentage is subdivided into scripts containing a single instance of that error, and those containing multiple instances. The results indicate that:

- Some of these errors are extremely common. Eight types of error (26, 30, 11, 35, 25, 14, 17 and 13) occur in more than half of all scripts.
- Many of these errors are likely to occur more than once per script. This suggests the errors arise from systematic misunderstandings, rather than being mere lapses or slips.

The small percentage figure for error 29 (number incorrectly expressed) arises because this error was much more likely in Task 1, where numerical data were being described. This error appears in the top 22 because it tended to occur many times per script (4.3 times per script on average, compared with 2.9 times for the top 10 error types). Only error 26 (singular/plural) occurred more often, at 4.5 times per script.

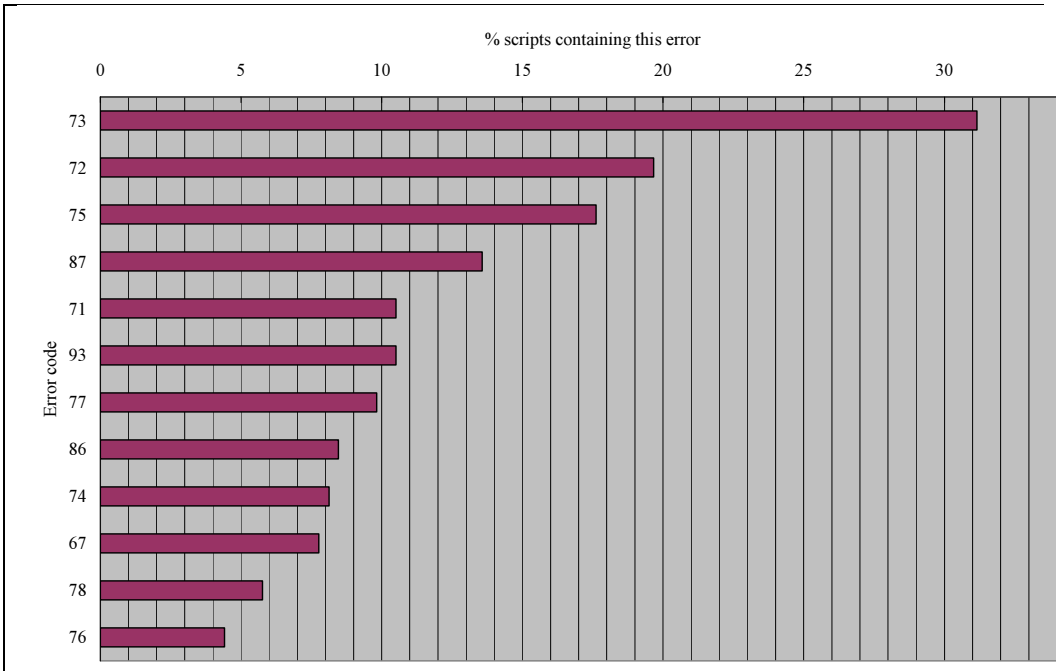

Figure 4. Global errors: Percentage of scripts in which the most common errors occur

As with the local errors, Figure 4 shows that some errors—particularly errors 73 (weak conclusion) and 72 (weak introduction)—are quite common.

APPENDIX B

ERROR-SPIKED TEXTS USED FOR THE PREDICTIVE TEST SYSTEM

Texts are shown here with each error underlined and followed by the error code (see Appendix 1) in brackets.

Task 1 Text

Nowadays, many men and women in British[20] receive further education. The number of students in the[36] further education was studied over a period of twenty years.

According to the researches[27], with regard to full time education, the number of male participants[14] was[32] increased slightly between 1970 and 1980 at[17] about 10%. From 1980 to 1990, the number of males increase[30] drastically, and the percentage of increasing is[39] about 40%. On [35] other hand, for females, [35] trend is different. From 1970 to 1980, the number of female[26] in further full time[56] education raised[13] rapidly from fifty thousands[29] to 200 thousands[29]. However, from 1980 to 1990, the number increased only slightly [25] approximately 10%.

As for part time education studies[39], the number of males slightly dropped[56] from about 900 thousands[29] to 600 thousand[29] in 1980 and 1990 respectively. On the contrary[61], the trend of[17] females which[45] enrolled in further education went the reversible[15] way.[51] the number showed large[15] growth over the years investigated.

[63] The table show[41] that in the first period, the number of males in further education were[41] higher than that of woman[26]. As time went by [51] the number of females gradually bigger than[20] that of males in[17] 1990.

Task 2 Text

In British[20], the first car appeared on roads[35] in 1888. At that time, people feel[30] very proud to own a car. But, it is surprised[15] that there are as many as 29 millions[29] vehicles on British roads 120 years later. As a result, car users do not feel any sense of proud[20] or distinctive[20]. More importantly[61], cars causes[41] many negative aspects.

These negative aspects like[13] the[36] pollution, the[36] parking area and chance of accidents. For[63] pollution, cars produce a great deals[26] of dirty gas, where[45] polluted[30] badly and heavily[56] our environment. They probably caused[30] global warming as well. Also, they produce noise pollutions[27], that[45] affects[13] people from having a silent[15] environment to study, sleep and play. On the other hand[61], as the[36] road space is limited, and the number of car[26] have increased[41] sharply, there will not have[13] enough space for parking. More seriously[61], as there are more cars in[17] the roads, the chance of car incidents[14] will probably be easier[20].

For the above disadvantages[61], we should encourage the[36] alternative forms of

transport, like bicycles,[51] and trains. Not only can bicycles provide more exercises[27] for car owners, but trains can also allow larger passengers[25] to quickly get to their places[14] at the same time. Besides[61], international laws should introduce[32] to control car ownership and use. For example, the government can suggest[13] that each family can buy only one car. Secondly[61], the government can suggest[13] when the car can use[32] on the road. For example, it might be permitted for the car to be used only three days a week. Then, at other time[26] alternative transports[27] should be used.

To conclude [51] the rapidly[20] increase in the number of vehicles are[41] undesirable. In order to protect our environment, we should do the above.

APPENDIX C

CALCULATION OF "CORRELATION SCORE" FROM PREDICTIVE TEST

For each error, we calculated a correlation score c ($-1 \leq c \leq 1$) as follows:

$$c = \frac{WP - (W'P + WP')}{WP + W'P + WP'}$$

where WP is the number of students who committed the error both in their mock IELTS writing script and in the predictive test; $W'P$ is the number who committed the error only in the predictive test; and WP' is the number who committed the error only in their writing script. A perfect correlation between W and P, meaning that the test perfectly predicts the students' actual errors, would give a score of +1; a total absence of correlation between prediction and committal of the error would give a score of −1.

Our definition of P, that is, committing an error in the predictive test, is that the student either gave no correction to the error, or gave an amendment that was not one of the predefined correct answers.

---

[i] "Graduands", different from "graduates", are those who finished their exam and coursework to receive a degree but haven't been conferred a degree yet. For a detailed definition, see the Oxford Dictionary entry: http://www.oxforddictionaries.com/us/definition/english/graduand.
For the origin of the word, see http://dictionary.reference.com/browse/graduand.
For a comparison between "graduands" and "grduates", see http://jchew01-cestlavie.blogspot.com/2010/09/graduand-and-graduate.html.

[ii] UGC press release: http://www.ugccepa.com/200708/0910result.asp retrieved on 6th October, 2010

[iii] In IELTS, Question 1 is about describing some figures and patterns in one of the many different format (e.g., pie charts, bar charts, flow charts, etc.), whereas Question 2 is pertinent to argumentative writing genre.

[iv] No data were collected for this error type, due to an oversight in devising the error-spiked text

[v] The c score for this error was distorted by the fact that only one student committed it in his/her writing—probably because the questions set for the IELTS mock writing test did not demand much use of numbers.