MarkWrite: Standardised feedback on ESL student writing via a computerised marking interface

Thesis submitted for the degree PhD in Applied Linguistics at the North-West University, Potchefstroom Campus.

Henk Louw

Supervisor: Professor A. J. (Bertus) van Rooy

January 2011

Potchefstroom

ACKNOWLEDGEMENTS

I would like to thank the following people for their assistance, support and encouragement in the completion of this thesis:

- Bertus van Rooy
- Susan Coetzee-Van Rooy
- My colleagues and friends at the School of Languages
- My family
- God
- All the participants in the experiments.

A warm thank you to everybody else who encouraged me or helped me in any way.

KEY WORDS

CALL

Second language writing

Writing pedagogy

Feedback

Error correction

Checklists

ABSTRACT

The research reported on in this thesis forms part of the foundation of a bigger research project in which an attempt is made to provide better, faster and more efficient feedback on student writing.

The introduction presents the localised and international context of the study, and discusses some of the problems experienced with feedback practice in general. The introduction also gives a preview of the intended practical implementation of the research reported on in this thesis.

From there on, the thesis is presented in article form with each article investigating and answering a part of two main guiding questions. These questions are:

1. Does feedback on student writing work?
2. How can feedback on student writing be implemented as effectively as possible?

The abstracts for the five individual articles are as follows:

Article 1

Article 1 presents a rubric for the evaluation of Computer-Assisted Language Learning (CALL) software based on international recommendations for effective CALL. The rubric is presented after a brief overview of the pedagogical and implementation fundamentals of CALL, and a discussion of what needs to be included in a needs analysis for CALL evaluation. It is then illustrated how the evaluation criteria in the rubric can be used in the design of a new CALL system.

Article 2

Providing feedback on student writing is a much-debated topic. One group of researchers argues that it is ineffective and another group remains convinced that it is effective, while at ground level teachers and lecturers simply carry on “marking” texts. The author of this article contends that both arguments have valid contributions to make and uses the arguments both for and against feedback to create a checklist for effective feedback practice. Adhering to this checklist should counter most of the arguments against feedback while supporting and improving the positive arguments in favour of feedback.

Article 3

This article reports on an experiment which tested how effectively standardised feedback could be used when marking L2 student writing. The experiment was conducted using a custom-programmed software tool and a set of standardised feedback comments. The results of the experiment prove that standardised feedback can be used consistently and effectively to a degree, even though some refinements are still needed. Using standardised feedback in a standard marking environment can assist markers in raising their awareness of errors and in more accurately identifying where students lack knowledge. With some refinements, it may also be possible to speed up the marking process.

Article 4

This article 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. The major problems with feedback on L2 writing are introduced and it is established 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; technique B: consciousness raising through generalised Boolean feedback; and technique 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, partly because a higher proportion of the instances of negative feedback on the first draft were corrected in the second draft (improvements), but more importantly 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 using such checklists to give feedback on student writing.

Article 5

This article describes an experiment in which a series of statements, answerable simply with yes or no (labelled Boolean feedback), were used to provide feedback on the introductions, conclusions and paragraph structures of student texts. A write-rewrite assignment (the same structure as in article 4) was used and the quality of the student revisions was evaluated. The results indicate that the students who received Boolean feedback showed greater improvement and fewer regressions than students who received feedback using the traditional method.

The conclusion provides a brief summary as well as a preview of the immense future research possibilities made possible by this project.

OPSOMMING

Die navorsing waaroor in hierdie tesis verslag gedoen word, vorm deel van die fondasie van ’n heelwat groter navorsingsprojek. Hierdie projek het ten doel om beter, vinniger en meer effektiewe terugvoer op studente se skryfwerk te lewer.

In die inleiding van die tesis word die plaaslike en internasionale konteks van die studie uitgestip, sowel as ’n aantal probleme rakende terugvoer op skryfwerk. Die inleiding gee ook ’n vooruitskouing van die praktiese implementering van die navorsing waaroor daar in die tesis verslag gelewer word.

Die tesis word in artikelformaat aangebied, met elke artikel wat ’n deel van die rigtende vrae ondersoek en beantwoord. Die twee vrae is:

1. Werk terugvoer op studente se skryfwerk?
2. Hoe kan terugvoer op studente se skryfwerk meer effektief geïmplementeer word?

Die opsommings van die vyf individuele artikels is as volg:

Artikel 1

Artikel 1 stel ’n rubriek bekend wat gebruik kan word vir die evaluering van sagteware-pakkette vir rekenaargesteunde-taalonderrig, oftewel “Computer-Assisted Language Learning” (CALL). ’n Kort oorsig oor die fundamentele pedagogiese beginsels van CALL word hier voorsien, sowel as die implementeringsbeginsels daarvan. Die behoeftebepaling wat met die implementering van die nuwe CALL gepaard moet gaan word ook bespreek. Die rubriek word dan bekend gestel en daar word geïllustreer hoe dieselfde rubriek gebruik kan word wanneer oorweging geskenk word aan ’n nuwe sagteware-pakket se ontwerp.

Artikel 2

Die lewering van terugvoer op studente se skryfwerk is ’n veelbesproke onderwerp. Een groep navorsers beweer dit is oneffektief; ’n ander groep navorsers beweer dit is funksioneel, en op grondvlak gaan onderwysers en dosente bloot voort om tekste na te sien. Die skrywer argumenteer dat sekere aspekte van beide argumente geldig is, en as gevolg hiervan word insigte van beide argumente in hierdie artikel gebruik om ’n oorsiglysie op te stel wat effektiewe terugvoer definieer. Deur te hou by die vereistes van hierdie oorsiglysie sal die meeste van die argumente wat teen terugvoer gemaak word teengewerk word, terwyl die positiewe aspekte met betrekking tot die voorsiening van terugvoer versterk en ondersteun sal word.

Artikel 3

Artikel 3 rapporteer oor ’n eksperiment waarin getoets word hoe effektief gestandaardiseerde terugvoer gebruik kan word wanneer tweedetaalstudente se skryfwerk nagesien word. Die eksperiment is uitgevoer deur van ’n pasgemaakte sagtewarepakket gebruik te maak, sowel as ’n stel voorafvervaardigde kommentaar. Die resultate van die eksperiment bewys dat gestandaardiseerde terugvoer tot ’n mate konsekwent en effektief

gebruik kan word, alhoewel sekere afronding steeds nodig is. Die gebruik van gestandaardiseerde terugvoer in ’n standaard-nasienomgewing kan die nasieners help om hul bewustheid van foute te verhoog en om meer akkuraat te identifiseer waar studente se kennis ontbreek. Die tegniek kan ook die nasienproses versnel.

Artikel 4

Die artikel beskryf ’n eksperiment waarin Booleaanse terugvoer (’n soort oorsiglysie) gebruik is om terugvoer te lewer op die paragraafstrukture van eerstejaarstudente in ’n module van Akademiese Geletterdheid. Die grootste probleme wat ten opsigte van terugvoer op tweedetaal-skryfwerk ervaar word, word uitgewys, waarna die fokus op paragraafstrukture regverdig word.

Die eksperiment was ’n werksopdrag in twee weergawes, waar drie soorte terugvoer gebruik is vir drie verskillende groepe studente. Die drie soorte terugvoer is handgeskrewe kommentaar, bewusmakingskommentaar deur algemene Booleaanse terugvoer, en spesifieke Booleaanse terugvoer. Die resultate toon dat spesifieke Booleaanse kommentaar meer effektief is as die ander twee tegnieke, omdat ’n groter gedeelte van die negatiewe kommentaar op die eerste weergawes van die tekste ná die tweede weergawes gekorrigeer is. ’n Belangriker aspek is egter dat die hersiene weergawes minder regressie toon. Vir algemene kommentaar was daar amper net soveel regressie as verbetering. In kombinasie met outomatiese analitiese tegnieke wat moontlik gemaak word deur sagteware, ondersteun hierdie eksperiment die stelling dat sulke oorsiglysies gebruik kan word om effektiewe terugvoer te lewer op studente se skryfwerk.

Artikel 5

Hierdie artikel beskryf ’n eksperiment waarin ’n aantal stellings, wat beantwoord kan word met “ja” of “nee” (genaamd Booleaanse terugvoer) gebruik is om terugvoer te lewer oor die inleiding, slot en paragraafstrukture van studente se tekste. ’n Skryf-herskryf-oefening is gebruik en die kwaliteit van die studente se hersiene weergawes is geëvalueer. Die resultate dui daarop dat die studente wat Booleaanse terugvoer ontvang het, beter kwaliteit hersiene weergawes kon lewer en minder regressie in hulle skryfwerk getoon het as studente wat op die tradisionele wyse terugvoer ontvang het.

Die tesis se gevolgtrekking bied ’n kort oorsig oor die bevindinge van die navorsing, sowel as ’n oorsig oor die verdere navorsingsmoontlikhede wat deur die studie moontlik gemaak word.

Contents

LIST OF FIGURES

LIST OF TABLES

CHAPTER 1
INTRODUCTION

1.1 Background

The work reported on in this thesis is part of a long-term project. The main aim of the project is to improve the efficiency and speed of providing feedback on student writing in a way that will assist lecturers, students and researchers. The research revolves around the development of the computerised marking interface, MarkWrite, in collaboration with the Centre for Text Technology (CTexT®) at the North-West University.

Research on this project commenced in 2004 with a Master's dissertation (Louw, 2006) which investigated whether it would be possible to standardise written feedback on second language student writing. The dissertation identified problems with the practice of feedback as elaborated on in international research. Reasons for continuing with the practice of feedback were investigated, and it is postulated in the dissertation that, despite the problems identified with the *practice* of feedback, the *principle* of feedback is sound. It was hypothesised that standardising feedback and incorporating it into a computerised marking system would improve the *practice*.

The dissertation then investigated in an experiment what markers typically focus on while providing feedback. Using that information, as well as a literature review, a list of standardised feedback "tags" was then created and tested in a revision exercise experiment on actual students. The results showed that students were more likely to improve texts marked with the standardised feedback than texts marked with normal, handwritten feedback. Both these groups also outperformed a "blank" group in which students were asked to identify and correct errors in a text that had no feedback marked on it. The study concluded that it was indeed possible to standardise written feedback on student writing, but the implementation thereof still had to be developed.

Feedback is, however, a two-way communicative process, and having established that standardised feedback could assist the student, the effectiveness of standardised feedback on the other role-players in this communication process (the markers or lecturers) had to be investigated. Also, the implementation of the standardised feedback necessitates the use of computers and software since some of the characteristics thereof are difficult to achieve by hand. Thismoves the research into the realm of Computer-Assisted Language Learning (CALL), which is a research field of its own. Furthermore, the standardised feedback was not effective in all areas as students experienced difficulty revising the higher-order (structural) elements of the texts.

This thesis therefore continues the research and development initiated in 2004. It refines and implements some of the findings from the previous study, by investigating lecturer behaviour when using standardised feedback, piloting a new technique for providing feedback on the structural elements of texts, and establishing whether the whole project adheres to best practice in CALL.

Firstly, the qualities of effective feedback, as established from international research, were reworked into a checklist which could be used to design and evaluate the effectiveness of any one specific feedback technique. Thereafter, an experiment was conducted to establish what markers typically focus on. In other words, the lecturer side of the two-way communication process was investigated. Based on problems identified in this experiment and those identified by Louw (2006), a subsequent experiment was conducted in which a new technique (aimed at structural qualities of texts) was tested. When positive results were obtained, a similar experiment aimed at more structural elements was commenced. At the same time, the software for the implementation of the standardised feedback was being programmed by the Centre for Text Technology (CTexT®) at the Potchefstroom Campus of the North-West University. It was therefore necessary to ensure that the program met the requirements of best practice in CALL, and an investigation into this best practice was begun.

The end result of the two studies was a software marking system (MarkWrite) incorporating a tag set of standardised feedback and a set of checklists for the effective marking of introductions, conclusions and paragraph structures. Both the tag set and the set of checklists can also be implemented effectively on their own in a limited fashion.

1.2 What seems to be the problem: global context?

There are many problems with current feedback practice, not least of which is the question "what exactly is the problem?" Writing is a very complex human action and research on feedback specifically has pointed out numerous problems. This is exactly what the problem is: there are many different problems, compounded by many different variables. Ferris (2004) claims that research on feedback on student writing suffers from a lack of consistency in both research methodologies and findings. The conclusions reached to date by various researchers do not agree (for example the "Grammar Correction Debate" discussed in the *Journal of Second Language Writing* by Truscott, 1996; Chandler, 2009 and Ferris, 2004).

As early as 1985, Zamel (in a much-quoted paper) bemoaned feedback practice with statements such as:

> ESL writing teachers misread student texts, are inconsistent in their reactions, make arbitrary corrections, write contradictory comments, provide vague prescriptions, impose abstract rules and standards, respond to texts as fixed and final products, and rarely make content-specific comments or offer specific strategies for revising the text. What is particularly striking about these ESL teachers' responses, however, is that the teachers overwhelmingly view themselves as language teachers rather than writing teachers (Zamel, 1985:86).

While attempts at improvements have been made, the problem (as stated above) is compounded by an astounding number of variables. Among these is the difference between first-language writers and second-language writers. This difference is considered to be so profound that a whole journal is devoted to the teaching of writing to L2 students – the *Journal of Second Language Writing*. Other variables include individual student preferences and learning styles, student differences across cultures, the influence of language acquisition

on writing, and many more, each of which can be considered a "problem" based on the effect it has on the efficiency of feedback.

In addition to the above variables, there are practical matters of concern as well. In some of the classes taught by the author, there were up to 450 students. While marking assistants were available in some cases, the time needed for marking assignments in this situation was still excessive. For example, to mark a single 500-word essay assignment takes about five minutes (if done fairly superficially). That amounts to 2 250 minutes (almost 38 hours) to complete the marking of the whole class's assignments, or practically a whole working week used up. More efficient ways need to be found, if not for the sake of saving time, then at least for efficiency.

Much research has gone into the teaching of writing to second-language students of English (for example Kroll, 2003). Different techniques are used in this research process, for example the process approach to writing (Krapels, 1990), error analysis (Ellis, 1996:48; James, 1997) and corpus linguistics (Granger, 2002; Wible, Kuo, Chien, Liu & Tsao, 2001). It is rare to find a study that attempts to incorporate elements from the different techniques into one approach. In addition, it is rarer still to find a study that attempts to *implement practically* findings from all the different methods. The aim of the present study is to contribute to the body of knowledge by utilising insights gained from Computer-Assisted Language Learning (CALL), the process approach to writing, writing across the curriculum, error analysis, academic literacy and corpus linguistics and implementing them practically.

Since the 1970s a movement called "Writing across the Curriculum" has been gaining ground. Its main aim is to promote general as well as discipline-specific learning through writing (Deng, 2009). Students need to be reminded of the importance of accurate and effective writing throughout their entire education (Snively, Freeman, & Prentice, 2006, quoted by Deng, 2009). Experience has taught us that it is not possible to adequately address students' lack of proficiency in writing in one or two semesters of writing instruction, since writing proficiency needs time to develop and students need to be able to practice writing often (Deng, 2009). Students need to receive feedback on their writing so that they know what to improve, but problems with feedback are numerous (cf. Ferris, 2002; Hyland, 2003 and 1998; Krapels, 1990; and Louw, 2006).

In addition, there has been an ongoing global debate since the 1980s about whether or not feedback is effective – i.e. does feedback lead to a demonstrable long-term improvement in student writing? The researchers participating in this debate are once again those involved in the Grammar Correction Debate mentioned above, but including Zamel (1985) and Askew and Lodge (2001). In Louw (2006) this debate is dismissed as irrelevant to the MarkWrite project since the practice of providing feedback is firmly established. The practice of providing feedback will not disappear because society, lecturers and students expect it, and the broad definition of *feedback* and *error* by Louw (2006) implies that the mere act of indicating improvement or assigning a mark is feedback in itself.

While Truscott (2007) argues that it is bad practice to provide feedback simply because learners expect it, others claim that it is indeed necessary to provide learners with what they

think they need as well. Feedback is, however, not just provided because students want it, it is also provided for the following reasons:

1. The students expect feedback (Chandler, 2009: 58).
2. Feedback is an old and established technique of teaching (although the practical application thereof will affect the effectiveness of the pedagogy).
3. Feedback is a communication process indicating to the writer how a reader interprets his or her text (Askew & Lodge, 2001).
4. Feedback can be effective although research (e.g. Truscott & Yi-Ping Hsu, 2008) has also illustrated that not always to be the case.

Although feedback has not always been found to be effective, it is still expected and can still have advantages. While a few variables and problems have been mentioned in passing in the discussion so far, it is not the purpose of this introduction to discuss the totality of feedback problems in detail. Rather, the problems with feedback *relevant to this study* are identified as follows:

1. The lack of consistency in technique and error identification by markers.
2. Incorrect focus by markers.
3. Unclear comments by markers.
4. Students' inability to understand and use feedback independently.
5. The amount of time it takes lecturers to comment effectively on students' texts.
6. Lecturers are not always consciously aware of how to provide students with effective writing pedagogy through feedback. This is especially relevant in content subjects where the lecturers are not trained in writing.

(Kasanga, 2004; Louw, 2006; Spencer, 1998; and Deng, 2009.)

1.3 Synthesising feedback techniques

Many different techniques have been developed in an attempt to improve on feedback, enhance learning gained from feedback, involve students more, speed up language acquisition and so forth. Some of these techniques are selective marking, audio feedback, feedback conferences, marking codes (correction codes), peer review and writing seminars. All of these have their advantages and disadvantages, but it is outside the scope of this thesis to discuss them in detail. What this thesis does, is to borrow some aspects of these techniques in order to improve feedback. If any of these techniques have some positive aspect (read: "it works"), it should not be neglected and a way should be found to implement it consistently to gain the maximum benefit from it.

To establish what is considered the most effective way to provide feedback, this thesis identifies the best and worst aspects of feedback. In Chapter 1 these aspects are used to construct a checklist for effective feedback. It is argued later in the thesis that it is not

possible to adhere to this checklist using conventional feedback techniques. The only way to provide feedback as effectively as prescribed is to use computer applications to compensate for human shortcomings.

1.4 Localised context of this study

Louw (2006) argues that by combining CALL with insights from research into feedback, the teaching of writing can be much more effective. He points out that this can be achieved by standardising feedback comments to an extent, and implementing this feedback in a computerised marking interface. He tested the first assumption and found that standardised feedback did indeed lead to greater improvement than traditional marking techniques when marking surface level errors. However, the standardised feedback tags used for issues of coherence, paragraph structure and textual cohesion were less effective and needed revision and re-evaluation.

For the experiment mentioned in Louw (2007), a first version of the electronic marking system was created. The first version was not user friendly and contained too many programming errors to be useful, so the Centre for Text Technology (CTexT®) at the North-West University, Potchefstroom Campus, commenced with programming the newer versions.

The ultimate aim of MarkWrite is to have a student side (MarkWrite Student*)* and a marker side (MarkWrite Marker*)*. The student side of the marking system will be a pre-warning system where students receive automated feedback on features which the computer can reliably identify such as spelling errors, commonly confused words, incorrect use of fixed expressions, and a host of other text features (cf. an early attempt by Trushkina, 2006). These still need to be developed or researched and fall outside the scope of this dissertation. They are merely mentioned as an indication of the scope of the MarkWrite project.

MarkWrite Marker is the focus of the current project (this PhD), and this study views feedback on student writing as the intersection point between knowledge from the different approaches used in the past to study this phenomenon. Weideman (2007) refers to the responsibility of applied linguistics in this regard, in which practical implementation and testing of solutions are attempted. The intention of testing and implementation is to alleviate real-world problems in society. The focus of this study (and the end product, MarkWrite Marker) is exactly that –testing of solutions and applying them practically to real-world problems.

Before continuing, it should be noted that the notion of marking and evaluating texts by hand is not an out-dated concept. Electronic assessment tools and rating tools such as *Criterion* and E-Rater from ETS, as well as commercial products such as Whitesmoke and other advanced grammar and style checkers have become available in the last 20 years. These products are not yet without problems and it may still be a number of years before the human marker is substituted by a fully automatic system.

Apart from being a practical solution, MarkWrite is also intended to be a research tool. The system contains tracking and other features which can be used to great effect for research.

It is long overdue that the immense amount of effort that goes into the marking of student texts be used for other purposes as well. When marking student texts, the marker is in fact doing work which is very similar to that of a corpus annotator annotating a corpus text. Techniques used in MarkWrite, such as radio button marking and assessment, can also generate vast amounts of data which could all be used for further research (see the work done by Wible *et al.*, 2001).

1.5 Research questions

With this background in mind, the research questions of this thesis are:

1. How does one evaluate a Computer-Assisted Language Learning (CALL) software package, and does MarkWrite qualify as effective CALL software?
2. What are the qualities of effective feedback on student texts?
3. What do lecturers focus on when marking student texts?
4. Can Boolean feedback (also called "radio button feedback" in this thesis) be used to provide useable feedback on paragraph structures quickly and efficiently? If the technique works, why does it work?
5. Can radio buttons be used to provide useable feedback on introductions, conclusions and overall textual coherence quickly and efficiently? If so, why?

1.6 Aims of the study

The aims of the study are to:

1. Establish a metric which can be used for the evaluation of CALL software to justify MarkWrite as acceptable CALL software.
2. Establish the qualities of effective feedback in such a way that it can be used to evaluate the effectiveness of feedback practices in student writing.
3. Determine what lecturers choose to focus on when marking student texts while using standardised feedback.
4. Evaluate the effectiveness of Boolean feedback on paragraph structure when students revise texts, and attempt to explain the findings.
5. Evaluate the effectiveness of Boolean feedback on introductions, conclusions and overall textual coherence, when students revise texts, and attempt to explain the findings.

This research project aims to demonstrate that it is possible to integrate and practically implement insights from writing pedagogy from a variety of approaches in a way that will

benefit both student and teacher and will solve (or at the very least, alleviate) some of the problems associated with the teaching of writing to students.

The research revolves around the partial creation and partial testing of MarkWrite. As such it forms only a part of MarkWrite and should not be seen in isolation, while at the same time most of the techniques can easily be applied in other teaching contexts without the use of a dedicated software system.

CHAPTER 2
OVERVIEW OF THE THESIS

The study is structured by answering the different questions in article format, with the articles flanked by an introductory and a concluding section. Each of the articles is briefly outlined below. The methodology for each article is introduced, but also explained in more detail in the article itself.

2.1 Article 1: Design considerations for CALL based on evaluation criteria for CALL

2.1.1 Main question

Since the implementation of the research on feedback in this study revolves around the computerised marking system, it is necessary to establish what the qualities of an effective Computer-Assisted Language Learning (CALL) system are. This article poses the question: what are the qualities of a good CALL system? This is necessary to evaluate whether MarkWrite has the best possible chance of working in a language learning environment.

2.1.2 Purpose of the article

The article has the main aim of establishing what are considered internationally to be recommended traits of CALL systems, so that an evaluation rubric can be created to assess the suitability of a CALL system for a specific pedagogical purpose.

2.1.3 Methodology

This article comprises two steps. In step one, a literature study on the evaluation of CALL systems was conducted to establish what the qualities of good CALL systems are. The overlapping and vague definitions used in the international literature are synthesised into a practically useful evaluation rubric which can be used by language lecturers or systems administrators to evaluate whether a software system is suitable for their situation. International literature referred to includes the works of the highly acclaimed Graham Davies of EuroCall (Davies and Hewer, 2009; and Davies, Hamilton, Weideman, Gabel, Legenhausen, Meus and Myers, 2009) as well as Ngu and Rethinasamy (2006).

Using this rubric, MarkWrite was then evaluated as CALL system in step two to demonstrate that it satisfies international requirements for effective CALL. Effective pedagogy was also identified as a vital part of CALL, which justifies determining the most effective methods for providing feedback if it is to be used in a CALL environment.

2.2 Article 2: Moving to more than editing – a checklist for effective feedback

2.2.1 Main question

What are the qualities of effective feedback?

Louw (2006) identified the qualities of effective feedback. A re-evaluation of these qualities in view of new, practical insights brought to light the fact that the qualities mentioned overlap and that the responsibility of the marker tends to be underestimated. The list of qualities was therefore tightened up and improved to provide a rubric for evaluating the effectiveness of a feedback technique.

2.2.2 Purpose of the article

Effective feedback has numerous characteristics. One technique of providing feedback on student writing may be more (or less) effective in a particular area of student writing than another technique. The differences between various techniques and theories are influenced by the practicalities of implementing the technique, such as the time available, the level of competence of the marker and the intended purpose of the feedback. Since the whole study attempts to create feedback which is clear, user-friendly, consistent and above all effective, it is necessary to establish what exactly is meant by effective feedback, in such a way that the relative effectiveness of a specific technique can be checked systematically. This article therefore serves as the central standard for evaluating the effectiveness of techniques experimented with in the later articles.

2.2.3 Methodology

This article is mainly a literature review, attempting to establish the *status quo* of research on feedback. It draws together information from various perspectives to establish what the qualities of effective feedback practice are. The various perspectives come from:

- researchers who believe feedback is effective (e.g. Hyland, 2003:219 and Askew & Lodge 2000:2)
- researchers who believe feedback is ineffective (e.g. Truscott, 1996)
- researchers who try to establish which technique of feedback is more effective than others (Spencer, 1998)
- researchers who attempt different kinds of writing instruction in order to facilitate better writing, such as the process approach, for example Matsuda (2003:21)
- researchers who use technology in various formats to enhance their pedagogy (Wible, Kuo, Chien, Liu & Tsao, 2001).

The argument is that, since these researchers base their findings on systematic research, there have to be points where they agree on what effective feedback entails. By establishing what those criteria are, the creation of standardised, computerised feedback can be done more effectively.

2.3 Article 3: Moving to more than editing – standardised feedback in practice

2.3.1 Main question

Can standardised feedback work in practice?

An experiment was conducted by Louw (2006) in which it was proved that standardised feedback is more effective than normal feedback when students revise their texts. However, the effectiveness of feedback is not just measured by whether students are able to use it, but also by the type of student problems on which feedback is presented by the lecturers. Excellent feedback on low-value problems is still largely a waste of lecturers' and students' time. What exactly do markers mark?

2.3.2 Purpose of the article

Article two therefore attempts to establish what *markers* focus on when using standardised feedback so that possible shortcomings in the feedback loop (from the lecturers' side) can be identified and addressed. Truscott (1996) indicated that feedback is ineffective because markers are often not capable of providing effective feedback. If this is the case, what do markers focus on, and what do they ignore? How can the marker be assisted in focusing on the important aspects or the aspects they choose to ignore?

2.3.3 Methodology

Four markers were asked to mark 400 L2 essays from the Tswana Learner English Corpus and the Afrikaans Learner English Corpus, using the standardised feedback incorporated into the very first version of MarkWrite. The intensive nature of the marking made it impossible to use more than four markers. Markers were shown how the system worked but were not given any additional instruction on how to mark or on what to provide feedback. Two of the markers were experienced at marking student texts, while two were relatively inexperienced. Once the marking had been done, all the feedback was tabulated and analysed. This was done to investigate what the "natural tendencies" in marking are.

A second step in the experiment was to interview the markers to establish their thoughts on the marking process and the effectiveness of computerised marking. The results were analysed and the findings indicated that lecturers were not consciously aware of what they focused on, were not able to identify their students' most frequently recurring errors, and tended to focus on surface level errors (Louw, 2007). The interviews also indicated problems with the marking interface which needed to be addressed, but confirmed that from a marker's perspective, the marking of student texts on computer is a feasible option. These problems prompted the next stage of the investigation, where an attempt had to be made to direct markers' attention to those aspects they tended to neglect.

2.4 Article 4: Yes/no/maybe – a Boolean attempt at feedback

2.4.1 Main question

Louw (2006) illustrated that feedback on surface structure elements is more effective than feedback on textual organisation. Feedback on matters of relatively lower importance was therefore more effective than feedback on matters of higher importance. This problem had to be addressed for standardised feedback to be considered effective overall. This raised the question of how to encourage or remind lecturers to focus on elements of higher importance, such as the effective structuring of a paragraph, or textual cohesion and coherence. Viewed in terms of Halliday and Matthiessen's (2004:24) strata of language, feedback tends to be provided only in terms of expression, while both strata of content and the context are ignored. This oversight on the part of markers needed to be rectified.

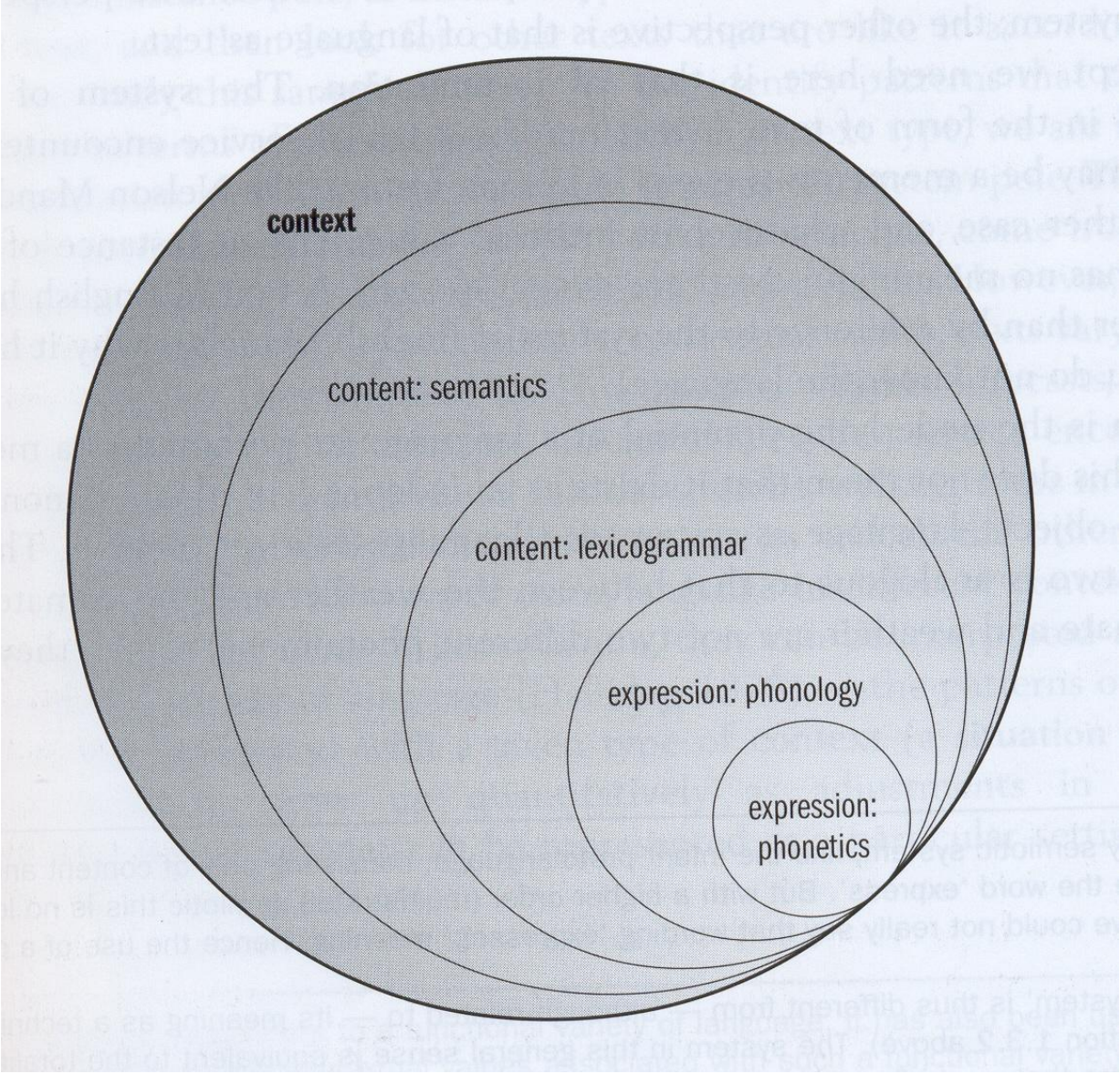

Figure 2.1: Halliday and Matthiessen's strata of language

The system used to provide feedback therefore needed to focus lecturers' attention on the other important issues and enable them to comment on those issues in a standardised way *without* burdening them unnecessarily with more work. The feedback should at the same time be effective in that students are able to use it to improve their writing.

Radio buttons (a type of checklist) was considered a good way to assist lecturers as they are one of the quickest and easiest ways to select options in a computer interface. If the qualities of a good paragraph can be captured by a finite list, a checklist of these statements can be used to provide feedback. The statements will cover enough of the characteristics of a good paragraph that feedback can be deduced from a combination of yes/no answers on these statements.

The main question investigated in this article is to what extent radio buttons can be used as feedback on paragraph structure to: (a) assist lecturers to focus on the important aspects, (b) generate assessment assistance to lecturers, and (c) help learners to understand and to revise their writing more effectively.

2.4.2 Purpose of the article

The aim of the article is to test a computer-replicable technique for providing more standardised feedback which should solve some of the problems identified in Article 2, as well as some of the earlier established problems with feedback. The technique encompasses elements from text linguistics, feedback research, consciousness raising, assessment research, composition training and reading and writing research.

2.4.3 Methodology

First-year students of Academic Literacy were assigned a topic on which to write two paragraphs. The paragraphs of three different groups were marked using three different techniques:

1. Marking the text with a single set of radio buttons for all paragraphs.
2. Marking each paragraph with radio buttons.
3. Marking the text by hand using the technique referred to as "hieroglyphic marking" by Louw (2006), which is in essence marking with scribbled notes.

The students then had to revise their paragraphs. The original and the revised paragraphs were retyped, randomly shuffled and then marked again by four different markers using radio buttons. The results obtained were used to compare the original and revised versions of the paragraphs to establish how effective the various marking techniques were.

A pre-intervention and post-intervention comparison was done with χ^2 to determine whether a statistically significant improvement had been obtained, and whether any one of the techniques had led to improvement. The proposed technique was compared to practices adopted by lecturers in Academic Literacy.

2.5 Article 5: Yes again – another case for Boolean feedback, or "how to mark essays with strategic 'yes' and 'no'"

2.5.1 Main question

Article 3 above focuses on the use of Boolean feedback for improving and partial standardising of feedback and assessment on paragraphing skills. A similar technique was used to provide feedback on overall textual cohesion, which includes feedback on introductions, conclusions and overall textual cohesion within the text. Once again, a pre-intervention and post-intervention comparison was done with χ^2 to establish the validity of the results.

The main question of this article is therefore to what extent radio buttons can be used to provide improved and more standardised feedback and assessment on *overall text structures, introductions* and *conclusions*, to the benefit of both the lecturer and the student. The argumentation structure of a text is in this sense operationalised as the relation between the introduction, the execution of stated intent and the conclusion.

2.5.2 Purpose of the article

The purpose of the article is to test and refine a series of radio button-based questions (Boolean feedback) which could be used to provide effective, relatively standard feedback on and assessment of text cohesion and structure. The article incorporates research findings from the overlapping areas of text linguistics (Halliday, 1976), composition training (Ferris, 2003; Gennrich, 1997), assessment (Bacha, 2001) and SLA (Katznelson, Perpignan and Ruben, 2001).

2.5.3 Methodology

Short essays by first-year students of Academic Literacy were marked using Boolean feedback dealing specifically with introductions, conclusions and textual coherence. The essays were then rewritten by the students. A control group received feedback using hieroglyphics. The results of the two groups were compared statistically.

2.6 Outline of argument

The argument in this thesis revolves around two main problems – what works as feedback, and how should one apply this knowledge? The two issues are investigated in article format, with some findings leading to new research questions.

Due to the article format of the thesis and the intended computerised consolidation of the techniques, considerable overlap occurs between the problem statements and literature reviews of the five articles. Article 1 is an attempt at situating the research within the broader context of CALL. Most of the overall literature review is covered by Articles 1 and 2. Article 2 (a checklist for effective feedback) then serves as a guiding principle for the design and investigation of the technique reported on in Articles 4 and 5. In other words, Article 2 refines the approach and establishes the benchmark for feedback practice. Article 2 identified additional problems which had to be covered by the technique reported on in Articles 4 and 5.

2.7 Bibliography for chapters 1 and 2[1]

Askew, S. & Lodge, C. 2000. Gifts, ping-pong and loops – linking feedback and learning. In Askew, S., (ed.) Feedback for learning, pp. 1-17. London: RoutledgeFalmer.

Bacha, N. 2001. Writing evaluation: what can analytic versus holistic essay scoring tell us? *System* 29 p. 371-383.

Chandler, J. 2008. Response to Truscott. *Journal of Second Language Writing* 18:57-58.

Davies, G. & Hewer, S. (2009) Introduction to new technologies and how they can contribute to language learning and teaching. Module 1.1 in Davies G. (ed.) *Information and communications technology for language teachers (ICT4LT)*, Slough, Thames Valley University [Online]. Available from: http://www.ict4lt.org/en/en_mod1-1.htm [Accessed 16 September 2009].

Davies, G., Hamilton, R., Weideman, B., Gabel, S., Legenhausen, L., Meus, V. & Myers, S. 2009. Managing a multimedia language centre. Module 3.1 in Davies, G. (ed.) *Information and communications technology for language teachers (ICT4LT),* Slough, Thames Valley University [Online}. Available from: http://www.ict4lt.org/en/en_mod3-1.htm [Accessed 4 August 2009].

Deng, X. 2009. The case for writing centres. *English Language Teaching World Online.* Website: http://blog.nus.edu.sg/eltwo/2009/08/12/the-case-for-writing-centres/html

Ellis, R. 1994. Understanding second language acquisition. Oxford: Oxford University Press.

Ferris, D.L. 2002. Treatment of error in second language student writing. Michigan: The University of Michigan Press.

Ferris, D.L. 2003. Response to student writing. Implications for second language students. London: Lawrence Erlbaum.

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

Gennrich, D. 1997. Teaching writing for 2005: new models of assessment and feedback. *The NAETE Journal* 12:16-31.

Granger, S. 2002. A bird’s-eye view of learner corpus research. In Granger, S. Hung, J and Petch-Tyson, S. (eds.) Computer learner corpora, second language acquisition and foreign language teaching. Amsterdam: John Benjamins Publishing Company.

Halliday, M.A.K. & Matthiessen, C.M.I.M. 2004. An introduction to functional grammar. Third ed. London: Arnold.

[1] The author inserted separate bibliographies for the separate sections of the thesis, since the article model has been followed. Each article already had its own bibliography. The additional sections (the two introductory chapters, the preludes to the articles and the postscripts to the articles) sometimes contain additional sources not in the articles, which necessitated separate bibliographies.

Halliday, M.A.K. 1976. Cohesion in English. Oxford: Oxford University Press.

Hattingh, K. 2009. The validation of a rating scale for the assessment of compositions of ESL. Potchefstroom: North-West University(PhD thesis).

Hyland, F. 2003. Focusing on form: student engagement with teacher feedback. *System,* 31:217-230.

Hyland, F. 1998. The impact of teacher written feedback on individual writers. *Journal of Second Language Writing,* 7:255-286.

James, C. 1997. Errors in language learning and use – exploring error analysis. Edinburgh Gate: Longman.

Katznelson, H., Perpignan, H. & Rubin, B. 2001. What develops along with the development of second language writing? Exploring the "by-products." *Journal of Second Language Writing,* 10:141-159.

Kasanga, L. 2004. Students' response to peer and teacher feedback in a first-year writing course. *Journal for language teaching.* 38(1):64-100.

Krapels, A.R. 1990. An overview of second language writing process research. In Kroll, B., *ed.* Second language writing – research insights for the classroom. Cambridge University Press. p. 11-23.

Kroll, B. ed. 2003. Exploring the dynamics of second language writing. Cambridge University Press.

Louw, H. 2006. Standardising written feedback on L2 student writing. North-West University (Potchefstroom Campus). MA dissertation.

Louw, H. 2007. Moving to more than editing: standardised feedback in practice. *Ensovoort*, 11(2):83-104.

Matsuda, P.K. 2003. Second language writing in the twentieth century: a situated historical perspective. In Kroll, B. ed. Exploring the dynamics of second language writing. Cambridge University Press.

Ngu, B.H. & Rethinasamy, S. 2006. Evaluating a CALL software on the learning of English prepositions. *Computers & Education* 47: 41–55.

Spencer, B. 1998. Responding to student writing: strategies for a distance-teaching context. Pretoria: University of South Africa. D.Litt thesis.

Truscott, J. 1996. The case against grammar correction in L2 writing classes. *Journal of Second Language Writing*, 46:327-369.

Trushkina, J. 2006. Automatic error detection in second language learners' writing. *Language Matters*, 37(2): 125-140.

Weideman, A. 2007. A responsible agenda for applied linguistics: confessions of a philosopher. *Per Linguam,* 23(2): 29-53

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

Zamel, V. 1985. Responding to student writing. *Tesol Quarterly,* 19: 1 March

CHAPTER 3
ARTICLE 1 – DESIGN CONSIDERATIONS FOR CALL BASED ON EVALUATION CRITERIA FOR CALL

3.1 Prelude to Article 1

While this thesis focuses first and foremost on the provision of effective feedback on student writing, the ultimate goal is to implement the knowledge gained in a computerised marking interface. This immediately broadens the study and places it within a much-researched field – Computer-Assisted Language Learning (CALL).

As part of the development and planning, an investigation was launched to answer the question: "What constitutes effective CALL?" The research followed an approach in which both positives and negatives associated with the concept of CALL were investigated in order to create a rubric which could be used to evaluate CALL software. Based on the common business practice of first finding out what the end user wants, and then designing the product, this set of evaluation considerations was then used as design considerations for CALL.

CALL is, however, an immense research field in itself, spanning numerous pedagogical and linguistic research fields. This article is only a very brief overview of the concepts.

Publication information for Article 1

The article was submitted for review and publication to the *Journal for Language Teaching*, but at the time of writing the reviewing process was still in progress. Minor editorial changes were necessary for adherence to the general format and layout of this thesis.

Abstract

This article presents a rubric for the evaluation of Computer-Assisted Language Learning (CALL) software based on international recommendations for effective CALL. After a brief overview of the pedagogical and implementation fundamentals of CALL, and a discussion of what should be included in a needs analysis for CALL evaluation, the rubric is presented. The author then illustrates how the evaluation criteria in the rubric can be used in the design of a new CALL system.

Keywords

Software evaluation, CALL, language laboratory, MarkWrite, writing across the curriculum, software development

3.2 Introduction

Computer-assisted language learning (CALL) came onto the scene of language pedagogy almost at the same time as the advent of the personal computer but, much as in the case of automatic translation, has not made as much headway as was once enthusiastically expected (Hémard, 2006). A disappointingly small number of lecturers espouse the use of

CALL. Just as new language learning books are continually published, new computer programs for learning and coaching languages are also produced annually. While the criteria for the creation of a text book are relatively fixed, the criteria for evaluating and designing CALL are not as cast in stone, largely because of the immense possibilities of the medium and the number of variables to take into account. Just as a book is written with the reader in mind, and evaluated accordingly, this article argues that a CALL system should be evaluated *as well as designed* with a set of detailed considerations in mind.

While many articles have been written on the evaluation of software (many of them are provided in the bibliography), none of them are complete enough to use in practice to make an informed decision on the best CALL package to purchase. Even worse, none of them makes explicit the link between what is evaluated and what is designed. To put it more bluntly: a complete evaluation grid could not be found for CALL software packages. Secondly, it seems as if designers and evaluators work from two different rule books – designers design what they think is needed, whereas evaluators evaluate according to their needs. It is easy to identify a few reasons for this. For example, the complex nature of language pedagogy and the individual needs dictated by different contexts may be crucial to why software is evaluated differently in specific situations.

The cost of both purchasing software (when compared to books) and developing software (also compared to books) simply increases the urgency of establishing a detailed set of considerations for evaluating and designing software.

The purpose of the article is twofold – to establish which considerations should be used to evaluate a CALL system and to illustrate how these considerations could be used in the design of a CALL system. The design of the new software program MarkWrite is used as an example to illustrate how these considerations may be used.

Although there are many similarities between selecting a CALL system and selecting a new course book for a module, selecting a CALL system is more difficult. One obvious difference is that it is easier to quickly page through a book to get an overview of its contents, while it is a considerably bigger and costlier task to buy, install and evaluate a piece of software. The more interactive nature of software also provides many variables which are not part of the evaluation of a textbook, such as more graphics, sound and audiovisual material, navigation, system requirements, and many other computer-specific considerations. The long-term implications of the choice of software may be more severe, since it is often an institution which has to buy software, whereas it is the students who buy new textbooks annually for themselves.

This article first provides general considerations for the use of CALL from national and international literature, after which a set of evaluation criteria for CALL is proposed. The evaluation criteria are then used as design criteria, using the MarkWrite software as an example.

3.3 Basic understandings and terminology of CALL from the literature

Computer-assisted language learning, as most things in the computer world, is a fast-developing discipline. Some terms and agreements have come into effect regarding the design and evaluation of CALL systems. Among these are eight fundamentals, falling into roughly pedagogical and implementation categories, each of which is elaborated on:

Pedagogical fundamentals

1. CALL-systems need to have a solid educational base and be integrated into the whole teaching curriculum. The effectiveness of CALL depends on an effective pedagogical base and not on the computer. Good pedagogical practice is just as necessary for CALL as it is in traditional teaching.

2. CALL should be seen as a teaching tool. A computer is not a human. Each may be better at their respective tasks, but CALL is and remains a teaching tool and cannot be a total solution by itself, even if the system is marketed as a stand-alone application.

3. There is a difference between a coach and an independent teaching tool. Although all CALL-systems are teaching tools at the disposal of humans, some CALL programs are designed to be utilised *only* as tools while others are stand-alone applications designed to teach *independently* of human teachers.

4. CALL systems are specialised teaching tools. There is no *single* computer program available to do comprehensive CALL. For completely computerised language learning to take place, it would be necessary to invest in more than one program, each focusing on a different aspect of teaching.

Implementation of CALL

1. Effective CALL is dependent on an effective policy. A policy must be created for the use of CALL in a language laboratory, classroom or module. Such a policy would have to indicate how the other considerations are to be handled.

2. CALL is dependent on a CALL environment. The whole context of the teaching situation influences the effectiveness of the CALL system. A CALL environment is not dependent simply on the quality of the computers or the software – the organisation of the physical classroom and the implementation of the programs are also important. This is just as relevant if CALL does not take place in a language laboratory, but is instead used as self-study tool.

3. All staff members need to be trained in the use of the program, but a dedicated CALL technician or CALL manager has to be appointed.

4. Software has to be evaluated annually to establish if it is still relevant to the specific educational situation. It is no use sticking to a program that does not perform or is not suited to the specific environment or application, or which does not deliver as promised.

Each of the above aspects is now dealt with in more detail.

3.4 CALL systems need to have a solid educational base and be integrated into the whole teaching curriculum

Teachers should assess software just as carefully as they scrutinise textbooks. Electronic resources can by their nature provide access to authentic language samples, but it is up to the teacher to structure activities and projects that promote meaningful interaction with these materials (Bradin, 1999:175; Buell, 1999:217).

Barr and Gillespie (2003:69) explain that "a computer-based environment needs to be carefully constructed in order to ensure that all the other components of learning are effectively integrated into it. It is important to ensure that the uses of computer technology in this type of environment are not seen as separate, but rather that they are integrated, working together to enhance the process of teaching and learning … CALL packages must not be seen as stand-alone creations." The technology must be used to integrate the learning and teaching methods with the resources available. Technology need not be used at all costs. "If other more conventional teaching and learning methods work well, then there is little point in using computer technology" (*ibid*).

Technology should be support for a total **environment for learning**, instead of a stand-alone tool or source of information only. Technology can change how, what and whom is taught, but it is more important to understand good pedagogy than to understand the technology (Egbert *et al.*, 1999:ix-1; Murray and Barnes, 1998). The pedagogical goals must be clear and the use of technology must further these goals (Bradin, 1999:160), while all CALL should be grounded in sound teaching methods (Levy, 1997).

Egbert *et al.* (1999:2-3) make it clear that despite the large number of language acquisition and language learning theories, researchers and teachers generally accept that:

- language acquisition is the result of an interplay between some kind of cognitive mechanism and environmental factors;
- not all language learners learn in the same way, at the same rate or for the same purposes;
- interaction between learners and other speakers is very important.

Very broadly, the following assumptions about learning are evident in the pedagogical theories currently in use:

- "Language learners must be involved not only in social interaction but in *purposeful* interaction, which includes a real audience that is actively involved with the learners" (Egbert *et al.*, 1999:4).
- Learners should have an authentic goal for their work. Authentic tasks have the same type of cognitive challenges as complicated real-world tasks. "It is important to design tasks so that students can use their current proficiency level to function in authentic communications" (Egbert *et al.*, 1999:5). The Internet and e-mail are useful tools for real, live communicative tasks.
- Learners should be exposed to a variety of sources of input (Egbert *et al.*, 1999:5).

- Educators must assist the learners by creating an environment with an optimal stress level. This is done by creating a learner-centred classroom in which learners have some control over their learning. The educators' expectations must be reasonable and the goals attainable (Egbert *et al.*, 1999:6).
- A learner-centred classroom is necessary. Such a classroom is one that develops learners' confidence and skills to learn autonomously. Learners should also be able to design and coordinate tasks in a variety of contexts (Egbert *et al.*, 1999:6).
- Peyton (1999:17) writes that all learners move in their speaking and thinking towards a stage where they can function alone.
- Written communication is important for learning. Peyton (1999:17) notes, "Most work on the dynamics of interaction and their effect on learning has focused on oral interaction. However, research on written interaction in dialogue journals with teachers and in letters exchanged with older students has shown that these interactions can also develop language, thought, and reading and writing abilities".
- Co-operative learning is important. Staton points out, "To be able to think in new situations – which is the real goal of all education – [learners] need a lot of experience in thinking with someone who is good at it. Just as we learn language by talking with someone who is good at it in specific situations concerning tangible, shared experiences, so we learn to think by thinking with someone to solve a joint task or problem" (quoted by Peyton, 1999:18).

Many of the above criteria can be met by the judicious use of computers in the classroom. Egbert *et al.* (1999:8) write that "… just as there seems to be no one right way to teach or learn language, there is most likely no one best way to use computers for language learning." Their argument encompasses five critical questions to ask about the computer-assisted classroom:

1. How effective is group work as an aid to L2 learning?
2. Should students drill and practice new structures?
3. What can be done to encourage participation among students who seldom ask questions or initiate interaction?
4. To what extent does the correction of errors assist in L2 learning?
5. Which technologies are best for supporting the best methods of teaching and learning?

All five of these questions are just as applicable to pedagogy in normal computerless classrooms.

3.5 Teaching tool

A computer cannot replace a human teacher. Bradin (1999:159) explains that any appraisal of a software package should be based on knowledge of what computers are capable of and what the inherent drawbacks of computers are. Teachers and computers may each be better at certain specific tasks (Egbert *et al.*, 1999:9). Ma and Kelly (2006: 21) trace CALL efficiency back to information available about the learner, which should influence *theory*, *computer technology* and *user actions*. Computer-assisted pedagogy is therefore much the same as

normal human interaction pedagogy, but a CALL classroom (teaching situation) cannot be directly compared to a normal classroom as the technology introduces many new variables. **Course content** should be emphasised as the focus of instruction – not the computer. The computer is a tool and not a human teacher (Sivert & Egbert, 1999:46), but since a computer communicates with learners, they can in fact learn communicative competence by simply using the computer (Chapelle, 2003:11). For example, the computer asks questions of the user such as: "Do you want to save the changes made to document X?" However, learners should already have basic computer literacy to enable them to use CALL software.

As a tool, computers have advantages. Peyton (1999:17) describes the following advantages of computers:

- It is a medium of communication that creates new opportunities for writing and learning.
- Computers provide synchronous (real-time) and asynchronous (time-delayed) interaction.
- One-on-one interaction between students and teachers or among students within classrooms is possible.
- Wider communication with individuals and groups around the world is possible as well.
- Text and talk are available in the classroom and in a rapidly expanding world.
- Resources are not bound by physical space.

Despite the above qualities, computers cannot be compared directly to a human teacher, but should be seen as very effective *tools* which can be utilised by a human teacher. Computers do not replace teachers, but simply change the nature of their work. Chapelle (2003:xiii) states that although the use of technology is regarded as the obvious (unmarked, normal and natural) way to go, the case to be made is that CALL should not be compared with classroom language learning. Rather, "the challenge is to provide evidence for the most effective ways to design software for CALL, to use the software effectively in tasks, and to help learners to take advantage of the electronic resources available to them" (*ibid.).* In other words, use the computer as a tool and not as a teacher replacement.

To elaborate on the difference between human teachers and computer tools, consider the common complaints against CALL systems. A 2003 survey of CALL systems done by the author identified common weaknesses in almost all the CALL software evaluated. For obvious reasons the programs surveyed cannot be mentioned here. These weaknesses are:

- Students can click through most of the screens without filling in something. Common pedagogical principles require that students should take some action. While many human teachers also allow students to be passive, this is not good practice.
- If the program requires something to be filled in, the user can simply type gibberish and the computer will accept it. Great advances have been made to counter the gibberish effect, but in some cases the software is still not able to detect it.

- The interfaces are in most cases very plain (which in itself is not a problem), but there is a lot of space that could have been used for additional information or supporting graphics.
- Most of the programs do have supporting graphics, but these vary in their effectiveness. The programs that the author found to be the most interesting and effective had very good supporting graphics.
- Most of the programs do not provide adequate feedback. Especially in a computer environment where one can basically "click until you get it right," feedback is needed on correct **and** wrong answers.

Students may also be sceptical of the computer's ability to judge their work effectively (Spencer & Louw, 2009). It should, however, be kept in mind that sometimes students are also sceptical of their human teachers' ability to judge their work effectively.

A final consideration to keep in mind is the distinction between software for language improvement (used by students who already know the language to an extent) and software aimed at teaching an unknown language from scratch. This article deals only with software for language improvement.

3.6 Coaches versus tools

When considering CALL for a university or school set-up, there are essentially two types of systems available in two different mediums: *coaches* and *tools* are available *online* (local area network or the Internet) and for *local* (stand-alone PC) units. A discussion of the differences and advantages/disadvantages of coaches and tools merits a whole separate article. Davies and Hewer (2009) distinguish between six different types of software and more than 20 different applications of Information and Communication Technology (ICT). The simplified distinction will be presented here:

1. **Coaches** are stand-alone units designed to be accessed by the students and worked through from start to finish. The programs are pedagogical (computerised "teachers"), providing lessons, examples, prompts and tests. The tests can comprise multiple choice and drag-and-drop quizzes. Many of the more modern systems are language coaches simulating the complete teaching experience. Someone can buy such a system from the local computer store or book store and independently start learning a language from scratch. In the South African context, an example would be the language learning software created by the Centre for Text Technology at the North-West University, such as Tsenang (Berg & Pretorius, 2003). More international language learning software, such as the Pimsleur or Rosetta Stone software, is sold over the Internet.
2. **Tools** are aids to teaching. They may also have lessons, examples, prompts and tests, but in addition they can be altered by the teacher to suit his or her own needs. In some cases the tools do not have any lessons or examples, unless the lecturer creates them. The most basic of these tool programs simply provide the students with drill-type exercises and are seldom if ever used independently of a structured curriculum. Standard corpus linguistics software used to analyse student writing may

also be considered a tool and its use as such is discussed by Cowan, Choi and Kim (2003) and Granger (2003).

Some programs have a composite nature (they are both tools and coaches). They aim at developing skills in some or all areas of language and provide tools to lecturers to assist them in customising the software to their specific needs.

Arnett (2009:27) however, warns against an over-dependence on computer tools. In his article he bemoans the poor ability of students to spell, and their blind reliance on their computer grammar and spelling checker. He implicitly warns against turning tools (such as a spelling checker) into a coach.

3.7 Specialisation

At present, most available computer programs are very specialised in content. For example, some are just aimed at improving writing, whereas others are just for improving vocabulary. To address more than one aspect of learning, more than one computer program would be necessary. One piece of software is never enough (Sivert & Egbert, 1999).

Related to this is the implementation of the CALL-software. One must establish whether the software will be used as bought for a specific module as a whole, or as support for one specific outcome in a module. One also needs to establish whether the CALL system will be used as a stand-alone module in a CALL laboratory where students have to work through it as part of their class activities, or whether it is simply support for them which they should work through at their own pace at home. As with normal classroom interventions, custom-made teaching materials are better than commercially available, generic material. CALL software which can be adapted to the specific situation is therefore a better option than generic software. See Chapelle (2003); Cowan, Choi and Kim (2003); and Granger (2003) on this topic. Specifically Granger (2003) and Cowan *et al.* (2003) use the computer to enhance their custom-made exercises.

3.8 Policy

Aligning module outcomes and their application is difficult whenever more than one person is working on the same module or set of modules. Adding a computerised section or computerised support for the module simply adds another variable and another participant to the mix which necessitates structured collaboration. Consider the view of Barr and Gillespie (2003:78):

> All learning environments are complex, but computer-based language-learning environments are particularly complex, so co-ordination is vital. We have found that the most successful integration of environments has occurred when there has been high-level management support for the development of such environments, not only in providing the finance, but also in shaping the direction and motivation of the system.

Bishop (1999) distinguishes between a mission for the CALL environment, policy for the classroom, and policy for the software. Guidelines for the policy of a CALL laboratory or classroom according to Bishop (1999:272) should state:

- when and how hardware and software are upgraded;
- who is allowed to download and upload files; and
- where information on policy is kept and what it is used for.

In addition, the mission of a CALL environment should ask the following questions:

- What is the instructional rationale for the use of CALL?
- Are the computer applications appropriate, effective and intelligently applied?
- What student results are expected?

(Bishop, 1999:272).

The following are Bishop's (1999:275-276) recommendations for a *software policy*:

- Recommendations for new software will have to be made at least annually or two or three times per year. Recommendations must be archived with the reasons for the recommendations.
- The software evaluation form can be developed in-house. The idea is that the form should prevent duplication of recommendations and also provide ideas for the use of the available software.
- A minimum 30-day trial period of any software considered for purchase is recommended. Personal experience and observation are very important.
- The linking of software with learning objectives is crucial: the software may be excellent, but if it does not fit your curriculum, you are wasting your money.
- No-one but the system administrator should be able to install software on the computers.

The various considerations for CALL software and the evaluation thereof could fill volumes, but the above points highlight some of the core principles. A next step in evaluating CALL software is to do a needs analysis.

3.9 Environment

Materials are not inherently better just because they are on the computer. CALL should ideally occur in an optimised learning environment. Sivert and Egbert (1999:41) describe the ideal technology-enhanced language learning classroom:

> The word *classroom* implies a place where different kinds of learning can take place and where technology use is subordinate to discovery and understanding. In this setting, learners enter a classroom designed for comfort and collaborative learning. They ideally find a cushioned seat equipped with casters in front of a large desk with a recessed monitor; each desk is part of a group of four desks facing one another. Books, papers, and pens are spread over the quad of desks. As learners begin their work, they move easily among other members of their quad and even among other members of the class and the instructor. Instead of concentrated silence, one hears

the lively discussion of learners working together on task-oriented and project-based assignments. The software available assists in driving the assignments, and several other media are used in developing and completing the tasks.

The resources available may not always permit such a classroom, but three types of technology-enhanced language learning classrooms exist:

1. The self-access lab
2. The computerised instructional classroom
3. The language development centre

(Sivert & Egbert, 1999:42).

Each of these has its own advantages and disadvantages which Sivert and Egbert (1999:42) discuss in more detail. As mentioned earlier, the use of CALL may in fact not be dependent on a specific classroom as students may work independently at home or simply learn from the computer as an additional source of input. The immersion principle suggests that learners are given enough exposure to the target language to develop their ability to comprehend the language. With the massive growth of the Internet, virtual immersion is possible for any student, but instruction would still be necessary to form and shape language acquisition (Chapelle, 2003:36).

A separate consideration under the environment is the level of interactivity which is available and the extent to which this interactivity is desirable. Some CALL systems (especially the online versions) introduce students to other students in other parts of the world as "chatting partners" to enhance their written communication skills in a real-world setting, and it appears to work (Fitze, 2006; Yuan, 2003). However, the poor quality of "chat room English", Mxit language and cellphone abbreviations (SMSs), may be the very reason why most teachers will discourage these practices despite research findings indicating that they do not have such a big adverse effect on students as is proclaimed in staff tea rooms and corridors (Plester, Wood and Bell, 2008).

Whichever model is chosen, it should be kept in mind that CALL is not a stand-alone activity. The specific outcomes of the computerised component should be discussed and settled on by all stakeholders, and the computerised component explicitly integrated into the curriculum. Everybody working with students will know that if something does not count for marks, very few students will work on it independently. This also means that a participation mark is not acceptable, since it degrades the perceived status of the computer component (Spencer & Louw, 2008).

3.10 Staff

A surprising number of academic staff (from all disciplines) are closet technophobes and are allowed to be so by their institutions. Technophobia may result in initial negative reactions to software (as also noted by Murray & Barnes, 1998:250). It may take some convincing to persuade especially older staff to (a) see the benefits of CALL and (b) once they have realised the benefits, to consent to being trained in its use. A single pundit of CALL at a campus is

simply not enough to successfully implement CALL, and this person will soon tire of continually having to motivate or nag his or her colleagues to use the available technology.

With regard to training, Bishop (1999: 278) makes it clear that **all staff members** should be explicitly trained in the use of the program or programs. It is no use having one expert whose absence causes the system to come to a halt. All teachers should know what the software is able to do to optimise its use. "Too many beautiful CALL centres have good software but only one person who knows about it. Such a centre is not a CALL environment; it is just a room full of computers." Barr and Gillespie (2003:77) have found that untrained, uninvolved or uninformed staff cause students to react negatively to CALL, as well as when there is limited, incomplete or inadequate equipment. It should be clear therefore that untrained staff (and staff who are unwilling to be trained) can scuttle the complete operation, and ignorance should therefore be avoided at all costs. Davies *et al.* (2009) also caution that staff training is an "ongoing and unending process" for which funding should be made available every year.

3.11 Relevance

CALL should continually be assessed to ensure its relevance to the situation. It is important to note the Hawthorne effect: "Any group that is being studied while doing a new or different activity usually performs better" (Egbert *et al.* 1999:10). Therefore CALL-systems should be evaluated annually to see if the programs are furthering the specific learning goals (Egbert *et al.*, 1999:11-12; Bradin, 1999:160). It is very important to ensure that a program is not just being used because it is available. "Every piece of software in the CALL centre must have a direct, positive impact on and relationship to what is being taught elsewhere in the school" (Bishop, 1999:274). Most programs create their own sets of data which can be analysed, but it is imperative to state in a CALL policy who gets access to this data in order to preserve the students' rights (Bishop, 1999:281-282). The evaluation should also be done by using a definite set of criteria stated in the CALL policy and it should be done *while using the system*. This is very important for research.

In the light of the above, Bishop (1999:273) provides some acceptable and some poor reasons for using CALL:

ACCEPTABLE reasons for using CALL:

1. Classes are too large to monitor **individual** progress.
2. Students in the same class are of **varying levels** and need more individual attention.
3. Students need CALL to prepare for their **real-life** business environments.
4. Students need more **one-on-one practice** than they can get in the classroom.
5. Students have to do **collaborative projects** as a means of enhancing communicative skills.
6. CALL provides for enriched, alternative means of **communication.**
7. CALL provides practice in a skill or an introduction to concepts that cannot be offered otherwise.

While these all seem like legitimate reasons, the tipping-point question is still most probably: “Does it work?” Davies and Hewer (2009) discuss the reasons for using CALL at length, starting with:

> **Concrete evidence** on the effectiveness of CALL is difficult to obtain. There is plenty of **anecdotal evidence** about the positive effects of CALL. Teachers often report on their students being "enthusiastic", "engaged", "motivated" and even "excited" in classes in which CALL is used, but are sceptical about **measuring** its effectiveness.

While anecdotal evidence is not overly convincing, the authors cite many different case studies and other research on the effectiveness of CALL which appear to point in the general direction of CALL being effective. Ngu and Rethinasamy (2006) however, found traditional teaching to be more effective in their context. Also compare Murray and Barnes’ (1998) discussion of how to get past the initial impression of software to evaluate its true pedagogical value. Tsiringa and Virvou (2004) tested their students with a pre-test and post-test, but they were only comparing two different kinds of software. The pre-test and post-test approach might be useful. For this approach to work, however, the software has to be purchased, and the question remains how to efficiently evaluate software without going to the trouble and expense of purchasing and implementing it. The evaluator should be able to determine whether a CALL system is likely to work before implementing and testing it.

Davies and Hewer (2009) also indicate that the immediateness of feedback is considered a big advantage of CALL by students and throughout their lengthy discussion of five different CALL centres in Europe Davies *et al.* (2009) mention many additional advantages of using CALL, some of which contradict Bishop (1999). It seems that the advantages and disadvantages vary depending on the specific types of activity and the structure of the CALL, but in each case there are definite advantages and disadvantages to using CALL in that situation.

POOR reasons for using CALL according to Bishop (1999:273) are:

- CALL is enjoyable (Davies *et al.* (2009) differ on this, though).
- It is available.
- Everybody is doing it.
- Students want to play on computers (once again, Davies *et al.* (2009) state the opposite – playing on computers may motivate students).
- Computers keep students busy.
- The teacher needs some extra preparation time.

Davies *et al.* (2009) mention another poor reason – saving money. They state quite clearly that “technology is much more expensive than ‘chalk and talk’…”, but that “unfortunately, administrators in schools and universities are prone to regard the use of technology as a means of cutting down on staff, in the belief that ‘throwing hardware at a problem’ will save money”. They refer specifically to a computerised language centre, but one can safely assume that their opinion also applies to CALL in general.

The issue of relevance has become even more evident with the advent of ICALL systems (*Intelligent* computer-assisted language learning systems) in which the computer learns what the students' specific needs are and adapts accordingly. While ICALL promises to be highly effective in individualising pedagogy, there is still a lot of research to be done before programs are made really intelligent. Most systems are intelligent on only a small part of the curriculum. The system evaluated by Tsiriga and Virvou (2004), for example, only focused on teaching the passive voice.

In short then, evaluating the effectiveness of CALL involves more factors than evaluating the effectiveness of a textbook or module. While the software is often more focused, there are many more variables to take into account.

3.12 Needs analysis

As with designing any new module, the first step in selecting software is to do a needs analysis with the parties involved. The findings of this needs analysis should be included in the evaluation rubric of the software.

The parties involved in CALL are, however, more diverse than those involved in simply designing a new module. In this case, three different sets of end users should be accommodated. Simultaneously, issues of integration, policy (already discussed) and budget are also "parties" to take into account. The three sets of users (in no particular order) are:

Set 1: the lecturer and the systems administrator

Set 2: the students and the systems administrator

Set 3: the university IT personnel and the systems administrator

The systems administrator is the only role player directly involved with all the other parties. Each of the identified parties is elaborated on below.

3.13 Lecturers

A needs analysis was done by the author by means of interviews at the Potchefstroom Campus of the North-West University for the purpose of setting up an evaluation rubric for the possible purchase of new software. The needs analysis found the following with regard to CALL: Lecturers want a program that is user friendly, does not require a lot of training, is intuitive and is adaptable to their needs. The program should decrease their workload and should not necessitate additional administration. The lecturer should also be in a position to indicate the specific outcomes required of the software.

These expectations are, however, overly optimistic of the abilities of the software. Lecturers may, for example, expect the same program to teach and correct grammar as well as teach argumentation and editing skills for various levels of student proficiency. This is obviously impossible for a single system to accomplish. Different linguistic skills are not directly equivalent to different levels of student proficiency and, unlike humans, a computer cannot hazard guesses as to the reasons why students make certain errors. It is important therefore that the lecturer should rank the requirements according to desirability. This ranking may

then be cross-correlated with the ability of the software to accomplish the outcome and the value it will add for the lecturer. Something which is easy to teach and mark, but is time-consuming, would therefore be ideal to delegate to a computer, on condition that it is done in a monitored, structured way.

3.14 Student needs

As indicated earlier, it is bad practice to use a system simply because it is available. The specific needs of the student should be taken into account. While the lecturers are often in a position in which they can indicate these needs, their intuition is not always accurate (Louw, 2007:96). It may be advisable to ask at least a few students to test a piece of software since their opinions are relevant seeing as they are the ones who will use the software (Lasagabaster & Sierra, 2003). It will further enhance the understanding of student needs if they write a standardised test to establish accurately what their exact language needs are.

A second important variable with regard to student needs is the time spent learning to use the software. Tsigira and Virvou (2004) found in their study that one system was more effective than another, but that the more effective one required more time to master by the students. Time is important to all, and students are quick to complain about anything which they deem to "waste" their time. A computerised language learning system, no matter how effective, should not be an add-on to an already full curriculum, otherwise students will see it as punishment. A cost/benefit analysis might be necessary to establish whether the time spent learning and using a new software package is worth the benefit provided by the package.

3.15 Systems administrator and IT personnel

It is preferable to have a separate systems administrator and IT support person, but in some instances the budget constraints will not allow for that.

3.15.1 Systems administrator

In a CALL system, systems administration is not a small job. Any system, no matter how advanced, has to be monitored and administered. This job should not just be dumped on the person with the most IT skill but should be assigned to a person who is aware of the module outcomes and who can make sure that the system is used to meet these outcomes. It is also advisable to establish beforehand the amount of time which could be expected to be spent administering the system. Some considerations here are:

1. Assisting students who are not computer literate enough to use the system
2. Answering e-mail queries about the system (Spencer & Louw, 2008:121)
3. Troubleshooting bugs in the software
4. Keeping the software updated
5. Setting assignments in the software
6. Selecting specific assignments to do
7. Drawing a report from the software
8. Calculating and assigning marks for exercises or assignments done on the software.

The above considerations come from personal experience as a systems administrator as well as interviews with other practitioners and the University IT personnel over a period of five years.

3.15.2 IT specialist

It should be clear from the above list that the responsibilities of the systems manager and the IT specialist may overlap. This largely depends on whether or not the system used is stand-alone software, a computer laboratory, or an online educational package. The job of the IT specialist will differ according to the same specifications. His job all but disappears if it is stand-alone software which the students take home. If the software is used in a LAN-based environment in a computer laboratory, he then has the most work, since he has to make sure that all the computers are working without a glitch. LAN-based software is sometimes loaded on a central server which may or may not be administered by the University's central IT administration. The IT specialist then has the job to liaise with them. Since these requirements and internal red tape will differ from university to university, a discussion of this is not warranted for the purposes of this article.

3.16 IT infrastructure and software costs

The university's IT expert should be in the position to advise on what the situation-specific infrastructure needs are and what the existing infrastructure can handle. While most universities nowadays provide computer access for students, Internet use is still expensive in South Africa, with South Africa ranking 66th of 114 surveyed countries in The Global Information Technology Report 2008-2009 (Dutta & Mia, 2010:350). Students are quick to complain if the software they are required to use requires them to go onto the Internet. This is especially true if they have to pay for their Internet use, or if they are required to work on their own at home on the software.

While engineering students are often required to purchase software packages, the purchase of electronic support for languages has apparently not yet caught on. Prescribing a set of software in addition to the normal coursework books will no doubt initially upset a number of students.

In South Africa, one can also not expect all students to have their own computers. Requiring an additional few hundred students to use computers for language learning may place an unbudgeted for strain on computer systems at the University. When planning to use CALL, it is therefore advisable to confer with the local IT infrastructure representatives.

Davies *et al.* (2009) present a lengthy discussion on the costs involved in setting up and managing a multimedia language centre, and they also issue a stern warning regarding copyright fines. If a computerised system requires workbooks to accompany it, the copyright laws must be adhered to. The number of variables to take into account when referring to the cost of a language centre are too numerous to discuss within the constraints of this article. Suffice it to say therefore that the cost of the software itself (as distinct from the complete CALL environment) should obviously be within reasonable limits.

Another important aspect to take into account when considering the infrastructure is quite simply where the students are. If the students are all on one campus, the campus layout and availability of computer laboratories may influence networking. One should also consider whether the students will have access to a lecturer or administrator if they need help. If distance students are somewhere in a remote location, completely cut off from human support such as other students or lecturers, they may be dependent on waiting for e-mail replies or telephone call-backs.

In short, any laboratory setting has the following variables regarding the infrastructure:

1. *Physical resources:* classrooms, laboratory space, libraries and academic offices.
2. *Technological resources:* the provision of up-to-date computers.
3. *Communication:* the management of information and its transmission to all involved in the learning process.
4. *Human resources:* staff who are trained to teach students and eager to adopt new methods and technologies. Significant technical support not only for the maintenance of hardware but the development of teaching materials is also required.
5. *Pedagogical strategies:* teaching strategies need to be drawn up and related to the delivery of the curriculum.
6. *Cultural context:* the approach to learning adopted by staff and students.

(Barr and Gillespie, 2003:69)

3.17 Budget

CALL is expensive and is not just a way to get by with less staff. Instead, it should be seen as a way to multiply the effectiveness of the available staff, even if that means paying extra initially. A decision should be taken beforehand on how much to spend although it is important to remain flexible in one's expectations. Different budget considerations are required for different types of software. Students are not likely to buy software costing more than a few hundred rand, while a server-based software package can easily cost hundreds of thousands of rand.

In addition, server-based software and Internet-based software may require the license to be renewed annually. It is necessary to establish how the rate increase will be calculated in order to avoid later problems. For example, it is possible that a good introductory price will later be raised to astronomical proportions (see Spencer & Louw, 2008). CD-based software may also require annual (or monthly) updates which (while often free) may require large downloads. Internet cost should therefore be taken into consideration, especially in Africa.

A needs analysis is not something to be taken lightly. The large number of variables and role players make this important step a time and resource-consuming activity.

3.18 Evaluating software

Based on the discussion above, and using the information from Barr and Gillespie (2003) and Bradin (1999), as well as discussions with local IT managers and lecturers, the following guidelines are provided to evaluate software. Just as a needs analysis is an extensive

process, exploring the software can be just as daunting a task. Bradin (1999) proposed an evaluation system to determine what software to use and how. In his system, exploring the software is a two-step process involving feasibility and quality (Bradin, 1999:162).

Feasibility considerations according to Bradin are:

1. Will the software run on the specific available computer platform?
2. Do teachers and students know how to use the specific platform?
3. Will the software run on your network? If the software crashes on one workstation, can the program be restarted without interfering with the rest of the network?
4. Can the software be made available to many students? Can it be installed on a Web-server or even taken home to be installed on personal computers of the students?
5. Does the software require Internet access? Some programs offer interactive lessons via the Internet. This requires a very fast Internet connection and lots of RAM. (While this consideration may have been relevant in 1999, it is most probably not relevant in 2009.)
6. Can you afford it?

Most of the above feasibility considerations actually form part of the needs analysis. As far as the quality of the program is concerned, Bradin (1999:164-165) mentions three specific areas of consideration:

1. Content
2. Format
3. Operation

3.18.1 Content

Bradin proposes the following considerations regarding content:

1. What is the goal of the software?
2. Is the level appropriate?
3. Is the content accurate and up to date? Has it been proofread carefully?
4. Is the material culturally appropriate?
5. Does the software accommodate the students' learning styles and preferences?
6. Is the software interesting?
7. How flexible is the software? It could be necessary to use more than one type of software to accommodate different learning styles.

The considerations regarding content are very much applicable to the evaluation of the contents of books, although CALL also has many more variables.

3.18.2 Format

Books, websites and software have increasingly more impressive and colourful layouts and graphics. However, an evaluator should be wary of smoke and mirrors and pay attention to the following aspects:

1. Is the interface consistent?

2. Is the screen display effective?
    a. Is text size sufficient?
    b. Are colours distracting or do they add to the attractiveness of the screen?
    c. Is the quality of graphics sufficient that they are clear?
    d. Do the pictures and graphics add to the pedagogical effectiveness of the program or are they just gimmicks?
3. Are the motivational devices effective?
    a. Can the sound be disabled?

The format and layout of any software should above all be functional.

3.18.3 Operation

Evaluating how easy the software is to operate should not just be left to the evaluator. In this case, the instructor, systems manager and students should be given an opportunity to evaluate how easy the software is to operate. Bradin proposes the following:

1. Is the software easy to use?
2. Are the tasks and directions clear?
3. Can the text and graphics be printed?
4. How much control are the learners allowed?
5. How interactive is the software?
6. Are the quality and degree of the feedback adequate?
    a. Is it appropriate to the age of the intended audience?
    b. Is it immediate?
    c. Do correct answers also get feedback?
7. How good is the HELP file?
8. What kinds of records does the software keep?
    a. Can the records be printed?

When software is being tested, some of the students should be on the test panel. It is also possible that software producers or companies will only direct the customer to their model schools (Bradin, 1999:172), which may necessitate an individual investigation to find additional (possibly negative) information. Tsiriga and Virvou (2004:412) also support the idea that the software should be tested empirically by quantitative and qualitative means, on real students. Regrettably, the time, money and personnel are not always available to adhere to this recommendation.

3.19 Writing a program evaluation rubric

Based on the information in the above discussion, a rubric was created to evaluate available CALL software packages with the specific aim of assisting students in their writing. The original version of the rubric was created in 2003 to find software for the Potchefstroom Campus of the North-West University.

The rubric, which has been adapted over time, is presented below. A discussion follows on how such a rubric can be used to evaluate how a new CALL tool which was developed at the North-West University measures up to international recommendations for CALL.

The column marked “status” indicates the relevant importance of the feature. A “very important” feature has 10 marks allocated to it and the evaluator must then assign a mark out of 10 to the feature. Less important features are awarded a maximum of 5 marks and the evaluator likewise needs to assign a mark to it. The criteria are also explained in enough detail that different people will know what is expected of the specific variables.

Evaluating software is a tedious and time-consuming job. The evaluation rubric is not exhaustive but it should provide a good starting point for somebody to evaluate CALL software.

SOFTWARE NAME: .....

NETWORK REQUIREMENTS

CRITERIA	**DEFINITION AND NOTES**	**STATUS**	**Mark**
1. Works on my available operating system	Which operating system are you using? Some software will not work on the latest operating system. If you cannot load the software, you cannot continue with the evaluation.	Very important. If the program cannot work on your system, you obviously cannot continue the evaluation.	…/10
2. Web based	Can be loaded on a central server. For our purposes, we needed the software to run from a central server. For other purposes, it may not be important, but it may be more important that the software can be taken home and installed on a personal computer.	Very important	…/10
3. Good support	A good HELP file or good online support is essential. A local (South African) distributor for the software is an added bonus. Online support is less optimal than a built-in HELP file.	Very important	…/10
4. Budget	Cheaper than Rx for individual packages (take-home packages) or cheaper than Rz for server-based or Internet-based software.	Very important	…/10
5. Upgrading	Preferably free (take-home packages) Fixed yearly rate increase (server-based or Internet-based software)	Very important	…/10
6. Autonomous	Can students access and work on the program unsupervised?	Advantage	…/5
Total for **Network Requirements**			

CONTENT (Pedagogy)

CRITERIA	DEFINITION	STATUS	MARK
7. Software outcome	*Note: As mentioned above, a complete discussion of pedagogy would not fit in the scope of this article. The reader is advised to list his or her own pedagogical requirements here, but must be sure to be as clear as possible. Two examples are provided below.*		
	Example 1 – General vocabulary: The software should assist students in acquiring new general vocabulary words. It should indicate usage in everyday life, the applicable register, and should "sound" the pronunciation.	Very important	**…/10**
	Example 2 – Vocabulary exercises: the system should provide diverse types of vocabulary exercises. It should not give the same exercises to all the students.	Advantage	**…/5**
8. Level	Advanced L2	Very important	…/10
9. Accuracy of content	Up to date	Very important	…/10
	Error free	Very important	…/10
10. Culturally appropriate	Many software programs are American in content and the topics and discussions are unknown or strange to South African students.	Advantage	…/5
11. Interesting	It is difficult to establish what students will experience as interesting; however, you are not looking for a textbook on a screen. If the software fails to make use of the available resources provided by a computer, it fails to make use of the pedagogical possibilities of the medium and as such may be considered not as well planned as	Very important	…/10

	one would hope.		
12. Authorable	Depending on the outcomes of the CALL-system, it may be necessary for the lecturer to change some parts of the program content.	Advantage	…/5
13. Graphics	Do they add to the pedagogical effectiveness?	Advantage	…/5
14. Motivational devices present?	Does the program have devices to motivate students to do better? Examples include games, token "trophies" or triumphant sounds.	Advantage	…/5
15. Workshop	Can the program be linked so that learners can use it in a workshop?	Advantage	…/5
Total for **Content**:			

NAVIGATION

CRITERIA	**DEFINITION**	**STATUS**	**Mark**
16. Directions clear	Are the instructions given to the students easy to understand?	Very important	…/10
17. Printable	Can a student print out a piece of work?	Advantage	…/5
18. Interactive	Are the students required to physically do something?	Very important	…/10
19. Feedback	Is it appropriate to the age and level?	Very important	…/10
	Immediate?	Advantage	…/5
	Do correct answers also get feedback?	Advantage	…/5
	Do wrong answers also get feedback?	Advantage	…/5
20. HELP file	Available	Very important	…/10
	Help on program issues	Very important	…/10
	Help on content issues	Advantage	…/5

21. Keep records	Store data	Very important	…/10
22. Print of records	Are stored records printable or only available digitally inside the program?	Advantage	…/5
Total for **Navigation**			

FORMAT

CRITERIA	**DEFINITION**	**STATUS**	**Mark**
23. Consistent interface	Do the screens look the same?	Advantage	…/5
24. Screen display	Text size sufficient	Very important	…/10
	Colours: adding value	Advantage	…/5
	Graphics clear	Advantage	…/5
Total for **Format**			

ADMINISTRATION

CRITERIA	**DEFINITION**	**STATUS**	**MARK**
25. Sifting	Grading or evaluating students, so that those who are more capable do not waste their time on easy exercises.	Very important	…/10
26. Pacing	Are students forced to work?	Very important	…/10
27. Active engagement	Students should not be able to simply click through all the screens without actively engaging in the activities.	Very important	…/10
28. Controlled by student	Can the student choose the sequence of the exercises?	Status depends on the outcomes of the software.	…/5
29. Controlled by lecturer	Can the lecturer choose the sequence of exercises?	Status depends on the outcomes of	…/5

		the software and the preferences of the lecturers.	

ADDITIONAL

CRITERIA	DEFINITION	STATUS	MARK
30. Manual for students	Online	Advantage	…/5
	Hard copy / In-program (Printed versions may be more expensive, but some students prefer them, while in-program manuals may be easier to navigate. Consider which is best for your students.)	Very important	…/10
31. Manual for administrator	Online	Advantage	…/5
	Hard copy / In-program	Very important	…/10
32. Manual for teacher	Online	Advantage	…/5
	Hard copy / In-program	Very important	…/10
33. Time spent on the software	A well-informed opinion is not possible if too little time is spent on the software.	Very important	…/10
34. Which activity seemed the most enjoyable?	Questions 34 and 35 are simply intended to test general perceptions of the evaluator. Be vigilant of big differences between lecturer and student perceptions. If the students hate the software, it will not be effective no matter what the lecturer's opinion.		
35. Which activity seemed the most effective?			

3.20 Can evaluation criteria be used in the creation of CALL?

While the above rubric was initially created with the intention of establishing which is the best software system for a specific purpose, being aware of these requirements make it easier to plan effectively the creation of a new system. Using the evaluation system, it was established that existing software did not fully meet the needs of the NWU students or lecturers, especially with regard to argumentation, and the implementation of the principles of writing across the curriculum, due to the segmented (specialised) nature of many programs. Initial research on feedback (Louw, 2006) proved the viability of using a computerised marking system (MarkWrite, now under development at CTexT® at the NWU), but the initial research snowballed and it is now considered appropriate to develop a much larger system than simply a marking system. MarkWrite shows tremendous potential for further development.

Based on the above categorisations, MarkWrite is a *tool* as it is not intended to be a stand-alone writing coach. Due to space constraints, the whole project cannot be described here, but suffice it to say that MarkWrite is an electronic tool for marking student texts faster and more accurately using standardised feedback. MarkWrite will have two parts in its final form – MarkWrite Marker and MarkWrite Student. In MarkWrite Marker, lecturers mark student texts and send them their feedback as an HTML file. MarkWrite Student will be a teaching tool in which students have to do exercises based on the feedback. MarkWrite Student will also systematically assist students to create their assignments.

By taking the evaluation criteria one by one, the table below illustrates how these design criteria are incorporated during the planning and programming stages of MarkWrite. Scores were not assigned since the purpose of scoring was to get an overall score for different systems to see which answered best to the needs, and MarkWrite is not being compared here to any other specific system, as it is custom built.

Also note that not all evaluation criteria are equally applicable to design and a different priority hierarchy will apply to design than to evaluation. For consistency though, the same rubric is presented here as above, illustrating how it can be used as design criteria.

Table 3.1: Design criteria taken into account during the planning and programming stages of MarkWrite

NETWORK REQUIREMENTS

CRITERIA	DEFINITION AND NOTES	STATUS	APPLICABILITY AND APPLICATION
Works on my available operating system.	Which operating system are you using? Some software will not work on the latest operating system. If you cannot load the software, you cannot continue with the evaluation.	Very important. If the program cannot work on your system, you obviously cannot continue the	MarkWrite was designed to work on the latest Windows systems and is therefore not

		evaluation.	backward compatible.
Web based	As a design criterion, this evaluation criterion differs according to the software purpose.	Very important	The intention with MarkWrite is that it can be used as a stand-alone application, or later integrated into a Web teaching platform such as WebCT or Sakai.
Good support	A good HELP file or good online support is essential. A local (South African) distributor for the software is an added bonus. Online support is less optimal than a built-in help file.	Very important	The developers at CTexT® went to great lengths to ensure that an accurate HELP file is shipped with the product. This includes a video illustrating how the software functions.
Budget	Cheaper than Rx for individual packages (take-home packages) or cheaper than Rz for server-based or Internet-based software.	Very important	Market research needs to be done to determine a fair price for the system. Since it is an own development for use at the NWU, this criterion is less relevant.
Upgrading	Preferably free (take-home packages) Fixed yearly rate increase (server-based or Internet-based software)	Very important	This consideration is dealt with by the marketing team and is not an actual design consideration, but upgrading

			can be done on a needs-driven basis.
Autonomous	Can students access and work on the program unsupervised?	Advantage	The initial vision of MarkWrite is not supposed to be used by students, but only by markers. This design consideration will be taken into account when the student part is being designed.

CONTENT (Pedagogy)

CRITERIA	**DEFINITION (as applicable to MarkWrite)**	**STATUS**	**APPLICABILITY AND APPLICATION**
Software outcome	The outcomes for a system such as MarkWrite cannot be directly related to the module outcomes of a specific module, since MarkWrite is meant to be used in writing across the curriculum. The outcomes applicable here are instead stipulated in the introduction to this thesis: the feedback provided with the system should work and it should be practically optimised.	Very important	

	Students should have a greater knowledge and awareness of the qualities which make for effective academic paragraphs, introductions and conclusions.	Important	Specific research has been done to test the techniques used in MarkWrite to ensure that students have a greater awareness of the qualities of good paragraphs, introductions, and conclusions. Research has therefore been done to ensure that the design outcomes match the teaching outcomes of the software.
	Students should have a greater awareness of the specific problems and recurring errors in language.		MarkWrite has been designed to count and calculate the number of errors which a single student or a class group make. This helps both the learner and the lecturer to identify recurring errors.
	MarkWrite should adhere to the qualities of good pedagogical feedback.	Very important	Standardised feedback and radio button feedback as utilised in MarkWrite, adheres to good pedagogical practice.

Level	Advanced L2	Important	MarkWrite is adaptable in the sense that the standardised feedback can be tailored to the specific level of the students, but since the intention is to use it in writing across the curriculum, the level needs to be at advanced L2.
Accuracy of content	Up to date	Very important	As no system can be up to date for more than a few days, there had to be commitment to continually develop MarkWrite. This will ensure that it stays up-to-date. During the internal testing phase of the program, user requests were considered and some were built into the system immediately.

	Error free	Very important	Having a program which is error free is virtually impossible, but stringent testing forms part of the design process to ensure as few errors as possible.
Culturally appropriate	Many software programs are American in content and the topics and discussions are unknown or strange to South African students.	Advantage	The standardised feedback in MarkWrite can be adapted to the specific situation, subject or language. The techniques used in MarkWrite have all been tried and tested on students from different cultural backgrounds and of different proficiency levels. The exercises used in MarkWrite will be tested under diverse situations as well.
Interesting	It is difficult to establish what students will experience as interesting; however, you are not looking for a textbook on a screen. If the software fails to make use of the available resources provided by a computer, it fails to make use of the pedagogical possibilities of the medium and as such may be considered not as well planned as one would hope.	Very important	MarkWrite is not a teaching coach and therefore this criterion does not influence the design process.

Authorable	Depending on the outcomes of the CALL system, it may be necessary for the lecturer to change some parts of the program content.	Advantage	As indicated before, some parts of MarkWrite can be adapted to the specific situation. However, much of the testing is done before implementation to reduce the need for subsequent rewriting of pedagogical content.
Graphics	Do they add to the pedagogical effectiveness?	Advantage	MarkWrite does not contain many graphics. It is aimed at being as functional as possible, much like a computer spelling and grammar checker. Much thought has gone into the layout of the page, the positioning of the tools, the text and window size and other functional layout issues.
Motivational devices present?		Advantage	The first version of MarkWrite is simply the teacher version. The student version of MarkWrite will

			have exercises for the students based on their feedback, and in this case the motivational devices will be designed into the system.
Workshop	Can the program be linked so that learners can use it in a workshop-environment?	Advantage	The initial version of the marker is not intended to be used by learners.

NAVIGATION

CRITERIA	**DEFINITION**	**STATUS**	**APPLICABILITY AND APPLICATION**
Directions clear	Are the instructions given to the students easy to understand?	Very important	The feedback given to the learner is standardised and the intention of the original research was to make it easy to understand (Louw, 2006).
Printable	Can a student print out a piece of work?	Advantage	Learners receive their feedback in HTML format. This is a "website" and the feedback can therefore be printed if necessary.
Interactive	Are the students required to do	Very important	In the first version of

	something physically?		MarkWrite, learners are not compelled to use the feedback. This is due to networking and system constraints. The intention is that a later version (MarkWrite Student) will force students to actively engage with the planning and editing of their texts, as well as with the feedback.
Feedback	Is it appropriate to the age and level?	Very important (The whole MarkWrite system at present is built around the concept of feedback)	This is not applicable to the current version of MarkWrite, but will be applicable to the exercises in MarkWrite Student.
	Immediate?	Advantage	This is not applicable at present.
	Do correct answers also get feedback?	Advantage	This is not applicable at present.
	Do wrong answers also get feedback?	Advantage	This is not applicable at present.
HELP file	Available	Very important	A help file with screen capture videos will be available.

	Help on program issues	Very important	A HELP file with screen capture videos will be available.
	Help on content issues	Advantage	Due to the variety of texts which can be marked with MarkWrite, this is not possible for the system.
Keep records	Store data	Very important	MarkWrite is able to store individual and class records.
Printing of records		Advantage	The output file of the records is printable.

FORMAT

CRITERIA	**DEFINITION**	**STATUS**	**APPLICABILITY AND APPLICATION**
Consistent interface	Do the screens look the same?	Advantage	MarkWrite currently has only two screens. The one screen is the marker interface which is always the same and the other is the output HTML file, which is also always the same.
Screen display	Text size sufficient	Very important	The text size can be adjusted at will.

	Colours: adding value	Advantage	Colours are used sparingly in MarkWrite, but when used the colours add value to the feedback. For example, feedback on different categories of errors is shown in different colours.
	Graphics clear	Advantage	Currently, graphics are not used in MarkWrite. It is possible to include graphics later on, but then with a clear pedagogical purpose.

ADMINISTRATION

CRITERIA	**DEFINITION**	**STATUS**	**APPLICABILITY AND APPLICATION**
Sifting	Grading or evaluating students, so that those who are more capable do not waste their time on easy exercises.	Very important	The intention with MarkWrite Student is that learners only receive exercises based on their individual feedback. This will ensure that all exercises are directly applicable to the specific student. This criterion is

			not applicable to MarkWrite Marker.
Pacing	Are students forced to work?	Very important	This is only applicable to MarkWrite Student and will be taken into account during development.
	Students should not be able to simply click through all the screens without actively engaging in the activities.	Very important	This is only applicable to MarkWrite Student and will be taken into account during development.
Controlled by student	Can the student choose the sequence of the exercises?	Status depends on the outcomes of the software.	This is only applicable to MarkWrite Student and will be taken into account during development.
Controlled by lecturer	Can the lecturer choose the sequence of exercises?	Status depends on the outcomes of the software and the preferences of the lecturer.	This is only applicable to MarkWrite Student and will be taken into account during development.

ADDITIONAL

CRITERIA	**DEFINITION**	**STATUS**	**APPLICABILITY AND APPLICATION**
Manual for	Online	Advantage	

students	Hard copy / In-program	Very important	MarkWrite's manuals are incorporated into the HELP files. At present the help file is only necessary for the lecturers or marking assistants who use it to mark, since the students will simply receive their HTML files via their e-mails as attachments, or via the web-based learning platform.
Manual for administrator	Online	Advantage	Currently MarkWrite is a stand-alone application. Once it is incorporated into a system such as Sakai, an administrator's manual will become necessary.
	Hard copy / In-program	Very important	See above.
Manual for lecturer	Online	Advantage	
	Hard copy / In-program	Very important	This is included in the help file.
Time spent on the software	A well-informed opinion is not possible if too little time is spent on the software.	Very important	This criterion has to do with the evaluation of the software. However, the amount of

			research time spent in development can apply here. The development of the software is a long process. Since it is not simply a series of drill-type exercises, much time has been spent on research and more is required to ensure a good product. Once the MarkWrite Student part of the software is being developed, research will be necessary to ensure high-quality, level-appropriate exercises adhering to the qualities of effective pedagogy and CALL.
Which activity seemed the most enjoyable?	This question and the one below are simply intended to test the general perceptions of the evaluator. Be vigilant of big differences between lecturer and student perceptions. If the students hate the software, it will not be effective no matter what the lecturer's opinion.		It is uncertain how exercises and feedback can be made enjoyable for students. At present, MarkWrite Marker is simply an advanced marking tool and this criterion

			applies even more so to MarkWrite Student
Which activity seemed the most effective?			The practice of providing standardised feedback has been proven to be effective. See Louw, 2006.

3.21 Conclusion

Integrating CALL into a language curriculum is not a decision to be taken lightly. It requires all parties involved to be educated about what is possible with CALL, will be aware that it is not just one person's job, and will take it seriously enough to evaluate software properly to fit the module outcomes. Before the evaluation can be done, certain decisions need to be made, clarified and discussed. Once a suitable software package has been decided upon, the specific modules have to be re-written in order to integrate the software into everyday teaching. Although this procedure may sound like a very daunting task, computerised language coaches and computerised language tools may in the end save lecturers and students a great deal of time. A cost-benefit analysis may prove that language learning software is still worth the trouble, but it is an absolute necessity to approach the process in a structured, well-thought-out manner.

The author has attempted to show that (as is the case with outcomes-based education) it is advisable to keep the final evaluation criteria and good pedagogical principles in mind when planning and creating a CALL system. It has been illustrated how many different variables and role players have to be taken into account during the evaluation of a CALL software package. It has also been illustrated that the same evaluation rubric can also be used effectively as a guiding principle for the design of a new system.

The rubric is in no sense hierarchical in nature, but by far the most important aspect of evaluation (which is emphasised in many of the sources consulted) is the fact that effective pedagogical practice is of paramount importance. No amount of features will compensate for poor pedagogical practice. Since MarkWrite is first and foremost a tool to provide effective (pedagogical) feedback on student writing, it is therefore vitally important to establish exactly what constitutes the most effective feedback on student writing.

Creating a CALL system or CALL tool is a daunting task requiring many hours of research and a lot of money. It may be more expensive to create a CALL system than to write a book, and in this case one should question if the benefits of using the computer will eventually really outweigh the costs. The author is of the opinion that the MarkWrite project, although it is

still in its infancy, will meet the requirements for effective CALL, and that the numerous further development possibilities of the system will prove to outweigh the cost of development by far.

3.22 Bibliography

Arnett, J. 2009. The way we write now. *Adults Learning* (20)5:27.

Barr, J.D. & Gillespie, J.H. 2003. Creating a computer-based language learning environment. *ReCALL* 15 (1): 68-78.

Berg, A.S. & Pretorius, R.S. 2003. Tsenang!: An interactive multimedia programme for learning Beginner Setswana. *Journal for Language Teaching* 37(1): 1-12.

Bishop, A.L. 1999. Setting policy for the evaluation of the CALL environment. In Egbert, J. & Hanson-Smith, E. (eds.) CALL environments – research, practice and critical issues. TESOL.

Bradin, C. 1999. CALL issues: instructional aspects of software evaluation. In Egbert, J. & Hanson-Smith, E. (eds.) CALL environments – research, practice and critical issues. TESOL.

Buell, J. 1999. Resources for CALL. In Egbert, J. & Hanson-Smith, E. (eds.) CALL environments – research, practice and critical issues. TESOL.

Cowan, R., Choi, H.E. & Kim, D.H. Four questions for error diagnosis and correction in CALL. *CALICO Journal* 20(3): 451-463.

Chappelle, C.A. 1999. Theory and research; investigation of "authentic" language learning tasks. (In Egbert, J. & Hanson-Smith, E. eds. CALL environments – research, practice and critical issues. TESOL.)

Chapelle, C.A. 2003. English language learning and technology. Amsterdam: John Benjamins Publishing Company.

Chen, J., Belkada, S., & Okamoto, T. 2003. How a web-based course facilitates acquisition of English for academic purposes. *Language learning and technology* (8)2: 33-49.

Davies, G. & Hewer, S. 2009. Introduction to new technologies and how they can contribute to language learning and teaching. Module 1.1. In Davies, G. (ed.) *Information and communications technology for language teachers (ICT4LT)*, Slough, Thames Valley University [Online]. Available from: http://www.ict4lt.org/en/en_mod1-1.htm [Accessed 16 September 2009].

Davies, G., Hamilton, R., Weideman, B., Gabel, S., Legenhausen, L., Meus, V., & Myers, S. 2009. Managing a multimedia language centre. Module 3.1. In Davies, G. (ed.) *Information and communications technology for language teachers (ICT4LT),* Slough, Thames Valley University (Online). Available from: http://www.ict4lt.org/en/en_mod3-1.htm Accessed 4 August 2009.

Dutta, S. and Mia, I. (eds.) 2010. The Global Information Technology Report 2008–2009 – Mobility in a Networked World. Geneva: SRO-Kundig.

Egbert, J., Chao, C. & Hanson-Smith, E. 1999. Computer-enhanced language learning environments: an overview. In Egbert, J. & Hanson-Smith, E. (eds.) CALL environments – research, practice and critical issues. TESOL.

Fitze, M. 2006. Discourse and participation in ESL face-to-face and written electronic conferences. *Language learning and technology* 10(1). January. pp.67-86.

Granger, S. 2003. Error-tagged learner corpora and CALL: a promising synergy. *CALICO Journal* 20(3): 465-480.

Hellar, I. 2005. Learner Experiences and CALL-Tool Usability — Evaluating the Chemnitz InternetGrammar. *Computer Assisted Language Learning* 18(1&2)119-142.

Hémard, D. 2006. Design issues related to the evaluation of learner–computer interaction in a web-based environment: activities v. tasks. *Computer Assisted Language Learning* 19(2&3): pp. 261 – 276.

Lasagabaster, D & Sierra, J.M. 2003. Students' evaluation of CALL software programs. *Educational Media International* 40:3/4: 293-304.

Levy, M. 1997. Theory-driven CALL and the development process. *Computer Assisted Language Learning* 10(1): 41-56.

Louw, H. 2006. *Standardising written feedback on L2 student writing*. Potchefstroom: North-West University. (MA dissertation).

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).

Murray, L. & Barnes, A. 1998. Beyond the "wow" factor: evaluating multimedia language learning software from a pedagogical viewpoint. *System* 26: 249-259.

Ngu, B.H. & Rethinasamy, S. 2006. Evaluating a CALL software on the learning of English prepositions. *Computers & Education* 47: 41–55.

Peyton, J.K. 1999. Theory and research: interaction via computers. In Egbert, J. & Hanson-Smith, E. (eds.) CALL environments – research, practice and critical issues. TESOL.

Plester, B., Wood, C. & Bell, V. 2008. *Literacy*, 42 (3). November. 137-144,

Sivert, S. & Egbert, J. 1999. CALL issues: building a computer-enhanced language classroom. In Egbert, J. & Hanson-Smith, E. (eds.) CALL environments – research, practice and critical issues. TESOL.

Spencer, B. & Louw, H. 2008. Theory versus implementation: First-World technology in a Third-World teaching context. *Language Matters* 39(1): 110-126.

Tsiriga, V. & Virvou, M. 2004. Evaluating the intelligent features of a web-based intelligent language learning system. *International Journal on Artificial Intelligence Tools* (13)2: 411-425.

Wachman, R. 1999. Classroom practice: autonomy through authoring software. In Egbert, J. & Hanson-Smith, E. (eds.) CALL environments – research, practice and critical issues. TESOL.

Wang, X. 2006. Evaluating a Web-based listening programme for Chinese University non-English majors. *US China Education Review* (3)5: 74-76

Yuan, Y. 2003. The use of chat rooms in an ESL setting. *Computers and composition* (20): 194-206.

CHAPTER 4
ARTICLE 2 – MOVING TO MORE THAN EDITING: A CHECKLIST FOR EFFECTIVE FEEDBACK

4.1 Prelude to Article 2

Article 2 is a brief literature review on the qualities of effective feedback. As indicated in Article 1, effective CALL will only be possible if it is based on effective pedagogical principles. The rest of the thesis therefore first establishes what effective pedagogy in this context will be, if it is possible to implement it practically (Article 3), and then explains the development and testing of a new technique (Articles 4 and 5).

Article 2 overlaps with Louw (2006) to an extent, but is a more refined version of its findings. The findings are rendered more operable in that the 13 qualities of effective feedback identified in this article are used later in the thesis both for the design and evaluation of new techniques attempted.

Publication information for Article 2

The article appeared as:

Louw, H. 2009. Moving to more than editing: a checklist for effective feedback. *Journal for Language Teaching*, 43(2): 86-102.

Minor editorial changes were made for the sake of consistency with the overall format of the thesis.

Abstract

Providing feedback on student writing is a much debated topic. One group of researchers argues that it is ineffective, and another group remains convinced that it is effective, while at ground level teachers and lecturers simply carry on “marking” texts. The author of this article argues that both sides of the argument are valid and uses the arguments both for and against feedback to create a checklist for effective feedback practice. Adhering to this checklist should counter most of the arguments against feedback while supporting and strengthening the arguments in favour of feedback.

Keywords

Feedback, second language writing, process approach to writing, consciousness-raising, error, input, correction, second language learning, language awareness

4.2 Introduction

Providing feedback on student writing is one of those teaching activities that is considered a given in most teaching institutions in South Africa. Students expect their texts to be marked without really considering why they expect it, and then often simply discard the feedback

since they do not have the skills to interpret and understand it and use it to improve their writing (Spencer, 1998:208; Hyland, 2003:218). Therefore teachers mark almost mechanically, apparently without considering why they do it, apart from knowing that it is expected of them. This in brief and general terms is the position on the ground in South Africa, despite a long-running global debate about the effectiveness and practical use of providing feedback on student writing (see Ferris, 2003:120 and Truscott, 1996).

Dealing with feedback is somewhat of a Catch-22 situation: the problems with providing meaningful feedback are numerous, but positive aspects attributed to feedback are also plentiful, and the arguments for and against feedback all cite studies to prove their positions. Some of the problems include a focus on form (Spencer, 1998:62; 76; Ellis, 1996:653), the generic nature of feedback (James, 1997:257), students' inability to spot recurring patterns of errors in their writing or to distinguish between more important errors and less important errors (Wible, Kuo, Chien, Liu & Tsao, 2001 and Spencer, 1998), and students' inability (Hyland, 2003:218) or unwillingness (Hyland, 1990) to use feedback. Feedback has also been found to confuse learners (Monyaki, 2001:66, 74), often does not have a clear purpose (Moletsane, 2002:27) and could lead to avoidance (Munchie, 2000:49). In some instances it does not lead to revision (Paulus, 1999:266), is sometimes insensitive (Spencer, 1998) and does not appear to lead to independent learners who are able to use the feedback to improve their writing (Munchie, 2000:49 and Monyaki, 2001:76).

On the other hand, feedback is expected by society and learners alike (Spencer, 1998; Storch & Tapper, 1997:245) and has been found to enhance learning (Hyland, 2003:219; Askew and Lodge 2000:2), improve writing (Hyland, 2003:218) and reduce errors (Ferris, 2002:17). It could also be used as a tool to counter fossilisation (Louw, 2006:59) and be used as a strong motivating factor (Moletsane, 2002:32-33), encourage communication and rewriting (Lyster & Ranta, 1997:41) and be used for consciousness-raising (James, 1997:257 ff; Louw, 2006:64).

Accepting as a premise that both sides of the argument have valid arguments grounded in empirical evidence, this article attempts to break out of the binary opposition between the two camps by creating a synergy of the two opposing views in order to present a "how-to" checklist for effective feedback, with an explanation of the criteria on the checklist. The net effect is that criticism against feedback could diminish when the practice of feedback is improved and used in a more meaningful manner.

4.3 Methodology

The checklist for effective feedback reported on here has resulted from a literature survey investigating the arguments both for and against feedback, as well as the various techniques with which feedback is provided. Effective marker practice was also investigated and techniques found to be effective were analysed to see why they are effective. Arguments against feedback were analysed to find the source of the problems raised by the objectors. Very often the problem turned out to be the *execution* of feedback rather than the *notion* of feedback as such, and the feedback checklist therefore aims to rectify poor practice in order to provide better and more effective feedback.

4.3 What exactly is feedback?

The first obstacle to overcome when discussing feedback on student writing is the definition of feedback, as no two researchers seem to have exactly the same idea when using the term. Since the definition of the term "feedback" and the use of the term are closely related to the concept of "error", and as both terms will be used throughout the article, it is important to define them at the outset of the argument.

There are numerous distinctions between the two definitions, but a close analysis reveals the distinctions to be mostly cosmetic or highly theoretical. In essence, the term "feedback" has two main definitions in the literature and is generally used to refer to any correction by a lecturer (implied by Moletsane, 2002), or any response to a text by any reader (Hyland, 2003, 1990, 1998; Lyster & Ranta, 1997; Askew & Lodge, 2000). In practice, both definitions aim at providing the learner with information that would enable him or her to adapt to a certain standard of language use, with the teacher as the "knower" of what that standard pertains to. The main difference between the two definitions lies in the amount of information provided. An all-out correction may provide less information than a well-structured explanation, and it may provide less learning opportunity as well. The amount of information provided should therefore assist in the communication process between the creator of the text and its audience. In this case, the communication aimed at the text creator also has the aim of facilitating his or her learning.

Communication is never a one-way process. In the "writing-as-process" approach, much emphasis is placed on the meaning of the text and the fact that the meaning of a text is not simply created by the writer *but co-created by the reader*. Feedback given by the reader should be indicative of the way an audience would experience the text. The problem with this is that the general audience of a text (average readers) would most probably not recognise all the errors in a text, and feedback is still a tool to provide learning as well.

In this approach, feedback is ideally a "ping-pong game" (Askew & Lodge, 2000) of comments going back and forth between the reader and writer until near perfect communication is created. Feedback may be provided in different formats and may differ in intensity and purpose, and one can therefore conclude that that feedback is part of a communicative process, in which a person knowledgeable about a *language* and a *subject* (teacher/lecturer) provides information on different levels of language use and content to a learner of the language or subject. The process of feedback can therefore be graphically illustrated as follows:

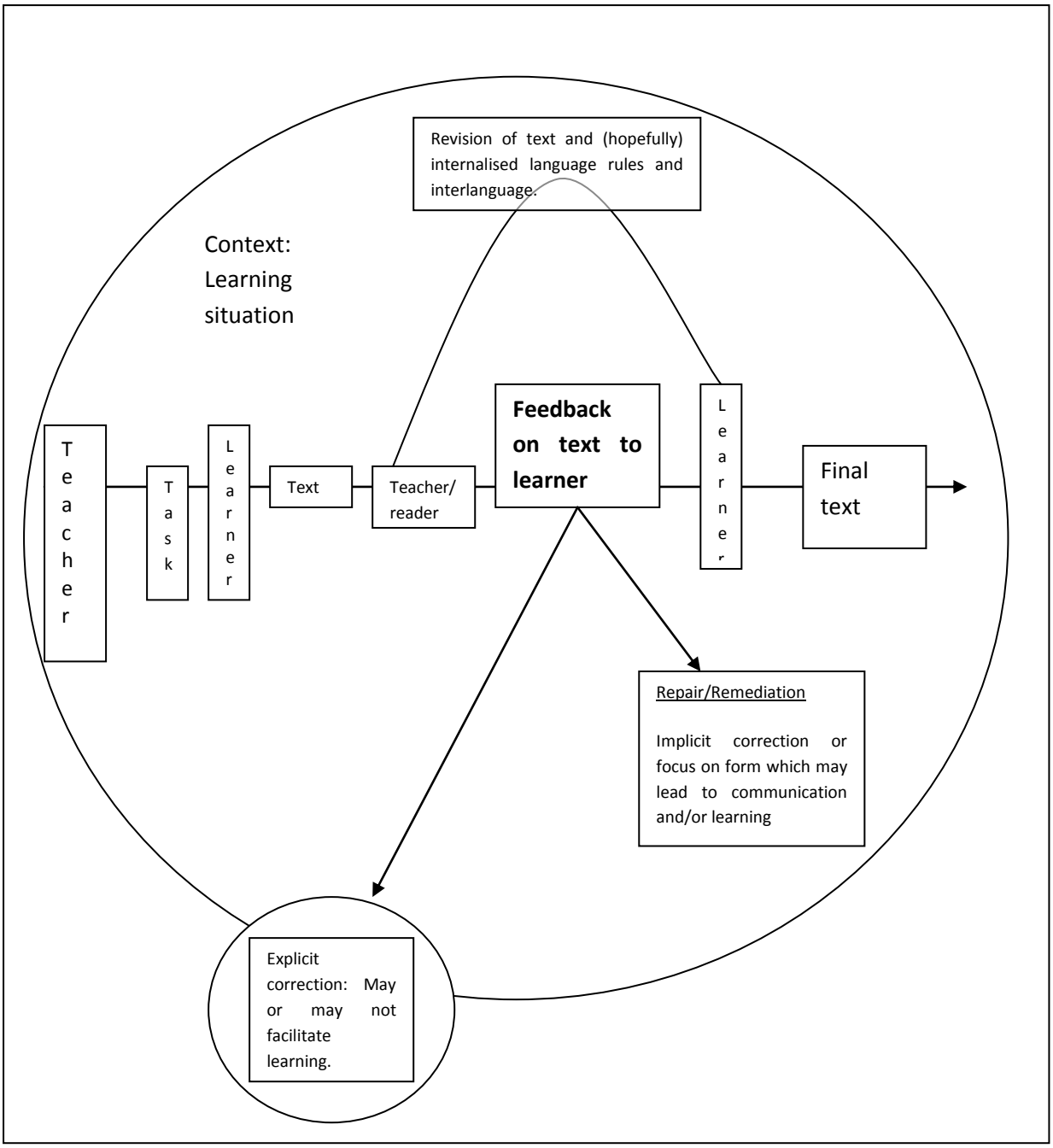

Figure 4.1: The communication timeline of feedback on writing in a learning context

In Figure 4.1, the communication timeline of feedback in a learning context is illustrated graphically. The circle around the timeline illustrates that the timeline is always situated inside a specific context.

Within the larger context of a learning situation, a teacher or lecturer will give a writing task to a learner. The specific situation will determine the goal of this writing task, which may or may not have a specific language-educational motive. It may only be to test a learner's knowledge of a specific topic, or it may be to evaluate the learner's writing competence, or

even to test a specific style of writing. It may also be to see how well learners have understood a learning objective such as mastering the passive voice for example.

The learner will receive the writing task either orally or in written form and may or may not know the specific goal of the exercise. The learner then creates a text, which may or may not contain certain errors (based on the definition of *error* below, almost all texts are bound to contain some errors), and returns it to the teacher, or perhaps to a fellow student (reader) for their comments.

The fellow student reader or teacher then provides feedback on the text. This feedback will differ in method and the specifics it focuses on. It may be either (or a combination of) *repairing feedback* leading to communication between the learner and reader, or an *explicit correction.* If explicit correction is considered to be the same as explicit instruction, the value thereof is contested (Spencer, 1998:69-75; Ellis, 1996:653).

In some teaching situations, this is where the process for this specific text will stop. The *first* version of the text will also be the *final* version. In other situations (specifically the process writing approach) the learner will have to revise the text and again return it to the student reader or teacher (see Krapels, 1990). During this revision process, the learner may learn something about the topic, about writing or even about his or her internalised rules of language (see Ellis, 1996:30 for an explanation of these interlanguage rules). Feedback in this context can be considered effective if the learner in fact *learns* something about his or her use of language, style or even the topic.

This does not mean that feedback can only be effective in a process writing approach. In situations where the first version of a text is also the last, the possibility still exists that students may use the feedback to revise their internalised rules of language. In effect, they will have to refer back to the feedback on a previous text to enhance a different future text. Applying this technique can get very difficult – especially if the student never looks at the feedback. In situations where a process approach to writing is not practiced, something should be done to ensure that students still utilise the feedback. Students should somehow be forced not only to look at but also use the feedback they receive.

Given this context a working definition of **feedback** is:

> **Feedback constitutes any mark by an external reader on the text. The mark may indicate something that is considered to be wrong, or something that is considered less than optimal. Feedback may also indicate instances where the reader is satisfied or impressed by something in the text.**

From this definition it is obvious that there is a focus on indicating instances where a text could be improved, but which are not necessarily wrong. One should be careful of simply focusing on incorrectness as this would not be effective feedback. There is a much greater opportunity for learning if the marker also focuses on parts of the text that could be improved, i.e. indicating something which may not be inherently wrong, but which could have been better. Only if “error” includes something that could be improved does real learning start to take place. Consequently it is essential to link the definition of feedback in

which the facilitation of learning is central to a working definition of “error”, defined for the purposes of this article as follows:

> **An error is any instance in a text in which language use is incorrect, or in which language use is not inherently wrong, but which could have been better.**

With these two definitions in mind, the checklist for effective feedback practice is presented.

4.4 What is effective feedback practice?

The checklist for effective feedback is work in progress as some of the requirements are at present still difficult to achieve. Others may need a change in attitude of teachers and markers alike.

In essence, effective feedback should:

1. be clear and understandable;
2. be consistent, 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 individualised;
12. be time effective; and
13. be searchable/archiveable/recordable and allow for research.

While the checklist is necessarily short and cryptic in order to function as a practical checklist, each of the criteria merits some discussion and explanation.

4.4.1 Feedback should be clear and understandable

That feedback should be understandable is really stating the obvious. The theory makes sense, but the practice is a different story. In Louw (2006) the term *hieroglyphics* is used to refer to the squiggles, lines, circles, question marks and ticks commonly used as feedback on student writing. While it is true that students are often able to identify an error if it is simply

indicated to them, research has shown that students are sometimes unable to identify an error which has been indicated (Louw 2006:134-146).

Likewise, what exactly does a tick indicate? A student can guess that a lecturer was satisfied with something, but the exact nature or object of satisfaction is not clear, so the action taken cannot be duplicated – was it a correct sentence structure, satisfaction that a difficult word was spelled correctly, or was it an indication of agreement with an argument?

Indicating more than one error in the same way is also not effective (Ellis 1996:585; Nwaila, 1996:83)**.** Hieroglyphic feedback often falls into this trap by, for example, circling more than one type of error simply to indicate that an error has been detected. In effect then, the student is in the same position as when an error is indicated but not identified.

In addition to an error being clearly identifiable, feedback markings should not be all over the text, making the feedback and the text illegible. Some texts are full of errors and adhering to this guideline can get tricky. If all errors are to be identified, some student texts do not leave enough room for all the feedback. This may be especially true for second language students. Difficult though it may be, this is one of the most important guidelines. If students cannot decipher feedback or if they fail to distinguish one comment from another, they cannot act on it. Due to the negative and uncritical attitude most students have towards feedback and rewriting (Spencer, 1998:73), it is possible that the least bit of confusion may be the final straw that results in students disregarding the feedback on a text completely.

4.4.2 Feedback should be consistent, complete and thorough

There is inconsistency in current feedback procedures and this can cause problems for students (simply compare the problem with the lack of conformity in the definitions for feedback and error in Louw, 2006). The method of providing feedback is not standardised – in fact there are at least 14 partially overlapping techniques for providing feedback (Louw, 2006:66). Different lecturers use different methods of providing feedback. In some instances *all* errors are indicated *all* the time and in some not. Certain types of error are more likely to be dealt with than others and the more often a particular type of error is made, the less likely the teacher is to deal with it. Teachers also sometimes indicate errors that had not been made or indicate more than one type of error in the same way (Ellis 1996:585; Nwaila, 1996:83).

Such inconsistency and frequent incompleteness are a problem. If all errors are not indicated, the students may assume that their language usage is correct, whereas the lecturer may simply neglect to indicate all errors, thinking that it is not necessary to mark a recurring error every time it occurs. Moletsane (2002:32-33) points out that comprehensively marking all errors, every time, counters fossilisation. He also warns that it can cause confusion if an error is marked in one place and not in another. On a positive note, some students are motivated to try and lower the occurrence of errors if they note the number of errors they have made (Moletsane, 2002:32-33).

It should be clear then that for feedback to be effective, no errors should be neglected, either deliberately or by chance. Errors should be indicated consistently, accurately and correctly.

4.4.3 Feedback should be correct

Some teachers mark language features as incorrect even though they are correct, because the teachers themselves have not completely acquired the language (Buthelezi, 1995; Van der Walt & Van Rooy, 2002:115). Given that feedback is a form of input, this is a grave problem since, as Krashen (1985:43-52) indicates, inappropriate input may contribute to the problem of fossilisation.

If the target language is simply learnt through communication-oriented instruction, communicative competence can be acquired before the grammatical structures of the language have been mastered, thereby increasing the risk of fossilisation (Moletsane, 2002:28). Feedback has to counter that risk. Krashen (1985:43-52) explains that fossilisation may be the result of:

- an insufficient quantity of input,
- an inappropriate quality of input,
- the affective filter,
- the output filter and/or
- the acquisition of deviant forms: in order to ensure that deviant forms are not acquired, they should be eradicated.

Feedback is seen as a way to provide input. However, comprehensible *input* alone is not sufficient for successful L2 learning. Comprehensible *output* is also needed, and that is mostly done through speaking or writing. If done through speaking, the learner can negotiate meaning with the listener (which can be seen as a form of feedback). If done through writing, however, the learner has no way of judging comprehensibility if he or she is not provided with feedback. Lyster and Ranta (1997:41) warn that subject-matter teaching does not provide adequate language teaching on its own, but that “language used to convey subject matter needs to be highlighted in ways that make certain features more salient for L2 learners”.

Feedback is therefore an important source of input and it is important that input be of the quality of language the students are required to acquire.

4.4.4 Feedback should indicate error status

Often in marking, no indication is given of the level of importance (status) of the error (Spencer, 1998). A student would have to guess how severe his or her problem is. A marker may also be fooled by plenty of typing errors in an otherwise good text. A typing error, for example, is not a high-level error, and is more of a nuisance than a hindrance to communication. Errors of syntax or word choice can, however, create much bigger problems by jeopardising the comprehension of the text.

4.4.5 Feedback should aim at improvement, not just correctness

In keeping with the definition of error, feedback should not only look at things which are wrong. Feedback is a form of input. If that "input" is only a comment on student output, a student is never encouraged to explore the language to get to know it better. An over-emphasis on incorrectness is therefore counter-productive for the learning process.

Feedback is also a type of consciousness-raising whereby learners are reminded of where they do not have the target language features under full control. If handled incorrectly, learners will not see the lecturers' feedback on their errors as a learning opportunity, but will instead strive for "perfect" language use. Instead of experimenting with the language, they then stick to what they know they are capable of, resulting in the undesirable effect of avoidance (Hyland, 1998:264).

This over-emphasis on correctness is not only an issue of language, but also of argument and style. Students seem to be conditioned to believe that "revision" refers to "correcting". Louw (2006:134) found that students often simply exchanged a word with a synonym when asked to support a statement. Students therefore tried to "correct" a surface element without understanding that they were supposed to improve the overall argumentative qualities of the text.

If students and lecturers alike are aiming at improvement instead of only correctness, more attention should be paid to the content of the writing instead of just focusing on the language (surface elements).

4.4.6 Feedback should be a learning opportunity

Despite the call for correctness and thoroughness of feedback, a marker should never just become a proofreader who focuses on incorrectness. An excessive focus on incorrectness is counter-productive for the learning process. Feedback is a type of consciousness-raising whereby learners are reminded of where they do not have the target language features under full control. If handled incorrectly, learners will not see the lecturers' feedback on their errors as a learning opportunity, but instead try to strive for "perfect" language use, which may once again lead to avoidance as explained above. This defeats the purpose that feedback tries to achieve. As mentioned earlier, feedback is part of input and as such should be an input opportunity.

As early as 1979, Selinker and Lamendella claimed that extrinsic feedback may assist in language learning. This is, however, a controversial statement, as there are almost equal numbers of researchers who have raised their voices in favour of and against feedback. It seems as if the tide has lately been turning in favour of feedback. This may have to do with a change in feedback techniques based on the growing attention given to the subject since the 1980s. Techniques are getting better and teachers are more aware of what NOT to do. The problem, it seems, has to do with *connecting* feedback and learning. Due to teacher awareness and techniques such as corpus analysis and error analysis, feedback can now enhance learning if used correctly (Wible *et al.*, 2001).

Lately more and more evidence has been pointing towards the effectiveness of feedback. Hyland (2003:219) mentions that "studies which measure student improvement longitudinally after error correction in terms of accuracy … suggest that students who receive error feedback over a period of time do improve their language accuracy". Askew and Lodge (2000:2) have come to a similar conclusion in their report on seminars on effective learning. They found that feedback is important in supporting learning at individual, group and organisational levels. They observed that a focus on feedback was popular and that the notion of feedback seems unproblematic. Askew and Lodge (*ibid.*) also observed that people had different perceptions of feedback and its functions and processes based on their perceptions of learning. Feedback should be seen as a crucial feature of teaching and learning processes and should be considered as an element in a repertoire of teaching techniques and connected strategies to support learning (Askew and Lodge, 2000:1).

The way to support this learning is by simply being a guide to recognising what has been done wrong (or what could be better) in order to prevent new errors. However, note that feedback on its own is not sufficient. Askew and Lodge (2000:1) warn that "learning is supported by a whole range of processes, one of which is feedback". Feedback for the sake of feedback is not effective teaching, but should be an integrated and on-going intervention strategy which is not simply aimed at the here and now (Monyaki, 2001:14-16). This trap of short-lived focus is easier to fall into than one may think. Hyland (1990:279) found feedback itself to become the focus of the action, instead of learners acting on it. When working with large class groups for example, teachers may also be tempted just to finish their marking in order to get a grade for all their students, without considering the teaching implications of the rushed job. Teachers willingly downgrade themselves to proof-readers and spelling checkers because that is all they have the time or energy for. Clearly then, it is not part of a teaching strategy, but just some mechanical action that goes with the territory. An ideal feedback technique should therefore rather remind the marker of his or her purpose and role in the communication creation process.

Many researchers have found that students fail to use feedback, mostly because they fail to understand it or do not know how to use it. In some cases it would not make any difference to their marks either, so they do not see any use in it. Feedback can only be a learning opportunity if *used* by the students, and they should therefore be encouraged, or better still, compelled, in some way or other to use the feedback they receive. A learner who simply "corrects" everything the teacher identifies as an error is not engaged with the negotiation of meaning or the "ping-pong game" (Askew & Lodge, 2000) of text creation. It also downgrades a teacher to a proofreader or spelling checker. Students/learners should use feedback for more than just correcting "bad text" or errors, as feedback should aim at creating not just passive "correctors", but thinking and learning students who can use this knowledge to produce better/more correct work in their written assignments.

Moletsane (2002:30) and Munchie (2000:49) suggest that students may be afraid to take a risk because of previous feedback (also see Hyland, 1998:264 on avoidance). Spencer (1998:56, 62, 109) also found that students are likely to omit an idea or construction if they are unsure about the correct action on a specific comment, even though the comment may

be easy to understand. This is clearly not making use of the learning opportunity, and the challenge is therefore to motivate (or force) students to use the feedback they have received to explore language and not just stick to what they already know.

4.4.7 Feedback should be purposeful

Marking for the sake of marking is ineffective, and lecturers and teachers should consciously decide on the purpose of marking and then use it in that way. Feedback should be an intervention strategy and not simply aimed at the here and now (Monyaki, 2001:14-16). As mentioned above, Hyland (1990:279) even found that feedback itself, rather than the learners acting on feedback, becomes the focus of the action. Moletsane (2002:27) also found that teachers tend to lose sight of the purpose of feedback, and seem to see their role as to simply identify errors. This is clearly losing sight of the purpose of feedback. Teachers should never reduce themselves to proofreaders; otherwise they are missing the purpose of teaching. The difference lies in focusing on the needs of students to provide them with a learning opportunity versus a lecturer focusing on his or her "duty" to finish marking a set of assignments.

A problem with lecturers' comments is that the lecturer reads the text expecting something specific – it could be that he or she understood the assignment much differently from the student. This is clearly a mistaken purpose of feedback. Feedback should have a clear pedagogical purpose and should not just be negatively inclined by looking for incorrectness or personal preferences.

4.4.8 Feedback should place responsibility on the learner

A lecturer provides input while providing feedback, but it is still up to the learner to use that input to his or her own advantage. Feedback should therefore aim at creating thinking and learning students, not passive "correctors", who are disengaged with the negotiation of meaning. Munchie (2000:49) and Monyaki (2001:76) found that learners often indiscriminately use a teacher's comment, which implies a lack of critical processing and evaluation of feedback. Munchie suggested that this lack of evaluatory and decision-making reasoning may reduce the impact and long-term effectiveness of feedback and revision. Spencer (1998:73) also found an uncritical attitude in students' reactions to feedback and also warned that it may be counterproductive. Learning a language is more effective when learners explore the language or are forced to think. The challenge to markers is therefore to provide feedback in a way that challenges learners to do more than "correct". Feedback has to be a tool for learners that will make them aware that their text is their own responsibility (Ferris, 2002: 78).

Students themselves experience a vague "rubber-stamp" type of feedback (or error correction as the only feedback) very negatively. This creates inattention to feedback and may be an additional reason why students fail to see patterns of errors in their writing (Hyland, 1990:282; Wible, Kuo, Chien, Liu & Tsao, 2001: 308-310). In effect then, poor feedback leads to poor motivation on the part of the learners. The ideal is for students to be motivated to try and do better and not to be demotivated to such an extent that they no longer even try.

4.4.9 Feedback should encourage communication and rewriting

Paulus (1999:266) indicates that an incorrect focus on feedback may contribute to ESL students' lack of writing and revision strategies. Feedback should encourage rewriting and guide students to see a text as a process and not a final product. This criterion is closely related to the criterion for student responsibility. The operative word here is "encourage". However, learners often fail to see rewriting as an essential part of the creation of a text; instead, they see an instruction to rewrite something as a punishment and not as an opportunity to improve their work (Monyaki, 2001:75; Moletsane, 2002:30). It seems that the problem lies partly in the way feedback is used, and partly in the attitude (both of teachers and students) towards rewriting. Students see rewriting as punishment, and no doubt teachers are reluctant to do "double marking".

However, as pointed out earlier, communication is never a one-way process. Feedback should encourage communication between the learner and the teacher. The meaning of a text is co-created by the reader. Feedback given by the marker should indicate the way an audience of average readers would experience the text, *but also* point out errors so as to be useful as a learning experience.

Once again, Askew and Lodge's concept of a ping-pong game of comments is relevant. The principle is sound and can be practised without sending a text back and forth time and again. Teachers should just be aware that one reading of a text is often not sufficient and should be willing to mark it at least twice. They should also not make the mistake of considering their reading of a text as the only correct one. A problem with lecturers' comments is that the lecturer reads the text expecting something specific – it could be that he or she understood the assignment very differently from the student. Students also resent comments on the content of the paper, especially if the content could be considered "personal opinion" (Spencer, 1998: 71-72).[2] Spencer (1998:55) states that the way lecturers read student texts is "upside down": a reader normally reads a text assuming that it has coherence and is structured in such a way as to convey the intended meaning effectively. A lecturer, on the other hand, approaches student writing with scepticism – going against the grain while reading. This type of scepticism is necessary on the one hand because the teacher is in a position to impart knowledge. On the other hand, this scepticism is not desirable when it results in communication failure or one-way (top-down) communication from the lecturer (Van der Walt, Van der Walt & Dreyer, 1994:14).

Top-down communication may result in feedback which is "nasty" to students. Negative feedback is discouraging in itself. Add to that the finding that students consider rewriting as punishment (Monyaki, 2001:75; Moletsane, 2002:30), and it should be self-evident that feedback should aim at being supportive. There is more to being supportive than just giving positive comments, though. Askew and Lodge (2000:7) indicate that feedback may encourage competition and comparison which could be negative as some learners would

[2] Conversations with the author's students have revealed that some of them dread "give your own opinion" questions. If the opinion of the student differs from that of the lecturer, they often get bad marks. Clearly the lecturer then did not want the students' opinions, but simply a rewrite of his or her own opinion.

simply give up trying. Askew and Lodge (2000:7) have coined the phrase “killer feedback” for situations where the receptive-transmission form of feedback blocks learning. Such feedback discourages all further drafting, is too much and feels overpowering, does not give any help and encourages no dialogue between learner and lecturer. It seems to boil down to the issue of a relationship between the learner and the lecturer, but with large classes it is not always possible for the lecturer to know his students well enough to know who will be discouraged by a certain style of feedback.

“Nice” feedback on the other hand, is not as easy as making a tick on the paper. Feedback that is simply indicative of satisfaction with a learner’s performance may prove to be unhelpful if it is given in a general or indiscriminate way. Askew and Lodge (2000:7) quote Brophy’s comment that “Infrequent but contingent, specific, and credible praise seems more likely to be encouraging … than frequent, trivial or inappropriate praise.”

These varying student reactions to feedback should therefore be monitored and if negative attitudes towards feedback, or negative competition among students, is observed, the situation should be addressed. That is why feedback should also be archiveable and searchable.

4.4.10 Feedback should encourage language awareness

As explained earlier, feedback may be used to counter fossilisation. Any attempt to highlight to students where they have not mastered the target language effectively can be seen as raising the students’ consciousness about language.

Consciousness-raising is a sensitive issue, since it touches on the question of whether or not formal instruction is useful. The term “formal instruction” refers to grammar teaching. It shows the importance and centrality attached to grammar in SLA (Ellis, 1996:611). Although there are researchers who claim that formal instruction is not useful, some research findings prove otherwise (Ellis, 1996; Ellis 1992: 232-241; James, 1997: 244). Formal instruction has been found to be advantageous for children and adults alike, and for intermediate and advanced students. Formal instruction has helped in acquisition-rich and acquisition-poor environments, as evaluated by means of different tests (Ellis, 1996: 613-614). Despite numerous arguments for formal instruction (Ellis, 1996; James, 1997: 246 and further for a detailed discussion), there is one limitation: formal instruction cannot alter the *route* of acquisition, but only the *rate* (Ellis, 1996: 631). In the end, learners who have had instruction demonstrate higher ultimate achievement (James, 1997:244; also compare Lightbown & Spada, 1999: 163).

Consciousness-raising is clearly linked to effective feedback and the trick is to provide feedback in such a way that it creates the opportunity for consciousness-raising to occur. Some researchers claim that consciousness-raising is not useful as it creates implicit knowledge, while learners use their explicit knowledge more. However, consciousness-raising does have a distinct advantage, as it has been found that once learners’ consciousness of a particular feature has been raised by formal instruction, they continue to be aware of that feature in subsequent communicative input (see James, 1997: 257 and further for a discussion). Learning is also not developmentally constrained, “which means

that there is no fixed order in which it must be learnt: it can be learnt and taught in any convenient order" (James, 1997: 257).

The other advantage of consciousness-raising is that it can go both ways. Moletsane (2002: 21-22) argues that if teachers are aware of the aims of marking and are conscientious in the application thereof, marking can be of invaluable assistance to them. This means that by indicating errors in students' work, the teachers themselves become aware of the language areas where their learners have trouble (see "Feedback should be purposeful"). If teachers are aware of the problem areas experienced by their students, they could adapt the input they produce for the learners in order to deal with these areas.

In order to highlight specific language features more effectively, feedback should sometimes focus only on specific errors without ignoring the others (see "Feedback should be thorough" for an explanation of why errors should not be ignored). Thoroughness and focus seem to cancel each other out. Focusing on specific errors does not, however, imply *not marking* others. Teachers should therefore decide on a method to emphasise a specific error by letting it stand out among the others (Ferris, 2002:4). This is not just selective marking; one should still mark everything, but should attempt to highlight a specific error.

4.4.11 Feedback should be individualised

When learners are expected to correct errors in exercise passages, the obvious problem is that "not every single learner in a class or group has committed all of the errors that are exposed. Those who have not made any or many of the errors might object to or be bored by the exercise" (James, 1997: 257). It is therefore important to know which students made which errors in order to provide personalised exercises or exposure (compare Ferris, 2002: 58-59). In addition, not all students react in the same way to feedback. While it is impossible to know how unknown students will react to feedback, lecturers should be aware that the same type of comment may have vastly different reactions from different students. Individual student reactions to feedback should be monitored in order to optimise teaching.

Feedback that is inapplicable to a student will not achieve the objectives of feedback, such as communication or rewriting. If a student cannot learn something from feedback, or does not need to learn something from it because it is simply too generic, the feedback will definitely not achieve the objective of being a learning opportunity. Archiving or tracing is one way of achieving this goal; see below.

4.4.12 Feedback should be time effective

Feedback will always *take* time from the lecturer, but it should never *waste* time. Feedback can be considered a waste of time when students discard it without looking at it or if the teacher does not have a clear purpose with it – in short, if feedback is not used properly and effectively to facilitate learning and improvement. The author believes that by adhering to the guidelines in this checklist, feedback would not waste the lecturer's time as it would be more effective and not just a futile "must-do" exercise.

4.4.13 Feedback should be searchable, archiveable or recordable and also allow for research

Feedback should be in a form that allows students to go back to it later.

A written form is not necessarily the only format in which students can return to feedback, but it is the most freely available form. The important issue here is not the format, but the fact that students and lecturers are able to return to the feedback later. One of the biggest problems in feedback practice is that students and lecturers alike are unable to identify recurring patterns of errors in student writing (Wible *et al.*, 2001: 308-310). If feedback is not in a format allowing for longitudinal analysis, it is even more difficult to identify recurring errors.

Many teachers of language are obviously not interested in doing a longitudinal analysis of their students' performance or development, but the aim here is not hard-core continued research. Rather, the aim is an opportunity for the teacher and student to step back and take a more objective look at their language learning experience. Individualisation is important for this as well to assist students in evaluating *their own* learning experience.

James (1997) makes a strong case for error analysis and the effectiveness of consciousness-raising. Unfortunately, most feedback ultimately disappears into students' waste baskets. A feedback system seldom provides the opportunity to study longitudinally the occurrence and change in feedback on a particular student's writing. Taking into account the immense amount of time spent on marking, this seems like a terrible waste of information. (Student portfolios are one way to counter this as students are at least asked to keep their written assignments a while longer). The problem here is that many quite important errors may disappear in a forest of other less important markings (Wible *et al.*, 2001: 308-310) with the result that neither a student nor a lecturer will see the importance of these errors. Students should also be in a position where they can do "research" on their own texts to find their personal weaknesses in order to attend to them. They will not do so, however, unless they see feedback as part of their learning experience and take responsibility for their own learning.

Every teaching situation differs. An analysis of different teaching situations will ensure better and more focused teaching. It is not feasible for teachers to do large-scale error analysis for every one of their classes, but if feedback is used correctly it may not be necessary to do large-scale research in order for it to be useful. Students may be able to do this themselves.

Ultimately, for it to be useful for research (both intensive research and self-evaluative research), feedback must be more than hieroglyphics and must be stored in some kind of way – preferably in a searchable format.

4.5 Conclusion

Feedback can be effective, but the way it is currently practised presents some problems. Feedback should be more than a mechanical action expected by students and society. It seems that what is needed is a drastic change in the attitude towards feedback (Moletsane, 2002:21) and perhaps more standardised techniques. If teachers and learners alike see and

use feedback as an important source of source language input, it would be possible to get much more teaching and learning out of something being done in any case every day.

The stance taken in this thesis is that the question should not so much be "Should feedback be provided?" since it is provided in any case. Rather, the question to be asked is "How can feedback be optimised?"

4.6 Acknowledgments

The author would like to thank Professor Bertus van Rooy and Professor A. M. de Lange for their comments on the ideas in this article and the its structure.

4.7 Bibliography

Askew, S. & Lodge, C. 2000. Gifts, ping-pong and loops – linking feedback and learning, pp. 1-17. In Askew, S. (ed.) *Feedback for learning*. London: RoutledgeFalmer.

Ellis, R. 1996. The study of second language acquisition. Oxford University Press

Ferris, D. 2003. Responding to writing, pp. 119-133. In Kroll, B., ed. *Exploring the dynamics of second language writing*. Cambridge: Cambridge University Press.

Ferris, D.L. 2002. Treatment of error in second language student writing. University of Michigan Press.

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, K. 1990. Providing productive feedback. *ELT Journal.* 44(4): 279-285

James, C. 1997. *Errors in language learning and use – exploring error analysis*. Edinburgh Gate: Longman

Krapels, A.R. 1990. An overview of second language writing process research, pp. 11-23. In Kroll, B. (ed.) Exploring the dynamics of second language writing. Cambridge University Press.

Krashen, S.D. 1985. The input hypothesis: issues and implications. London: Longman.

Lightbown, P.M. & Spada, N. 1999. How languages are learned – revised edition. Oxford University Press.

Louw, H. 2006. Standardising written feedback on L2 student writing. Potchefstroom: North-West University. (Dissertation – M.A.).

Lyster, R. & Ranta, L. 1997. Corrective feedback and learner uptake. *SSLA* 20: 37-66.

Moletsane, J.R. 2002. Selective error correction. In ESL narrative compositions. Potchefstroom: PU for CHE. (Mini-dissertation – Magister Artium).

Monyaki, B.S. 2001. A comparison of teachers' and learners' views on feedback strategies used in the writing process. Potchefstroom: PU for CHE. (Mini-dissertation – Magister Artium).

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. University of Pretoria (D.Litt thesis).

Paulus, T.M. 1999. The effect of peer and teacher feedback on student writing. *Journal of Second Language Writing* 8(3): 256-289.

Selinker, L. & Lamendella, J.T. 1979. The role of extrinsic feedback in interlanguage fossilisation – a discussion of "rule fossilisation: a tentative model". *Language Learning* 29(2): 363-75.

Spencer, B. 1998. Responding to student writing: strategies for a distance-teaching context. Pretoria: University of South Africa (D.Litt thesis).

Storch, N. & Tapper, J. 1997. Student annotations: what NNS and NS university students say about their own writing. *Journal of Second Language Writing* 6(3): 245-264.

Truscott, J. 1996. The case against grammar correction in L2 writing classes. *Journal of Second Language Writing* 46:327-369.

Van der Walt, J.L. & Van Rooy, B. 2002. Towards a norm in South African Englishes. *World Englishes* 21(1): 113-128.

Van der Walt, J.L., Van der Walt, I. & Dreyer, C. 1994. The effect of error correction on the grammatical competence of ESL students. *Journal for Language Teaching* 28(1): 7-17.

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

CHAPTER 5
ARTICLE 3 – MOVING TO MORE THAN EDITING: STANDARDISED FEEDBACK IN PRACTICE

5.1 Prelude to Article 3

Adhering to the standards set for effective feedback is more difficult than it seems at first glance. As in many other situations, what seems ideal on paper is often difficult to put into practice. Also, theory may predict an ideal solution, but in practice a technique often simply does not work that way. Hypothetically, by using the semi-standardised feedback tags developed earlier (Louw, 2006) one would be adhering to most of the requirements for effective feedback. The feedback showed improvement in the students' writing, but within the context of the two-way communication situation which is feedback on writing (Askew & Lodge, 2000), will it be effective with the other role players (the markers)?

One of the first big hurdles for the MarkWrite project was therefore to establish whether lecturers (markers) can use it consistently. A first version of MarkWrite (at this stage still simply called "Essay marker") was programmed with the intention of testing the standardised feedback in practice from the lecturers' side. In other words, the test was intended to establish whether standardised feedback could be implemented practically.

This article reports on this experiment. The results were reported at the SAALA conference in Potchefstroom in July 2006.

Publication information for Article 3

This article was published in 2007 in the journal *Ensovoort*. Minor editorial changes were made for the sake of consistency with the general format of the thesis.

The complete bibliographical entry is:

Louw, H. 2007. Moving to more than editing: standardised feedback in practice. *Ensovoort* 11(2): 83-104.

Bibliography

Askew, S. & Lodge, C. 2000. Gifts, ping-pong and loops – linking feedback and learning. In Askew, S., (ed.) Feedback for learning. London: RoutledgeFalmer, pp.1-17)

Louw, H. 2006. Standardising written feedback on L2 student writing. Master's dissertation, North-West University (Potchefstroom Campus).

Abstract

This article reports on an experiment which tested how effectively standardised feedback could be used when marking L2 student writing. The experiment was conducted by using a custom-programmed software tool and a set of standardised feedback comments. The results of the experiment prove that standardised feedback can be used consistently and effectively to a degree, even though some refinements are still needed. Using standardised feedback in a standard marking environment can assist markers to raise their awareness of errors and in more accurately identifying where students lack knowledge. With some refinements, it may also be possible to speed up the marking process.

Keywords

Feedback, second language writing, student writing, standardisation, error

5.2 Introduction and background to the project

The process of providing feedback (marking) on student essays is usually very time-consuming. Considering the amount of time spent on it by teachers and the amount of attention paid to it by students, it may be regarded as one of the least effective duties of language teachers (Moletsane, 2002:21; Hyland, 1990:282).

In 2004 a project commenced at the North-West University to investigate the possibility of getting more "teaching" out of marking. The main objectives of the study were to:

A. Establish whether standardised feedback would ensure more clarity for the student (Louw, 2006)

B. Create a system to keep effective records of student development (cf. Wible, Kiu, Chien, Liu and Tsao, 2001; Louw 2006)

C. Establish whether standardised feedback could be used consistently

D. Establish whether standardised feedback would ensure ease of use for the marker

E. Force students to pay attention to the feedback.

Spencer (1998) researched strategies for responding to student writing, while Wible *et al.* (2001) created an electronic marking system to keep track of student development. Wible *et al.'s* (2001) system did not work with standardised feedback, while Spencer found that with current working limitations, certain marking strategies work better than others. The project at the North-West University aimed to integrate these findings into one project. The first step was to establish whether or not feedback could be standardised to an extent and if it would actually benefit the student. Louw (2006) found this to be the case. This article reports on objectives C and D as discussed above. In addition, the feasibility of the marking system is also addressed.

For background purposes, a brief overview of the findings on feedback is presented in the next section.

5.3 What is effective feedback?

To provide standardised feedback, it was first necessary to establish exactly what constitutes feedback as well as the nature of effective feedback. The different classifications of feedback are too numerous and intricate to discuss here (Louw, 2006),[3] but some of the important facts about feedback can briefly be summarised as follows:

- The interpretation and use of the concept "feedback" is closely related to the user's definition of and attitude to "error".
- Feedback is not just error correction, but any response (positive or negative) on a student text by any reader of the text (Hyland, 2003, 1990, 1998; Lyster & Ranta, 1997; Askew & Lodge, 2000). By definition then, feedback can be provided in many different ways.
- Depending on the purpose or background of the "feedback giver", feedback can be classified as performing many different functions:
    - evidence (linguists)
    - repair (discourse analysis)
    - correction (L2 teachers)
    - focus on form (SLA researchers) (Lyster & Ranta, 1997:38).
- Learners expect feedback, but often neglect to look at it (Ferris, 2002:13-14).
- There are many advantages and disadvantages to feedback, but there are conflicting research findings regarding these (Ferris, 2003:127; Louw, 2006).

Because there are so many overlapping and contrasting definitions of feedback (Louw, 2006),[4] the following working definitions will be used in this article:

*An **error** is any instance in a text which is incorrect language use or language use which is not inherently wrong, but which could be improved.*

***Feedback** is any reaction to a text by any reader of the text, for the purpose of pointing out errors to the writer. In keeping with the definition of "error", feedback could also indicate satisfaction with something in the text.*

[3] At the time of publication of this article, Article 1 had not yet been published due to delays in the reviewing process, but the information is equally relevant to the definition of effective feedback. See Louw, 2009. Moving to more than editing: a checklist for effective feedback. *Journal for Language Teaching*, 43(2).

[4] Once again, Article 1 is applicable: Louw, H. 2009. Moving to more than editing: a checklist for effective feedback. *Journal for Language Teaching*, 43(2).

5.2 Can feedback be standardised?

In Louw (2006) a standardised set of feedback tags was created. The tag set contains a list of "popular" error tags used by markers (as established by research and in an experiment by Louw, 2006) and in corpus linguistics (Granger & Meunier, 2003). The tag set is constantly being updated and refined. An example of the current version (at the time of writing) is attached as Addendum A. An experiment proved that marking, using this set of standardised feedback comments, is more effective than the normal squiggles, lines, strike-throughs and question marks often used by lecturers (Louw, 2006). For reference purposes, these squiggles and other feedback marks are called hieroglyphics. See Figure 5.1 for an example of hieroglyphics.

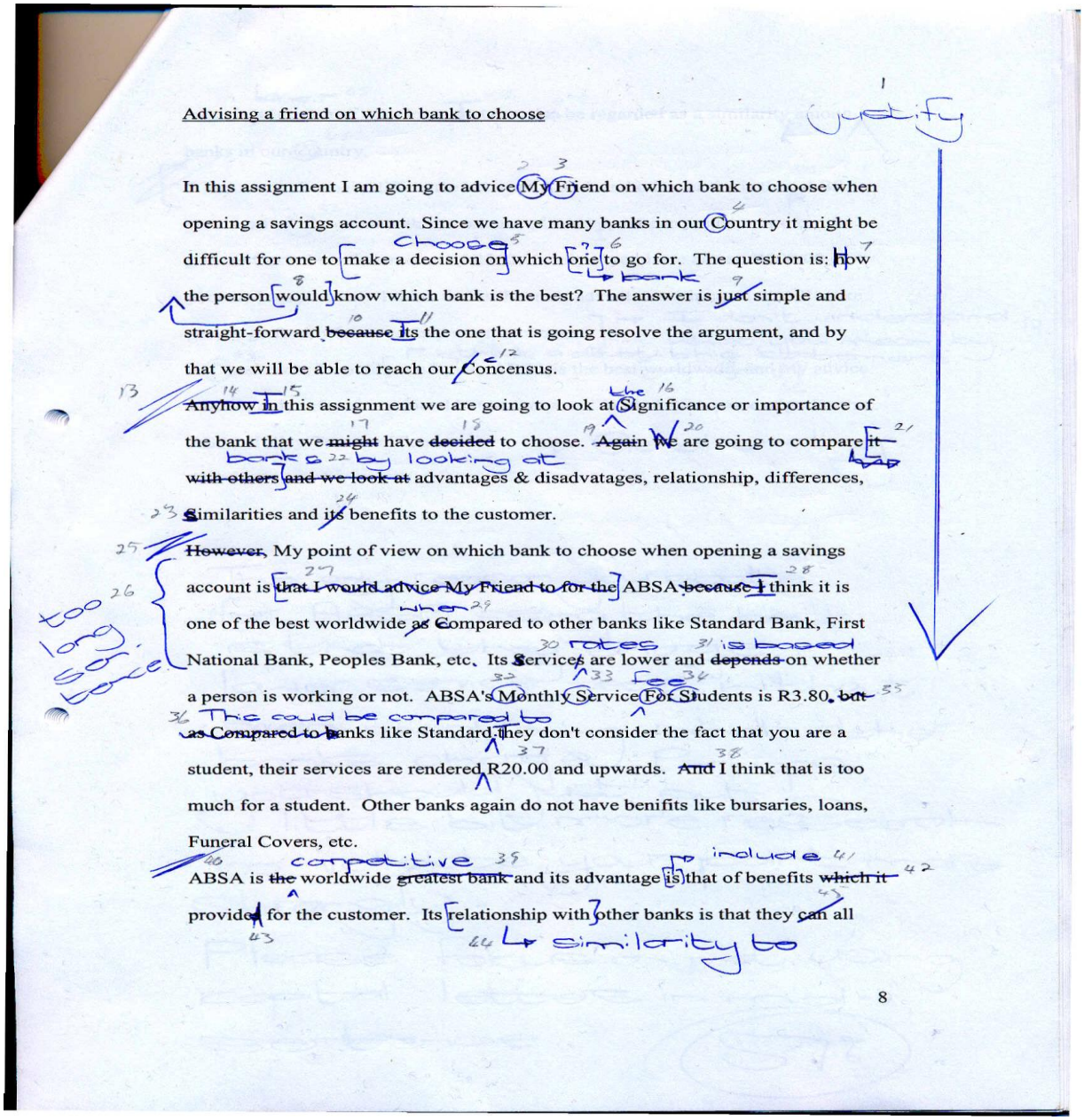

Figure 5.1: Example of hieroglyphic feedback

In the experiment in Louw (2006), a hieroglyphic marking technique was pitted against standardised feedback and a blank text on which the number of errors were indicated, but no errors were marked. The results indicated that students were seldom able to identify errors in the blank texts, much less correctly revise them. On the other hand, students were able to correctly revise errors marked with the hieroglyphic feedback, but the standardised feedback proved to deliver the greatest improvement in all categories tested. This shows that students are often able to revise errors once they are indicated to them, but in order to facilitate maximum improvement in writing, standardised feedback is more effective.

Based on these findings, the question then arises whether standardised feedback could be implemented in practice with consistency. Another experiment was therefore conducted to investigate the following questions:

A. Can standardised feedback be used consistently by markers?[5]

B. Will standardised feedback make it easier for markers to mark texts effectively?

The rest of this paper reports on this experiment.

5.4 Methodology

Four markers (two experienced and two inexperienced) were instructed to mark a number of L2 learner essays using the feedback tag set as a tool. The essays for the experiment come mostly from the Tswana Learner English Corpus (TLE) (Van Rooy & Schaefer, 2002), with a minimum number of essays from the Afrikaans Learner English Corpus which is still under construction.

The tag set was incorporated into a custom-built software package. The software package imported the text to be marked into a marking window and had the entire standardised feedback tag list in a drop-down tree view on the right of the text (see Figure 5.2). When finding an error, markers had to simply highlight the error and click on the relevant error category. The computer would then insert the feedback. Markers were shown how the system worked but were not given any additional instructions on how to mark or on what to provide feedback. Additional feedback comments not covered by the tag set could also be added by making use of the comment box. It should be noted that the drop-down tree view simply indicates the “name” of the specific error tag used by the markers, and is not the (more complete) feedback a student would see (compare Addendum A and Figure 5.3).

[5] With reference to the introduction of this thesis, these two questions are directly related to the questions of what works in feedback and how this knowledge can be applied.

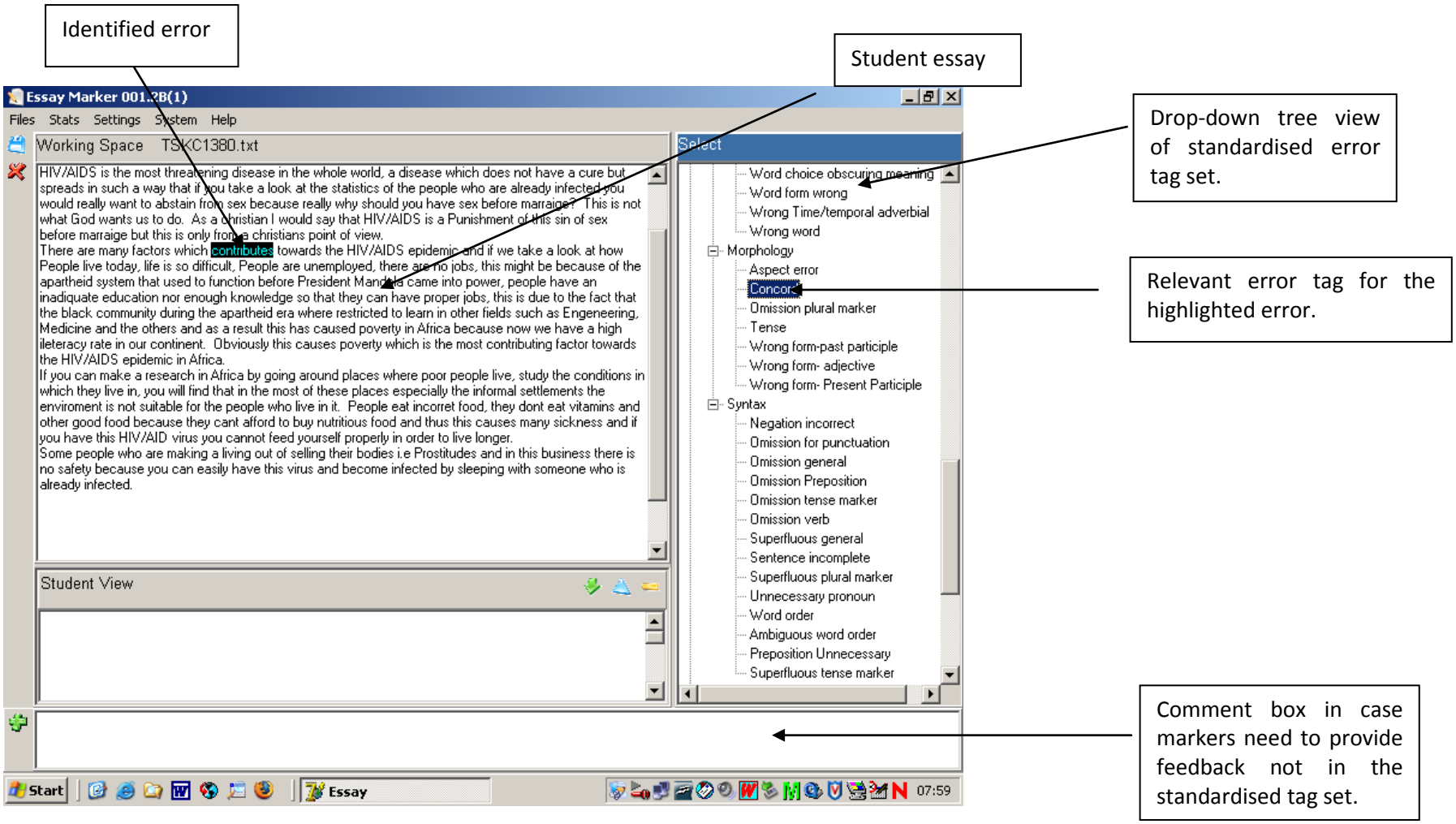

Figure 5.2: Essay marker screenshot

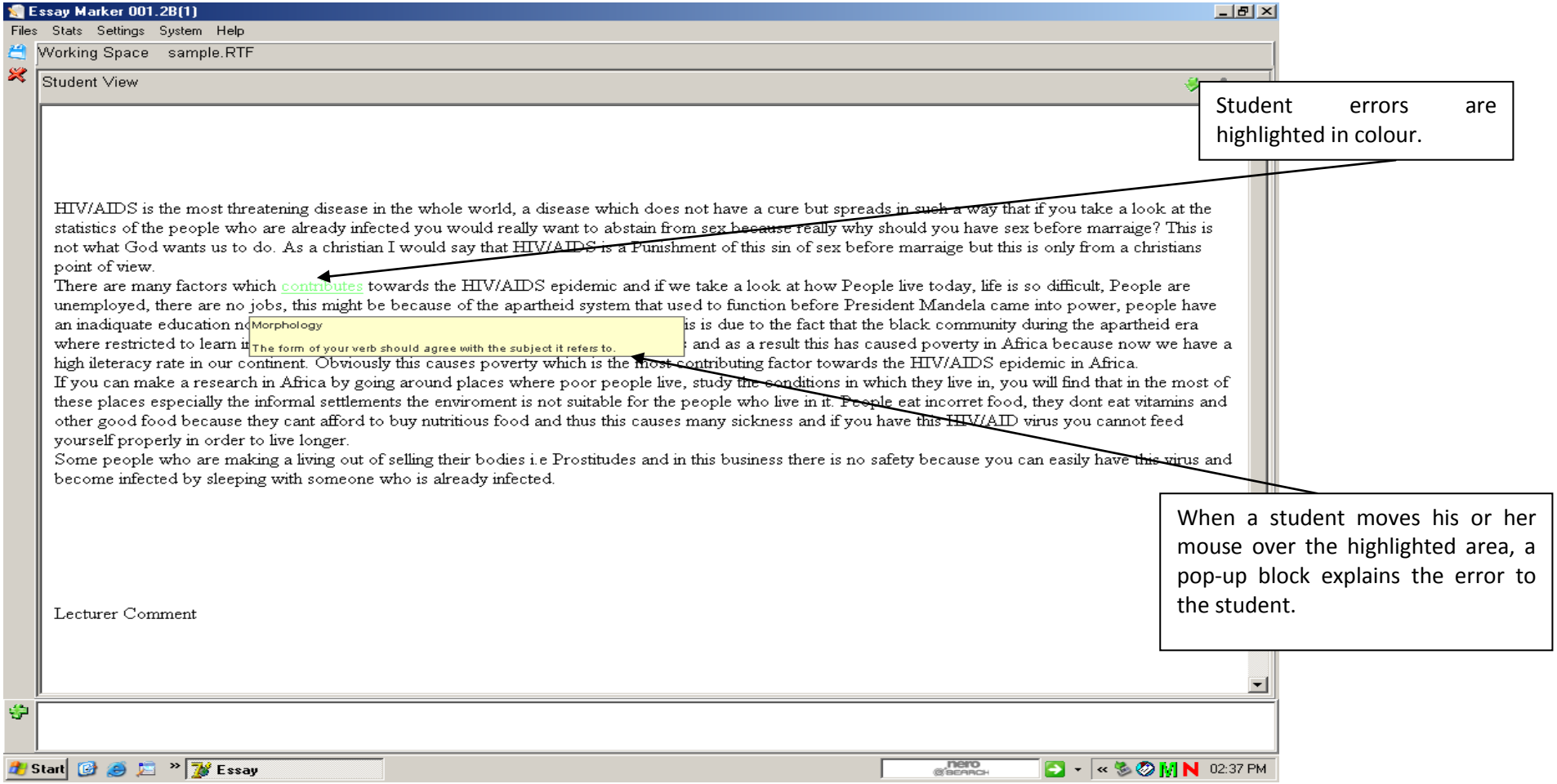

Figure 5.3: Student view illustrating how students will receive feedback

Figure 5.2 illustrates the view of the marking system as the marker would see it. In this figure, the marker has identified a concord error and highlighted the error. The marker has also identified the relevant error tag (in this instance “concord”) in the tree view on the right-hand side of the screen.

Figure 5.3 illustrates how the student would receive this specific feedback. The student will get his or her text back in an HTML file which can be opened by any standard web browser such as Internet Explorer or Mozilla Firefox. Errors are highlighted in different colours, for example red indicates a grammatical error. In this case, the student has moved his or her cursor over the highlighted word and got the pop-up message: “The form of your verb should agree with the subject it refers to”.

The marked essays were stored electronically with their error tags for analysis. The error tags the markers used were extracted from the database. The markers did not mark an equal number of essays. To counter this problem and in order to compare apples with apples, the data were normalised. The number of times a tag was used was reworked to the number of times it was used per 1 000 errors tagged. All numbers reported in this paper therefore refer to the normalised totals.

In addition to analysing the tags used, the markers were also sent a questionnaire with 18 questions asking them about their experiences with the marking system. These questions are included in Addendum B.

As mentioned earlier, the research aimed to answer two questions:

1. Can standardised feedback be used consistently?
2. Is the system easy for the markers to use and if not, how can it be improved?

In order to answer question 1, two types of analyses were done on the marked essays.

1. An analysis of all the tags was done to establish marker tendencies.
2. A close analysis of the way the markers used the tag “better word” was conducted. (This tag was chosen because all four of the markers used it as one of their top ten favourite tags.)

The answers to the questionnaire were also used to judge the consistency with which the markers marked. To answer question 2 regarding the ease of use of the marker system, the answers provided on the questionnaire were used.

5.5 Results

5.5.1 Can standardised feedback be used consistently?

5.5.1.1 Marker tendencies

Table 5.1 indicates the top 20 tags used by the markers. The first three columns identify and explain the error tag (see Addendum A for additional clarification), while “Knorm”, “Mnorm”, “Pnorm” and “Tnorm” show the number of times the specific tag was used by the

different markers. The column “Normed Total” indicates the total number of times a tag was used out of a total of 4 000 marked errors.

Table 5.1: Top 20 tags used by the markers

	Subordinate	Domain	Set (Error Tag)	Knorm	Mnorm	Pnorm	Tnorm	Normed Total
1	Presentation	Spelling	Spelling/typing error	105	92	150	62	409
2	Grammar	Lexis	Wrong word	92	40	87	47	266
3	Presentation	Capitalisation	Capitalisation	18	130	33	25	206
4	Grammar	Syntax	Superfluous general	53	35	26	59	173
5	Grammar	Morphology	Concord	19	46	79	25	169
6	Presentation	Punctuation	Punctuation missing	23	44	47	51	166
7	Grammar	Lexis	Better word	8	16	81	34	139
8	Grammar	Syntax	Omission general	40	26	17	32	115
9	Grammar	Morphology	Tense	42	29	5	38	114
10	Discourse	Coherence	Reasoning inconclusive	73	2	8	30	113
11	Grammar	Lexis	Word form wrong	21	36	40	16	112
12	Grammar	Morphology	Omission plural marker	14	61	6	29	109
13	Grammar	Lexis	Article missing	2	62	7	19	90
14	Grammar	Lexis	Preposition wrong	15	16	35	23	89
15	Discourse	Style	Sentence vague	53	2	23	10	88
16	Grammar	Lexis	Determiner incorrect	18	39	2	26	85
17	Discourse	Factual correctness	Facts wrong	14	2	8	44	69
18	Presentation	Punctuation	Punctuation wrong	25	5	28	10	68
19	Discourse	Cohesion	Reference vague	9	18	0	33	61

20	Grammar	Lexis	Word choice obscuring meaning	15	28	9	7	59
20	Grammar	Syntax	Omission verb	4	31	10	14	59

With regard to the top 20 tags, the following points are interesting to note:

1. Four of the top 20 are errors which are only present in writing: Punctuation wrong, Punctuation missing, Capitalisation, Spelling/typing error. These are surface element errors only.

2. Seven errors have to do with lexis: Wrong word, Better word, Word form wrong, Article missing, Preposition wrong, Word choice obscuring meaning, Determiner incorrect.

3. Three error tags have to do with morphology: Concord, Tense, Omission plural marker.

4. Three errors have to do with syntax: Superfluous general, Omission general, Omission verb.

5. Only two error tags deal with coherence or cohesion and both of these are coherence on a small scale – within paragraphs or within sentences. The relevant tags are: Reasoning inconclusive and Reference vague.

The results indicate that the markers did not only focus on surface elements. Style, coherence and the accuracy of facts feature, but only on a small scale – within the paragraph or sentence level. These results are the same for all four markers. Although it is not ideal to focus on surface structure errors, these results are no different from those found in previous studies (Louw 2006:103).

One can deduce that:

1. The markers were relatively consistent in focusing more on surface level errors, even though they did not actively work together and even though they focused on different surface level errors.
2. Errors other than grammar, spelling and punctuation are markedly more difficult to identify.
3. Surface structure errors bothered the markers – to such an extent that they even ignored other errors. Admittedly, there were so many errors in some of the sample essays that it was extremely difficult to mark for argument. See example texts one and two in Addendum B.
4. Writing seems to be an effective way to notice a poor lexicon.

5.5.1.2 The markers' personal favourites

For purposes of comparison, the inter-marker consistency (top ten tags used) is included in Addendum D. In Table 5.2 the tags which occurred in all the markers' top ten are presented.

A count of four therefore indicates that all four markers used the tag as one of their most frequently used tags, while a count of three indicates that three of the four markers used the tag as one of their most frequently used tags.

Table 5.2: Tags that occurred in the makers' personal favourites

Error tag	Occurrence in top ten of marker favourites
Punctuation missing	4
Spelling/typing error	4
Wrong word	4
Superfluous general	3
Word form wrong	2
Better word	2
Capitalisation	2
Concord	2
Omission general	2
Punctuation wrong	2
Tense	2

Very low inter-marker consistency is evident here. Although it can be argued that all the markers should have marked the exact same essays, the focus of the experiment was not just on consistency, but also on ease of use for the marker. A broader spectrum of essays to mark generated a broader spectrum of possible uses of the tag set and was hence a more thorough test of the marking system. In addition, these essays were all written by students of more or less the same competency (first or second-year undergraduate L2 users of English), so the comparison (although not perfect) can still be seen as legitimate.

Only three tags occur in all four markers' top ten. These are "punctuation missing", "spelling/typing error" and "wrong word". The only tag to occur in three of the four top ten lists is "superfluous general" which is used to indicate superfluous words.

If markers have to focus constantly on incorrect word choice, punctuation, overuse of words and incorrect spelling, it is likely to point to two interacting issues:

1. Students have a very poor ability to make themselves understood, which forces markers to indicate these errors in an attempt to point out that they were unable to understand the text.
2. Markers are overly finicky with regard to surface level errors or simply find it easier to comment on them.

5.5.1.3 Least used tags

Table 5.3 presents the tags least used by the markers. Although these tags were available, the markers tended not to use them much.

Table 5.3: Least used tags

Subordinate	Domain	Set	Knorm	Mnorm	Pnorm	Tnorm	Normed Total
Discourse	Style	Active voice	0	0	0	1	1
Grammar	Syntax	Superfluous tense marker	1	0	0	0	1
Grammar	Syntax	Omission tense marker	0	0	0	1	1
Grammar	Syntax	Preposition unnecessary	0	0	0	2	2
Discourse	Style	Passive voice	0	0	0	2	2
Discourse	Positive Comments	Good reasoning	2	0	0	0	2
Grammar	Morphology	Wrong form – adjective	2	0	0	0	2
Grammar	Syntax	Negation incorrect	1	0	1	0	2
Discourse	Structure	Paragraph jumbled	1	1	0	0	3
Discourse	Style	Construction overuse	0	1	0	2	3
Grammar	Lexis	Quantifier error	1	0	2	0	3
Discourse	Structure	Paragraphing: Relate or move	0	0	3	2	5
Discourse	Style	Gender bias	0	0	0	5	5
Grammar	Lexis	Wrong time/temporal adverbial	0	3	2	0	5
Discourse	Factual correctness	Unsupported argument	0	0	0	6	6
Grammar	Lexis	False friend	3	0	2	1	6
Grammar	Lexis	Problem with conditional	2	0	3	1	7
Presentation	Layout	Layout inhibits reading	0	0	7	0	7
Discourse	Structure	Introduction weak	0	0	4	3	8
Discourse	Structure	Paragraph: Start new	3	0	3	2	8
Grammar	Morphology	Wrong form - past participle	2	2	2	3	9
Discourse	Style	Register too formal	8	0	0	1	9
Discourse	Positive comments	Interesting point	8	0	2	0	10
Grammar	Morphology	Aspect error	5	0	5	0	10
Discourse	Structure	Paragraph: weak opening sentence	3	4	0	3	10
Grammar	Syntax	Preposition unnecessary	2	4	0	5	11
Grammar	Syntax	Sentence incomplete	0	1	10	1	12
Grammar	Lexis	Pronoun wrong	8	0	1	6	15
Grammar	Lexis	Inappropriate word	13	0	0	2	15

Discourse	Style	Verbosity	1	2	2	13	18
Discourse	Factual correctness	Opinion	3	2	3	11	20
Presentation	Punctuation	Apostrophe error	1	12	0	7	20
Discourse	Factual correctness	Reference omitted/wrong	0	0	18	1	20
Grammar	Syntax	Unnecessary pronoun	7	7	0	7	21
Discourse	Factual correctness	Unbalanced statement	7	0	1	14	22
Grammar	Lexis	Wrong modal	4	2	2	15	23
Grammar	Morphology	Wrong form – Present participle	8	9	5	3	25
Grammar	Syntax	Omission for punctuation	0	15	3	9	27
Discourse	Coherence	Inconsistency	12	0	9	6	27

Regarding the least used tags, the following issues appear to be significant:

1. Five of the six tags in the domain "structure" occur among the least used tags. The markers therefore seldom touched upon the issue of paragraph structure. The domain "structure" falls under errors of discourse and refers mainly to errors concerned with paragraphing.
2. Positive comments are also amongst the least used tags with only two of the markers ever using positive comments, and then only extremely sparingly.
3. The rest of the least used tags are issues of grammar that are either low frequency ("superfluous tense marker") or more difficult to identify ("aspect error").

The results indicate that the markers still sometimes had editing or spell checking in mind when marking the student texts. However, one should keep in mind the virtual incomprehensibility of some of the student texts (see Addendum C). With some of these texts it would be very difficult to comment on structure since it is difficult to understand the text in the first place. Most of the texts are not that difficult to comprehend though, so in spite of some difficulties the question remains: How can markers be assisted to be more than spell checkers who simply look for instances of surface level incorrectness?

5.5.2 Close analysis of the use of one tag

The tag "wrong word" was selected for close analysis since it was in the list of top ten used tags of all four markers. The results indicate that intra-marker consistency was relatively good, while inter-marker consistency needs some work. Markers will need some training in order to be consistent *with one another,* or will need to work together more closely. The close analysis of this specific feedback tag highlighted the following problems:

1. Favourite generics: the markers tended to use the tag as a generic term instead of using more specific available tags.

2. Doubles: in some instances more than one tag may have applied, but the markers reverted to the one they used previously.

3. Personal preference resulted in tags of “wrong word” where “better word” would have been a better option.

4. Incorrectly tagged.

5. Errors that were difficult to classify.

Each of these will be discussed briefly.

The markers tended to use the tag as a generic term instead of using more specific available tags.

Examples:

1) They must go to the streets to beg for <Wrong word>money to eat</Wrong word>.

In example (1), the tag “Sentence Ambiguity” could have worked better. “They” do go to the streets to beg for money *to buy food* to eat. The problem therefore lies much more with the sentence construction than with the word choice.

2) …an <Wrong word>infinitive</Wrong word> tapestry of green maize and yellow sunflower

In example (2), the word “infinitive” should be “infinite”. A better tag would therefore have been “word form wrong”. It is therefore a morphological error rather than a lexical or semantic error. On the other hand, one can make an argument for the tag “better word” also in order to use “seemingly endless” or “never-ending” if the idea of an “infinite” farm proves problematic. Part of the problem therefore lies in the interpretation of the error.

The problem of using a favourite tag as a generic tag can be overcome by some collaboration between the markers and more specialised training of the markers. Presumably as markers become more used to the system and get to know the tags better, they will be more aware of additional tags they can use instead of their favourite generic tag. The process of providing feedback continuously on multiple drafts written by students could also assist in this.

5.5.3 Doubles: More than one possible tag

3) … what they do to keep <Wrong word>this</Wrong word> <Word form wrong>tradition/s</Word form wrong> from**…**

In example (3), the difficulty is that there are many possible ways to correct this sentence. Should it be “these traditions” or “this tradition”? The context will normally dictate the answer. In this specific case, however, one can ask if the problem is a “wrong word”, a “wrong form” or an “omission plural marker”.

Another example of several possible interpretations of an error is the word “irregardless” as used in one of the essays. The dictionary classifies it as informal so it is a question of style rather than a blatantly “wrong word”. The tag “better word” could work as well.

There is a similar occurrence of double errors with run-on sentences. Is it classified as "omission punctuation" or "run-on sentence"? The initial idea was to simply tag double errors with both applicable tags, but due to a technical limitation in the marking program, that was not possible. With the system as it is now, the question facing the marker is which tag to use. It seems that the markers usually opted for their "generic favourite".

5.5.4 Personal preferences

4) In this essay <Wrong word>one</Wrong word> is going to try to <Wrong word>prove</Wrong word> that the prison system is outdated…

In example (4), the word "prove" could also be tagged as "better word" if the marker was of the opinion that "argue", "show" or "demonstrate" would have been a better choice. This is, however, a harsh judgement by the marker which is clearly indicative of a personal preference.

5.5.5 Incorrectly tagged

5) …how the Romans whipped and <Wrong word>cruxivied</Wrong word> them…

The word "cruxivied" in example (5) is definitely not a wrong word. It is the *correct* word that had been spelled incorrectly so the tag should have been "spelling/typing error". The student could now be under the impression that "crucified" is not the correct word to use and will struggle in vain to find the "correct" word to use.

6) Every bank has different options regarding a savings account to <Wrong word>consider.</Wrong word>

In example (6), the word "consider" is not such a big problem as some omitted words: "Every bank has different options regarding a savings account **one has to** consider," is a possible correction.

7) We all know this catchy tune and I am sure a lot of us actually gave it some <Wrong word>tough</Wrong word> one time or the other.

This is a spelling error in example (7). The student meant "thought". Because of the incorrect tag the student still does not know that he or she simply misspelled "thought" but is instead under the impression that "thought" is the wrong word to use.

5.5.6 Errors that were difficult to classify

Some errors were difficult to classify because it was not immediately evident what the learner wanted to say. It is only possible to tag an error once it can be established what the intended meaning was. This is the case especially when learners write very long "sentences" without any verbs or where their use (or lack) of punctuation obscures meaning, and also where the sentence constructions do not follow convention. Example (8) illustrates this problem.

8) Compulsory modules some of them are good but some of the they are full of nonsense because we gain nothing from them while others are very good because they prepare us for the working enviroment and also to become good professionals. Modules like sociology especially if you want to do community work is great and also koms because we learn how to communicate or improve our communication skills since were are preparing to be professionals. Modules like entr and wtll I don't really know why we should do them cause according to me is total waste of time and money. Any way we just have to do them because we don't have choice at all.

In example (8), there are so many possibilities for improvement that it is difficult to decide how to go about it. The urge would be to simply rewrite the whole section to illustrate to the student where he or she erred, but that would not be effective feedback (although it might be considered "input").

5.5.7 Consistency: Marker comments

The questions asked of the markers brought the following to light:

Only one marker was able to correctly identify the tag he or she used most. The others indicated error tags that were not even in their top ten. This indicates that the markers were often not consciously aware of what they focused on even though they were under the impression that they paid attention to more than "editing errors". However, one should keep in mind that three of the four markers indicated that they often had difficulty in understanding what exactly the students intended to say. This makes it understandable that they focused on the surface level errors instead.

All the markers indicated that they found the students' ability to present an effective argument underdeveloped. One marker explained, "Sometimes students showed great insight and had impressive ideas, but they were unable to incorporate them into the argument. Usually any insight was lost in a sea of words."

On the other hand, one marker indicated that he or she consciously decided to ignore spelling or typing errors since he or she found the other errors of more importance. From these comments, it seems that the marking system should also have been tested on texts other than the Tswana Learner Corpus. It is possible that if markers are able to understand the text better, they will mark it with more care and more comprehensively. This will, it is hoped, also result in greater consistency. .

As far as consistency is concerned, all the markers turned out to be more than just spelling checkers, but there were differences in what they focused on. This problem can probably be overcome if the system is used as part of a well-structured writing lesson where all the markers know what the aim of the exercise is, and therefore focus on the same issues in unison.

5.5.8 Ease of use for the marker

An integral part of the project to provide better feedback was to try and make it easier and faster for the marker to provide more (and more thorough) feedback. Unfortunately the

software used for this experiment was a prototype version and contained many bugs and system limitations which hampered the process.

Despite the bugs, the markers indicated that they could get a turn of speed from the system, especially once they got to know the tag set. Unfortunately, it still took between 10 and 30 minutes per 500-word essay. Before any harsh judgements on the time effectiveness of the system are made, it has to be compared to normal manual marking. Even if it turns out that it takes just as long to mark an essay with the marking system as traditional manual marking takes, the system has more advantages than traditional marking.

It seems that the main reasons for the slow marking were the following:

1. Bugs and limitations in the system.
2. It takes a while to classify an error; a simplified tag set (less elaborate) may streamline the marking process. A balance needs to be found between the explicitness of the feedback and the value students get from it.
3. The markers were not used to reading text on a computer screen, which slowed their reading speed.

Solving these problems will speed up the marking. In addition, the following plug-ins are being considered for the system, which could greatly assist markers to speed up the marking process:

1. A custom spelling and grammar checker that can identify and mark surface level errors before the teacher even gets the text.
2. A “focus” function which allows the teacher to use specific tags only, enabling him or her to focus on certain aspects at a time.
3. A teacher prompt function to remind the teacher to use a greater variety of tags.
4. A voice prompt to enable markers to use their voice to insert an error tag instead of with mouse clicks.

However, these possible solutions are time-consuming and expensive to develop, test and implement.

5.6 Conclusion

This article commenced with the questions of whether standardised feedback could be used consistently and whether it would make the life of the marker easier. The answers are that although there is still much work to be done, the initial findings are very positive and that the system still needs a lot of refinement.

The first testing of the system indicated that it has certain advantages:

1. It can assist markers to raise their awareness of errors.
2. A regular analysis of the error tags used could assist the teacher to identify where students lack knowledge.

3. A regular analysis of the error tags used could assist the teachers to identify where they are overly sensitive to a specific error, or fail to pay attention to important errors.

In addition, the experiment emphasised the following problems:

1. It is difficult to provide effective feedback since it entails tiring thought processes and error analysis.
2. Detailed feedback on surface level errors is possible in a standardised way, but markers will need some assistance to consciously move away from merely editing students' work.
3. Some errors can be classified in more ways than one. This makes it difficult for markers to be consistent in how they mark. The problem may be solved with adaptations to the error tag set.
4. At present, feedback remains a time-consuming activity.
5. A single standardised set of feedback tags does not seem to be useful for different levels of students, since the weaker students make so many surface level errors that the text is difficult to mark.

Addressing these problems will require lengthy research and plenty of computer programming, but at least a start has been made and the data show that the project is heading in the right direction.

5.7 Author's note

The financial support of the National Research Foundation (grant number FA2004043000051) is gratefully acknowledged. All opinions expressed in this article are the author's, and should not be attributed to the National Research Foundation. The author also wishes to thank Professor Bertus van Rooy for his assistance as well as three anonymous reviewers for their comments.

5.8 Bibliography

Askew, S. & Lodge, C. 2000. Gifts, ping-pong and loops – linking feedback and learning. In Askew, S., (ed.) Feedback for learning. London: RoutledgeFalmer, pp.1-17).

Ferris, D. 2003. Responding to writing. In Kroll, B. (ed.) Exploring the dynamics of second language writing. Cambridge University Press, pp.93-114).

Granger, S. & Meunier, F. 2003. Error tagging project – revised guidelines. (Unpublished document).

Hyland, F. 2003. Focusing on form: student engagement with teacher feedback. *System* 31: 217-230.

Hyland, F. 1998. The impact of teacher-written feedback on individual writers. *Journal of Second Language Writing* (7)3: 255-286.

Hyland, K. 1990. Providing productive feedback. *ELT Journal* 44(4): 279-285

Louw, H. 2006. *Standardising written feedback on L2 student writing*. Potchefstroom: North-West University. (MA dissertation).

Lyster, R. & Ranta, L. 1997. Corrective feedback and learner uptake. *SSLA* 20: 37-66.

Moletsane, J.R. 2002. Selective error correction in ESL narrative compositions. Potchefstroom: PU for CHE. (Mini-dissertation – Magister Artium).

Spencer, B. 1998. Responding to student writing: strategies for a distance-teaching context. Pretoria: University of South Africa. (D.Litt thesis).

Van Rooy, Bertus and & Schaefer, Lande, 2002. The effect of learner errors on POS tag errors during automatic POS tagging. *Southern African Linguistics and Applied Language Studies* 20: 325–35.

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.

5.12 Addendum A: Extract from tag set

Please note that the full tag set could not be included due to space constraints.

Superordinate	**Domain**	**Tag**	**Feedback**	**Example/Explanation**	**XML Label**
Grammar	Lexis	Repetition	You use the same words repeatedly. Find different words that may convey your message more clearly.	Use this tag when you realise that a student keeps on using the same word,. e.g. If a student used the word "Good" to mean "excellent" and "strong" and "hard" and "pretty" etc. This will be context sensitive. If you have to use the "Word: better word" tag a lot for the same word, rather start using the "Word: repetition" tag.	GLRE
Grammar	Lexis	Word choice obscuring meaning	This word is not clear enough. Find a better word to say what you want to say.	Use this when another word would make the intended meaning much clearer, e.g. "Only third-year students were **able/allowed** to go." All were able to go, but not all were allowed to go.	GLWC
Grammar	Lexis	Word form wrong	This word should have been in a different form for this context.	Use this for words in the wrong form, not covered by the other labels below.	GLWF

Discourse	Style	Active voice	The passive voice might be more appropriate here.	Use this where the student used the active voice, but in your opinion, the passive voice would have worked better.	DSAV
Grammar	Morphology	Wrong form - present participle	This the wrong form of the word. Use the "-ing" form of the word.	I am busy work in the garden.	GMWO
Grammar	Morphology	Wrong form - adjective	Wrong word form: use the correct form of the word.	Use this when the learner should have used the adjectival form of the word, e.g. "He gave me **rot** apples."	GMWF
Grammar	Lexis	Wrong time/ temporal adverbial	This time word does not fit the rest of the essay.	Use this when a student uses e.g. a word in the past-tense when the whole essay is written in the present tense.	GLWT
Grammar	Lexis	Wrong word	This is the wrong word. Find and use the correct word for the context.	Use this when a student should have used another word instead, e.g. "Students should be learned (taught) to…" or "injury" (damage) to property." Property cannot hurt.	GLWW

5.13 Addendum B: Questions to markers who used the marking system

Please answer the following questions regarding your experience using the marking system last year. You may answer in the document and just e-mail it back. Don't be shy to make positive or negative comments.

In the questions, I distinguish between:

A) System: the computer program.

B) Error tags: the error categories (the buttons you used and the list of categories you had.)

1. What is your definition of "error"?
2. Which error tag do you think you used the most?
3. Which errors were the most difficult to identify?
4. Were you pressed for time when marking these essays?
5. How much longer do you think it took you to mark an essay with the system than without it?
6. Can you read a text and ignore (or fail to notice) spelling errors or grammatical errors?
7. What was your overall impression of the quality of the students' writing?
8. What was your overall impression of the students' ability to present an effective and clear argument in their texts?
9. Do you read for spelling and grammar errors separately from reading for the argument in a text, or do you pay attention to both at the same time?
10. Were there any error tags in the system you did not understand or did not know how to use?
11. Do you think the tags available in the system raised your awareness of possible errors? If so, please give an example.
12. What were the most common errors students made in their writing?
13. How often would you say was it difficult to decide which tag to use? Give an example if you can.
14. How often did you use the system? Did you use it regularly or now and then for a big batch?
15. Did it get easier to use the system after using it for a while?
16. What are your recommendations to improve the error tag set?

17. What are your recommendations to improve the marking system?

18. Any other comments you wish to make?

5.14 Addendum C: Examples of student writing

1. Compulsory modules some of them are good but some of the they are full of nonsense because we gain nothing from them while others are very good because they prepare us for the working enviroment and also to become good professionals. Modules like sociology especially if you want to do community work is great and also koms because we learn how to communicate or improve our communication skills since were are preparing to be professionals. Modules like entr and wtll I don't really know why we should do them cause according to me is total waste of time and money. Any way we just have to do them because we don't have choice at all.

2. Poverty is short fall of consumption or income if somebody can not meet the basic needs he or she is regarded as a poor-man. It has Found that African countries are under developed so is where the poverty is highly located.As poverty is highly concentrated in rural area, or town outskirt and women and children and tenagers. Aids goes hand in hand with poverty because those women not working had a lot of children and they are straving. So women as parents has to find food a clothing and school for their children therefore the only altarnative is to practice prostitution or forced to be married by those who can help them. In this way teenagers would go out with elders especicaly businessmen to facilitates funds.

3. Our South African players are not paid well they are being underpaid they dont get the salaries that befit their job our players are putting the country in to high places they are proving it to the world that they can compete with other strong countries Our officials must start thinking properly the high salaries that are being paid to the officials who are doing nothing just sitting the whole day in their offices and attending alot of meetings making business contacts for them selfs the real heroes are being underpayed the reason for our players to leave the country to go and play in foreign countries is that they are paid well they get the money that they are playing for those the reason when our players are in foreign countries they play the sport with pride in them they get proper treatment our players must be paid correctly because the sport that they are playing is their career they have families to feed they are not only in the sport for the sake of money but because of the love for the sport they are proffessionals they master the sport that they are playing it is of no use representing your country but you are not paid according to the job that you are doing if you do a job correctly you expect to be rewarded accordingly so our players are our pride and they are putting the country into greater heights so if they can start getting decent salaries than there wont be a need for them leaving the country; The reason why they are leaving the country is because they are offered better opportunities they make alot of money in a short space of time and they get alot of expoture when they are playing in foreign countries; If they can be paid as the foreign countries are paying them than there wont be a need for them to leave the country given the same opportunities as those given to them by the foreign countries; So our officials must start thinking properly and try to improve the way our players are treated by giving them the correct salaries or else our country will endup with no players all the player will leave for European countries the europeans will take all the good players and at the end the country will be left with nothing it will be unable

to compete with their counterpaths because all the good player will be playing for European countries so let us start taking this thing into consoda ration and pay our player more so that they can stay at home and make us proud of them. After all "Home brewed is best".

5.15 Addendum D: Personal favourites

K PERSONAL FAVOURITES

	Subordinate	Domain	Tag	KK normed number of occurrences
1	Presentation	Spelling	Spelling/typing error	105
2	Grammar	Lexis	Wrong word	92
3	Discourse	Coherence	Reasoning inconclusive	73
4	Discourse	Style	Sentence vague	53
4	Grammar	Syntax	Superfluous general	53
6	Discourse	Cohesion	Sentence cohesion	46
7	Grammar	Morphology	Tense	42
8	Grammar	Syntax	Omission general	40
9	Presentation	Punctuation	Punctuation wrong	25
10	Discourse	Coherence	Relevance to topic	23
10	Presentation	Punctuation	Punctuation missing	23

M PERSONAL FAVOURITES

	Subordinate	Domain	Tag	MB normed number of occurrences
1	Presentation	Capitalisation	Capitalisation	130
2	Presentation	Spelling	Spelling/typing error	92
3	Grammar	Lexis	Article missing	62
4	Grammar	Morphology	Omission plural marker	61
5	Grammar	Morphology	Concord	46
6	Presentation	Punctuation	Punctuation missing	44
7	Grammar	Lexis	Wrong word	40
8	Grammar	Lexis	Determiner Incorrect	39
9	Grammar	Lexis	Word form wrong	36
10	Grammar	Syntax	Superfluous general	35

P PERSONAL FAVOURITES

	Subordinate	**Domain**	**Tag**	**P normed number of occurrences**
1	Presentation	Spelling	Spelling/typing error	150
2	Grammar	Lexis	Wrong word	87
3	Grammar	Lexis	Better word	81
4	Grammar	Morphology	Concord	79
5	Presentation	Layout	Layout error	57
6	Presentation	Punctuation	Punctuation missing	47
7	Grammar	Lexis	Word form wrong	40
8	Grammar	Lexis	Preposition wrong	35
9	Presentation	Capitalisation	Capitalisation	33
10	Presentation	Punctuation	Punctuation wrong	28

T PERSONAL FAVOURITES

	Subordinate	**Domain**	**Tag**	**T norm number of occurrences**
1	Presentation	Spelling	Spelling/typing error	62
2	Grammar	Syntax	Superfluous general	59
3	Presentation	Punctuation	Punctuation missing	51
4	Grammar	Lexis	Wrong word	47
5	Discourse	Factual correctness	Facts wrong	44
6	Discourse	Style	Register too informal	42
7	Grammar	Morphology	Tense	38
8	Grammar	Lexis	Better word	34
9	Discourse	Cohesion	Reference vague	33
10	Grammar	Syntax	Omission general	32

CHAPTER 6
ARTICLE 4 – YES/NO/MAYBE: A BOOLEAN ATTEMPT AT FEEDBACK

6.1 Prelude to Article 4

Articles 2 and 3 established qualities for effective feedback and illustrated that markers can use standardised feedback with a degree of efficiency. To relate this to the two main questions posed in the introduction to this thesis, it has been illustrated that semi-standardised feedback on computer can work in theory and practice. However, some shortcomings were identified which had to be addressed.

In the article "Standardised feedback in practice" (Article 3) a number of problems have been noted, among which was the finding that markers tended to focus on the surface structure elements. It is not immediately obvious why this is the case, but possible reasons could be:

1. The tag set was insufficient.
2. The markers forgot to pay attention to the issues of organisation.
3. The markers found the surface structure elements a distraction.
4. The markers found it easier to mark surface elements only and then felt justified in ignoring the organisational issues.
5. The markers did not know what to comment on when marking for organisational issues.

Results from Louw (2006) also suggested that the learners failed to improve on their paragraphs when asked to revise texts. In other words, the pedagogy in these instances had failed and needed addressing. The issues in need of addressing are therefore identified as:

1. Improve on the standardised feedback specifically with reference to organisational issues.
2. Assist lecturers by reminding them what to pay attention to when providing feedback on organisational issues.

A technique had to be found which adhered to the qualities of effective feedback, and which was above all fast and user friendly. Radio buttons (a checklist selection procedure often used on computers) was identified as a possible solution and the decision was made to test this with paragraph structure first, using a set of statements which could be answered with a simple "yes" or "no" (Boolean feedback). If positive results could be obtained, the experiment would be extended to other areas of concern as well.

Publication information for Article 4

Article 4 was accepted for publication in the *Journal for Language Teaching* with the following bibliographical details:

Louw, H and Van Rooy, Bertus. 2010. Yes/No/Maybe: A Boolean attempt at feedback. *Journal for Language Teaching.* 44(1) 107-125.

Minor editorial changes were necessary for adherence to the general format and layout of this thesis.

Note: The article has been published jointly by the author and Professor Bertus van Rooy. The latter was the statistician, but the theory, concepts and interpretation of data were the author's responsibility and any errors or misconceptions can only be attributed to the author. The statistical assistance is much appreciated.

Bibliography

Louw, H. 2006. Standardising written feedback on L2 student writing. North-West University (Potchefstroom Campus). (MA dissertation).

Abstract

This article 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; technique B: Consciousness-raising through generalised Boolean feedback, and technique 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.

6.2 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 is not 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".[6]

With large workloads and tight time constraints, it is understandable why hieroglyphics are used, 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 Section 6.3, followed by the research method in Section 6.4 and the results from the experiment in Section 6.4.4. The merits of the solution are considered in Section 6.4.5, before turning the attention to possible criticisms in Section 6.4.6. A few conclusions are offered in Section 6.7.

6.3 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 regard to feedback on student writing. Extending the work of Louw (2006), we have identified 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 individualised (James, 1997: 257), and students and lecturers find it difficult to effectively recognise recurring patterns of errors (Wible *et al.*, 2001: 308-310). In addition, there is the 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 & Cimasko, 2007: 126). Moletsane (2002: 27) also found that feedback often lacks a clear purpose and teacher expectations are unclear (also see 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),

[6] One anonymous reviewer objected to the use of the term "hieroglyphic" feedback, stating that many markers take great pains to write out clear feedback and ensure that their 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 crabbed handwriting.

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) also found that feedback does not lead to independent learners and Truscott (2007) claims error correction of grammar may even be detrimental. Finally, feedback is very time consuming (Moletsane, 2002: 21).

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

6.3.1 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 computerised 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 that version was still in its infancy, and as a result was very basic. The improved version of the Essaymarker software has been renamed MarkWrite.

The effectiveness of the partially standardised feedback in MarkWrite is discussed by Louw (2006, 2007, 2008). While the data indicated that 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*. Some possible reasons may be that the 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 problem related to the difficulty in adhering to the qualities of effective feedback mentioned above is the consistency of utilising 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 already have in a systematic, practical manner. This article therefore aims to demonstrate how already available knowledge can be used in a systematic way.

6.3.2 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, focusing on introductions, conclusions, paragraph structure[7] 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 writing an effective

[7] Article 5 in this thesis discusses the experiment in which Boolean feedback was used on introductions, conclusions and paragraph structure combined.

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 it 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 students are not able to write a focused paragraph, they are not able to write according to the plan they 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 students' ability to write clear, focused paragraphs in the examination will assist them to communicate better, and thereby to obtain better marks.

6.3.3 What are the qualities of effective 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 of a paragraph. Based on a number of sources (Du Toit, Heese and Orr, 2002; Emory, 1995; Hannay & Mackenzie, 2002; Henning, Gravett and Van Rensburg, 2002; McClelland & 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 of 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 six 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 in academic writing 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 being intelligible to the student and not regressing into hieroglyphics or "paint by numbers". To meet 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 6.1.

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	This paragraph links up with the paragraph above or below it.	YES	NO

Figure 6.1: Example of marking grid

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 positive aspects.

6.4 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.

6.4.1 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. The 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 had received any formal instruction in effective writing apart from what they had been taught at school.

6.4.2 Design of the experiment

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

Two additional instructions were added to highlight the focus on paragraph structure. Firstly, the students were told to underline the main idea in every paragraph so 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 type to emphasise 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 6.1) which 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 which was 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 listed above, the assignments were returned to the students in class. General feedback was also provided in class on the first draft (the two paragraphs), after which the students were asked to revise the original two paragraphs based on the feedback they received and resubmit them with their original two paragraphs as a single assignment. Figure 6.1 below is an example of a typical student text after revision.

ORIGINAL

Paragraph 1

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 roads crisises.
4. **Furthermore** vehicles need to be checked if they are roadworthy or not.

Paragraph 2

1. Some traffic officers take bribery from road rules offenders.
2. **As a result**; government can not reach its goal of maintaining safety in the roads.
3. **Surprisingly**; it could be licensed drivers who are reckless on the roads.
4. **Moreover** roads needs to be maintained in good condition.

REVISED

Paragraph 1

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.

Paragraph 2

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.

Figure 6.2: A typical student text

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 or 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 standardised feedback on cohesive devices is a topic for another study.

The general impression among the lecturers responsible for groups B and C, was that the 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 the students were still unsure about the specific use and implementation of the eight characteristics of an effective paragraph. To determine whether feedback of any sort had a significant effect on the quality of the revised versions, and specifically whether Boolean feedback led to more improvement in the revised versions than the alternative, the data from the experiment were subjected to statistical analysis.

6.4.3 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 6.1 above) to give feedback on a random selection of original and revised paragraphs from the classes. By comparing their feedback on the original paragraphs to the revised ones, it was possible to determine whether feedback in general had beneficial effects on the students' work. In addition, it was 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 randomised. The 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 four times by four different markers, 360 different responses were collected, and in each response, six different questions were answered.

The 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 6.1:

Table 6.1: 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.

One response by one marker on one essay pair (from feedback technique A) had to be discarded, leaving a total of 2 154 classifications that were subjected 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.

$H1_0$: The general null hypothesis is that feedback does not lead to improvement after revision.

$H1_A$: The alternative hypothesis is that feedback leads to improvement after revision.

$H2_0$: The specific null hypothesis is that Boolean feedback (feedback technique C) does not lead to more improvement than the other two techniques after revision.

$H2_A$: The alternative hypothesis is that Boolean feedback (feedback technique C) leads to more improvement than the other two techniques after revision.

The first hypothesis was 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 six) 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 were 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 was evaluated by computing the χ^2 statistic for the distribution of the four response types in Table 6.1. Assuming a 95% confidence level, with 6 degrees of freedom (3 marking techniques and 4 classifications, thus df = (3-1) x (4-1) = 6), the critical value of χ^2 for rejecting $H2_0$ is 12,59. If, in a 3 x 4 contingency table of all classifications for the feedback techniques, there is not only compelling evidence that the data are not distributed similarly across the three feedback techniques, but also that improvements for feedback technique C exceeds its expected value, while regression for feedback technique C is lower than its expected value, we will have found support for $H2_A$. [8]

6.4.4 Results

6.4.4.1 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 not for the

[8] All statistical information was 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 χ^2 test statistic. Computations of t-tests were done in *Statistica*, while the χ^2 statistic calculations were done by hand, using the procedures for computing expected values and the χ^2 itself set out by McClave and Sincich (2000: 721-726).

other two feedback techniques individually. The summary of the statistical analysis is presented in Table 6.2.

Table 6.2: 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 6.2. 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 a minimum of 0.

The overall improvement after feedback across the techniques is 0.25, which means that, on average, in one in every four paragraphs a student showed a net gain of 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 shows 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 to 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 to 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 the method that has the best possible yield, even if the difference is a small one. Furthermore, the averages presented in Table 6.2 hide an even more important aspect of the

feedback, viz. the fact that all three feedback techniques are relatively effective in helping students with revision, if measured as Improvements (as defined in Table 6.1). However, when Regressions are considered, it becomes clear that feedback technique C is much better than the other two (see Table 6.3 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.

6.4.4.2 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, was 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 6.3 represents the data in terms of the four-way classification presented in the methodology section (specifically Table 6.1).

Table 6.3: 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 who received feedback with technique C had written better originals in the first place, while those in the class who received feedback with technique A had written poorer originals. Using a χ^2 distribution, the uneven distribution is statistically significant (χ^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 of regressing back to a 0 upon revision or maintaining 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 χ^2 statistic on improvement versus no-improvement data is 2,4, which is below the 95% confidence limit of χ^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 6.3, 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.

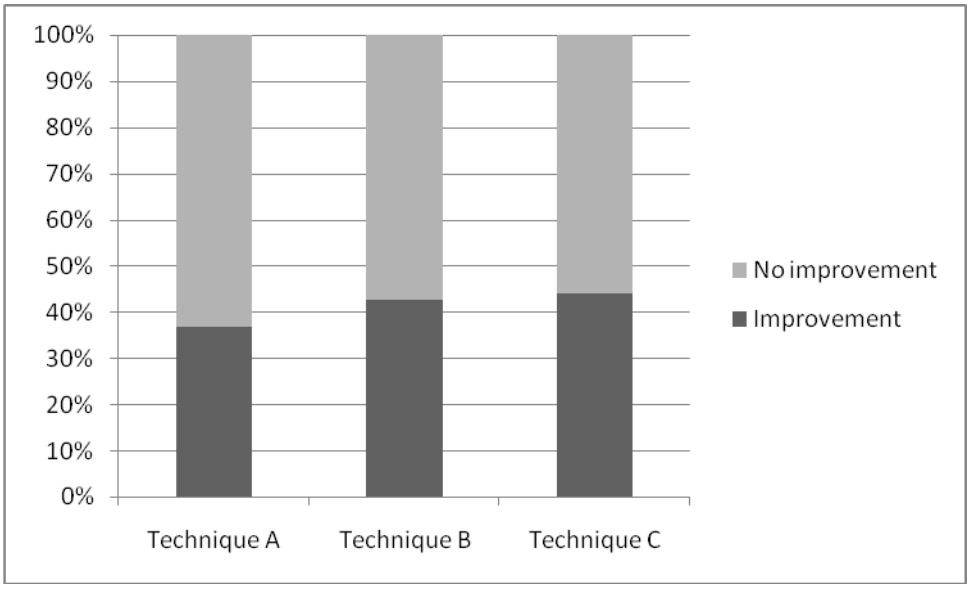

Figure 6.3: Percentage improvement per feedback technique

Given that the difference between the feedback techniques is not located in the potential to prompt improvement, we had to look elsewhere, and indeed found that the real difference between the techniques was in the extent to which regression was observed. Feedback technique C, which prompted 104 improvements from the original version to the revised version, also unfortunately led 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 (χ^2 = 27,6, df = 2, $p<0.05$). The comparison is visualised in Figure 6.3.

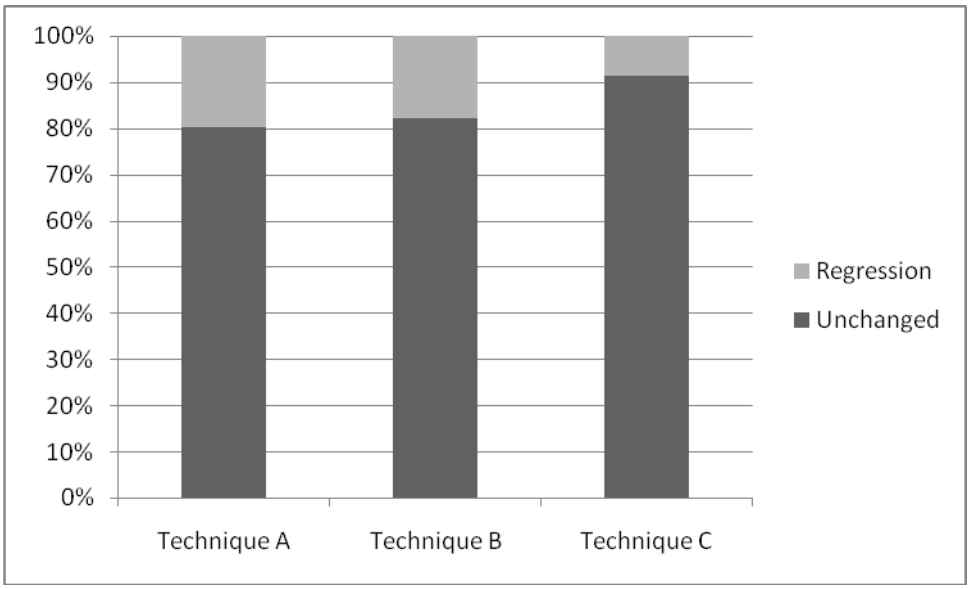

Figure 6.4: Percentage regression per feedback technique

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 37% and 43% respectively of the original NO scores on the feedback grids (Figure 6.4), 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.

6.4.5 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).

Table 6.4: 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
11. be individualised	No, although individualised comments may be added if necessary. It is individualised to the extent that the specific yes and no answers refer to the specific students' work.	0	Yes	1

12. be time effective and	Yes, but with caveats: the idea is not to comment on the structure of *all* paragraphs in the text, unless that was the focus of the specific assignment. The 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 similar amounts of feedback 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%

One may argue that some of the qualities of feedback as explained in Table 6.4 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, except for 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 of what to focus on, their awareness of the desired outcomes will increase. 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 & 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 (2007) 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. It is possible 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 standardised marking system and 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 is asked to answer the questions. The purpose of 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 randomly tested on paragraphs 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 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 again scrutinising the checklist for effective feedback. 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 standardised, written feedback goes, this is quite good although it will still not compare well with other, more labour and time-intensive techniques such as structured, personal interviews.

Since standardised feedback is more effective than hieroglyphics, the question is rather why it is 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.

6.4.6 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 such as the Criterion and E-Rater services of ETS are available on the market (Chodorow & Burstein, 2001; Chodorow &Burstein, 2004).

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 standardisation. Anybody can create a list of questions (even a well-researched list) and use the answers to 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 meet 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 regard to fully automatic feedback, there are many problems. The first problem is students' immediate distrust of fully automatic feedback (see Spencer & 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 computerised 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, which allows 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 only to be 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 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 ironically 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 receives only YES ticks on the feedback grid can still be improved.

Secondly, the statements need not necessarily be followed as a 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 free writing or brainstorming

sessions have been completed, but such editing guidelines are seldom targeted with accusations of prescriptivism.

In the third place, writing 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.

6.7 Conclusion

The experiment has shown that it is possible to improve on feedback on paragraph structure 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 greater than either hieroglyphics or generalised Boolean feedback on the entire assignment rather than individual paragraphs separately. However, an unforeseen risk of feedback is that, when students are prompted 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 level 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.

6.8 Acknowledgements

The authors would like to thank the three anonymous reviewers of the *Journal for Language Teaching* for their valuable input regarding the article.

6.9 Bibliography

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. 2001. Enriching automated essay scoring using discourse marking. Report available online.: http://www.eric.ed.gov/ERICDocs/data/ericdocs2sql/content_storage_01/0000019b/80/19/5d/82.pdf]. Accessed: 19 October 2009.

Chodorow, M. & Burstein, J. 2004. Beyond Essay Length: Evaluating e-rater®'s Performance on TOEFL® Essays. *TOEFL REPORT* 73. February.

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. pp. 119-140 In B. Kroll (ed.), *Exploring the dynamics of second language writing*. 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. Masters dissertation, North West University (Potchefstroom Campus).

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).

McClave, J.T. & Sincich, T. 2000. *Statistics* (8th ed.). 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. Potchefstroom: PU for CHE. (Mini-dissertation –Magister Artium).

Monyaki, B.S. 2001. A comparison of teachers' and learners' views on feedback strategies used in the writing process. Potchefstroom: PU for CHE. (Mini-dissertation – Magister Artium).

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. University of Pretoria. (D.Litt thesis).

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. Pretoria: University of South Africa (D.Litt thesis)

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, & researchers. *Computers & Education* 37: 297-315.

6.10 Postscript to Article 4

Due to space constraints in the journal, the full implementation of the Boolean feedback could not be explained in Article 4. For purposes of clarity, the explanation is included here. Note that this formalised implementation will improve the score of the Boolean feedback when checked against the characteristics for effective feedback.

6.10.1 Proposed implementation of radio buttons in MarkWrite: "Automatic discussion"

In MarkWrite, students do not simply receive a yes and no as their feedback. The computer is set up to rewrite the combination of yes/no answers into a coherent "discussion" of the paragraph. A student can therefore receive a discussion of the selected paragraph such as in the following example:

> Of the SEVEN[9] characteristics of a good paragraph, you have FOUR **right.**
>
> 1. √ Your paragraph has a sentence or part of a sentence which can function as the main idea for the whole paragraph. Well done.
> 2. × You treat more than one idea in this paragraph. A paragraph should have only one main idea. Rewrite your paragraph so that you only deal with one main idea in the paragraph.
> 3. √ The main idea of your paragraph is supported with evidence in the other sentences. Well done.
> 4. × Not all the information in this paragraph is relevant to the main idea of the paragraph. Find the irrelevant information and delete it, or move it to another paragraph.
> 5. × The sentences in your paragraph do not follow each other in a logical manner. Restructure your paragraph.
> 6. √ Your paragraph links up with the paragraph above or below it. As such, it fits in with the rest of your essay.
> 7. √ This paragraph is well placed in the text.

The student therefore knows exactly what has been done right (so he or she could hopefully repeat it) and exactly what has been done insufficiently, with a hint on how to correct it. This will presumably enhance the clarity of the feedback.

[9] The experiment mentioned in the article used only six characteristics, since the paragraphs were used in isolation and the seventh statement was not relevant. The "missing" statement in the article is, "Your paragraph links up with the paragraph above or below it."

CHAPTER 7
YES AGAIN: ANOTHER CASE FOR BOOLEAN FEEDBACK, OR "HOW TO MARK ESSAYS WITH STRATEGIC 'YES' AND 'NO'"

7.1 Prelude to Article 5

Having established that radio button-based Boolean feedback could be considered effective when providing feedback on paragraph structure, it was decided to extend this process to test the effectiveness thereof on introductions and conclusions as well. Before presenting the article itself, it might be appropriate to explain the reasoning behind the specific focus on paragraphs, introductions and conclusions.

7.2 Why the focus on paragraphs, introductions and conclusions

As mentioned in the prelude to Article 3, two of the identified problems were that markers tended to focus on the surface structure elements, and even if they provided feedback on higher-order elements such as paragraphing, students failed to revise those sections effectively (Louw, 2006; 2008). Therefore a way had to be found to provide effective feedback on issues of organisation. As also mentioned before, paragraphs form the base for short-answer questions in the examination, and the building blocks for larger texts. The interrelation between paragraphs, introductions and conclusions may also have a significant impact on the perceived coherence and quality of a text.

The functional theories of Givón (1989), Halliday and Hasan (1980) and Halliday and Matthiessen (2004) attempt to describe coherence and cohesion in texts as experienced by readers. When considering cohesion, it seems that most sources first decide what a text is and then which units they want to analyse – i.e. working at sentence level or paragraph level. In this thesis, it is unnecessary to establish what a text is, since the presupposition is that when a student hands in an assignment to be marked, they consider it to be a text. (Random, decontextualised grammar exercises will not be marked with the marking system, and therefore do not apply here.)

A marker confronted with a student text will automatically look for a pattern of coherence in the text and subconsciously start making predictions about the direction a text is heading, especially after reading the introduction. Since most markers will mark a text on a pre-set topic, they will also probably know the relative direction a text will take and the relative information they can expect to find. Linking the patterns of information in a text together effectively causes coherence. Linking these patterns together in an unorthodox or surprising way causes humour since the brain is forced to jump rapidly from the predicted path to an unexpected conclusion. If it does not cause humour, it causes irritation or confusion and triggers feedback like "vague". In undergraduate classes and writing classes, instructors have also called these consistent patterns in a text the "golden thread".

Based on an interpretation of the work by Halliday and Hasan (1980), Halliday and Matthiesen (2004) and Givón (1989), the golden thread may be briefly formalised as follows:

At sentence level cohesion exists which is not just confined to a specific sentence level, but also spans across the whole text.

At paragraph level we observe cohesion and coherence, and at text level we have cohesion and coherence with more emphasis on coherence. Inside all of this we have the argument structure as well. There are therefore overlapping categories as if there were three levels of cohesion and four levels of coherence; and they are interwoven like a DNA strand.

- Level 1 cohesion: sentence
- Level 2 cohesion: sentence to sentence
- Level 3 cohesion: the overlap/link between paragraphs
- Level 1 coherence: the links between sentences, words or ideas which transgress paragraph boundaries and may “skip” or “jump” over one paragraph to another.
- Level 2 coherence: the link between paragraphs which may skip or jump over one paragraph to another.
- Level 3 coherence: overall textual coherence within the boundaries of the text, encompassing the argument.
- Level 4 coherence: situating the text within the context: this study will not investigate that, since it falls outside its scope.

In Figure 7.1 below these relations are illustrated. Theoretically it should be possible to superimpose such a sketch on a coherent and cohesive text. In other words, if all seven levels are hand-drawn on a well-written text, a picture similar to Figure 7.1 should emerge.

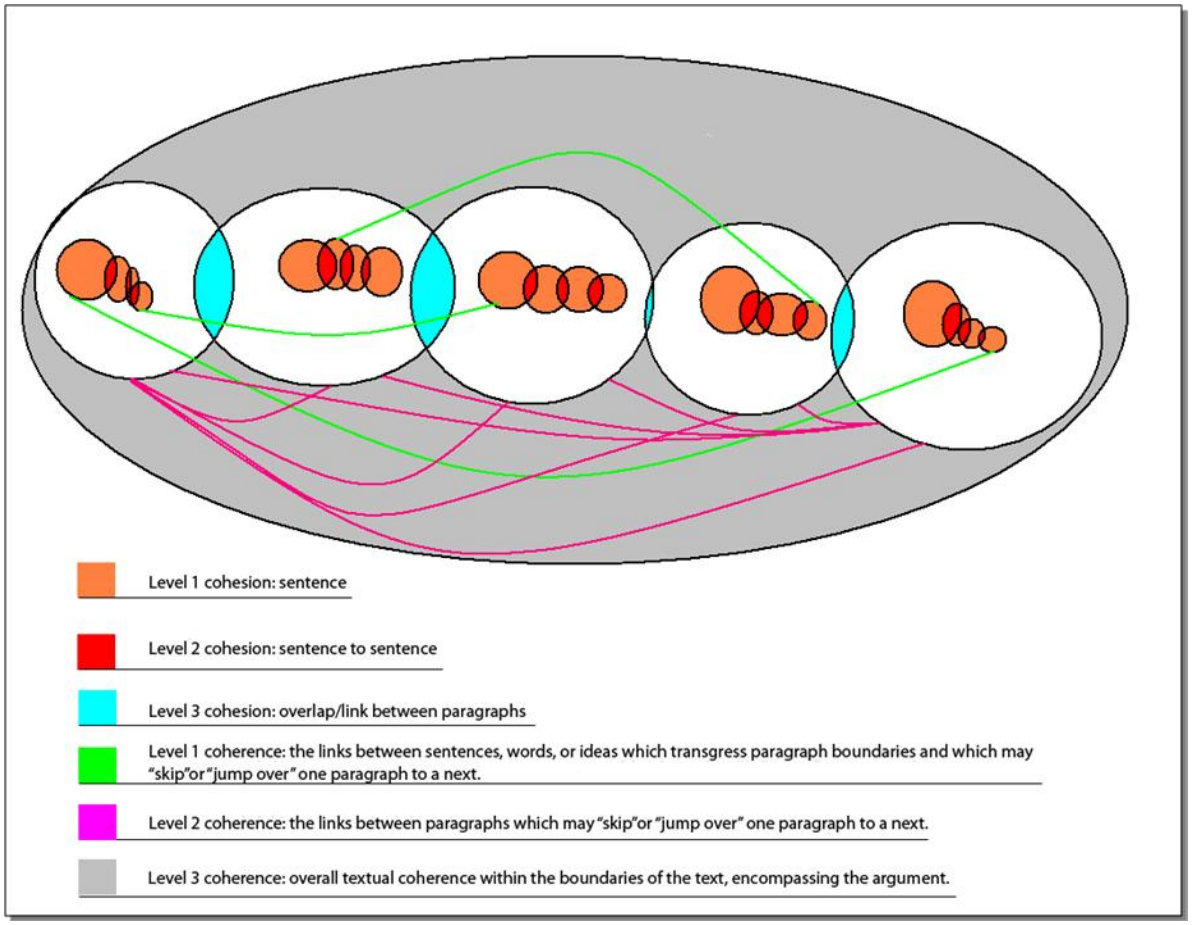

Figure 7.1: Graphic illustration of cohesion and coherence in a written text

Level 1 cohesion is at the sentence level and is where most of Halliday and Hasan's (1980) analyses lie. In extreme cases, such as a complete lack of sentence control, a complete lack of understanding, very vague sentences, and ambiguous sentences, level 1 cohesion may be deficient. Beyond that, however, level 1 cohesion is present in the writing of students with even elementary language skills.

Level 2 cohesion is only possible if the paragraph is actually a paragraph adhering to the standards that qualify it as a paragraph, i.e. dealing with one main idea. All the sentences in a paragraph should work together to build a whole.

Level 3 cohesion is where much teaching is focused, as the link between paragraphs often depends on linking devices in the first and last sentences of the adjoining paragraphs.

Level 1 coherence is obtained by referring to previously mentioned facts or using definitions or words explained previously in the text. It is the classic case of negotiation of meaning – agreeing to use a specific sequence of letters to mean a specific thing. In other words, if the author explains a self-made abbreviation in an early paragraph, he "negotiates" with the reader that from there on, he may use the abbreviation instead of the full explanation. In this specific text for example, the term "hieroglyphics" was negotiated to mean unintelligible feedback in the form of symbols.

Level 2 coherence refers to the effective structuring of a piece of writing. If the introduction is structured effectively each of the subsequent paragraphs will link up with the introduction. The conclusion (if structured as an effective conclusion) should also link up with each of the previous paragraphs (as indicated in the diagram), also in a show of level 2 coherence. It is these overlapping strands which weave a text together (the golden thread) and the diagram provides a visual picture of this.

Note that the sketch does not make provision for the quality of the argument or the quality of the information in the text. This is simply because it is possible to write a perfectly structured (coherent and cohesive) text using absolute gibberish as information and arguments.[10]

This discussion and sketch illustrates how a good start may influence the direction and quality of a text. As will be seen in Article 5, the quality of the introduction has a direct influence on the quality of the rest of the text. A second reason is that the theories mentioned above (although useful and plausible to academics) are hard to explain or illustrate to students, but a simple "connect the dots" exercise with a text and an overhead transparency in class, can illustrate this to students in a graphic, practical manner.

Publication information for Article 5

Article 5 will shortly be submitted in revised form to the *Journal of Second Language Writing*. An earlier version of the chapter was presented as a paper at the SAALA/LSSA/SAALT joint conference in 2010. The feedback from the audience is greatly appreciated.

Bibliography

Givón, T. 1989. Mind, code and context – essays in pragmatics. London: Lawrence Erlbaum Associates, Publishers.

Halliday, M.A.K. & Hasan, R. 1980. Cohesion in English. London: Longman.

Halliday, M.A.K. & Matthiessen, C.M.I.M. 2004. An Introduction to Functional Grammar. 3rd ed. London: Arnold.

Louw, H. 2006. Standardising written feedback on L2 student writing. Masters dissertation, North-West University (Potchefstroom Campus).

Louw, H. 2008. The effectiveness of standardised feedback when L2 students revise writing. *Language Matters* 39(1): 88-109.

[10] Examples of these include most political speeches, propaganda, and other forms of comic relief.

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 used 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.

7.3 Introduction

Written texts are incredibly complex, and as a result 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 the differences between first language writing and second language writing, the relationship between writing and SLA, the relationship between writing and reading comprehension, and numerous others. 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, any readers 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).

7.4 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 the problems of providing 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 (a) more thorough feedback (b) faster, (c) without burdening the marker unnecessarily, and (d) to provide 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?

7.4.1 Marking scheme as a checklist?

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 established by Louw (2009). Also, feedback on a text is not always directly related to the specific marking scheme.

7.5 Why the focus on introductions and conclusions?

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. The experiments are increasing in levels of difficulty, with the next step focusing on argument structure in argumentative texts.

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.

7.6 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 Heasly (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.

7.7 The experiment

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

7.7.1 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.

7.7.2 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.

7.7.3 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. Facebook[11]

[11] Two examiners 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.

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 were 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 7.1.

Table 7.1: 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
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 the 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.[12]

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.

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 7.2. Note that due to space constraints, a full table has not been included.

Table 7.2: 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.

[12] 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.

7.7.4 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, we 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 we therefore conducted further analysis of the data 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 7.3 below.

Table 7.3: 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, we explored the relationship between feedback on argument structure and the mark by a conducting 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.

7.8 Results

7.8.1 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 all the data used in this experiment, we nevertheless expect a small but statistically significant improvement in the marks in order to reject the null hypothesis. The average mark of the originals and the added improvement are represented in Figure 7.2:

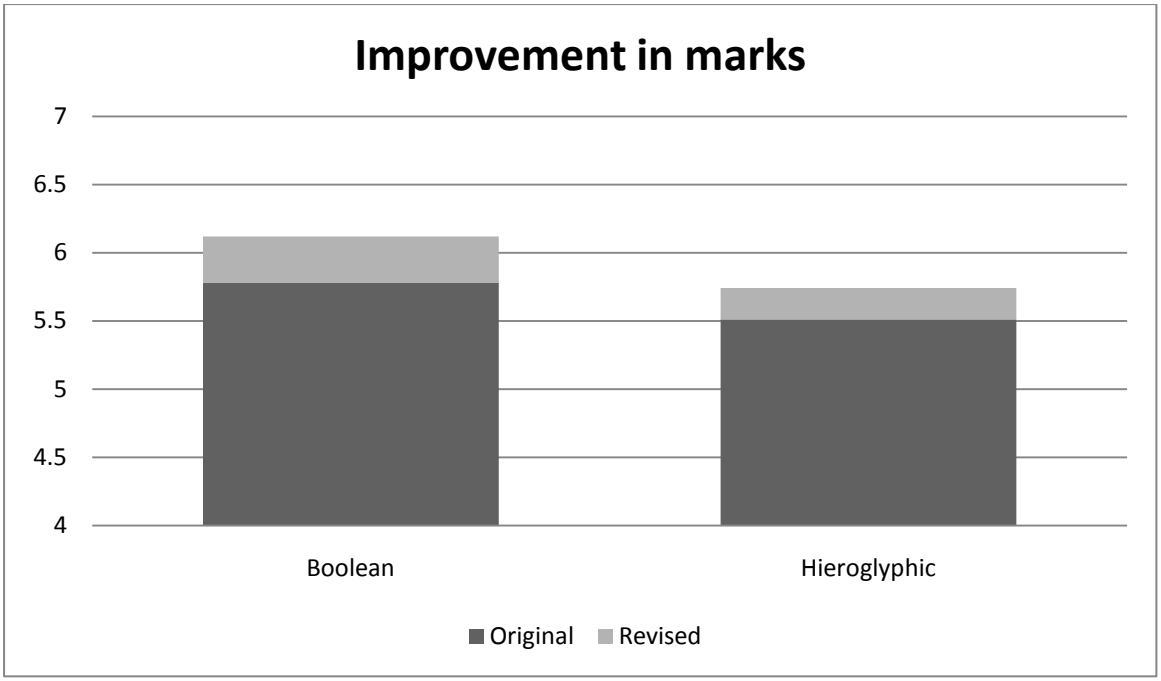

Figure 7.2: Original average marks out of ten for two groups of assignments, with average improvement after revision, adding up to an average mark for revised versions

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, if we separate the essays that received Boolean feedback 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 the MarkWrite interface, this is enough reason to advocate the use of the technique.

7.9 Contribution of feedback checklist

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 were 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 7.4. 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 7.4 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 7.4: 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 7.5, 7.6 and 7.7. 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 7.5: 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 7.6: 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 7.7: 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 7.5–7.7 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 7.8.

Table 7.8: 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. We are not sure why this is the case, but tentatively advance two reasons: 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. The intention is that markers should use it to comment on one or two randomly selected paragraphs in a text, using the computerised marking interface, MarkWrite.

7.10 Relationship between marks and sections from feedback 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. We undertook statistical modelling with multiple linear regression to determine how good 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. We 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 almost automatically.

7.11 Why does it work?

Although it is not a complete revolution in the struggle to improve student writing though 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 definition was sought for "checklist". After consulting numerous dictionaries (both online and offline) the most thorough definition found was the one in Wikipedia:[13] "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

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

checklist would be a schedule, which lays out tasks to be done according to time of day or other factors."
en.wikipedia.org/wiki/Checklist

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%B4s_new-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. 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 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 the computer interface MarkWrite.

In essence then, the individual categories of the 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.

7.12 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 that 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 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.

7.13 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.

The radio buttons are already implemented in the MarkWrite electronic feedback system as one of the features which cannot be altered by the user. The intention is that any lecturer in any subject will be able to comment on these specific text qualities, but can also add their own comments using the "comment" function. In MarkWrite, students do not simply receive a yes or no statement. The yes or no status of the statement is translated into specific feedback as illustrated in Table 7.9. In this way correct actions by students are reinforced – "you have done *this* right, so be sure to do it again", or incorrect actions are specified very clearly with instructions on what to change or avoid.

Table 7.9: Interpretation of radio buttons in MarkWrite

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	**Your introduction gives a clear background about the topic to your reader.**

statement 2	
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.

For example, if a conclusion does not link up well with an introduction, the student will receive this instruction: "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 the conclusion." The student is informed exactly what is incorrect and instructed what to do about the situation.

7.14 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 the computerised marking system, MarkWrite, 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 such as the MarkWrite interface and using radio buttons in MarkWrite, 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.

7.15 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. Mini-dissertation, PU for CHE.

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 ed. London: Arnold.

Hamp-Lyons, L. & Heasly, 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. 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).

Louw, H. & Van Rooy, Bertus. 2010. Yes/No/Maybe: A Boolean attempt at feedback. (Accepted by *Journal for Language Teaching*, forthcoming).

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.

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

CHAPTER 8
CONCLUSION

When considering the findings of this study, one has to keep in mind that the research conducted in this study is part of a much larger project. The MarkWrite project opens an immense number of doors for further development and for utilising the research potential of the everyday, tedious, but important activity of marking student texts.

This specific study came from a number of different approaches aiming at one goal – to provide more effective feedback in as efficient and practical a way as possible. These approaches include Computer-Assisted Language Learning (CALL), English Second Language teaching, language acquisition, writing pedagogy, corpus linguistics, error analysis, applied linguistics, and of course, real-life practicality.

The inherent aim of this research was not to establish a brand new theory of feedback or effective teaching, or even a new method of teaching composition. The study grew from an honest attempt to ease the burden of marking which is yoked on all lecturers and teachers, while improving the impact marking can have on student performance.

While there are many thousands of books and articles on "how to write", the fact of the matter is that the teaching of anything has so many variables and so many contrasting ideas of "do's and don'ts" that it is absolutely impossible to implement all the available knowledge. For example, this study has not even touched upon theories and research on human motivation, behavioural change, or human memory.

In addition, future generations will probably wince at the thought of the number of man-hours spent marking student texts if one considers that all that data usually go to waste.

8.1 The findings of this study

This study should be seen as part of a larger, ongoing study. It started in 2004, with the commencement of an MA thesis on the standardisation of feedback on L2 writing (Louw, 2006). The results of the MA indicated that in experimental situations, certain feedback categories can be standardised effectively. However, the standardised feedback was not effective in all areas, and the practical application of this standardised feedback still had to be established. This opened the door for further research on the topic.

The use of computer software was considered the most effective way to deliver standardised feedback on student writing. This immediately situated the study in the field of Computer-Assisted Language Learning (CALL), which necessitated an investigation into what could be considered effective CALL. However, CALL is an immensely broad field, straddling pedagogy, programming, language acquisition and many other fields, so a specific focus was sought. The focus fell on the design of CALL since the results of this thesis will be incorporated into a brand new software package. A literature review, combined with practical experience, was used to establish a rubric for

the evaluation of CALL packages. Then, adopting such an “end-user perspective”, these same criteria were used as design considerations for the planning of the system.

The research into CALL evaluation yielded a few findings of relevance to the rest of the study. In the first place, there is a distinction between CALL tools and CALL coaches. Feedback software would fall primarily under the tool category, but based on the pedagogical nature of feedback, some coaching characteristics would be present as well. Future possibilities for the software would yield even more coaching situations. The marking software therefore has to adhere to qualities for both tools and coaches.

The second important aspect of CALL is the caution that excellent software will not compensate for poor pedagogy. The content and techniques used in a CALL package should therefore be tried and tested and well researched.

Article 2 then established what is considered effective feedback, effectively establishing best practice for feedback in pedagogy. This was done by critically analysing international research to establish what works and what does not work. Based on these findings, a checklist was drawn up which could be used as a scorecard to estimate the relative usefulness of a specific feedback technique. The 13 qualities identified for feedback to be effective were not ranked hierarchically, and although some appear self-evident, it would be extremely difficult to adhere to all 13 qualities. The abilities of computers will definitely be necessary to enable a marker to observe the qualities more efficiently.

Next, an experiment was conducted with the first rudimentary version of MarkWrite, then still called Essaymarker. Louw (2006) illustrated that standardised feedback does assist learners in revision, but the qualities of effective feedback require that it should assist the marker as well and not waste his or her time. The purpose of the experiment was therefore largely to test whether actual users could use an electronic marking system consistently, and to establish what they focused on when marking. The experiment pointed out, amongst other things, that markers need to be assisted in moving away from the editing mentality of focusing on surface level errors. In addition, the earlier study (Louw, 2006) also pointed out that students had less trouble revising surface level errors than revising paragraphs, introductions and conclusions. This prompted the next stage of investigation.

A way had to be found to provide feedback quickly, effectively and accurately on issues of structure in student texts. An attempt was then made to provide more helpful feedback on paragraph structures in a quick, thorough manner by means of checklists. Firstly, the qualities of effective academic paragraphs were identified and turned into a series of carefully worded statements. These statements were then used as a checklist marking scheme upon which a marker could simply answer yes or no. It was hypothesised that students would be able to use these statements and the answers to them to effectively revise their paragraphs. The technique proved to be effective to a statistically significant degree, especially in enabling students to avoid regression. In other words, while the technique did allow some students to improve on their writing, it also prevented students from editing their texts into a poorer state. The experiment proved enough of a success to warrant the continued exploration of this means of providing feedback.

A subsequent experiment added introductions and conclusions, after having first established from the literature what the qualities of effective introductions and conclusions are. Once again the results were positive enough to justify using the technique.

The feedback checklist (while not a new concept) assists lecturers by reminding them of all the features they should take into account when providing feedback on student texts. It also assists the learners in that they effectively receive complete feedback in full sentences. The technique can of course be applied by hand, but numerous further research possibilities are possible with this technique, including assisting the lecturer in assessment instead of just feedback. It would also be possible (with a larger data set) to establish exactly which features of a text have the greatest influence on the readers' perceived quality evaluation of a text. Although Article 5 touched upon this issue, the applicable data set was too small for accurate investigations.

The regression analysis in Article 5 also illustrated that there is a high correlation between the combination of Boolean feedback and the mark provided by the markers. In other words, if a writer's text contains the elements checked for with the Boolean checklists, the writer stands a better chance of obtaining a good mark. This indicates that the elements checked for with the Boolean feedback are indeed important in the structural organisation of texts.

The electronic nature of the radio button feedback (Boolean checklists on computer) ensures that future research is indeed possible. There are a number of future possibilities, some of which are elaborated on below.

In the meantime, while the experiments were under way, the Centre for Text Technology (CTexT®) commenced the programming of a newer, easier-to-use version of MarkWrite. While MarkWrite is at present just a computer tool with very few CALL functionalities, the ultimate goal is to have MarkWrite as a combination of two or three different software functionalities with advanced CALL functions. Research on the MarkWrite project is still far from complete, since one of the goals of the creation of such a system is continuous research. The research project has evolved since 2004 in an incremental kind of way, with research questions identified based on previous successes and failures. The data generated by the long-term use of MarkWrite will also be used to improve grammar checkers, spelling checkers and other automatic applications in MarkWrite. As new technologies emerge and new problems become evident, research can continue on the system itself and on the data generated from its use.

While the study established what effective feedback entails, in reality it is not applied regularly and consistently due to practical constraints. It is hoped that MarkWrite and the techniques incorporated into the system will be able to alleviate at least some of the practical problems.

The findings discussed above confirm that all five of the research questions mentioned in the introduction have been answered and the aims reached. The qualities of effective feedback have been established and used for evaluative and design purposes. Shortcomings have been identified in the marking practice of lecturers and the technique of Boolean feedback has been tested to correct this problem, with the technique proving effective. In addition, a rubric was created which can be used to evaluate CALL systems. The rubric can also be used effectively when designing new systems.

8.2 The contribution made by the study

As stated earlier, the MarkWrite project does not have the inherent aim of establishing a new theory of writing, or even a new theory of writing pedagogy. Instead, it is a real-world attempt to integrate the immense body of (mostly theoretical) knowledge available in writing pedagogy into a practical tool useful in the real world. The contribution should therefore be seen as a synthesis of existing knowledge to shift the boundaries of the implementation of knowledge. The result is a new door that has been opened to research possibilities. The study also illustrates that the conflict about whether feedback is useful or not can be resolved by utilising insights from both sides of the argument and operationalizing them. In addition, the study highlights the possibilities provided by computer technology to put the "data annotation" which is marking, to much better use, thereby greatly increasing the value of the work done. In other words, a new level of efficiency of an old activity can be obtained.

In essence, the thesis emphasises that the principle of feedback is sound, and by applying some innovation, the practice of feedback can be more effective. It is a vast untapped research resource.

8.2.1 The contribution of the MarkWrite Project in national and international context

The MarkWrite project should not be seen as an isolated South African attempt to solve isolated South African problems. Based on the number of international conferences and journals dedicated to the topic, the effective teaching of writing to students (either first or second language) is evidently a topic worthy of research. In the international context, research is continually conducted on writing pedagogy, feedback, corpus linguistics and the like, which is the reason why so many international sources are quoted in this thesis. MarkWrite will provide a platform for integrating more of that research with practice.

However, MarkWrite should also be seen within the context of the South African linguistic landscape. Researchers such as Weideman and Van Dyk (2004) do research on the improvement of academic literacy, a large component of which consists of appropriate writing skills, textual organisation skills and vocabulary – all aspects which could find applicability in future developments of MarkWrite. Also, Butler's (2006) research on highlighted the need for writing pedagogy as part of university level academic literacy within the South African context. Likewise, Hattingh (2009) argued for the need for an updated, easy to use assessment scale for the marking of student essays in matric examinations – also in essence research on writing pedagogy. The integration of Hattingh's research has already been discussed with CText® for future versions of MarkWrite. It should be clear from these examples that there is a national urgency in developing solutions to the perceived problem of student writing proficiency in South Africa and MarkWrite is one step toward that solution. The future developments mentioned in section 8.4 should be seen as possible solutions to problems identified in both national and international contexts.

8.3 Current state of MarkWrite

At the time of writing, MarkWrite Beta II has already been field tested by CTexT® at the North-West University, Potchefstroom Campus. CTexT® is busy with final revisions to the system based on user reviews. The reviews were positive overall, with a few practical suggestions made by lecturers in the field. CTexT® is also busy with negotiations with the local Sakai representatives to convert the free-

standing version of MarkWrite into a web-based version for the Sakai interface. The free-standing version will be available for purchase and download from the CTexT® website in early 2011, while the availability of the Sakai version will depend on external time schedules and cannot be determined at the time of writing.

The accompanying CD contains video illustrations of how MarkWrite Beta II functions, as well as a trial version of the software.

8.4 Future developments and further research

Designers of software applications and technology devices have coined the phrase “feature creep”. It refers to the human tendency to over-innovate something with more, newer and better features creeping into the development to the extent that there are so many options that the user becomes lost in a myriad of options. The experience of total bafflement when faced with a new cellphone or a new software application is due to unchecked feature creep. While feature creep is a legitimate concern, systematic and purposeful development will ensure that it is possible to expand on the MarkWrite capabilities with sensible and useful future developments. There are many different possibilities all in need of research, and these will be briefly discussed to provide the reader with a sense of the magnificent scope possible for future MarkWrite versions.

8.4.1 Future developments: MarkWrite Marker and MarkWrite Student

The two biggest and most important of these developments is the distinction between MarkWrite Marker (lecturer side of the interface) and MarkWrite Student. As mentioned in the introduction, the aim is to have a student side of the system on a web-based learning environment and even as a free-standing application, which will take the student step by step through the whole writing process. The text which then arrives on the desktop of the marker will, it is hoped, be a much better version than what he or she would have received in the absence of MarkWrite Student. The lecturer would therefore have a less complicated marking job as most of the surface level problems and even some of the structural deficiencies in the students’ texts would have been dealt with. The lecturer would then be required to focus more on the higher-order qualities of the texts, instead of fixating on the surface level problems.

8.4.2 Future development: innovation and technologies

A number of future developments are possible in the MarkWrite system. These include, but are not limited to:

1) Global development

2) Automatic error identification

    a) Batch scanning

    b) Nosey thesaurus and phrase analyser

    c) Style analyses

    d) Text comparison and plagiarism detection

3) Additional radio buttons: Boolean argument analysis

4) Student prompts in MarkWrite Student

5) Customised remedial exercises

6) No exercise, no mark

7) Selective marking

8) Improved feedback categories and shared feedback sets

9) Assessment assistance

10) Voice recognition

11) Audio feedback

12) Peer review

13) Order of development feedback

14) Free up lecturer's time

15) Type/token ratio

16) Reading ease score

17) Level of importance of various feedback categories

18) Research on user friendliness

19) Pre-checks for students

20) Type/token ratio feedback

21) Style analyser

22) Teacher check-ups

23) Mobi site and cellphone usage

24) Effects of reading comprehension on feedback interpretation

25) Screen capture

Each of the above will be explained very briefly.

8.4.2.1 Global development

The size and scope of MarkWrite is currently very limited, but with the continual growth of web-based learning platforms such as Blackboard or the Sakai community (www.sakaiproject.org), global acceptance is not an unrealistic goal. Sakai-based learning platforms are used by the North-West

University, UCT and Unisa in South Africa, as well as 235 other universities worldwide, including Cambridge, Yale, Columbia, Cornell, Lancaster and Stockholm universities.

Negotiations have already commenced to integrate MarkWrite into the Sakai interface for local implementation and it is hoped that with an international foothold, research on the project can commence at a faster pace.

8.4.2.2 Automatic error identification

The more surface errors the computer can identify, the more time the human marker has to concentrate on the textual structure and argument. The Beta version of the software already contains a freeware spelling checker. Funding is currently the hurdle to upgrading to a more advanced, more accurate spelling checker in numerous different languages. Spelling and grammar are not, however, the most important errors to focus on and the computer could also be instrumental in identifying and eradicating other common student errors by means of automatic assessments. These include a nosey thesaurus, a style checker, argument analysis, text comparison and plagiarism detection as mentioned above. These features would be implemented in the MarkWrite Student interface to enable students to identify and eliminate common errors before they occur – it is like a pre-emptive strike on commonly occurring errors.

8.4.2.3 Batch scanning

One of the features already in MarkWrite is that it can identify recurring spelling errors. If a word is identified as a spelling error early in the text, MarkWrite asks the marker if the other instances of that error should be tagged as errors as well. It is hoped that this feature will not be necessary once MarkWrite Student is operational. However, it would be possible to identify other "strings" of errors (exact error matches) throughout a whole batch of student assignments. This will eliminate the tedious and frustrating task of identifying an error that recurs in many different essays.

8.4.2.4 Nosey thesaurus and phrase analyser

The concept of a "nosey" thesaurus is quite simple. A thesaurus-like feature should "nose around" in the student text to find commonly misused words and phrases, and even overused words and phrases, and then suggest alternatives to the student. To establish this software will no doubt require intensive research. There are a number of different techniques to accomplish such a feat. For example, a comparative score of words and phrases used in the native language and the learner language would allow the computer to find errors of overuse and provide the student with alternatives even without the student asking for them. For example, the author's students frequently start a sentence with the word *also*. Ideally, the computer should identify "also" at the beginning of a sentence and then automatically suggest alternative linking words to the student, like "in addition".

8.4.2.5 Style analyses

Certain features of academic writing style are relatively easy to identify with the computer. It would be possible for the MarkWrite Student side of the system to use a style checker to identify common student problems and to teach through the feedback. For example, experience has shown that first-year students have a tendency to be overly informal in their writing.

8.4.2.6 Text comparison and plagiarism detection

It is a given that plagiarism detection needs to be a feature of both MarkWrite Student and MarkWrite Marker. What is possible in addition to this is to find instances of "in-class" plagiarism. It often occurs in the author's marking situation that students copy from each other to the same extent that they copy from sources. For instance, if a student writes a plagiarism-free text and then shares it with a class mate, the software should be able to pick that up within the same batch of texts.

8.4.2.7 Additional radio buttons: Boolean argument analysis

At the time of writing, an experiment is already under way to establish the possibility of providing Boolean feedback on argument structures as well. If successful, the feature will be programmed into MarkWrite.

8.4.2.8 Student prompts in MarkWrite Student

MarkWrite Student will simply not allow a student to submit his or her essay without first editing it. One way of going about this is simply for the computer to ask "Have you checked your …?". An alternative would be for the computer to run the student through the whole planning process of writing a text. Many different variants of this planning process are explained in many books on student writing. It would therefore even be possible to include different techniques of the writing process into the system to cater for individual preferences between students. This could include a mind map tool, a scrap bin feature for saving unused pieces of text or ideas, a time estimate tool, assistance with referencing, and a multitude of other features. Substantial research would be required to establish which are more effective, popular and user friendly.

8.4.2.9 Customised remedial exercises

If nothing is required of the students after receiving their feedback, the effectiveness of the teaching situation is probably halved. Students should be required to engage actively with the feedback – they should be required to do something so that they take the feedback seriously and learn from it. They should receive remedial exercises on their specific problems which will ensure that they receive exercises only in the areas where they have problems. This database of exercises should obviously not contain only surface level exercises such as "place the comma in the correct place". There have to be exercises that test and teach text linguistic features as well. Setting up these exercises is a gargantuan task, and establishing which types of exercises work best for certain features would require intensive research. Despite these obstacles, customised remedial exercises are probably the most important and the most exciting future development possibility of the whole MarkWrite project. More daunting for second language acquisition is to turn the information obtained from MarkWrite not into exercises as such, but into focused input to speed up the acquisition process of a second language.

8.4.2.10 No exercise, no mark

One of the biggest problems with providing feedback is that students simply do not care. Hundreds of marked papers are never retrieved so students are not even aware that they have a problem. A way to overcome this is to program the system so that students' marks are only accepted into the grade book *after* they have done their remedial exercises. By returning the papers electronically, the

computer system should also be able to monitor whether or not students did in fact check the feedback.

8.4.2.11 Selective marking

To guard against overkill and feature creep, the system should be able to provide opportunities for selective marking. This will entail using separate parts of the feedback tag set one at a time and allowing the student to see it like that as well. In Article 5, it is contemplated that the students did not use all the feedback because there was too much of it. This will have to be investigated, and a selective marking tool may be necessary. The optimal combination and amount of feedback can be investigated as well.

An alternative method is to use a marking team, where one marking assistant focuses on surface level errors, then passes the essay on to the next one who analyses the argument/facts and passes it on to the third who will comment on coherence and cohesion. It is possible that this production-line technique of marking may also speed up the process.

8.4.2.12 Improved feedback categories and shared feedback sets

In order to provide selective marking, improved feedback categories are necessary. These improvements will occur all the time as tag sets are adapted for the specific needs of the lecturers. CTexT® has already mentioned the possibility of providing a series of tag sets for various subject areas as a starting point for lecturers. This will allow for a true "writing across the curriculum" application. With the initial testing of MarkWrite Beta on a small scale at the NWU, Potchefstroom, lecturers already created their own feedback tag sets. Once MarkWrite is incorporated into a standardised platform, such as the Sakai interface mentioned above, it will be possible to create a "share platform" where different users worldwide can upload and share their tag sets for various different subject areas.

8.4.2.13 Assessment assistance: Radio buttons and marker assistance

Feedback is not the only thing that needs to be standardised. There is a need for standardising assessment, especially for something with higher stakes (e.g. matric examination). For marking English papers, it is possible to incorporate the well-researched assessment grid created by Hattingh (2009) into MarkWrite.

In addition to this assessment grid, the radio button patterns can be analysed and correlated with the marks assigned, as explained in Chapter 5. With enough data, it may even be possible for MarkWrite to suggest a possible mark to the lecturer for features such as "paragraphing" or "argument structure" based on the combination of yes/no answers.

This has a few advantages:

- Standardisation of assessment
- Ease of assessment (marking assistants should be able to do this without supervision)
- No more calculations to do – the computer calculates these marks.

Markers have to concentrate on so many variables when marking that it gets very difficult to give a dependable mark every time. It should be possible to have the computer highlight certain text elements for the marker to assist in marking. For example, when the radio button task pane asks questions about cohesion, the computer could highlight all the possible linking devices. If the question is about the relatedness of the introduction and conclusion, the marker should have the ability to hide the rest of the text for a few moments (this is already possible with the "split screen" view).

8.4.2.14 Voice recognition

It takes time to highlight and click. It should be faster to highlight and speak. Limited voice recognition abilities specifically aimed at the feedback categories should not be too difficult to incorporate, especially if programmed for the specific tag set with which one is working.

8.4.2.15 Audio feedback

Comments at the end of the essay are currently typed. These might just as well be on audio files. Audio feedback is quicker than typing so more information can be conveyed to the students in the same amount of time. An experiment is under way at the time of writing using a feature entitle "Backchat" to test the effectiveness of audio feedback. It is even possible to incorporate audio feedback at various specific places in the text, as a small "audio" feedback tag. New technology allows audio to be compressed to very small sizes so it should not be too much of a drain on IT resources.

8.4.2.16 Peer review

With a local network and a computer laboratory, peer review sessions should be a possibility. Since the system has assessment assistance, this will count as **guided peer review** so both the reviewer and the reviewed should learn more from the exercise. The reviewer is reminded continually what to look for in the text, while the reviewed will get somebody else's opinion on how effective his or her text is.

8.4.2.17 Order of development feedback

Research has shown that learners acquire a language in stages. It should theoretically be possible to tailor feedback to the specific stage the learner is in, in order to get him or her to move to the next one more rapidly. This option will take a few years to develop and will need intensive research. The possibility even exists for combining this option with a test of academic language ability or some other standardised test such as the TALL (Weideman & Van Dyk, 2004).

8.4.2.18 Free up lecturers' time

One of the main purposes of MarkWrite is to increase the efficiency of marking. This is a three-fold goal – to speed up the marking, make feedback more effective and to use the generated information more effectively. Of course, if feedback is more effective and the generated information is used efficiently, the overall effect would be to speed up marking as less time will be spent on recurring errors. The three goals feed into one another.

A second way in which MarkWrite will free up lecturer's time is that more trust could be placed in marking assistants. With the "example" feature in MarkWrite indicating to the marker when to use a

tag, and with the assessment assistance in the marking schemes, it is possible that greater inter-marker reliability will ensue. Marking assistants will have more guidance, resulting in them marking more in line with the lecturers' intended outcomes.

8.4.2.19 Type/token ratio

At present, MarkWrite indicates the type/token ratio in any marked text, but no feedback or remedial exercises are associated with that as yet. Some research will be necessary to generate feedback on a type/token ratio. A type/token ratio can be used to estimate the level of vocabulary proficiency of students, but this will differ according to subject field, topic, length of texts and language used. The type/token ratio should be correlated with the style checker as it can identify overused words or overly long or short sentences.

8.4.2.20 Reading ease score

A reading ease score, cross-correlated with the assessment and feedback, might also be useful to establish a general quality goal for specific year groups. It should also be possible to establish feedback based on the reading ease score. This will require a lot of research to accomplish though, since reading ease scores are controversial and they differ from language to language.

8.4.2.21 Level of importance of various feedback categories

The danger of "feature creep" has been mentioned already, with the warning that a "truckload" of feedback on a short text will be demotivating and ultimately useless to the student. "Feature creep" has another dimension and that is the simple, but hard to answer question, "which feedback is more important than other feedback?" It is easy to use theories of language and communication (and this study has done so) to argue the case that the surface structure of texts is less important than the underlying elements, such as argument structure and the quality of the information. One should, however, be wary of simply working towards the obvious. Additional research may indicate that to the final reader, the quality of the argument is of less importance than the presentation.

This observation was brought to the author's attention by the following: some of his students receive very good marks for their content subjects, but they are unable to write complete sentences. In other words, their lecturers in the content subjects identify key words and concepts in their test or examination scripts and mark them as correct, without establishing if the sentence is actually legible. The same thing happens in a normal spoken conversation. Full, well-structured sentences do not occur throughout natural, spoken texts, with interruptions, interjections and other "noise" marring the sentence structures. Nevertheless, meaning is still transferred in spoken texts. The question to be investigated therefore is the level of importance connected to these features. Likewise, are good introductions really that important? In other words, what should ideally be commented on? What are the few key features in texts which make the biggest difference?

These questions touch upon the issue of key behaviours. In the highly touted book, *Influencer– the power to change anything* (Patterson, Grenny and Maxfield, 2007), the authors provide many examples of situations where the focus is on changing only a few key behaviours. By changing these few behaviours, unrelated behaviours are subsequently changed as a sort of "by-catch". The authors also make the distinction between *outcomes* and *behaviours* and claim that an outcome cannot be reached without changing a key behaviour or set of behaviours. In other words, the question to

investigate is which key behaviours of students could be focused on, manipulated or changed in order to reach the desired outcomes faster.

8.4.2.22 Research on user friendliness

The biggest hurdle for MarkWrite to overcome is user friendliness. It is vitally important that MarkWrite be as user friendly for the markers as possible. As with any software development, user friendliness cannot be pre-programmed, but must be evaluated and adapted. This is obviously not inherently research on feedback or student writing, but it is very important for the future of the MarkWrite project. Based on anecdotal evidence, lecturers often mark student texts while sitting in front of the television, on their patios, in bed, or other comfortable places. Despite all the advantages offered by MarkWrite, it is more comfortable to mark sitting in bed than behind a computer. While MarkWrite will never be "comfortable", it should at least be user friendly enough that its ease of use cannot be mentioned as an excuse not to use it. Once users are convinced of the additional benefits as well, *comfort* should no longer be such a big problem.

8.4.2.23 Pre-checks for students

As mentioned above, MarkWrite Marker will be supplemented with MarkWrite Student. The plan with MarkWrite Student is that it should take the student through all the necessary steps for writing a good essay. The student should also be taken through a number of pre-checks before even being allowed to submit his or her essay for final evaluation by a marker. These will include a spelling check, style check (as explained above), plagiarism check, nosy thesaurus and a self-assessment incorporating the radio buttons for paragraph, introductions, conclusions and argument structure where applicable.

8.4.2.24 Type/token ratio feedback

At the moment, MarkWrite provides the type/token ratio for every text. No feedback connected with this ratio has as yet been established. It should be possible to provide students with feedback on their vocabulary usage. It should also be possible to measure cross-correlations between the type/token ratio and established lists of academic vocabulary such as the one compiled by Coxhead (2000).

8.4.2.25 Style analyser

Academic writing style is a very difficult concept. What exactly constitutes an academic style and what makes it different from the journalistic or informal styles? While student writing textbooks are quick to list differences, it is more difficult to identify these differences in practice. As technology advances, and as data from real-life marking with MarkWrite become available, it should be possible to set up a style analyser to identify and eradicate common errors of poor style. This is another example where bootstrapping will make it possible to improve the software and feedback with long-term application.

8.4.2.26 Teacher check-ups

One of the big advantages of the marker system is its ability to count feedback. This could be used to identify which marking assistants habitually focus on only certain aspects of the text. Marking assistants are also continually learning to mark more effectively – MarkWrite can speed up their acquisition of marking experience. If the computer can assist in any way with that, it should be done.

At present, the check-up function only reminds the marker when he or she has not used a specific feedback superordinate. While this is already an advantage, this option allows for much more advanced marker reminders.

8.4.2.27 Mobi sites and cellphone usage

While many people in Africa do not have access to computers, the use of cellphones is proliferating and more and more people are accessing the Internet through their cellphones. It would therefore be an important development to create the MarkWrite output files in such a way that they would be suitable for mobi sites which are better supported by cellphone browsers. In this case, it would also be possible to send detailed feedback to distance students via mobile technology.

8.4.2.28 Effect of reading comprehension on feedback interpretation

Much research has been done on the effects of reading comprehension and its influence on different aspects of teaching and learning. Despite this, the author has failed to unearth research specifically comparing the effectiveness of feedback with reading comprehension. Also, what are the effects of reading comprehension on the editing of the authors' own texts? To put it bluntly: can students understand their own writing and to what extent? Based on anecdotal evidence and experience with students in class and one-on-one interviews, the author suspects that some students actually have difficulty in understanding their own writing; they fail to see that what they have written does not convey their intention. While this is not new information, the effect that this has on text production in the process approach to writing has not been established.

It is possible to use eye tracking software as a diagnostic tool in combination with MarkWrite. This will enable researchers to establish how learners read and interpret feedback *in relation to* specific parts of the texts. It can also be used on markers to see what *exactly* they read while deciding to add specific comments.

8.4.2.29 Screen capture

At the moment, MarkWrite provides written feedback with a mild visual stimulus in that the parts of the text commented on are coloured. To enhance the audio-visual effectiveness of the feedback, it would be possible to make screen capture videos while the marker is marking the text. These videos can then be supplied to the students as well to illustrate to them how a reader interacts with a text. The practical feasibility of this suggestion will need to be investigated as well as its effectiveness, but it may prove to be relatively easy to implement.

8.5 Implementing MarkWrite in writing across the curriculum

Initially, MarkWrite was intended for language studies alone, but that would unnecessarily limit the scope of the project. At the time of writing, MarkWrite has already been tested by lecturers in other writing-intensive courses and requests have been received from lecturers in Economics, Statistics and Mathematics, among others, to adapt MarkWrite to their specific needs. For example, lecturers in Statistics will want to mark spread sheets, and lecturers in the Faculty of Education will want to mark PowerPoint presentations. No matter which subject field, an effective paragraph has the same qualities, and in probably all fields except creative writing, introductions and conclusions will have similar qualities.

It is hoped that with consistent implementation in many different fields, students will continually be reminded of what they should focus on when communicating through writing. This will be writing across the curriculum in a way that makes it easier also for non-linguists to comment on the writing aspects of the assignments their students produce. As such, implementation of MarkWrite will create strong links to Content-based Instruction.

8.6 Problems to overcome

As with any project of this scope and with these lofty ideals, there are many difficulties to overcome, some of which are discussed here.

8.6.1 Funding

It is expensive to program and reprogram a system. The current MarkWrite system required a very big capital investment from CTexT® and took a year to develop. All the further advantages which are possible will require intensive research and many more man-hours of programming. Both the research and reprogramming are expensive.

8.6.2 Research volume and time

Research itself is expensive. At least 10 of the topics in the list of future developments have the potential to be a complete Master's thesis and a few of the topics may even require a PhD or post-doctoral studies. The volume of research available requires additional researchers and funding.

8.6.3 Theoretical/pedagogical inconsistencies in techniques

The research literature on the teaching of writing, and especially on feedback techniques, is not consistent in its findings. This has been extensively discussed by Ferris (2003, 2004). In practice, it remains to be seen how these conflicts will influence the acceptance and usage of MarkWrite. It may be possible that MarkWrite can contribute to the debate.

8.6.4 Standardisation is difficult

It is difficult to write the feedback tags clearly enough to be of use to the student, but still generic enough to apply to a range of different errors. In addition, the option allowing anybody to write their own tags negates some of the advantages of using a standardised feedback tag set. It is hoped that once MarkWrite has gained a global foothold, it will be possible to get a greater degree of standardisation of the different tag sets for different subject fields.

8.6.5 Technophobia

Some people simply prefer to mark on paper and many of the author's senior colleagues are computer shy. Technophobia will have to be overcome if MarkWrite is to be adopted in mainstream educational settings.

8.6.6 Lack of access to computers

MarkWrite cannot be implemented in schools at present, simply because many schools in South Africa do not have sufficient access to computers, the Internet and related equipment.

8.6.7 Mindset shift

The idea of marking a text really comprehensively is intimidating, since it takes a long time. The mindset of markers should shift from “marking to get a mark” or “marking to get my work done” to really “marking for learning”. It is not clear what the optimal length of a text should be to get benefit from the feedback on it. It is possible that a shorter text, marked more comprehensively, and requiring students to actually use the feedback, will result in faster and better learning, than using traditional marking on much longer texts. This would make for interesting research, but once again would be a daunting question to investigate and would require funding.

8.6.8 What about fully automatic marking?

Criticism against this technique is that it appears to be outdated before it is implemented, because “multiple choice feedback” has been tried already with many lecturers simply using macros in MSWord to provide feedback. Fully automatic feedback such as the Criterion and E-Rater services of ETS (Attali and Burstein, 2006) seem to be the way forward.

Multiple choice feedback or the use of macros is very similar to this technique, but the difference here is the scale of implementation and the level of standardisation. Anybody can create a list of questions (even a well-researched list) and use the answers to provide feedback. However, if this feedback is not implemented in a system aimed at holistically providing feedback on multiple areas of a student’s text, it does not meet all the requirements of effective feedback as illustrated 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.

With regard to fully automatic feedback, there are multiple problems. The first problem is students’ immediate distrust of fully automatic feedback. The second is that fully automatic feedback (while getting better all the time) is still not fully trustworthy (and neither are human markers). This technique, if incorporated into a computerised marking support interface and linked with assessment assistance, is therefore an attempt to find the intersection point between fully automatic and fully manual, much like the manumatic transmission (also called Tiptronic transmission) in cars, allowing improved performance without the loss of user control.

The third and fourth problems with fully automatic feedback become apparent when regarding 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 really facilitate writing across the curriculum. 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 intention of automating more of the system. In addition to that, MarkWrite is not intended to be only an English marking system. The technology in MarkWrite can (with some adaptation) be used for other languages and subjects other than language subjects within the South African context.

8.6.9 Beware "work to rule"

As "work-to-rule" strikes have illustrated so many times, strict adherence to all the rules can cripple a system. With the number of features available in MarkWrite, this kind of rebellion could be triggered in users.

8.6.10 Different styles of learning

The theory of multiple intelligences (Gardner, 1993) explains that different students acquire and remember information in different ways. While this is still controversial, it is an interesting avenue for further research: the relation between the theory and the effects of different feedback techniques on the improvement of writing. At the moment MarkWrite does not explicitly cater to the different needs of students, in that feedback is mostly written. MarkWrite does, however, provide opportunities for further development, and the *Backchat* feature mentioned above, as well as the screen capture feature, might provide a more holistic learning experience.

8.7 Final remarks

As has been illustrated above, with all the possible future developments the MarkWrite project is still in its infancy. With new developments in computer technology, natural language processing and a new generation of "computer-raised" students, it is hoped that MarkWrite will mature into a tool which will become an everyday thing to use as is a normal word processor. It is hoped that MarkWrite can be used in conjunction with other technological advances to speed up and enhance the learning experience of students, and reduce and alleviate the tediousness of marking for lecturers, while at the same time greatly increasing the efficiency of an activity which is as old as the hills, but which was in serious need of rethinking. MarkWrite can ensure that the correlation between research and teaching will become easier, but that will only happen some time in the future. It is hoped that funding will be approved to enable the continued development of the project.

8.8 Bibliography

Attali, Y. & Burstein, J. 2006. Automated scoring with E-rater V.2. The *Journal of Technology, Learning, and Assessment* 4(3):3-30.

Butler, H.G. 2006. A framework for course design in academic writing for tertiary education. Pretoria: University of Pretoria (PhD thesis).

Coxhead, A. 2000. A new academic word list. *Tesol Quarterly* 34(2):213-238.

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.

Gardner, H. 1993. Frames of mind – the theory of multiple intelligences. New York: Basic Books.

Hattingh, K. 2009. The validation of a rating scale for the assessment of compositions of ESL. Potchefstroom: North-West University (PhD thesis).

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

Patterson, K., Grenny, J. & Maxfield, D. 2007. Influencer – the power to change anything. New York: McGraw-Hill.

Weideman, A. & Van Dyk, T. 2004. Switching constructs: on the selection of an appropriate blueprint for academic literacy assessment. *SAALT Journal for language teaching* 38 (1): 1-13.