

Yes/No/Maybe: A Boolean attempt at feedback

A B S T R A C T This paper describes an experiment in which Boolean feedback (a kind of checklist) was used to provide feedback on the paragraph structures of first year students in an Academic Literacy course. We begin by introducing the major problems with feedback on L2 writing and establishing why a focus on paragraph structures in particular is of importance. The experiment conducted was a two-draft assignment in which three different kinds of feedback (technique A: handwritten comments, B: consciousness-raising through generalised Boolean feedback, and C: specific Boolean feedback) were presented to three different groups of students. The results indicate that specific Boolean feedback is more effective than the other two techniques, in small part because a higher proportion of the instances of negative feedback on the first draft was corrected in the second draft (improvements), but much more substantially because in the revision a much lower number of changes to the text resulted in negative feedback on the second draft (regressions). For non-specific feedback, almost as many regressions occurred as improvements. In combination with automatic analytical techniques made possible with software, the results from this study make a case for the use of such checklists in giving feedback on student writing.

Keywords: Boolean feedback, checklist, academic literacy, analytical techniques, improvements, regressions, L2 writing

1. Introduction

The teaching of writing to second language students is a labour-intensive task. It includes large amounts of tedious, boring, and ineffective marking (Hyland, 1990, 1998, 2003; Louw, 2006; Moletsane, 2002; Spencer, 1998; Truscott, 1996, 2007). Teachers frequently find themselves in situations where there simply isn't time to provide effective feedback on all aspects deemed in need of comment. Regrettably, as shown by Louw (2006), some teachers then resort to circles, tick marks, exclamation marks, and comments like "unclear" – generally unhelpful forms of feedback, labelled "hieroglyphics".¹

¹ One anonymous reviewer objected to the use of the term "hieroglyphic" feedback, stating that many markers take great pains in writing out clear feedback and ensuring that their (cont. p108)

With large workloads and strenuous time constraints, it is understandable where hieroglyphics come from, but they do not make a meaningful contribution to the development of the student, since they are mostly ineffective. The question remains, how is it possible to provide effective, clear, usable, user-friendly feedback on student writing without devoting unreasonable amounts of time to it?

We will consider problems with feedback and possible solutions in the remainder of section 1, followed by the research method in section 2 and the results from our experiment in section 3. The merits of the solution are considered in section 4, before turning the attention to possible criticisms in section 5. A couple of conclusions are offered in section 6.

1.1 Problems with feedback

There are a number of important debates on feedback in the current literature. Most notable is the “grammar correction debate” to which Truscott (1996, 2007) and Ferris (2003, 2004) have made important contributions. The debate concerns the question whether all the painstaking grammar corrections done by teachers on student texts actually work. Truscott (2007) is adamant that, while these corrections do lead to improvement after revision, they do not lead to *learning*. Truscott notes that in subsequent assignments, students simply revert to their old error patterns.

Apart from the inability of students to turn *correction* into *learning*, there are numerous other problems identified with regards to feedback on student writing. Extending the work of Louw (2006), we want to identify the following concerns: Firstly, research has found that a focus on form is ineffective (Ellis, 1996:653; Spencer, 1998:62, 76), feedback is often not individualized (James, 1997:257), and students and lecturers find it difficult to effectively recognize recurring patterns of errors (Wible *et al.*, 2001:308-310). In addition, there is a question as to which errors carry more importance than others (Spencer, 1998; Truscott, 2007), but the research on error gravity has not been conclusive (Roberts and Cimasko, 2007:126). Moletsane (2002:27) has also found that feedback often lacks a clear purpose and teacher expectations are unclear (also compare Hyland, 1990:279; Monyaki, 2001:14-16 on the purposefulness of feedback). Feedback may confuse learners (Hyland, 1998; Hyland and Hyland, 2001; Monyaki, 2001:66, 74; Moletsane, 2002:31), learners often do not know how to use the feedback (Hyland, 2003:218) or most often, students simply do not get the chance to use the feedback since they only receive feedback on a final draft (Monyaki, 2001:63, 65), resulting in a lack of revision (Munchie, 2000:50-51; Paulus, 1999:266). Even if feedback is presented early enough, students are often unwilling to use feedback. Moletsane (2002:30), Munchie (2000:49), Spencer (1998:56, 62, 109) and Truscott (2007) found that feedback may lead to avoidance. Monyaki (2001:76), Munchie (2000:49), and Spencer (1998:73) have also found that feedback does not lead to independent learners and Truscott (2007) claims error feedback on grammar may even be detrimental. Finally, feedback is very time-consuming (Moletsane, 2002:21).

This brings us to the question what effective feedback then entails. Louw (2009) maintains that many current feedback practices are akin to editing and fall short of effective feedback

feedback does not regress into unintelligible hieroglyphics. This was also the case in this specific experiment, but for ease of reference, the term “hieroglyphic feedback” will be used to refer to handwritten comments – no doubt aided by the authors’ acute awareness of their own crablike handwriting.

for teaching and learning. From international and local research on effective and ineffective feedback practice Louw (*ibid*) distilled the following checklist for effective feedback. Effective feedback should:

1. be clear and understandable.
2. be consistent and complete and thorough.
3. be correct.
4. indicate error status.
5. aim at improvement, not just correctness.
6. provide a learning opportunity.
7. be purposeful.
8. place responsibility on the learner.
9. encourage communication and rewriting.
10. encourage language awareness.
11. be individualized.
12. be time effective.
13. be searchable/ archiveable/ recordable and allow for research.

In practice, adhering to these qualities is very difficult without technological assistance, especially in a situation where class sizes continue to grow and more demands are made of lecturers. In other words, while the qualities in the above checklist may appear self-evident (obviously feedback should be clear for example) the hard reality is that few teachers, faced with the daunting task of 80 hours of marking ahead of him/her, will have the time or energy available to adhere to these qualities all the time.

The checklist above is part of ongoing research, and some of the ideas may sound unattainable at present. Others may need a change in attitude from teachers and markers alike. However, with continual improvement and implementation, it is possible to achieve many of the above requirements for feedback. Even with small increments in effectiveness in the above categories, the overall effectiveness of feedback will improve. This article reports on one attempt at improving one part of feedback on one part of student writing.

1.2 Earlier attempts at improving the effectiveness of feedback

The technological assistance necessary to adhere to the requirements for effective feedback is being developed in the form of *MarkWrite* – a computerized marking interface developed by the Centre for Text Technology (CTeXt) at the North-West University. The aim of the software is to allow lecturers to provide partially standardised feedback on student writing in a fast and efficient way. The project was initially called *Essaymarker* and is explained in more detail in Louw (2007), although the version therein was still in its infancy, and as a result, very basic. The improved version of the essay marker software has been renamed to *MarkWrite*.

The effectiveness of the partially standardised feedback in *MarkWrite* is discussed by Louw (2006, 2007, 2008). While the data indicated it is possible to standardise at least some aspects of feedback, the areas in which standardised feedback so far turned out to be ineffective were *cohesion*, *paragraph structure*, and *argumentation*. Possible reasons include that lecturers in the experiments did not focus enough on these features, or the feedback categories provided

were not adequate. It is also possible that learners were not consciously aware of how to engage with the concepts practically, or were not aware of the qualities of good paragraphs and arguments in the first place (Louw, 2006:164; 2007).

This problem prompted further investigation of how a marker can provide fast and relatively standardised feedback on paragraph structures in a way that is as effective (or hopefully more effective) as normal marking.

A related problem to the difficulty in adhering to the qualities of effective feedback mentioned above is the consistency of utilizing commonly known techniques. There are numerous books promising to teach effective writing to students. Most of these make mention of effectively combining sentences, writing clear paragraphs, writing good introductions and conclusions, and structuring an argument. These guides are generally quite similar in what they offer, leaving us to wonder why new ones are published so often. We contend that a new guidebook will not solve the problem unless a way can be found to implement the knowledge we have already in a systematic, practical manner. This article therefore aims to demonstrate how already available knowledge can be used in a systematic way.

1.3 The focus on paragraphs

In the larger research project of which the current study forms part, the focus extends beyond the structuring of paragraphs only. We are already working on a follow-up experiment to apply the technique described in this article to complete texts, focussing on introductions, conclusions, paragraph structure, and overall textual coherence. However, in this article, we narrow our focus to paragraphs, and not coherence and argumentation, for a number of reasons.

Students find it difficult to write well structured and focused paragraphs consistently. Our own experience of working with students in class suggests that they have less trouble in writing an effective introduction and conclusion than writing a focused paragraph after instruction. This may be due to the fact that students in the study population received more training in introductions and conclusions than in paragraph structure. Colleagues at other universities have also indicated that their students have more trouble with paragraphs. It is possible that students lose focus when writing paragraphs since there are more paragraphs in a text than there are introductions and conclusions. It is therefore easier to focus attention on the structure of a short segment of text (the introduction and conclusion) than to keep sustained focus on various segments – the many individual paragraphs making up the text.

The question may be raised why introductions and conclusions are linked so closely to paragraphs in our view. An adequate answer to this requires a lengthy discussion that goes beyond the scope of this article. Suffice to say that in the school-type paragraph essay so popular in writing courses, most introductions and conclusions are actually only one paragraph in length and students are taught to have a sustained focus from their introduction to their conclusion. It is a serious problem that students find it difficult to write effective paragraphs, since paragraphs are the building blocks of any text. If a student is not able to write a focused paragraph, the student is not able to write according to the plan he or she established in the introduction. Furthermore, the basic paragraph forms the basis of the answers for many of the three and four mark questions in the student examinations. The student's ability to write

clear, focused paragraphs in the examination will assist them in communicating better, thereby obtaining better marks.

1.4 What are the qualities of effective academic paragraphs?

Having established that good paragraph writing skills are important, the next issue to consider is what exactly constitutes a good paragraph. There are many definitions for a paragraph. Based on a number of sources (Du Toit, Heese and Orr, 2002; Emory, 1995; Hannay and Mackenzie, 2002; Henning, Gravett and Van Rensburg, 2002; McClelland and Marcotte, 2003), we propose that a good paragraph displays the following characteristics:

1. The paragraph deals with only one main idea.
2. The paragraph has a single sentence, or part of a sentence, which clearly stipulates the main idea for the paragraph.
3. The main idea is supported with evidence, which may take many different forms.
4. Irrelevant information is left out of the paragraph.
5. The sentences in the paragraph should follow each other in a logical manner.

Paragraphs seldom function in isolation and should rather be assessed within the overall textual context. The following characteristics are relevant to paragraphs in context:

6. The paragraph should link up with the paragraph above and/or below it.
7. The idea in the paragraph should support the main argument of the text.
8. The paragraph should be in the right position in the text to support the logical flow of the text.

Depending on the *function* of the specific paragraph and its position, some of the proposed qualities may not be relevant. For example, if a student writes a single paragraph in answer to a question in the examination, none of the last three characteristics would be relevant. The marking scheme used for this specific experiment only had 6 questions since questions 7 and 8 are only applicable when the paragraph forms part of a larger text.

We are aware that these characteristics of a paragraph are very simplified. McClelland and Marcotte (2003), in their book on writing and grammar, explain many different types of paragraphs. For example, they touch upon a *descriptive* paragraph, a *narrative* paragraph, and an *example* paragraph. However, the basic characteristics mentioned above should be present in all good paragraphs regardless of the specific purpose of the paragraph.

Effectively marking a paragraph is more difficult than meets the eye. It seems that neither students nor lecturers are always consciously aware of all eight of the above characteristics of effective paragraphs. In this regard, Truscott (1996) indicated that language correction is often ineffective as teachers lack the skill to analyse and explain the problems which students experience, while Hyland and Hyland (2001) found that teachers' indirectness may lead to incomprehension and miscommunication.

To raise the awareness of these features for *both* the marker and the student, a marking technique should be found in which all eight of these characteristics are addressed without imposing an additional burden on an already overworked marker. The ideal technique should be quick and easy to implement, while at the same time intelligible to the student and not

regressing into hieroglyphics or “paint by numbers”. To meet to all these requirements, we propose to evaluate the feasibility of feedback based on Boolean principles. This proposal for marking paragraphs quickly, easily, and above all *clearly*, centres on a set of simple statements which could be answered by a *yes* or *no*. The marker only has to indicate whether a condition has been met or not (which is what the ones and zeros do in Boolean mathematics as well). This can be done manually using a marking grid such as shown in figure 1.

Figure 1: Example of marking grid

1	This paragraph has a sentence (or part of a sentence) that can function as the main idea for the whole paragraph.	YES	NO
2	This paragraph deals with one main idea only.	YES	NO
3	The main idea is supported with evidence in the other sentences.	YES	NO
4	This paragraph contains only relevant information.	YES	NO
5	The sentences in the paragraph follow each other in a logical manner.	YES	NO
6	The paragraph links up with the paragraph above or below it.	YES	NO

The more technically correct paragraph would have an evaluator answering “yes” to each of the questions. “No” answers are what we strive to avoid. Do keep in mind, once again, that the intention is not to use this kind of feedback in isolation from other feedback and that additional comments may obviously be necessary to clarify specific shortcomings in student texts, or to highlight certain positives.

2. Research method

A write/revision experiment was designed to test the effectiveness of the Boolean feedback. The aim of the experiment was to test whether a set of statements highlighting certain features of paragraphs could be used effectively to provide feedback on student writing.

2.1 Study population

The population in this quasi-naturalistic experiment was three groups of first-year students taking the compulsory course, *Introduction to Academic Literacy* (AGLE 111) at the North-West University, Potchefstroom Campus, in the first semester of 2009. Students were divided into three groups, based on the class they attended. The classes were divided alphabetically without reference to academic performance. The experiment was conducted early in their first year, before they have received any formal instruction in effective writing apart from what they have been taught at school.

2.2 Design of the experiment

Students in all three groups were given the same assignment. They were instructed to write two paragraphs on a particular topic. One paragraph should argue for a specific topic, and the other should argue against it. (The instructions and topics were intentionally relatively vague in order for students to pick a topic they knew something about).

Two additional instructions were added to highlight the focus on paragraph structure. Firstly, students were told to underline the main idea in every paragraph in order that the markers could establish whether the students’ impression of their main idea and the actual focus of the paragraph correlated. Secondly, they had to use bold typeface to emphasize the connectives

they had used. This was designed to check if students understood the concept of connecting devices and if they used them correctly in their text production.

The paragraphs were marked in three different ways by the lecturers:

- Group A assignments were marked in the conventional way using normal handwritten comments and symbols (hieroglyphics).
- Group B assignments were marked with a single marking grid (see figure 1) that was stapled to the assignment. No attempt was made to indicate the grammar errors, spelling errors, or other surface level errors (generalised consciousness raising).
- Group C assignments were marked using one marking grid per paragraph, stapled to each paragraph. No attempt was made to indicate grammar errors, spelling errors, or other surface level errors (specific Boolean feedback).

After giving feedback using one of the three techniques just listed, the assignments were returned to students in class. General feedback was also provided in class on the first draft (the two paragraphs), after which students were asked to revise the original two paragraphs based on the feedback they received and resubmit it with their original two paragraphs as a single assignment. Figure 2 below is an example of a typical student text after revision.

Figure 2: A typical student text

<p><u>ORIGINAL</u></p> <p>Paragraph 1</p> <ol style="list-style-type: none">1. Even though traffic officers ensures that safety is maintained in our roads.2. Yet, there is still high numbers of road accidents in our country.3. Probably is due to unlicensed drivers; hence we have these <u>roads crises</u>.4. Furthermore vehicles need to be checked if they are roadworthy or not. <p>Paragraph 2</p> <ol style="list-style-type: none">1. Some traffic officers take bribery from road rules offenders.2. As a result; government can not reach its goal of maintaining <u>safety in the roads</u>.3. Surprisingly; it could be licensed drivers who are reckless on the roads.4. Moreover roads needs to be maintained in good condition. <p><u>REVISED</u></p> <p>Paragraph 1</p> <ol style="list-style-type: none">1. Even though traffic officers ensure that safety is maintained in our roads.2. Yet; there is still high numbers of roads accidents in our roads.3. This result from unlicensed drivers who use the roads unlawful.4. In addition vehicles conditions need to be checked for roadworthiness. <p>Paragraph 2</p> <ol style="list-style-type: none">1. Traffic officers need to be strict on road rules offenders.2. As a result government will reach its; goal of maintaining safety in the roads.3. Furthermore it is everyone's responsibility to obey road rules and signs.4. Government should also ensure that roads are maintained in good conditions.5. This will make our country safe in terms of transportation on roads.

The student did not follow all the assignment instructions, since no topic sentence was underlined. It should be evident from the example that this student in particular failed to correct many of his/her language errors between the two submissions and was still unsure about the use of cohesive devices. Keep in mind however, that the purpose of the exercise was not to focus on grammatical correctness or cohesive devices, but on the argument structure of the paragraph. Providing standardized feedback on cohesive devices is the topic of an additional experiment.

The general impression amongst the lecturers was that students submitted improved paragraphs after revision. This concurs with the finding of Cho (2003), that students improve their writing simply by following a process approach. However, we found that students were still unsure about the specific use and implementation of the eight characteristics of an effective paragraph. To determine if feedback of any sort had a significant effect on the quality of the revised versions, and specifically whether Boolean feedback lead to more improvement in the revised versions than the alternative, the data from the experiment was submitted to statistical analysis.

2.3 Measuring improvement

Four independent markers (not the three lecturers whose students were exposed to the three different feedback techniques), were asked to use the six-question Boolean feedback checklist (Figure 1 above) to give feedback on a selection of original and revised paragraphs from the classes. By comparing their feedback on the original paragraphs to the revised ones, it is possible to determine if feedback in general has beneficial effects on the students' work. In addition, it is also possible to compare the relative effectiveness of the three different feedback methods used in class.

The original pairs of student paragraphs, as well as the revised pairs of paragraphs, were retyped and completely randomized. Markers did not know if they were marking an original pair or a revised pair, or which technique, A, B, or C, was used in the first place to mark the pairs of paragraphs. The complete data set consisted of 45 original sets of paragraphs and 45 revised sets. Nine of the original assignments received type A feedback while two groups of 18 received types B or C feedback. Because each essay was marked 4 times by four different markers, 360 different responses were collected, and in each response, 6 different questions were answered.

Markers ticked YES or NO on the feedback checklists on each of the six questions. All the responses were entered as 1 for YES and 0 for NO in an Excel spreadsheet. The outcomes were classified into one of four possible classifications, based on the responses by the markers, as set out in Table 1.

One response by one marker on one essay pair (from feedback technique A) had to be discarded, leaving a total of 2154 classifications that was submitted to statistical analysis.

Two null hypotheses, with alternative hypotheses complementing them, were formulated for the analysis, dealing respectively with the general possibility of improvement after feedback and with the relative effectiveness of the individual feedback techniques.

H10: The general null hypothesis is that feedback does not lead to improvement after revision.

H1A: The alternative hypothesis is that feedback leads to improvement after revision.

Table 1: Classification of data

Feedback on original version	Feedback on revised version	Classification
0 (NO)	0 (NO)	No improvement: the feedback did not help the student to improve.
0 (NO)	1 (YES)	Improvement: the revised version shows improvement in respect of the original.
1 (YES)	0 (NO)	Regression: the student had had a particular aspect right in the original, but after revision, this was changed in such a way that it was poorer.
1 (YES)	1 (YES)	Maintained: the student had had something and maintained that in the revised version.

H20: The specific null hypothesis is that Boolean feedback (Feedback technique C) does not lead to more improvement than the other two techniques after revision.

H2A: The alternative hypothesis is that Boolean feedback (Feedback technique C) leads to more improvement than the other two techniques after revision.

The first hypothesis is evaluated by comparing the number of 1-responses in the revised versions to the number of 1-responses in the original versions for the three feedback techniques. This was done by conducting a t-test for dependent samples (a paired different test) on the total (out of 6) for each pair of paragraphs (original and revised), and setting the confidence interval to 95%. A significant improvement on the number of 1-responses per paragraph will indicate that feedback has worked (for a particular technique). The statistical procedure of a t-test on dependent samples is somewhat more sensitive than a more typical t-test on independent samples. In the case of our data, this is justified, because the revised versions are indeed dependent on the originals, and hence any movement upwards from the number of 1-scores on the original versions must be detected.

The second hypothesis is evaluated by computing the X^2 statistic for the distribution of the four response types in Table 1. Assuming a 95% confidence level, with 6 degrees of freedom (3 marking techniques and 4 classifications, thus $df=(3-1) \times (4-1)=6$), the critical value of X^2 for H_{20} rejecting is 12,59. If, in a 3x4 contingency table of all classifications for the feedback techniques, there is not only compelling evidence that the data is not distributed similarly across the three feedback techniques, but also that improvements for feedback technique C exceeds its expected value, while regressions for feedback technique C is lower than its expected value, we will have found support for H_{2A} .²

3. Results

3.1 Hypothesis 1: Effectiveness of feedback

There was a significant improvement in the number of YES (or 1) scores per paragraph for the data set in its entirety, as well as for feedback technique C, the specific Boolean feedback, but

² All statistical information is taken from McClave and Sincich (2000), specifically section 9.2 for the assumptions of a paired difference experiment and section 13.2 on contingency tables and the X^2 test statistic. Computations of t-tests were done in *Statistica*, while the X^2 statistic calculations were done by hand, using the procedures for computing expected values and the X^2 itself set out by McClave and Sincich (2000:721-726).

not for the other two feedback techniques individually. The summary of the statistical analysis is presented in Table 2.

Table 2: Differences in Mean number of YES-scores for original and revised paragraphs per feedback technique

	Mean	Std.Dv.	N	Diff.	Std.Dv. Diff.	t	df	p
Original all data	4.02	2.02						
Revised all data	4.27	1.84	359	0.25	1.85	-2.59	358	0.01
Original technique A	3.48	2.24						
Revised technique A	3.72	2.07	71	0.24	2.25	-0.90	70	0.37
Original technique B	3.94	1.98						
Revised technique B	4.12	1.86	144	0.17	1.90	-1.09	143	0.28
Original technique C	4.36	1.88						
Revised technique C	4.70	1.59	144	0.34	1.57	-2.59	143	0.01

The basic results from the overall comparison of paragraphs before and after receiving feedback are presented in Table 2. The column Diff. represents the difference in the before and after score, which shows that the assessment of an essay on the same criterion improved after receiving feedback. The scores represent the total number of YES-scores on a question, with a maximum of 6 and minimum of 0.

The overall improvement after feedback across the techniques is 0.25, which means that, on average, in one in every 4 paragraphs, a student showed a net gain one YES for the paragraph. Given that each student submitted two paragraphs, this translated into a new improvement of one YES response for every second student. A closer look at the three separate feedback techniques show that the specific Boolean feedback, technique C, was the most effective in helping students along, such that one in every three paragraphs showed a benefit from this type of feedback.

Statistical significance is obtained for the overall data set, as well as for feedback technique C, but not for techniques A (hieroglyphics) or B (generalised Boolean feedback not tied to specific paragraphs). Thus, in strict statistical terms, we find evidence to reject $H1_0$, and by implication support $H1_A$. It is, of course, a reasonable question whether improvement on one in every three or one in every four paragraphs is substantial enough to warrant further investment and possible implementation of the feedback technique. Two possible arguments can be offered to answer this question in the affirmative, and therefore offer further support for $H1_A$. If feedback is expected by students and is going to be provided in any case, as was pointed out in the introduction, then one may as well adopt that method that has the best possible yield, even if the difference is a small one. Furthermore, the averages presented in Table 2 hide an even more important aspect of the feedback, viz. the fact that all three feedback techniques are relatively effective in helping along students when revision, if measured as Improvements (as defined in Table 1). However, when Regressions are considered, it becomes clear that feedback technique C is much better than the other two (see Table 3 below). The actual gain of feedback technique C is therefore not so much the possibility of improvement, but the much lower probability of regression, as will be shown in the evaluation of hypothesis 2.

3.2 Hypothesis 2: Relative merit of individual feedback techniques

The students who submitted their assignments received feedback given by means of either technique A, B or C. They took the assignments home and revised them, after which they resubmitted the final assignment for marks. In the previous section, improvement, as measured not by marks but by the scores on the feedback grid, has been assessed in terms of global improvement in the score. It emerges that there was rather moderate improvement for all three techniques, but this was statistically significant only in the case of feedback technique C, the Boolean feedback on specific paragraphs. To understand how this technique produced a different outcome from the other two, it is necessary to consider the data in a more nuanced way. Table 3 represents the data in terms of the four-way classification presented in the methodology section (specifically Table 1).

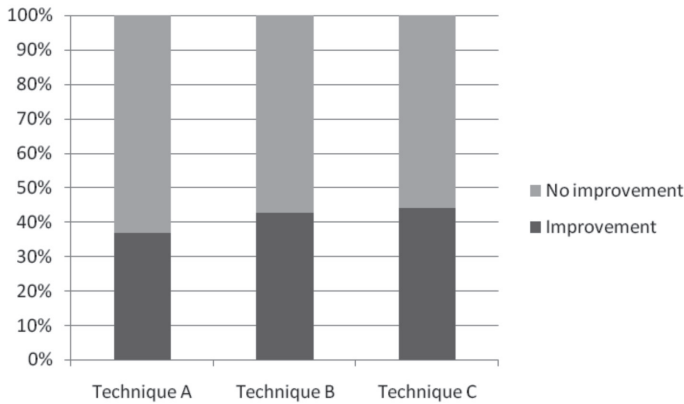
Table 3. Classification of individual responses per marking technique

	Improvement	No improvement	Unchanged	Regression	TOTAL
Technique A	66	113	198	49	426
Technique B	126	169	468	101	864
Technique C	104	132	573	55	864

Before examining the effect that the different techniques had on possible changes in the revised versions, it has to be noted that the original essays were not equally well written by the three groups. Those students in the class that received feedback with technique C had written better originals in the first place, while those in the class that received feedback with technique A had written poorer originals. Using a X^2 distribution, the uneven distribution is statistically significant ($\chi^2=28,8$, $df=2$, $p<0.05$). This is not such a serious obstacle, since the statistical analysis of the data simply factors the original distribution into the equation. For each essay, every original 0 (ticked as NO on the grid) represents an opportunity to either improve (returning a score of 1 on the revised version), or not to improve, while every original 1 (ticked as YES on the grid) represents the risk to regress back to a 0 upon revision or to maintain the performance unchanged. The statistical analysis from here on does not consider the scores in terms of global successes (1 scores), but very specifically considers improvement and regression in their own terms. The null hypothesis simply means that there is an equal probability of improvement and regression respectively for each feedback technique, judged in terms of the baseline established by the originals for each group.

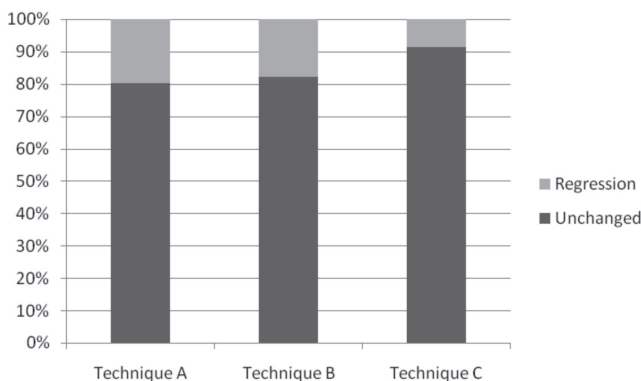
Given this background, the results on improvement versus no improvement do not show any statistically significant difference. The value of the X^2 -statistic on improvement versus no-improvement data is 2,4, which is below the 95% confidence limit of $X^2=6,0$ for two degrees of freedom. This means that students receiving feedback improved in almost equal measure irrespective of the type of feedback they received. Another way to visualise the improvement is presented in Figure 3, where the degree of improvement (NO-ticks on the feedback grid of the original paragraphs that become YES ticks in the revised version) is expressed as a percentage of the total number of NO-ticks on the feedback grids of the original versions. Feedback technique C (specific Boolean feedback) did lead, very marginally, to more improvement than B and both C and B were somewhat more effective in prompting improvement than A, but below the level of statistical significance.

Figure 3. Percentage improvement per feedback technique



Given that the difference between the feedback techniques is not to be located in the potential to prompt improvement, we had to look elsewhere, and indeed found that the real difference between the techniques was with the extent to which regression was observed. Feedback technique C, which prompted 104 improvements from the original version to the revised version, also, unfortunately, lead to 55 regressions, where an original YES-tick was changed to a NO-tick in the revised version. Such regression is dwarfed, however, if one compares the corresponding numbers for feedback techniques B and A, where the regressions almost completely cancel out the improvements: 101 regressions against 126 improvements for feedback technique B, and 49 regressions against 66 improvements for feedback technique A (bearing in mind that there were twice as many essays receiving feedback techniques C and B than A). Statistically, the differences in the distribution is significant ($X^2=27,6$, $df=2$, $p<0.05$). The comparison is visualised in Figure 4.

Figure 4. Percentage regression per feedback technique



The second null hypothesis can be rejected, because the three feedback techniques did not make similar contributions to the improvement from the original to revised versions. Feedback techniques A and B were good for improvement of respectively 37% and 43% of the original NO-scores on the feedback grids (Figure 3), but at the same time, most of the improvements were cancelled out by regressions from original YES-scores to No-scores in the

revised versions, leaving a net improvement of around 3% of the total number of YES-scores for both techniques. By contrast, and in support of the second alternative hypothesis, feedback technique C prompted improvement of 44% of the original NO-scores, and with significantly fewer instances of regression, the net improvement was about 5%. The reason for the better effect of feedback technique C is therefore fewer regressions, rather than more improvements.

4. The effectiveness of specific Boolean feedback

The purpose of the experiment in this study was to test a technique to provide feedback on paragraph effectiveness (which is often neglected), and to do so with a technique which is the optimal compromise between speed, clarity and efficiency. The effectiveness of this feedback technique can be measured using the qualities for effective feedback as explained in Louw (2009).

One may argue that some of the qualities of feedback as explained in Table 4 (page 120), carry more weight than others and the mark allocation system should therefore be adapted. This could be a valid argument, but even if the mark system changes, Boolean feedback will still score well enough to qualify as effective feedback. It might even score better. While simplistic, the above score sheet is rather generous towards handwritten marking. As mentioned in the introduction, numerous research articles have indicated clarity, correctness, consistency, etc. as definitive problems in the provision of feedback, so in many instances a mark of 0.5 might just as well be a full zero, or a full one. This kind of scoring is situation dependent, which is exactly the point – using Boolean feedback will ensure greater consistency regardless of the specific marker, excepting point 3 above. The effectiveness of the proposed Boolean feedback technique will improve even more once it is implemented within the broader framework of the *MarkWrite* interface. By contrast, the more detailed, specific and helpful the handwritten feedback becomes, the more time consuming it becomes as well.

A system that is so simple and easy to implement begs the question, “Why does it work?” It seems that by systematically and constantly reminding students what to focus on will increase their awareness of the desired outcomes. Research on the continual use of spelling checkers by students has indicated that constantly reminding students of how to correct their language, can lead to improvements (Potter and Fuller, 2008). One can therefore anticipate that the same will hold true for paragraph structure, but this still needs to be tested. The Potter and Fuller findings do create the positive expectation that Boolean feedback will have longer lasting influence if implemented consistently although Truscott and Yi-ping Hsu’s (2008) findings cast doubt on that. In our experiment, however, the improvement was over the short-term in a once-off situation, similar to Truscott’s (2008) findings.

However, Chamberlain, Button, Dison, Granville and Delmont (2004) found that it is indeed possible to stimulate higher order thinking in students by making use of “short answer questions” during testing. One can speculate that these short statements therefore fulfil the same function.

The idea with this technique is not to use it only when students are aware they are being tested on their paragraph structure, but to use it as part of the larger standardized marking system and to use it in writing across the curriculum. When any text is being marked in any discipline, the computer or lecturer will randomly select two or three paragraphs on which the marker

Table 4: *The Effectiveness of specific Boolean feedback*

Feedback should	Radio buttons	Points	Handwritten comments	Points
1. be clear and understandable;	Yes (some refinement and teaching may be necessary)	1	This depends on the marker.	0.5
2. be consistent and complete and thorough;	It is consistent. Since it only focuses on one aspect of the paragraph, it cannot be expected to be complete and thorough.	1	This depends on the marker. Consistency will be a problem	0.5
3. be correct;	This depends on the marker.	1	This depends on the marker.	
4. indicate error status;	The technique does not indicate error status just yet, but with additional research it will be possible to identify which of the characteristics of an effective paragraph, are more important than others.	0.5		0.5
5. aim at improvement, not just correctness;	Yes	1	This depends on the marker.	0.5
6. be a learning opportunity;	Yes	1	This depends on the marker.	0.5
7. be purposeful;	Yes	1	This depends on the marker.	0.5
8. place responsibility on the learner;	Yes	1	This depends on the marker.	0.5
9. encourage communication and rewriting;	The technique encourages rewriting. Communication between lecturers and students is however, dependent on more variables.	0.5	This depends on the marker.	0.5
10. encourage language awareness;	No. Since the technique focuses on one aspect of the text only (paragraph structure), it does not touch upon language awareness except for influencing the use of cohesive devices. This is why the technique should not be used in isolation.	0	No. Feedback on the structure of a paragraph will most probably not refer to language issues; if it does, that will be a different feedback category.	0

is asked to answer the questions. The purpose with this is to remind students constantly to be clear and structured in their writing.

The effectiveness of these feedback statements therefore still needs to be tested on paragraphs randomly taken from larger pieces where students are not aware that they should be focusing on paragraph structure and cohesive devices. The long term effectiveness of this technique needs to be established especially in situations where students do not have the liberty to rewrite

11. be individualized;	No, although individualized comments may be added if necessary.	0	Yes	1
12. be time effective; and	Yes, but with caveats: the idea is not to comment on the structure of <i>all</i> paragraphs in the text, unless that was the focus of the specific assignment. Markers in the experiment also indicated that, once they knew the statements, it was a quick way to mark.	1	This will depend on the amount of feedback provided by the lecturer. To provide feedback in similar amounts as is possible by using the Boolean feedback, will not be time effective.	0.5
13. be searchable/ archiveable/ recordable and allow for research	Not if done by hand. If done on computer, yes.	1	No.	0
TOTAL		10/13 = 77%		5.5/13 = 42%

their papers (as is the case in most situations), but as the Potter and Fuller (2008) experience shows, continual reminders may eventually have positive effects. Lee (2002:1) also found that her explicit teaching of coherence structures directed the learners' attention "to the discourse level of the texts while revising".

The short-term improvements evident in this experiment could be explained by scrutinizing the checklist for effective feedback again. Of the 13 characteristics, 11 focus on how the feedback assists the learner, and on these criteria, the technique is judged effective on 77%. As far as standardized, written feedback goes, this is quite good although it still will not compare well with other, more labour intensive and time intensive techniques such as structured, personal interviews.

Since standardised feedback is more effective than hieroglyphics, the question is rather why is it that technique C (a Boolean grid for every single paragraph) is more effective than technique B (a single Boolean grid providing a general impression).

The fact that marking technique C was the most effective can indicate something of value: simply knowing which features to look for in a text does not assist learners as much as having these features directly linked to a specific part of the text in a consistent manner. This provides support for Spencer's (1998:88-90) finding that students want all errors to be indicated and recurring errors should not just be indicated the first time they appear. Moletsane (2002:32-33) also warns that it can cause confusion if an error is marked in one place and not in another and Ellis (1996:585) and Nwaila (1996:83) warn against indicating the same error in two different ways. This is demonstrated most spectacularly by the fact that the non-specific feedback of technique B prompted almost as many regressions (percentage-wise) as the hieroglyphic feedback technique A. If the feedback is not specific, students seem to look for areas where they can improve along the lines of the advice, but may actually change an aspect that was acceptable and overlook an area more in need of correction.

5. Possible criticism

We are aware that various kinds of criticism can be levelled against the solution we propose here. Within the broader context of the *MarkWrite* system, the idea of fully automated marking may be regarded as old hat and little new work can be done in this area. Researchers have already tried using the comment function in word processing software, macros have been tried, and fully automatic feedback systems like the Criterion and E-Rater services of ETS are available in the market (Chodorow and Burstein, 2004; Chodorow and Burstein, 2001).

Macros and multiple choice types of feedback are similar to the technique we propose, but the difference here is the scale of implementation and the level of standardization. Anybody can create a list of questions (even a well-researched list) and use the answers on them to provide feedback. However, if this feedback is not implemented in a system aimed at providing feedback holistically and systematically, on multiple areas of a student's text, it does not answer to all the requirements of effective feedback as developed by Louw (2009). Furthermore, one of the aims of the *MarkWrite* project is to gather large amounts of student writing data which can be used for the creation of a partially annotated corpus for further research.

With regards to fully automatic feedback, there are multiple problems. The first problem is students' immediate distrust of fully automatic feedback (see Spencer and Louw, 2008). The second is that fully automatic feedback (while getting better all the time) is still not accurate enough. While human markers are not infallible either, the Boolean feedback technique we propose, if incorporated into a computerized marking support interface and linked with assessment assistance, can overcome the limitations on both fronts. It represents an attempt to find the intercept point between fully automatic and fully manual, much like the manumatic transmission (also called Tiptronic transmission) in cars, allowing for improved performance, without the loss of user control.

The third and fourth problems with fully automatic feedback become apparent when keeping in mind the intended implementation of this specific technique within the bigger context of *MarkWrite*. As mentioned above, any marker, in any discipline should be able to comment on a student's paragraphing in any text in order for the system to facilitate feedback in a writing-across-the-curriculum situation. Fully automatic marking systems make use of textual comparison techniques within a specific discipline and genre, severely limiting their usability across the curriculum. It will, however, be possible to use the user generated data from this system to train computer systems in future with the intended application of automating more of the system. In addition to that, *MarkWrite* is not intended to be only an English marking system. The technology and techniques in *MarkWrite* can be used (with some adaptation) for other languages and subjects other than language subjects within the South African context.

A further possible criticism, raised by reviewers and members of the audience at SAALA 2009, is the question whether the implementation of such a technique will not regress into a "write by numbers" (read: "prescriptive") recipe for student writing. The answer (in an ironic un-Boolean way) is both yes and no. In the first place, in writing (as with cooking) adhering to a certain recipe does not necessarily dampen creativity or personal interpretation. However, it is important to follow general guidelines, which is what these statements are. Choosing to adhere

to the statements will render the writing better, but it is still up to the personal interpretation of the writer how this will be done. A paragraph which received only YES-ticks on the feedback grid, can still be improved upon.

Secondly, the statements need not necessarily be followed as recipe during the initial drafting and writing stages, but will actually function better during the editing of the texts. Most of the writing guides mentioned above contain hints on editing after the initial freewriting or brainstorming sessions have been completed, but such editing guidelines are seldom taken as targets for the accusation of prescriptivism.

In the third place, write by numbers can save time and can be of great assistance, especially to the weaker students. Radecki and Swales (1998) have found that as students become better, they assign a more restricted role to the language teacher in their writing, but the weaker students need more help. This may happen with this technique as well.

6. Conclusion

The experiment showed that it is possible to improve on paragraph structure feedback by standardising it to an extent, without placing an additional burden on the marker. All three feedback techniques were effective in prompting improvement upon revision, and the advantage of feedback technique C is not significantly more than either hieroglyphics or generalised Boolean feedback on the entire assignment rather than individual paragraphs separately. However, an unforeseen risk of feedback is that, upon prompting the student to revise an assignment, they may actually change aspects that were relatively acceptable into less acceptable formulations. Such regression is significantly less with the specific feedback given through standardised YES/NO-questions, as proposed in this article. In the trade-off between improvement with fewer regressions, the specific Boolean feedback of technique C had a statistically significant advantage over the two alternatives considered in this article.

The results of the experiment should not be seen in isolation. It forms part of a bigger project aiming to provide *more standardised, more effective, faster, more user-friendly feedback* on student writing. It is also hoped that using such a technique will counter the problem of students simply focusing on their surface errors during revision, as was found by Kasanga (2001).

Further research is dependent on the implementation of the system. It includes the possibility of using the information from the radio buttons to establish students' general level of paragraph awareness and to provide assessment assistance to lecturers based on the pattern of yes/no answers.

REFERENCES

- Chamberlain, C., Button, A., Dison, L., Granville, S. & Delmont, E. 2004. The role of short answer questions in developing higher order thinking. *Per Linguam*, 20(2): 28-45.
- Cho, Y. 2003. Assessing Writing: are we bound by only one method? *Assessing Writing*, 8: 165-191
- Chodorow, M. & Burstein, J. 2004. Beyond Essay Length: Evaluating e-rater®'s Performance on TOEFL® Essays. *TOEFL REPORT*, 73. February.

- Chodorow, M. & Burstein, J. 2001. Enriching automated essay scoring using discourse marking. [Available Online]: http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/19/5d/82.pdf. [Date of access]: 2009/10/19.
- Du Toit, P., Heese, M. & Orr, M. 2002. *Practical guide to reading, thinking and writing skills*. Cape Town: Oxford University Press.
- Ellis, R. 1996. *The study of second language acquisition*. Oxford University Press.
- Emory, D. 1995. *Improve your essays*. New York: Glenco/McGraw-Hill.
- Ferris, D. 2003. Responding to writing. In: Kroll, B. (ed.). *Exploring the dynamics of second language writing*. Cambridge: Cambridge University Press.
- Ferris, D. 2004. The "Grammar Correction" Debate in L2 Writing: Where are we, and where do we go from here? (and what do we do in the meantime . . .?). *Journal of Second Language Writing*, 13: 49–62.
- Hannay, M. & Mackenzie, J.L. 2002. *Effective writing in English – a sourcebook*. Bussum: Uitgeverij Coutinho.
- Henning, E., Gravett, S. & Van Rensburg, W. 2002. *Finding your way in academic writing*. Pretoria: Van Schaik.
- Hyland, K. 1990. Providing productive feedback. *ELT Journal*, 44(4): 279-285.
- Hyland, F. 1998. The impact of teacher written feedback on individual writers. *Journal of Second Language Writing*, 7(3): 255-286.
- Hyland, F. 2003. Focusing on form: student engagement with teacher feedback. *System*, 31: 217-230.
- Hyland, F & Hyland, K. 2001. Sugaring the pill – praise and criticism in written feedback. *Journal of Second Language Writing*, 10: 185-212
- James, C. 1997. *Errors in language learning and use – exploring error analysis*. Edinburgh Gate: Longman.
- Kasanga, L.A. 2001. Students' response to peer and teacher feedback in a first-year writing course. *Journal for language teaching*, 38(1): 64-99.
- Lee, I. 2002. Teaching coherence to ESL students: a classroom inquiry. *Journal of Second Language Writing*, 11: 135-159.
- Louw, H. 2006. Standardising written feedback on L2 student writing. Unpublished Masters dissertation. Potchefstroom: North-West University.
- Louw, H. 2007. Moving to more than editing: standardised feedback in practice. *Ensovoort*, 11(2): 83-104.
- Louw, H. 2008. The effectiveness of standardised feedback when L2 students revise writing. *Language Matters* 39(1): 88-109.
- Louw, H. 2009. Moving to more than editing: a checklist for effective feedback. *Journal for Language Teaching*, 43(2): 86-101.
- McClave, J.T. & Sincich, T. 2000. *Statistics* (Eighth edition). Upper Saddle River, N.J.: Prentice Hall.
- McClelland, L.D. & Marcotte, P.H. 2003. *Writing matters! Introduction to writing and grammar*. New York: McGraw-Hill.
- Moletsane, J.R. 2002. Selective error correction in ESL narrative compositions. Unpublished MA mini-dissertation. Potchefstroom: North-West University.
- Monyaki, B.S. 2001. A comparison of teachers' and learners' views on feedback strategies used in the writing process. Unpublished MA mini-dissertation. Potchefstroom: North-West University.

- Muncie, J. 2000. Using written teacher feedback in EFL composition classes. *ELT Journal*, 54: 47-53.
- Nwaila, C. 1996. Black English and education in South Africa – an investigation. Unpublished DLitt thesis. Pretoria: University of Pretoria.
- Paulus, T.M. 1999. The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing*, 8(3): 256-289.
- Potter, R. & Fuller, D. 2008. My New Teaching Partner? Using Grammar Checkers in Writing Instruction. *English Journal*, 98(1): 36-41.
- Radecki, P.M. & Swales, J.M. 1998. ESL student reaction to written comments on their written work. *System*, 16(3): 355-365.
- Roberts, F. & Cimasko, T. 2007. Evaluating ESL: Making sense of university professors' responses to second language writing. *Journal of Second Language Writing*, 17: 125-143.
- Spencer, B. 1998. Responding to student writing: strategies for a distance-teaching context. Unpublished DLitt thesis. Pretoria: University of South Africa.
- Spencer, B. & Louw, H. 2008. Theory versus implementation: First-world technology in a third-world teaching context. *Language Matters*, 39(1): 110-126.
- Truscott, J. 1996. The case against grammar correction in L2 writing classes. *Journal of Second Language Writing*, 46: 327-369.
- Truscott, J. 2007. The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing*, 16: 255-272.
- Truscott, J. Yi-ping Hsu, A. 2008. Error correction, revision and learning. *Journal of Second Language Writing*, 17(4): 292-305.
- Wible, D., Kuo, C., Chien, F., Liu, A. & Tsao, N. 2001. A Web-based EFL writing environment: integrating information for learners, teachers, and researchers. *Computers & education*, 37: 297-315.
-

ABOUT THE AUTHORS

Henk Louw

School of Languages
Potchefstroom Campus
North-West University
Potchefstroom
2520
Email: Henk.Louw@nwu.ac.za

Bertus van Rooy

School of Languages
Vaal Triangle Campus
North-West University
PO Box 1174
Vanderbijlpark
1900
Email: Bertus.VanRooy@nwu.ac.za