

Henk Louw

North-West University (Potchefstroom Campus)

Marking for structure using Boolean feedback

Abstract

This paper presents evidence that marking student texts with well-considered checklists is more effective than marking by hand. An experiment conducted on first-year students illustrated that the checklists developed to mark introductions, conclusions and paragraphs yielded better revision results than handwritten comments. Additional benefits made possible by

the technique make a strong case for the use of such a technique in the marking of student texts. The marks assigned to the student texts also make a strong case for focusing on these specific textual features.

Keywords: Feedback; Second Language Writing; Checklist feedback; Text structure; Cohesion

1. Introduction

Written texts are incredibly complex, and as a result providing feedback on texts is a very daunting task indeed. Since the 1980s on-going research has been done to investigate various aspects of feedback on writing (Truscott, 1996:329; Truscott & Yi-Ping Hsu, 2008:292-293; Ferris, 2004). Research branched out into matters such as the differences between first language writing and second language writing, the relationship between writing and SLA, the relationship between writing and reading comprehension and other related topics. A relatively small body of research (when compared to reading comprehension and writing research) focused on feedback on L2 writing. However small, this body of research has contributed its fair share of controversy, with arguments over the relative effectiveness of feedback taking centre stage. Both sides of the argument find instances of misinterpretation in the techniques and interpretations of the others. The so-called “grammar correction debate” published in the *Journal of Second Language Writing* is the best example of such a controversy, with Truscott (1996), Truscott and Yi-ping Hsu (2008), Ferris (2004) and Chandler (2009) being the main role players. As pointed out by Ferris (2004), a lack of consistency in research on this topic is one of the greatest barriers to overcome.

In addition, researchers or teachers who immerse themselves in the research on feedback on writing will find the lack of shared understanding of terminology a barrier to the interpretation of the research. Not all researchers mean the same thing by commonly used terminology such as “feedback”, with some referring to “any response” and others referring to “any correction” (Louw, 2006:21-29).

2. Human fallibility and checklists

In the meantime, while academics battle to obtain replicable conclusions, teachers, lecturers and marking assistants at ground level still continue marking ever-increasing volumes of student texts, despite all the known problems with feedback (discussed in more detail in Ferris, 2003 and 2004; Spencer, 1998; Truscott & Yi-ping Hsu, 2009; Louw, 2009). Louw (2009) identified 13 qualities for effective feedback, but held that it is virtually impossible to adhere to these 13 qualities without the use of computer assistance. While the practice of providing feedback is difficult in itself due to the complex nature of texts and human communication, human limitations while marking also influence the effectiveness of the feedback. In other words, bias, boredom, concentration lapses and the fallibility of human memory are additional variables thrown into the already crowded mix of problems with the provision of feedback.

One area of feedback where many variables come into play is text structure, and research is necessary in this area to assist markers in providing better feedback. Louw and Van Rooy (2010) reported on an experiment in which radio buttons (a kind of checklist) were used to provide feedback on paragraph structures – eight qualities were identified to which an effective academic paragraph should answer. The purpose of using the

checklist was four-fold – to provide feedback which (a) is more thorough (b) is faster, (c) does not burden the marker unnecessarily, and (d) provides the marker with reminders of what to focus on while marking. The results of the experiment proved that there is merit to the idea of using checklists while marking, although the authors stated that this kind of feedback should not be used in isolation.

The use of a checklist is motivated by observing other areas of human endeavour where large numbers of variables need to be taken into consideration. Two of the best-known examples of the use of checklists are the World Health Organisation checklist (discussed in more detail later) and the CAA checklist (Civil Aviation Authority). While extensive research has been conducted on assessment and marking schemes, the author has not been able to locate any research on the use of checklists for feedback, although it is often mentioned with regard to editing (cf. Currie, 1998; Carstens & Van de Poel, 2010). This is odd, since if two of the most respected industries in the world see the need for (and effectiveness of) the systematic application of checklists to their industry, why do writing educators not make consistent use of the same technique but instead opt for holistic or impressionistic feedback?

One may argue that a marking scheme (assessment scheme) is a kind of checklist in that a marker has to work systematically through steps to award a specific mark for the student text. Louw (2006) also explains that any assessment mark (grade) given on a student text is implicit feedback, but the difference here is that a final grade or even a grade in a specific position in a marking scheme does not necessarily translate into feedback for the student. In order for a checklist to function as feedback, it should answer to the qualities of effective feedback as examined in Louw (2009). Also, feedback on a text is not always directly related to the specific marking scheme. (For the sake of clarity, the qualities of effective feedback are added in addendum A).

An experiment on the standardisation of feedback on student writing (Louw, 2006) indicated that it could be standardised to an extent with positive results during student revision. The experiment failed, however, in areas of paragraph structure and cohesion. A follow-up experiment was then conducted (Louw & Van Rooy, 2010) which showed that even non-computerised implementation of a checklist feedback strategy can be more effective in helping students to revise paragraphs than normal, handwritten feedback. The next logical step in the process was to test whether the results could be extended to introductions and conclusions in combination with paragraphs – i.e. moving towards textual organisation.

The structures of paragraphs, in combination with effective introductions and conclusions, assist in creating meaning. Nightingale (1988:278) explains that the complexity of structuring content in students' texts may be more likely to lead to student failures than grammatical errors, even though grammatical errors may in some cases obscure meaning. And is this not how it should be? According to Functionality Theory (Givón, 1989; Halliday & Matthiesen, 2004), language use should in the first place be aimed at communication. An overemphasis by lecturers of focusing on surface level errors does not necessarily lead to better communication. Or, to put it more bluntly, grammatically

perfect sentences may still “communicate” gibberish, as has been so amply illustrated by Chomsky’s famous line “Colorless green ideas sleep furiously.” Louw (2006:98) has also found that lecturers tend to focus more specifically on surface structure elements, probably because they are easier to identify, so it is necessary to remind lecturers to focus on structural components. Assisting them to do so by means of a checklist simply makes sense.

3. Effective introductions and conclusions

As mentioned above, effective introductions and conclusions have many characteristics. A survey of numerous books on “how to write better” revealed the characteristics of effective introductions and conclusions. The books surveyed included, but are not limited to the following:

- Du Toit, Heese and Orr (2002)
- Emory (1995)
- Greetham (2001)
- Hamp-Lyons and Heasley (2002)
- Hannay and Mackenzie (2002)
- Henning, Gravett and Van Rensburg (2002)
- McClelland and Marcotte (2003).

Based on information from these and other books, the qualities of effective introductions and conclusions in academic writing were established to be:

Introduction

1. An introduction should clearly state the question to be investigated in the rest of the text. Alternatively, it should make a clear statement that could be defended, explained or refuted in the text.
2. An introduction should clearly explain the background of the topic to the reader.
3. An introduction should explain to the reader why the student is writing about the specific topic.
4. An introduction should give a clear preview of the contents of the rest of the paper.
5. An introduction should link up with the conclusion.

6. An introduction should have a novel angle of approach to the topic in order to catch the attention of the reader.

Conclusion

1. A conclusion should efficiently recapitulate the main points of the paper without repeating them verbatim from the text.
2. A conclusion should provide the final answer to the question stated in the introduction. Alternatively, it should provide the final verdict on the statement given in the introduction.
3. A conclusion should indicate the relevance of the findings in the text to the reader.
4. A conclusion should never provide brand-new information.
5. A conclusion should link up with the introduction.

These statements about the structure and content of introductions and conclusions are not all of equal importance. For example, many introductions fail to catch the reader's attention with a novel angle of approach, but the introduction can still function as an introduction. Likewise, the degree to which a conclusion recapitulates the main points of the text might not be as important as actually coming to a genuine conclusion (called a "final answer" above to avoid confusion.)

The qualities of effective introductions and conclusions were then incorporated into a checklist marking scheme for the purposes of conducting an experiment.

4. The experiment

A write/revision experiment was designed to test the effectiveness of the Boolean feedback.

4.1. The test group

The student population, on which the experiment was conducted, consisted of two groups of first-year students taking the compulsory course, *Introduction to Academic Literacy* (AGLE 111), at the North-West University, Potchefstroom Campus in 2010. The students were divided into two groups, based on the class they attended. The classes were divided alphabetically without reference to academic performance.

It should be noted that the experiment was conducted very early in their first year, before the students had received any formal instruction in effective writing apart from what they had been taught at school.

4.2. Aim of the experiment

The aim of the experiment was simple: to test whether a set of statements highlighting certain features of introductions, conclusions and paragraphs could be used effectively to provide feedback on student writing.

4.3. The structure of the experiment

Before the students received any formal training in the writing of introductions, conclusions or paragraphs, they were instructed to write a short essay on a specified topic. The instructions were:

1. Write a short argumentative essay on one of the following topics.
 - a. Facebook¹
 - b. This sport (pick one) is being neglected/overemphasised to our detriment.
 - c. Obesity
 - d. Lecturers expect too much/too little of first-year students.
2. The essay must be no more than 500 words in length.
3. The essay must have a clear introduction and conclusion and at least three separate, clear paragraphs.
4. Your essay needs a clear title.

The students were also warned that they would receive a flat zero for the assignment if any error was left in the text, which would have been identified by the computer spelling checker. This (false) warning was intended to force the students to make use of the available proofing tools. It was also hoped that this instruction would weed out most of the surface structure errors which could negatively affect lecturer perceptions of the texts.

The first drafts of the assignments were marked in two different ways. One half of the assignments were marked by hand, using conventional marking (hereafter referred to as “hieroglyphics”). The other half of the assignments was marked with a Boolean feedback checklist. A marking sheet with 32 questions was attached to every assignment and the relevant box was simply ticked; “yes” if the criterion had been met, or “no” if the criterion had not been met. The marking scheme is shown in Table 1.

1 Two reviewers pointed out that neither Facebook nor Obesity elicits argumentation. That is true. Students were taught in class that there is a difference between a topic and a title and were thus expected to create their own argumentative title for the texts.

Table 1: Marking Scheme

INTRODUCTION			
1	Your introduction clearly states the question to be investigated in the rest of the text, or makes a clear statement you wish to defend, explain or refute in the text.	YES	NO
2	Your introduction gives a clear background about the topic to your reader.	YES	NO
3	Your introduction explains why you are writing about the specific topic.	YES	NO
4	Your introduction gives a preview of the contents of the rest of the paper.	YES	NO
5	Your introduction links up with your conclusion.	YES	NO
6	Your introduction has a novel angle of approach on the topic which can catch your readers' attention.	YES	NO
PARAGRAPH 1			
7	This paragraph has a sentence (or part of a sentence) that can function as the main idea for the whole paragraph.	YES	NO
8	This paragraph deals with one main idea only.	YES	NO
9	The main idea of this paragraph is supported with evidence in the other sentences.	YES	NO
10	This paragraph contains only relevant information.	YES	NO
11	The sentences in the paragraph follow each other in a logical manner.	YES	NO
12	The paragraph links up with the paragraph above or below it.	YES	NO
13	This paragraph is in the right position in the text.	YES	NO
PARAGRAPH 2			
14	This paragraph has a sentence (or part of a sentence) that can function as the main idea for the whole paragraph.	YES	NO
15	This paragraph deals with one main idea only.	YES	NO
16	The main idea of this paragraph is supported with evidence in the other sentences.	YES	NO
17	This paragraph contains only relevant information.	YES	NO
18	The sentences in the paragraph follow each other in a logical manner.	YES	NO
19	The paragraph links up with the paragraph above or below it.	YES	NO
20	This paragraph is in the right position in the text.	YES	NO
PARAGRAPH 3			
21	This paragraph has a sentence (or part of a sentence) that can function as the main idea for the whole paragraph.	YES	NO
22	This paragraph deals with one main idea only.	YES	NO
23	The main idea of this paragraph is supported with evidence in the other sentences.	YES	NO
24	This paragraph contains only relevant information.	YES	NO
25	The sentences in the paragraph follow each other in a logical manner.	YES	NO
26	The paragraph links up with the paragraph above or below it.	YES	NO
27	This paragraph is in the right position in the text.	YES	NO

INTRODUCTION			
CONCLUSION			
28	Your conclusion effectively recaps the main points of your paper without repeating them exactly as they were in the text.	YES	NO
29	Your conclusion gives the final answer on the question in the introduction, or the final verdict on the statement in the introduction.	YES	NO
30	Your conclusion indicates the relevance of your findings to the reader.	YES	NO
31	Your conclusion does not provide brand new information	YES	NO
32	Your conclusion links up with the introduction.	YES	NO

Note that questions 7-13 deal with paragraph structures as used in Louw and Van Rooy (2010). These seven questions are repeated three times, making allowance for three paragraphs. The data generated by these serves as an additional validation of the findings by Louw and Van Rooy and could also be used to investigate the interaction between paragraphs, introductions and conclusions.

Based on the results of a previous experiment (Louw, 2006), a “blank” group was not included because the students fared poorly in revising unmarked texts. After the first draft, all the students received further instructions urging them to:

1. use the computer proofing tools
2. pick a side in their argument
3. try to focus on one idea per paragraph
4. pick a descriptive title
5. write an introduction that is more than just a definition.²

The students then had two weeks in which to revise their essays. Twenty-two pairs of essays (first and revised drafts) per marking technique were randomly selected from both groups. These essays contained no feedback marks, since the students also had to submit digital copies of their essays. The essays were randomised using a computerised randomiser and then marked by six experienced markers using the original Boolean feedback marking scheme. Five of the markers (one was unavailable) later gave a mark out of 10 to each text in a separate process. The markers were also asked to write down a few brief comments on how they experienced the use of the Boolean feedback.

2 The audience at SAALA 2010 questioned the rationale behind numbers 2 and 5. The reason for urging the students to pick a side was that most of them were so diplomatic in their approach to the topic that they ended up writing expository essays and never actually came to any sort of conclusion on the topic. They also failed to identify a problem, and many introductions were simply a definition of obesity or Facebook.

The results were digitised for all 32 questions to allow statistical analyses to be done. The raw data (a series of “yes” and “no” answers) were fed into a spreadsheet, with the number one assigned to a “yes” answer and a zero assigned to a “no” answer as illustrated in Table 2. Note that due to space constraints, a full table has not been included.

Table 2: Extract from raw data sheet

Original number	Shuffled number	Marking technique	Version	Marker	Q1	Q2	Q3	Q4	Q5	Q6	Q7	Q8	Q9	Q10	Q11	Etc.
24	1	Buttons	revised	T	1	1	0	0	1	0	1	1	1	1	0	
63	2	Hand	original	T	0	0	0	0	0	0	1	0	1	0	0	

With an analysis system such as this, the original and second draft versions of the same text will appear randomly interspersed among the different texts. The two versions will then also be marked with the same 32 statements and the better of the two versions will have a larger number of ones on the marking sheet than the other.

The raw data were then used to do statistical analyses to establish whether the improvements or regression in the texts could be ascribed to chance.

In order to determine whether the assignments had improved after revision, and secondarily whether the feedback categories related meaningfully to the marks, the markers were asked, four months later, to re-look at the assignments and award a mark out of 10. This was done to ensure that the marks had not been awarded on the basis of the checklist, but instead to determine their general (if somewhat intuitive) sense of the quality of the particular assignment.

The analyses were guided by the following thesis, which is operationalized as a null hypothesis.

4.4. Thesis

By answering a series of strategically chosen “yes” and “no” questions (Boolean feedback), effective feedback can be provided on the structure and purpose of introductions and conclusions in combination with paragraphing. Due to the checklist nature of this feedback, students as well as lecturers will be reminded of all the qualities of effective introductions and conclusions.

H_0

The null hypothesis, which this study sets out to reject, is that the Boolean feedback does not lead to greater improvement after revision than handwritten feedback.

H_a

The research hypothesis is therefore that Boolean feedback will lead to more improvement after revision than handwritten feedback.

To operationalize this statistically, it was attempted to reject the null hypothesis by examining the marks that the markers awarded to the assignments. A dependent t-test was done on the difference between a mark awarded for a specific assignment before and after revision by an individual marker.

As will be shown below, the null hypothesis can indeed be rejected, and further analysis of the data was conducted to determine whether and how the individual components of the feedback checklist related to improvement in the essays. A χ^2 statistic was computed separately for the distribution of the changes from the original to the revised version of the introduction, individual paragraphs and conclusion. A multiple regression model was also extracted to determine whether there was a significant relationship between some of the five sections and the actual mark obtained.

The χ^2 analysis provided more direct information on the extent to which improvement in the revised versions could be attributed to sub-components of the feedback, and closely paralleled the analysis of Louw and Van Rooy (2010) on paragraph structure. This analysis was extended, however, by considering the effect of revision on the introduction and conclusion as well. Like Louw and Van Rooy (2010), we classified the responses into four possible categories: If the original version of the essay was deemed unsatisfactory by a marker on a particular feedback category, and was thus awarded a NO (or 0 score), then the revised version may show no improvement or may improve to a YES (or 1 score). By contrast, if the original version was deemed satisfactory (and thus awarded a YES or 1 score), it may potentially be maintained upon revision or regress to unsatisfactory if the revision did not improve the quality but rather detracted from it (in the view of an individual marker). The classification categories are set out in Table 3 below.

Table 3: Classification of the data

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

Given that marks for the assignment as a whole were also available the relationship between feedback on argument structure and the mark achieved was also investigated by conducting a multiple regression analysis. Taking the marks for the original and revised versions separately as dependent variables, the analysis tried to find the best predictive model from the five groups of variables to account for the mark. Only the average score for an entire section was taken and not the individual items of the five sections of the questionnaire, since the answers to individual items were discrete (either 0 or 1), whereas the average scores form a numerical scale from 0 to 1 (e.g. 2/6 on a section translates to an average of 0.33 for that section). Such data satisfy the assumptions of multiple regression, which requires numerical rather than ordinal/discrete data. The question here is not so much hypothesis testing, but exploring whether the kinds of categories in the feedback system are meaningfully related to the marks.

5. Results

5.1. Improvement of marks after revision

If feedback has served its purpose, the assignments should be better after revision based on the feedback than the originals that were first submitted. While not all students would have engaged with the feedback with equal diligence, and while markers may have been somewhat inconsistent when marking, a small but statistically significant improvement in the marks may nevertheless be expected in order to reject the null hypothesis. The average mark of the originals and the added improvement are represented in Figure 1:

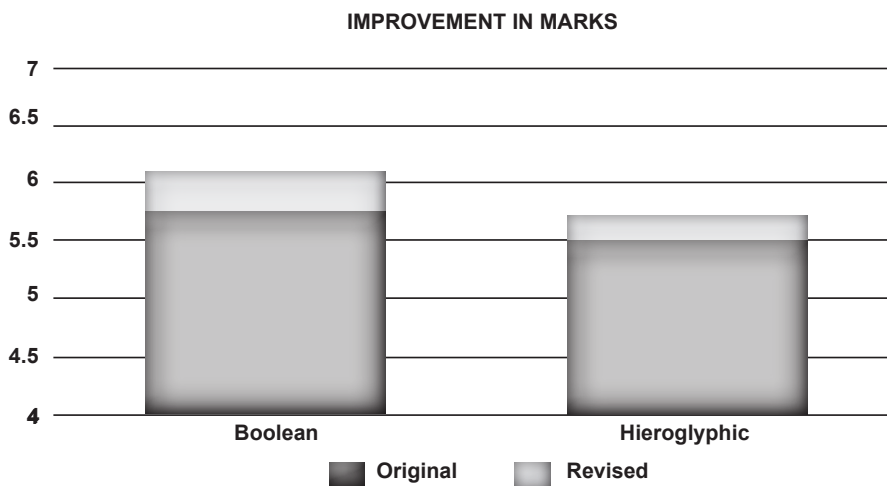


Figure 1: Original average marks out of ten for two groups of assignments, with average improvement after revision, adding up to an average mark for revised versions

Using a dependent t-test, which directly compares the marks for each individual essay per marker with its revised version, we find an improvement of 0.29/10 for the entire data set. Thus, feedback and revision in general lead to improvement in the mark, at a statistically significant level ($t = 2.84$, $df=219$, $p<0.05$). However, the essays that received Boolean feedback are separated from those that received hieroglyphic feedback, only the Boolean feedback improved the essays to a statistically significant degree ($t = 2,30$, $df = 109$, $p<0.05$; improvement 0.32/10), while the hieroglyphic feedback did not yield a statistically significant improvement ($t= 1,72$, $df = 110$, $p>0.05$; improvement 0.25/10).

While the improvement is admittedly small, the reader is reminded that the purpose of this technique is to empower both students and lecturers, and it is hoped that with consistent use of the technique, the cumulative effect over time will be greater. Also, these checklists can be utilised by lecturers in other subject areas as well, effectively making a small contribution to writing across the curriculum. In addition, Boolean feedback is not intended to be used in isolation (the experiment was a bit artificial in that sense) but in combination with a series of other feedback techniques. The cumulative effect thereof cannot be estimated at present. Suffice it to say then that *even in isolation*, use of this technique can refute the null hypothesis. With the additional advantages presented by consistent use of semi-standardised feedback, this is enough reason to advocate the use of the technique.

Revision in response to feedback contributes to improved writing, as has been demonstrated by an improvement in marks noted above, and also with reference to the micro-level of argumentative features in paragraphs by Louw and Van Rooy (2010). To determine the nature and extent to which the feedback checklist proposed in this article contributes to the improvement, a further statistical analysis of the data was undertaken using the χ^2 statistic. By looking at the effect of each of the five sections of the checklist,

namely the introduction, three paragraphs and conclusion, and determining whether there is a difference in the patterns of improvement or regression, we can establish whether the checklist is effective.

As was already shown by Louw and Van Rooy (2010), it is necessary to examine separately the data relating to improvement of aspects that were not satisfactory in the original version, and data relating to regression of aspects that were satisfactory. The χ^2 values indicate whether the proportion of improvements or regressions in the two data sets (Boolean or hieroglyphic) is similar (low χ^2) or different (high χ^2) by comparing the observed number of improvements or regressions with the expected number, based on a null hypothesis of no difference in distribution. Overall, only one analysis, i.e. the distribution of improvements in the introductory paragraph, yielded a statistically significant difference, but all the other analyses also showed that the number of improvements was proportionally higher in revisions that received Boolean feedback, and likewise regressions were proportionally lower in revisions that received Boolean feedback. This finding confirms the results of Louw and Van Rooy (2010) for paragraphs, if less conclusively.

The data for introductory paragraphs are presented in Table 4. The improvements were significantly more likely in the assignments that received Boolean feedback ($\chi^2 = 8.99$, $df = 1$, $p < 0.05$), but the very slight advantage for Boolean feedback on regressions in the introductions is not significant ($\chi^2 = 0.31$, $df = 1$, $p > 0.05$). The data presented in Table 4 show that there were 162 instances of improvement in essays receiving Boolean feedback, which is considerably higher than the value of 140, which is the expected value if the two feedback methods were equally good at prompting improvement upon revision. Thus, of necessity, the essays that received hieroglyphic feedback showed only 134 improvements, lower than the value of 156 that was expected in terms of a null hypothesis of no difference. This also makes it clear why the regressions were not significantly different: there were only three fewer regressions than the expected value for Boolean feedback, thus not much better than the essays that received hieroglyphic feedback.

Table 4: Distribution of differences between original and revised versions for all responses to elements from the introduction checklist, with observed numbers followed in brackets by expected values

	No improvement	Improvement	Regression	Maintained
Boolean	310 (332)	162 (140)	85 (88)	235 (231)
Hieroglyphic	389 (367)	134 (156)	77 (74)	192 (195)

Data for the three paragraphs are presented in Tables 5, 6 and 7. It is clear that the Boolean feedback consistently does a little better, because the values for improvement are always a little higher than the expected values, and the values for regression are always a little lower than the expected values – with the differences being generally bigger for regressions than for improvements, as was also the case in the study by Louw

and Van Rooy (2010). However, the advantage for Boolean feedback remains below the 95% confidence level of a χ^2 value of 3,84.

Table 5: Distribution of differences between original and revised versions for all responses to elements from the paragraph 1 checklist, with observed numbers followed in brackets by expected values

	No improvement	Improvement	Regression	Maintained
Boolean	201 (209)	155 (147)	108 (116)	460 (452)
Hieroglyphic	229 (221)	148 (156)	119 (111)	427 (435)

Table 6: Distribution of differences between original and revised versions for all responses to elements from the paragraph 2 checklist, with observed numbers followed in brackets by expected values

	No improvement	Improvement	Regression	Maintained
Boolean	210 (208)	157 (159)	121 (131)	436 (426)
Hieroglyphic	195 (197)	153 (151)	146 (136)	430 (440)

Table 7: Distribution of differences between original and revised versions for all responses to elements from the paragraph 3 checklist, with observed numbers followed in brackets by expected values

	No improvement	Improvement	Regression	Maintained
Boolean	269 (273)	233 (229)	124 (114)	298 (308)
Hieroglyphic	223 (219)	179 (183)	131 (141)	391 (381)

One issue that emerges from comparing the data from Tables 5-7 is that the paragraphs became increasingly weaker as the essays progressed for both groups of students. This is shown by the gradual increase in the values for **Improvement** and **No Improvement**, and the gradual decrease in the values for **Regression** and **Maintained**. The gradual decline in writing quality does not seem to impact on the degree to which the students managed to revise their work successfully, but just indicates that they tended to present their best/clearest argument first, and then resorted to what was left as they carried on.

Revisions to conclusions were more like the revision to introductions, in the sense that Boolean feedback held a bigger advantage for improvements than for avoiding regressions. Once again, the differences remained below the 95% level of confidence and are therefore not conclusive, as was the case with the three paragraphs. The data are presented in Table 8.

Table 8: Distribution of differences between original and revised versions for all responses to elements from the conclusion checklist, with observed numbers followed in brackets by expected values

	No improvement	Improvement	Regression	Maintained
Boolean	249 (258)	106 (97)	86 (91)	219 (214)
Hieroglyphic	305 (296)	110 (110)	81 (76)	172 (177)

The closer analysis of feedback categories from the checklist is not as supportive of the technique as were the results from Louw and Van Rooy (2010). While differences remained, and always in the right direction, they were only statistically significant on the introductions. It is unclear why this is the case, but two reasons may be tentatively advanced: fatigue and lack of specificity. It has already been noted that the students did progressively worse from paragraph 1 to 3, irrespective of the feedback method or original versus revised version. It may also be that they were more enthusiastic about revising their introductions, but increasingly paid less attention to their feedback and just revised in general. This was exacerbated by the amount of feedback in the case of the students who received Boolean feedback: they received ticks on all of the 32 categories, and in the case of those on the introduction at the top of the list, it was easier to link the feedback specifically to the introductions. The list perhaps became just too long for sustained attention throughout, and the students aligned their reading of the feedback with the specific paragraph they were about to revise. Fortunately, the intended application of the Boolean feedback is not to use it for a whole text, but to select specific problematic paragraphs to comment on with the purpose of eliciting improvement.

6. Discussion

6.1. Relationship between marks and sections from feedback

The checklist

An assumption that underpins much of the work presented here is that there is a relationship between the quality of an essay (as measured by the mark awarded to it), and the characteristics of a good introduction, paragraph and conclusion (captured in the checklist). This is not necessarily self-evident. It is also not necessarily true that all aspects contained in the checklist are equally important. In the current experiment, where marks and the scores from the checklists are available, it is possible to shed some light on the issue. Statistical modelling with multiple linear regression was undertaken to determine how effective a model can be derived to predict the marks, using the feedback from the checklist for building the predictive model.

The nature of the individual elements of the checklist, which is binary data, makes it unsuitable for regression modelling, which requires data of a more continuous nature. It was therefore opted to compute the average number of YES ticks from the feedback checklist for each of the five sections, namely the introduction, each of the three paragraphs, and the conclusion. These five scores were the independent variables in the model, with the mark as the dependent variable. If the data formed continuous scales on each of the 32 individual feedback items, a more complex model utilising all 32 items would have been possible, and more informative at micro-level. Nevertheless, the degree to which a global fit is obtained between the checklist sections and the mark should still reveal whether the concepts contained in the checklist have a bearing on the marks.

Models for the original and revised versions were computed separately, but they had an almost identical overall fit (as measured by the Multiple R value), and were both statistically highly significant. For the original essays, the model had a Multiple R = 0.66 (F (5, 167) = 25.64, $p < 0.001$), and for the revised essays a Multiple R = 0.67 (F (5, 184) = 30.64, $p < 0.001$). Thus the combined correlation values for the two models are almost similar and very high – in more concrete terms, using the R^2 values (0.43 and 0.45 respectively), the model is able to predict very close to 50% of the variance in the marks. This is really helpful, bearing in mind that the actual content (substance, factual correctness or depth) and the surface form (“grammar”) were not factored into the analysis at all. This result shows very clearly that the elements of good writing captured by the checklist form a very significant component of the assessment of essays by markers.

The results also allow a more refined look at the relative contribution of the five sections of the checklist. Besides the Multiple R value, the computations also include a β (beta) value for each of the components, with an assessment of statistical significance of each component in terms of its contribution to the overall predictive power of the model. For all components, whether statistically significant or not, the β values were positive, which implies that the relationship between all components and the marks is positive: the more yes marks in any section of the checklist, the higher the mark. Furthermore, the introductory paragraphs had the highest β values in the regression models of both the original and revised versions (Original: $\beta = 0.47$, $t(167) = 6.45$, $p < 0.001$; Revised: $\beta = 0.31$, $t(184) = 4.47$, $p < 0.001$). The difference between the original and revised versions lies in where the rest of the predictive power comes from. For the original essays, paragraph 1 was the other statistically significant component of the prediction ($\beta = 0.19$, $t(167) = 2.61$, $p < 0.05$), whereas the situation was more evenly balanced in the revised version, with significance for the conclusion ($\beta = 0.26$, $t(184) = 3.69$, $p < 0.001$) and paragraph 2 ($\beta = 0.14$, $t(184) = 2.01$, $p < 0.05$), with paragraph 1 not far outside the cut-off point for significance either ($\beta = 0.11$, $t(184) = 1.60$, $p = 0.11$).

The regression model points to two very important conclusions. Firstly, the elements of the feedback checklist correlate significantly with the marks for assignments, and can therefore be taken to represent a real aspect of student writing. This provides global confirmation for the type of approach advocated here, and specifically the constructs included in the feedback checklist. If students do indeed manage to abide by the implied

guidelines in the checklist, they will do well. Secondly, the introduction is perhaps the most important predictor of the mark of an assignment, and sufficient attention should be given to the introduction. It may well be, in any case, that other elements take their lead from the introduction. One can venture to state that if a text is well planned and the introduction effectively structured, the rest of the text should fall into place effectively.

6.2. Why does it work?

Although it is not a complete revolution in the struggle to improve student writing through feedback, the feedback technique proposed in this article does show enough improvement to make it useful. But why does it work?

As is frequently done when trying to explain something, a dictionary definition was sought for “checklist”. After consulting numerous dictionaries (both online and offline) the most thorough definition found was the one in Wikipedia:³ “A checklist is a type of informational job aid used to **reduce failure by compensating for potential limits of human memory and attention**. It helps to **ensure consistency and completeness** in carrying out a task. A basic example is the ‘to-do list.’ A more advanced checklist would be a schedule, which lays out tasks to be done according to time of day or other factors.” (Wikipedia/Checklist, 2012)

Other definitions that contained relevant information were the following (all Internet based):

- A list used to ensure that **no tasks are omitted, no important aspects are forgotten, and all key functions are checked**. www.actano.com/20911_EN-What%20is%20new-Glossary.htm
- An instrument used to record the **presence or absence** of something in the situation under observation. (102) www.mhhe.com/socscience/psychology/shaugh/ch03_concepts.html
- A list of usability and **quality assurance** questions (for example, “Does each chapter have a clearly defined goal?”) that require a yes or no answer. www3.sympatico.ca/bkeevil/tapuser/gloss.html

Key information in the definition was highlighted in bold by the author.

Some of the only other scientific studies specifically mentioning checklists which could be found were from medical science. These include a study by the World Health Organisation on their Surgical Safety Checklist (Haynes, *et al.* 2009) and a recommendation by Lyons (2010) that checklists be implemented as standard practice in surgical procedures.

3 We are aware of the academic bias against extensive reliance on Wikipedia, but in this specific instance, this simply was the best definition.

Comparison of the results of the current study with the WHO results provided some interesting insights, although this does not completely explain why checklists are effective.

The World Health Organisation implemented a checklist at a number of hospitals to great effect. The WHO *Safe Surgery Saves Lives Checklist* uses 19 items and managed to reduce deaths in its eight pilot hospitals by 36%. Unfortunately, the authors of the WHO study are not sure exactly why such a drastic improvement occurred with the implementation of the checklist. They write, “Whereas the evidence of improvement in surgical outcomes is substantial and robust, the exact mechanism of improvement is less clear and most likely multifactorial” (Haynes, et al, 2009:496). They note that the implementation of the checklist created a change in *systems* and individual *behaviour* and also found that some steps in the checklist were omitted in some cases. “Although the omission of individual steps was still frequent, overall adherence to the subgroup of six safety indicators increased by two-thirds. The sum of these individual systemic and behavioural changes could account for the improvements observed (2009:497).” Lyons (2010) claims that checklists simply raise awareness. To establish exactly how checklists function in complex situations would require additional research.

Similar results were found in the current feedback experiment in that the overall average of all five categories of the checklists improved more consistently than with the non-specific type of feedback through conventional marking. The World Health Organisation study authors and Lyons (2010) postulate that the observer’s paradox could have influenced the results, but in the current experiment there was no observation. Both of the medical studies also pondered the practical feasibility of implementing a checklist at various sites. Their conclusion on the matter was that it is an easy technique to implement. In the current study, checklist feedback is also easy to implement manually or through a computer interface.

In essence then, the individual categories of a checklist combine with the situation to create a change in systems and behaviour, the overall synergistic result being greater than the sum of its parts. If checklists are then consistently implemented in writing pedagogy, it is hoped that new, effective habits for writing may be formed in students.

6.3. Marker comments

The markers were required to write a few comments on the experience they had with the checklist. Apart from providing hints on improvement, one marker did indicate how it helped her, which could explain the effectiveness of the system to some extent: “Using the tables and questions definitely helped me stay **consistent** in marking a single essay, especially because it provides a kind of structure or ‘recipe’ for marking and because certain questions repeat.”

The markers had the following to say about the technique (direct quotes):

1. Not all questions can be answered by a simple yes or no.

2. What if a quality is only met partially?
3. The content of some paragraphs is so marginal that the questions can hardly be applied to it.
4. In the paragraph tables, include a question that addresses the length or content of the paragraph. Many paragraphs were only one or two sentences long and lacked substance and I was not able to indicate this using the questions in the table.
5. Include a separate table with questions that focus on the essay's title (a very important structural component).
6. Some of the words could be interpreted differently, for example link up, logical manner and relevant.
7. Make grammar and language usage the focus point. Grammar should not cost the student marks, but when grammar and language usage make it impossible to follow the argument, should it not be addressed?
8. Marking various versions or even exact copies of the same text made me question my own judgment and I am uncertain whether I was consistent in my marking or not.
9. Marking a single essay using the system took more or less 3 to 4 minutes.

Some of these comments need to be addressed:

Comments 1–4 are easy to address, especially since the idea with the technique is not to use it in isolation. Where needed, the marker can obviously add additional comments. The purpose of the checklist was to be applicable to *most* situations; not *all* situations.

Comments 4 and 5 are actually requests for the use of checklists to be extended, so should be seen as positive.

Comment 6 is valid, but difficult to address as is often the case when dealing with abstract pedagogical concepts. It is believed that training the markers before they use the system will largely eliminate this problem.

Comment 7 shows a tendency to focus on the surface structure (as mentioned above) which is a misconception on the part of the marker. Focusing on surface structure will not make a difference to the organisational structure of the text. In agreement with the marker though, surface structure should not be ignored, but as has already been proven in Louw (2006), that can be dealt with in other ways.

Comment 8 touches upon marker consistency. While the findings of the two medical studies seem to indicate greater consistency in their situations, research will be necessary to see if consistency is in fact improved in language pedagogical settings as well.

Comment 9 indicates that this technique can save time, which should be obvious. The table itself contains about 500 words of text, and it would take substantially longer to provide that amount of handwritten feedback. In a non-experimental marking situation, the marker will probably also choose to focus on one or two paragraphs instead of marking all the paragraphs.

6.4. Proposed implementation

The intention is not to mark a whole text using just the radio buttons. Although they were used on their own in this study, the ideal is to use them as part of a more thorough feedback process. In other words, where the radio buttons are not as effective as conventional explanatory notes, they should be supplemented with additional comments – in other words, the checklist should be supplemented by making use of the effective conventional marking techniques.

If students are supplied with an explanatory table as in Table 9 below, the pedagogical value of the technique increases in that students are supplied with instructions and reminders as well. In other words, the student is informed exactly what is lacking and instructed what to do about the situation, bridging the frequent problem of students seeing there is an error, but not knowing how to correct it.

Table 9: Interpretation of radio buttons

Original statement 1	Your introduction clearly states the question to be investigated in the rest of the text, or makes a clear statement you wish to defend, explain or refute in the text.
IF YES:	Your introduction makes it clear to the reader which question you want to investigate, or which statement you want to address. Make sure that you do indeed treat this question or statement in the rest of the paper.
IF NO:	Your introduction does not have a clear question to guide the rest of your text or it does not make a clear statement which you can treat in the rest of your text. Read the rest of your paper and then rewrite your introduction to fit it better.
Original statement 2	Your introduction gives a clear background about the topic to your reader.
IF YES:	Your introduction gives sufficient background about the topic to the reader.
IF NO:	Your introduction does not give sufficient background about the topic to your reader. Expand on it.

Original statement 3	Your introduction explains why you are writing about the specific topic.
IF YES:	Your introduction explains sufficiently well why you are writing about the specific topic.
IF NO:	Your introduction does not explain well enough why you are writing about the specific topic. Indicate why the topic is important enough for you to write about it and for your reader to read about it.
Original statement 4	Your introduction gives a preview of the contents of the rest of the paper.
IF YES:	Your introduction gives a sufficient preview of the contents of the rest of the paper.
IF NO:	Your introduction does not give a sufficient preview about the contents of the paper. Rewrite your introduction to give your reader an indication of what he or she can expect to find.
Original statement 5	Your introduction links up with your conclusion.
IF YES:	Your introduction links up with your conclusion.
IF NO:	Your introduction does not link up well enough with your conclusion. The questions or statements in your introduction should be answered, supported or refuted in conclusion.
Original statement 6	Your introduction has a novel angle of approach to the topic which can catch your readers' attention.
IF YES	Your introduction has a novel angle of approach to the topic.
IF NO	Your introduction does not have something in it that will interest your readers by catching their attention. It is always a good idea to draw your readers' attention to your writing with an interesting introduction.

7. Conclusion and future research

With the time-saving features and the added advantages of radio button feedback in an electronic environment, a good case exists for the use of this technique in practical everyday feedback practice.

Three areas for further research on this technique have been identified:

1. A new experiment is already under way to test the effectiveness of radio button feedback against voice feedback (audio-taped feedback).
2. The inter-marker reliability has not yet been established. With a sample of only 88 texts, the inter-marker reliability cannot be tested reliably. In addition to inter-marker reliability, another very interesting variable has not been tested – what exactly the handwritten comments commented on. It is an almost 100% certainty that the markers did not comment on all the features covered by the Boolean feedback.
3. It is possible that some of the Boolean feedback may be more effective if combined with some kind of graphic such as dragging and dropping a word to its correct place in a sentence, or dragging and dropping a sentence to the relevant paragraph. The common marking technique of circling a word and drawing an arrow to its correct position in a sentence will definitely be clearer than simply reading a statement about it, for example.

In summary, radio button feedback can be implemented manually or electronically to the benefit of both the marker and the student. For students, the radio buttons allow them greater accuracy in revision with resulting bigger improvements. For lecturers, it is a relatively quick way to provide large quantities of feedback and it reminds them what to focus on while evaluating student texts.

The information provided above also illustrates the importance of focusing on introductions and conclusions in writing pedagogy, since the data clearly illustrate the effect these features have on the mark assigned. If implemented in a computerised marking system, checklist feedback may lead to even bigger gains in accuracy than illustrated here, although its effectiveness in the manual environment already warrants its use.

When marking student texts, markers are in fact annotating data, and at present most of these data are simply going to waste. By consistently marking with semi-standardised techniques and using radio buttons, it is hoped that the data generated by the everyday activity of providing feedback can one day be connected to even more detailed feedback on student writing. It creates tremendous possibilities for research, possibilities which are at present not being realised. Much more needs to be done to realise the true potential of the everyday activity of marking student texts.

Acknowledgements

The author would like to thank Professor Bertus van Rooy for his work on the statistical analyses.

Bibliography

- Carstens, W.A.M. & Van de Poel, K. 2010. *Teksredaksie*. Stellenbosch: Sun Media.
- Chandler, J. 2009. Response to Truscott. *Journal of Second Language Writing* 18:57-58.
- Chomsky, N. 1957. *Syntactic Structures*. The Hague: Mouton.
- Currie, J.R.C. 1998. *Composing strategies of successful and less successful ESL essay writers – a comparison*. Unpublished mini-dissertation, Potchefstroom University for Christian Higher Education.
- Du Toit, P., Heese, M. & Orr, M. 2002. *Practical guide to reading, thinking and writing skills*. Cape Town: Oxford University Press.
- Emory, D. 1995. *Improve your essays*. New York: Glenco/McGraw-Hill.
- 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.
- Ferris, D. 2003. *Response to student writing – implications for second language students*. London: Erlbaum.
- Givón, T. 1989. *Mind, code and context – essays in pragmatics*. London: Lawrence Erlbaum Associates, Publishers.
- Greetham, B. 2001. *How to write better essays*. New York: Palgrave.
- Halliday, M.A.K. & Matthiessen, C.M.I.M. 2004. *An Introduction to Functional Grammar*. 3rd edition. London: Arnold.
- Hamp-Lyons, L. & Heasley, B. 2002. *Study writing*. Cambridge: Cambridge University Press.
- Hannay, M. & Mackenzie, J.L. 2002. *Effective writing in English – a sourcebook*. Bussum: Uitgeverij Coutinho.
- Haynes, A.B., Weiser, T.G., Williams, R.B. Lipsitz, S.R., Breizat, A.S., Dellinger, E.P., Herbosa, T., Joseph, S., Kibatala, P.L. Carmela, M., Lapitan, M., Ferry, A.F., Moorthy, K., Reznick, R.K., Taylor, B., & Gawande, A.A. 2009. A Surgical Safety Checklist to Reduce Morbidity and Mortality in a Global Population. *New England Journal of Medicine* 360(5): 491-499.

- Henning, E., Gravett, S. & Van Rensburg, W. 2002. *Finding your way in academic writing*. Pretoria: Van Schaik.
- Louw, H. 2006. *Standardising written feedback on L2 student writing*. Unpublished Masters dissertation, North-West University (Potchefstroom Campus).
- Louw, H. 2009. Moving to more than editing: a checklist for effective feedback. *Journal for Language Teaching* 43(2): 86-100.
- Louw, H. & Van Rooy, Bertus. 2010. Yes/No/Maybe: A Boolean attempt at feedback. *Journal for Language Teaching*, 44(1): 107-125.
- Lyons, M.K. 2010. Eight-year experience with a neurosurgical checklist. *American Journal of Medical Quality* 25: 285-289.
- McClelland, L.D. & Marcotte, P.H. 2003. *Writing matters! Introduction to writing and grammar*. New York: McGraw-Hill.
- Nightingale, P. 1988. Understanding processes and problems in student writing. *Studies in Higher Education* 13(3): 263-283.
- Spencer, B. 1998. *Responding to student writing: strategies for a distance-teaching context*. Pretoria: University of South Africa. (Thesis – D.Litt.)
- Truscott, J. 1996. The case against grammar correction in L2 writing classes. *Journal of Second Language Writing* 46: 327-369.
- Truscott, J. & Yi-ping Hsu, A. 2008. Error correction, revision and learning. *Journal of Second Language Writing* 17(4): 292-305.
- Wikipedia Foundation. 2012. Checklist. <https://en.wikipedia.org/wiki/Checklist> Date of access: 12 November 2012.

Addendum A: Qualities of Effective Feedback

From: Louw, H. 2009. Moving to more than editing: a checklist for effective feedback. *Journal for Language Teaching*, 43(2): 86-100.

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. be 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; and
13. be searchable/ archiveable/ recordable and allow for research.

About the author

Henk Louw

Centre for Academic and Professional Language Practice, School of Languages

Faculty of Arts
North-West University (Potchefstroom Campus)

Private Bag X6001. Potchefstroom,2520

Email address: henk.louw@nwu.ac.za

Henk Louw is a senior lecturer in the Centre for Academic and Professional Language Practice at the North-West University, Potchefstroom Campus. He presents Academic Literacy and works in close collaboration with the university writing laboratory. His main research foci are the optimization of feedback on student writing and the effective implementation of blended learning in a language environment.