

How Learners Use Automated Computer-Based Feedback to Produce Revised Drafts of Essays

Jonny Laing^{a*}, Khaled El Ebyary^b, and Scott Windeatt^c

a. INTO, Newcastle University, Newcastle upon Tyne, UK

b. Department of English, Alexandria University, Alexandria, Egypt

c. School of Education, Communication & Language Sciences, Newcastle University, Newcastle upon Tyne, UK

Abstract. Our previous results suggest that the use of *Criterion*, an automatic writing evaluation (AWE) system, is particularly successful in encouraging learners to produce amended drafts of their essays, and that those amended drafts generally represent an improvement on the original submission. Our analysis of the submitted essays and the feedback provided on the first drafts suggests, however, that the students use a variety of quite different strategies when using the automated computer-based feedback to produce amended drafts. These include simply accepting a suggested correction, interpreting a feedback comment to modify the text, and avoidance strategies such as leaving out text that was highlighted as incorrect or problematic. Our data suggest that the strategies the students use are at least partly influenced by the confidence they have in the feedback, and therefore in the system itself, but may also be influenced by their interpretation of how marks are awarded by the system. This presentation will discuss the findings of an in depth analysis of the changes made in second drafts submitted to the system, linking the changes to the automatic feedback provided on the first draft, and exploring the reasons for the changes made by the students. We will suggest ways in which teachers can explore the utility of various strategies with their learners.

Keywords: reflective practice, assessment and feedback, automatic writing evaluation.

1. Introduction

One of the more difficult tasks that learners face is developing proficiency in writing, and it is generally assumed that timely and appropriate feedback is important in developing such proficiency (Black & Wiliam, 1998a, 1998b; Hyland & Hyland, 2006). There is, however, less agreement on how feedback can be most effectively targeted (on grammar, lexis or organisation/structure), on whether feedback should be explicit or

* Contact author: jonathan.laing@ncl.ac.uk

implicit, and on whether feedback is best provided by tutors, or peers, or a combination of the two. Research does not provide a clear answer to these questions, and teachers have developed a variety of pragmatic solutions, usually involving provision of at least some feedback themselves. However, this is inevitably time-consuming, especially if they attempt to provide feedback which is individualised, content-related, and timely, and if they encourage the production of multiple drafts (Grimes & Warschauer, 2010; Lee, Wong, Cheung, & Lee, 2009).

Peer feedback is a widely used technique that can help to reduce the teacher's workload by shifting the focus in feedback from just the teachers' to both the teacher's and the learners' actions and opinions (e.g., Ferris, 2003). Research suggests, however, that learners see peer feedback as serving a different purpose from instructor feedback (Jacobs, Curtis, Brain, & Huang, 1998). Whatever approach is adopted, therefore, at least some teacher feedback is likely to be desirable and probably necessary, although this becomes increasingly difficult to provide as the number of students a teacher has to deal with increases. One possible solution is to exploit advances in technology such as computer applications which are claimed to be capable not only of assessing written work, but also of generating feedback for the learners – “intelligent CALL” which can interact with the material to be learned, including (providing) meaningful feedback and guidance (Warschauer & Healey, 1998).

There is published research on the use of such applications for assessing writing (e.g., Rudner & Liang, 2002), comparing human scoring to computer scoring (e.g., Wang & Brown, 2007), and validating computerised scoring systems (e.g., Powers, Burstein, Chodorow, Fowles, & Kukich, 2001), and it is claimed that such applications match the reliability of human raters in assessing writing (e.g., Dikli, 2006). However, there is still relatively little research that has investigated the value of computer-based feedback (CBF) on students' written work (e.g., Attali, 2004; Coniam, 2009), and much of it relates to L1, or English as an additional language (EAL) rather than English as a foreign language (EFL), writers of English. This paper reports the result of using one such automatic writing evaluation system – Criterion – with four different classes of EFL students in a variety of contexts over the last four years.

2. Method

2.1. Aims and participants

This paper is based on the results of four studies that were conducted in Alexandria University, Egypt ($N = 24$), Hail University, Saudi Arabia ($N = 23$), and Newcastle University, UK ($N = 11$ and $N = 15$) between 2008 and 2012. The participants were all university students studying academic English, though learning English in order to study a variety of different subjects (the Alexandria students were training to be English teachers, for example, while the Newcastle students were planning to study a variety of other subjects at postgraduate level).

Each study had its own particular focus, but aims that were common across the studies were to investigate learners' attitudes towards the computer-based feedback they were given, the nature of the feedback that Criterion provided, and what actions learners took as a result of the feedback. The aim of this paper is to use data from the four studies to investigate the ways in which AWE systems can be used, either on their own or together with teacher feedback, the actions of the learners once they received feedback, i.e., to study the changes made in second drafts submitted to the system, the content of the automatic feedback provided on the first draft which is linked to those changes, and, where possible, the reasons for the changes that learners made.

2.2. Results

There was a very high re-submission rate for the essays, i.e., almost all learners in the studies submitted a second revised draft for each title, using feedback provided by Criterion on what it categorises as *grammar, usage, mechanics, style and organisation*. The accuracy of the feedback in these categories varied in our studies with, for example, feedback on organisation and development tending to be rather unpredictable. Comments in this category which referred to missing “thesis statements”, for example, sometimes accurately highlighted a problem but at other times simply failed to correctly identify that the essay did contain such a statement. Feedback in other categories sometimes correctly identified a problem, but not necessarily the cause (a missing auxiliary was sometimes the reason for Criterion highlighting a verb and labelling it as “ill-formed”, for example). Criterion also had difficulty – not unexpectedly – in correctly identifying where the use of, or lack of, an article was a problem.

Nevertheless, according to Criterion's own marking system, the second draft submitted by a learner was almost always better, or at least at the same level as the first, and examination of some sample essays confirmed that this did indeed appear to be the case (in some of the studies there was some teacher correction of second drafts as well as computer feedback). One possible explanation for this is that Criterion managed to correctly identify sufficient surface-level errors that the learner was able to correct and produce a second draft that was at least better in terms of those features than the first draft. A second explanation is that the simple fact of receiving feedback on a first draft encouraged the learners to reflect not only on the highlighted problems, but on other aspects of their draft, before revising and resubmitting the essay. A third explanation is that even feedback that is ambiguous or inexplicit may, by encouraging reflection, lead a learner to find a correct, or more acceptable, alternative to a highlighted problem, suggesting that learners may be able to benefit from such feedback if they already have the required linguistic resources at their disposal. An example of the latter was a student who was observed reading a Criterion comment that referred to a “fragment, or a subject or verb missing”. In fact the highlighted problem was a verb in the present simple, which should have

been in the progressive. The learner was observed to consider the comment, and the highlighted problem, at some length, eventually correctly changing “begins” to “is beginning”.

3. Discussion

Our preliminary analysis of the results suggests the following:

- Criterion proved useful in the variety of contexts in which it was tried, and especially in the situations where learners would normally be offered little, if any, teacher feedback;
- In classes with more proficient learners, where regular teacher feedback was expected, it was received positively, though with some reservations;
- It seemed to be most useful for learners at or below intermediate or upper-intermediate level;
- There was a high rate of submission of second drafts among all groups (for practical reasons learners were limited to submitting two drafts);
- Where teacher feedback was also available, learners found the process of receiving automatic feedback on drafts useful in helping them produce an improved final draft which they hoped would be well received by the teacher;
- The accuracy of the feedback provided by Criterion varied, as did its specificity, and its apparent value to the student. There is nevertheless evidence that the feedback encouraged the learners to reflect on their writing, to act on their reflections, and to produce improved drafts;
- There is also some evidence that reflection on even ambiguous feedback could result in successful correction, perhaps of “mistakes” rather than, in error analysis terminology, “errors”.

4. Conclusions

There is still much to analyse in the data, but our tentative conclusion is that Criterion appears to be most suited to EFL learners at, or below, an intermediate or upper-intermediate level. It is especially effective at encouraging learners to reflect on their writing, and to produce second drafts. Given the nature of the feedback that Criterion provides, and the focus of the feedback, it is likely to be most useful when used in conjunction with teacher feedback. Work on the first two drafts can help learners eliminate some of the surface level errors, and encourage them to evaluate the structure and organisation of their writing, allowing the teacher more time to comment on content in subsequent drafts.

There are a range of strategies that remain to be explored for combining computer and teacher feedback, including the possibility of integrating computer- and teacher-feedback with peer-feedback. In addition, although our studies were carried out as an

integral part of normal language courses, each lasted no more than a few weeks. We therefore have, as of yet, no data that would allow us to be confident that we have progressed beyond the possible influence of a novelty effect, and to investigate long term changes in attitudes towards computer-based feedback, and the long term effect on writing.

References

- Attali, Y. (2004). *Exploring the feedback and revision features of Criterion*. Paper presented at the National Council on Measurement in Education, San Diego, CA.
- Black, P., & Wiliam, D. (1998a). Assessment and classroom learning. *Assessment in Education: Principles, Policy & Practice*, 5(1), 7-47.
- Black, P., & Wiliam, D. (1998b). Inside the Black Box: Raising Standards through Classroom Assessment. *Phi Delta Kappan*, 92(1), 81-90.
- Coniam, D. (2009). Experimenting with a computer essay-scoring program based on ESL student writing scripts. *ReCALL*, 21(2), 259-279. doi: [10.1017/S0958344009000147](https://doi.org/10.1017/S0958344009000147)
- Dikli, S. (2006). An Overview of Automated Scoring of Essays. *The Journal of Technology, Learning, and Assessment (J.T.L.A)*, 5(1), 1-36.
- Ferris, D. (2003). *Response to Student Writing: Implications for second language students*. Mahwah, New Jersey: Lawrence Erlbaum Associates.
- Grimes, D., & Warschauer, M. (2010). Utility in a Fallible Tool: A Multi-Site Case Study of Automated Writing Evaluation. *JTLA*, 8(6), 1-43.
- Hyland, K., & Hyland, F. (2006). Feedback on second language students' writing: State of the Art. *Language Teaching*, 39(2), 83-101.
- Jacobs, G. M., Curtis, A., Brain, G., & Huang, S. (1998). Feedback on student writing: taking the middle path. *Journal of Second Language Writing*, 7(3), 307-317. doi: [10.1016/S1060-3743\(98\)90019-4](https://doi.org/10.1016/S1060-3743(98)90019-4)
- Lee, C., Wong, K., Cheung, W., & Lee, F. (2009). Web-based essay critiquing system and EFL students' writing: A quantitative and qualitative investigation. *Computer Assisted Language Learning*, 22(1), 57-72. doi: [10.1080/09588220802613807](https://doi.org/10.1080/09588220802613807)
- Powers, D., Burstein, J., Chodorow, M., Fowles, M., & Kukich, K. (2001). *Stumping e-rater: Challenging the validity of automated essay scoring*. Princeton, NJ: Educational Testing Service. Retrieved from www.ets.org/Media/Research/pdf/RR-01-03-Powers.pdf
- Rudner, L., & Liang, T. (2002). Automated Essay Scoring Using Bayes' Theorem. *The Journal of Technology, Learning, and Assessment (J.T.L.A)*, 1(2), 1-22.
- Wang, J., & Brown, M. (2007). Automated Essay Scoring Versus Human Scoring: A Comparative Study. *The Journal of Technology, Learning, and Assessment (J.T.L.A)*, 6(2), 1-29.
- Warschauer, M., & Healey, D. (1998). Computers and language learning: An overview. *Language Teaching*, 31(2), 57-71. doi: [10.1017/S0261444800012970](https://doi.org/10.1017/S0261444800012970)



Published by Research-publishing.net
Dublin, Ireland; Voillans, France
info@research-publishing.net

© 2012 by Research-publishing.net
Research-publishing.net is a not-for-profit association

CALL: Using, Learning, Knowing
EUROCALL Conference, Gothenburg, Sweden
22-25 August 2012, Proceedings
Edited by Linda Bradley and Sylvie Thouésny

The moral right of the authors has been asserted

All articles in this book are licensed under a Creative Commons Attribution-Noncommercial-No Derivative Works 3.0 Unported License. You are free to share, copy, distribute and transmit the work under the following conditions:

- Attribution: You must attribute the work in the manner specified by the publisher.
- Noncommercial: You may not use this work for commercial purposes.
- No Derivative Works: You may not alter, transform, or build upon this work.

Research-publishing.net has no responsibility for the persistence or accuracy of URLs for external or third-party Internet websites referred to in this publication, and does not guarantee that any content on such websites is, or will remain, accurate or appropriate. Moreover, Research-publishing.net does not take any responsibility for the content of the pages written by the authors of this book. The authors have recognised that the work described was not published before (except in the form of an abstract or as part of a published lecture, or thesis), or that it is not under consideration for publication elsewhere. While the advice and information in this book are believed to be true and accurate on the date of its going to press, neither the authors, the editors, nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher makes no warranty, expressed or implied, with respect to the material contained herein.

Trademark notice: Product or corporate names may be trademarks or registered trademarks, and are used only for identification and explanation without intent to infringe.

Typeset by Research-publishing.net
Cover design: © Raphaël Savina (raphael@savina.net)
Aquarelle reproduced with kind permission from the illustrator: © Sylvi Vigmo (sylvi.vigmo@ped.gu.se)
Fonts used are licensed under a SIL Open Font License

ISBN13: 978-1-908416-03-2 (paperback)
Print on demand (lulu.com)

British Library Cataloguing-in-Publication Data.
A cataloguing record for this book is available from the British Library.

Bibliothèque Nationale de France - Dépôt légal: décembre 2012.