Students' Conceptions of Tutor and Automated Feedback in Professional Writing

RAFAEL A. CALVO AND ROBERT A. ELLIS
The University of Sydney

BACKGROUND
Professional writing is an essential outcome for engineering graduates and hence a vital part of engineering education. To provide a successful learning experience for students engaged in writing activities, timely feedback is necessary. Providing this feedback to increasing numbers of students poses a major challenge for instructors. New automated systems work towards providing both timely and appropriate writing feedback, but students' views on automated feedback, and feedback in general, are not well understood.

PURPOSE (HYPOTHESIS)
To contribute to a deeper understanding of students' conceptions of feedback from tutors and an automated system called Glosser, and how these conceptions are related to achievement.

DESIGN/METHOD
Students in an engineering course worked in pairs to write an engineering report on e-business. The design of the study involves in-depth interviews and the analysis employs an approach in which student conceptions of automated feedback are investigated in relation to related feedback from their tutor, perceptions of automated feedback in general, and their academic achievement.

RESULTS
Students' conceptions of feedback vary and can be grouped into cohesive and fragmented, which is consistent with other theoretical models. Close associations were found between more cohesive conceptions of feedback and better academic performance.

CONCLUSIONS
A student's conception of traditional and automated feedback is similar, being either cohesive or fragmented. Changing one may change the other. Deep learners see feedback as a way of learning about the topic whereas shallow learners see them as a way to improve the communication aspects of writing. Design considerations based on these results are discussed.

KEYWORDS
automated feedback, phenomenography, professional writing

I. INTRODUCTION

Communication, especially written communication, is a critical part of engineering practice and an indispensable graduate outcome for engineers. Research shows that engineers spend 40–60 percent of their time communicating (Silyn-Roberts, 1998; Tenopir and King, 2004). Engineers report spending between 20 percent and 40 percent of their workday writing, a figure that increases with the responsibility of the position (Kreth, 2000). ABET, Inc., an accreditation body for applied science, computing, engineering, and technology, requires evidence that graduates have the "ability to communicate effectively" as essential for accreditation. There is no doubt that an ability to write effectively is a necessary competency for engineers. Helping engineering students to learn to write effectively, however, poses many challenges.

A number of approaches are used for teaching writing skills in engineering curricula internationally. In the United States of America, the Writing Across the Curriculum (WAC) movement uses writing as a way to develop students' abilities to write clearly for their discipline as well as a learning task for complex problem-solving situations (Yalvac et al., 2007). In Australia, approaches derived from an applied linguistic perspective works at the level of genres of professional texts (Bonanno and Jones, 1997; Prosser and Webb, 1994). This approach involves carefully detailing the linguistic structures of valued disciplinary texts, such as engineering reports, in ways that allow apprentice writers to develop a mastery of the structures. This approach has been complemented with relational student learning theory research which has shown that the way students think about their writing, and the qualitative nature of the strategies they adopt in the writing process, are closely related to the levels of achievement and the patterns of language they use (Ellis, 2004; Ellis and Calvo, 2006).

Reviews conducted by Haswell (2006), Leydens (2008), Shute (2008), and others illuminate the complexity of student responses to feedback, the variation in existing findings from studies of feedback, and the need for additional work in this area. One of the main challenges highlighted by these reviews is the provision of sufficient feedback for students during and at the end of the writing process so that their linguistic and disciplinary understanding develops appropriately during writing tasks.

Whether based on constructivist, cognitive, or linguistic models (Connolly and Vilardi, 1989; Herrington, 1981), all approaches to teaching writing in engineering demand timely feedback. Within present-day universities, particularly those with large undergraduate student populations, the provision of timely feedback to large cohorts can be difficult if not impossible to sustain. This has led academics to look for shortcuts or for automated tools that reduce their workload (Haswell, 2006). Furthermore, successful approaches to engineering education should not only help students to write well, but to also better understand disciplinary ideas. Thus, feedback is required on both the academic style and the quality of the content.

Due to its significance, formative feedback has been studied for decades. However, the variables that make feedback effective are

not well understood. In an extensive review, Shute (2008) recently concluded that despite the large body of research, studies of the effects of formative feedback have often been inconsistent. For example, several authors have argued that feedback complexity (in length and content) is not significantly related to feedback effects, but others have found significant negative effects. Shute argued that the inconsistencies may be due to "individual differences among motivational prerequisites (e.g., intrinsic motivation, beliefs, need for academic achievement, academic self-efficacy and metacognitive skills)" (Shute, 2008, p. 24). Automated feedback may offer new opportunities, but studies like Shute's suggest that new approaches to study its effects are required. The approach followed here concurs with Narciss and Huth (2004), Shute (2008), and others who have argued that to study the effects of feedback, its function, content, and mode should be studied separately and interactively with learner characteristics and instructional variables.

Researchers have studied ways to provide timely feedback in computer-based learning environments. This type of feedback is provided during or at the end of a learning activity, and can be generated in three ways (Gouli et al., 2005): (1) automatically, by a software system, (2) by tutors, or (3) by peers. The first type of feedback typically checks the correctness of students' answers, and often provides students with an assessment and information to correct their mistakes and misconceptions. More recently, particularly in Intelligent Tutoring Systems (ITS), the focus has been on finding ways of adapting the predefined feedback to students' preferences, certainty of their answers, and learning styles (Gouli et al., 2005). A challenge involved in this approach is that compositions are typically neither totally correct nor totally incorrect. Several projects have used computational approaches to assessing and providing automated feedback on writing to address this challenge (Foltz, Gilliam, and Kendall, 2000; Kakkonen, Myller, and Sutinen, 2004; Landauer, 2003). Despite a variety of strategies to improve the quality of feedback in the student experience of writing, providing timely and appropriate feedback to students at key stages of the writing process remains a serious challenge for university teachers.

Beyond the technical challenges (not discussed here) of building automated feedback tools, an important research question that needs to be addressed relates to how students perceive, use, and benefit from such feedback. The literature has focused on instructor or peer-feedback and how students relate to it, or on the automated feedback tools (typically used outside class-related activities) and the effects they have on learning. Despite the increased popularity of automated feedback tools, to our knowledge there is no previous research on how students view these tools and how these are related to their perception of expert feedback. Despite some controversy, there is evidence (Hyland and Hyland, 2006) that students value their teachers' corrections and generally use them to make accurate changes to their texts. However, there is no study that investigates if students view (and use) automated feedback in the same way.

In this study, the authors investigate a blended use of automated systems to complement traditional forms of feedback. To do this, they examine the benefits, efficacy, and limitations of an automated feedback system known as Glosser in relation to students' writing experiences. Glosser uses textual data mining and computational linguistics algorithms to quantify features of the text to provide students with descriptive information about patterns in the text.

II. Prior Research

A. Tutor and Peer Feedback

Feedback on writing and the way students respond to this feedback have been studied from different perspectives (Haswell, 2006; Hounsell 1987; Weaver 2006). In this article feedback is conceived as input that a reader provides so the writer can revise the composition (Keh, 1990). Several researchers have produced taxonomies of the components that can be included in such feedback, and discussed why there is no consistent evidence that labor-intensive necessarily leads to learning gains (Haswell, 2006; Shute, 2008).

One reason why students do not react as expected to instructional feedback is miscommunication between student and instructors. The unexpected ways in which students understand feedback and reasons for not engaging with the feedback are also discussed by Haswell (2006). In general, many authors agree that students' reactions to feedback are very complex, affecting identity (Haswell, 2006), confidence, and self-esteem (Young, 2000).

A meta-analysis of feedback in the workplace (DeNisi and Kluger, 2000) showed that feedback may produce reduced performance in up to one-third of cases. This evidence highlights the importance of understanding how the feedback provided either by peers, instructors, or computers (or combinations of these) affects students' learning.

According to Lizzio et al. (2002), students assess the quality of feedback from peers or tutors in three core dimensions: developmental, encouragement, and fairness. The developmental dimension can be thought to include the suggestions, comments, and questions that a reader provides that help the writer identify problems in the composition. In their study, developmental feedback showed the strongest association with quality feedback attributes (Lizzio, Wilson, and Simons, 2002) and it is the focus of the Glosser system described later. As a form of feedback, Glosser provides trigger questions and visual representations of the document but does not try to provide an assessment based on an ideal composition.

There are many difficulties in providing effective feedback in writing, even more so through using computational approaches. Among the most significant is the difficulty in identifying reliable textual features of good writing that can be implemented as a computer model. For example, many authors assume that cohesion (or coherence) is a feature that identifies good writing. This seems to be supported by some evidence (Witte and Faigley, 1981), most of which comes from research on second language writers (L2) (Liu and Braine, 2005) . However, highlighting the difficulty of identifying unequivocal quality features of texts, even this commonly held assumption has been questioned in a recent study (McNamara, Crossley, and McCarthy, 2010) that showed no significant correlations between cohesion measures and essay ratings.

Glosser is designed around the principle that precise quality judgments on text features are difficult (if not impossible), so rather than being directive, the feedback should provide different perspectives (Flower, 1994) on the writing and help students engage with it.

B. Automated Feedback

Haswell (2006) reviewed systems for automated feedback dating back to the 1950s. These systems are similar to (if not inclusive of) the four major systems of Automated Writing Evaluation software (i.e., Project Essay Grade [PEG], Intellimetric, Intelligent Essay

Assessor [IEA], and E-rater) evaluated by Keith (2003). Despite being increasingly used in higher education, there is still a strong resistance to automated feedback, particularly amongst composition instructors. Haswell (2006) and others (Shermis and Burstein, 2003) have discussed arguments both in favor and against using such tools. Many of the positive and negative claims made about how these feedback tools support learning remain largely unsubstantiated. However, a detailed discussion of this issue is not the goal of this paper.

Systems like *Writers Workshop*, developed by Bell Laboratories, and *Editor* (Thiesmeyer and Thiesmeyer, 1990) focus on grammar and style with limited pedagogical benefits (Beals, 1998). While these automated systems faced problems of accuracy, the main difficulty identified was that correcting surface features of student texts did not help them to improve the quality of their ideas or the knowledge of the topics that they were studying.

SaK, a more recent writing tutoring system developed at the University of Memphis (Wiemer-Hastings and Graesser, 2000), is based on Flower's (1994) notion of voices that speak to the writer during the process of composition. SaK uses avatars to give the impression of giving each of these voices a face and personality (Wiemer-Hastings and Graesser, 2000). Each avatar provides feedback on a different aspect of the composition, identifying strengths and weaknesses in the text but without offering corrections. Glosser adopts a less explicit form of Flower's idea, using tabs rather than avatars.

SaK, like Glosser, uses Latent Semantic Analysis (LSA) to calculate the average distance between consecutive sentences and to provide feedback on the overall coherence of the text. LSA is a technique used to measure the semantic similarity between texts and has been described thoroughly elsewhere (Landauer et al., 2007). SaK can also analyze the topic of a sentence, identifying clusters of topics so when a new topic arises the student can be asked for an explanation or reformulation.

Other automated systems which have become commercial products, such as Summary-Street (Wade-Stein and Kintsch, 2004), tend to focus on drills. Their success tends to be skills-based, such as learning to summarize, rather than driven by disciplinary concepts.

While some automated systems have been developed to be more discipline-informed, such as those designed for science and engineering students (Andeweg and De Jong, 1996; Hayes et al., 2007), the real value of automated feedback remains contested. The techniques of Natural Language Processing (NLP) and Machine Learning used for automated writing tutors are similar to those used in Automated Essay Scoring (AES). The increasing use of these AES in many institutions has sparked debates about accuracy and pedagogical value. Two recent books discuss advances in automatic essay scoring, one taking a very supportive approach (Shermis and Burstein, 2003) and one providing a more critical perspective (Ericsson and Haswell, 2006).

This study extends and deepens the research described above by considering the structure and purpose of Glosser, a new automated feedback system, and by investigating how students relate its use to their views on other forms of feedback and to their learning outcomes. In doing so, this paper discusses some of the issues in developing an automated student feedback system that draws on data-mining techniques, machine learning, and software engineering methodologies. These techniques are similar to those used in automated essay scoring systems (Ericsson and Haswell, 2006; Shermis and Burstein, 2003) but used here for feedback in the form of suggestions that trigger reflection and inquiry. The techniques themselves are not discussed in detail as they have been covered elsewhere (Ericsson and Haswell, 2006; Shermis and Burstein, 2003).

To understand the student experience of automated feedback in writing, this study looks at their conceptions of feedback and of automated feedback in general, how students perceive automated feedback when provided, and how these conceptions and perceptions relate to learning outcomes. To investigate the student experience of automated feedback in their writing, a relational student learning research framework is adopted (Prosser and Trigwell, 1999; Ramsden, 2002).

III. The Theoretical Framework

Key studies investigating the quality of university students' experiences of learning have focused on what students think they are learning, their conceptions of learning, their perceptions of the context in which their learning is taking place, and how these aspects are related to their academic achievement (Biggs, 1999; Prosser and Trigwell, 1999). For writing activities, this is especially relevant, as prior research has shown a close relationship between the quality of conceptions of writing reported by students and their academic performance.

Student learning research (Prosser and Trigwell, 1999; Ramsden, 2002) have identified qualitatively different conceptions of learning reported by students. Some conceptions retain the idea of learning at the centre: that the development of comprehension and deep understanding is the main purpose of the activity and this underpins the entire experience. These are usually referred to as cohesive conceptions. Other conceptions tend to be partial, fragmented into unrelated aspects in which there is no real awareness of understanding of the development of the key ideas being studied. These conceptions are usually referred to as fragmented or multistructural.

Students' experiences of writing have also been studied using this phenomenographical approach (Marton and Booth, 1997). In general, this line of study shows that students have qualitatively different conceptions of writing and that these variations affect their learning outcomes. For example, the views of writing of eighth grade ($N = 97$) students (Levin and Wagner, 2006) were studied through the metaphors they used when asked to reflect about the writing-to-learn task. The grouping of these views was similar to other theoretical models and provides an insight on how qualitatively different views are possible, and how that impacts learning outcomes. These views are also similar to those in other studies that explore the conceptions of writing in college students (Hounsell, 1997; Lavelle, 1993).

This study focuses on the students' conceptions of feedback on writing activities rather than on writing itself. Furthermore, the theoretical framework used focuses on how students experience feedback, not on the quality of the feedback itself. With this focus the study does not make assertions about the accuracy of Glosser's algorithms or the learning gains that automated and other forms of feedback may or may not support.

This theoretical framework, used in countless studies, divides the experience of the phenomena (i.e., learning through feedback) into the conception the student has about the phenomena (what is feedback) and the approach to be taken (how to use it). At the

core of this division is the intuitive idea that in order to learn from feedback, the student needs to have a conception of what feedback is. Interestingly, understanding these conceptions is not always straightforward. Furthermore it would be naïve to assume that students view automated feedback in the same way they view feedback from a tutor or a peer. Therefore, if they conceive feedback in different ways, the implication is that they will use it in different ways.

From the perspective of phenomenography, each of these forms of feedback (tutor, peer, and automated) is a different phenomena. Learning within this framework is about developing better ways to experience each phenomenon; in our context this means developing new ways to understand what the different forms of feedback are and how they can be used.

In order to help students learn from feedback, faculty (and software developers) need to first understand how students conceive and use the different forms of feedback. This is the goal of this study. The methodology aims to build an understanding that can be used to adapt the different forms of feedback (e.g., the software) or the way it is presented to students.

This study adds to previous research by investigating the conceptions area of the model; that is, student conceptions of feedback in the writing process, including feedback from the tutor and automated feedback from Glosser. A key purpose of the study is to help instructors develop a better understanding of the different conceptions that students retain while they write and how this is related to their experience and performance.

IV. Aims

The aims and hypotheses of the research are:

1. How do students conceive learning from feedback on writing from tutors?
2. How do students conceive learning from automated feedback on their writing? Our first hypothesis, related to questions 1 and 2, was that students would have qualitatively different conceptions of learning through feedback and automated feedback (Hypothesis 1).
3. How do students perceive the value of feedback from automated systems? Our hypothesis was that student perceptions of the potential contribution of automated systems to learning would vary (Hypothesis 2).
4. Are there associations among conceptions of feedback, automated feedback and academic achievement? Our hypothesis was that there would be associations among qualitatively different conceptions, perceptions, and outcomes such as academic achievement (Hypothesis 3).

Before discussing the methodology and results arising from the research questions, it is worth looking at the structure and purpose of Glosser in more detail.

V. What is Glosser?

Glosser is an automated feedback system which provides contextualized feedback to students about their professional texts. It uses textual data mining and computational linguistics algorithms to quantify features of the text and produce feedback for the student. This feedback is in the form of descriptive information about different aspects of the document. For example, by analyzing the words contained in each paragraph, it can measure how "close" two adjoining paragraphs are. If the paragraphs are too "far" this can be a sign of a lack of flow, and Glosser flags a small warning sign.

The version of Glosser (1.0) described here provides feedback on four aspects of the writing: structure, coherence, topics, and keywords. It also has a section that analyzes participation of individuals in a group. Each of these areas is identified by a tab on the homepage of the Web application as shown in Figure 1. The *Questions* section contains triggers to prompt student reflection on a particular set of features of the composition and it can be customized for each course or activity. The *Instructions* section describes how to interpret the feedback, which is provided as "*gloss*" on the lower half of the page. The feedback is an alternate type of visual or textual representation of features that have been automatically sourced from the text. The history of *versions* provides a record of the writing process, which includes information on the time of each revision and the author responsible for it.

Glosser is not a writing tool. Students can write in Microsoft Word, a wiki page, or Google Docs. Glosser integrates these tools through file uploads or the use of their Application Programming Interfaces (APIs).

A. Structure

The tab labeled "structure" is designed to provide students with an analysis of the key textual patterns of text. The focus is on ideas such as cohesion, topic sentences, and the flow of a key argument across the whole text (Bonanno and Jones, 2000). According to the textual pattern that is expected in the document, Glosser presents what are statistically the "key" sections of the document, and with a series of trigger questions the student is expected to reflect on whether or not they have followed the guidelines they have received. For example, for a standard essay—with an introductory paragraph followed by a few paragraphs elaborating key ideas, and a closing paragraph summarizing the argument-the system pulls out just the key ideas and displays them alongside a series of trigger questions such as: Does the conclusion follow from the argument? Does each point contribute to the argument?

The many textual patterns commonly found in technical writing make it challenging to make representations that are meaningful to the student. New visualizations of the composition's "patterns" are being explored (O'Rourke and Calvo, 2009).

B. Coherence

The purpose of the Coherence section is to help students consider issues of flow and how each paragraph and sentence logically follows from the previous one. In this section, the system's "gloss" calls attention to spots of possible breaks in coherence within the essay by inserting a "broken" icon between problem paragraphs. The system does so by identifying consecutive paragraphs that display a low level of lexical cohesion, related to coherence (Foltz, Kintsch, and Landauer, 1998), meaning that two paragraphs are far apart in the latent semantic space. This technique, introduced by Foltz, Kintsch, and Landauer (1998), has been applied to successfully identify problems in writing. There is evidence that design documents that show better coherence according to this measure are considered better designs (Dong, 2004).

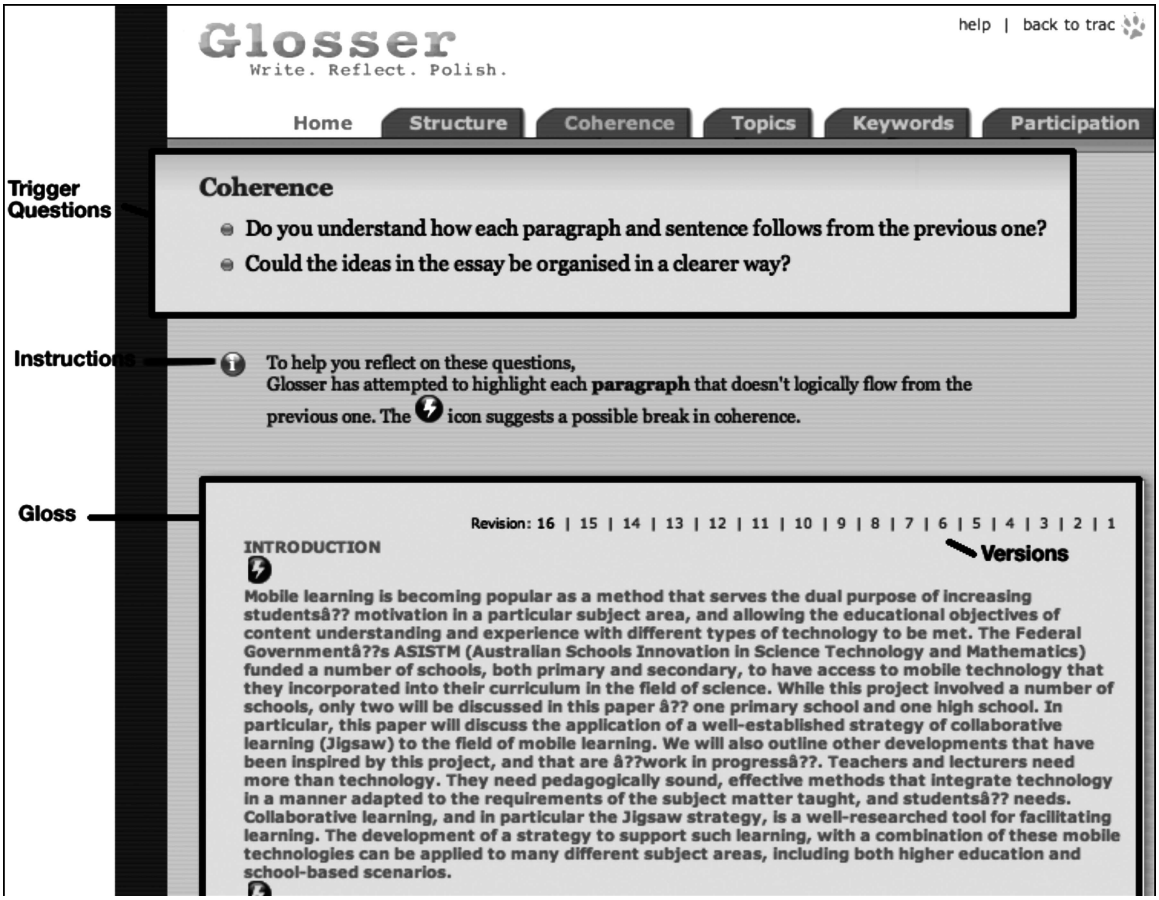

Figure 1. An example of the Glosser interface.

In order to calculate these distances, the semantic space formed by the paragraphs of the document is used. After reducing the dimensionality, distances are calculated between all consequent paragraphs; if the distance surpasses a predefined threshold, the two paragraphs are marked with a potential problem. In addition, hovering over a paragraph with the mouse in either the structure or coherence sections produces a tooltip with the name of the author associated with this particular paragraph and the date of the last revision.

C. Topics

In the Topics section shown in Figure 2, the feedback takes a different form. Rather than presenting an annotated version of the essay, it presents a list of topics extracted from the writing. This is calculated using the semantic space formed by the sentences of the document.

The idea is to provide a different perspective of the document, so that writers can view, at a glance, what someone else (i.e., a reader) might interpret as the set of topics that have received most attention in the document. This, in turn, may trigger reflection on issues such as "Was that what I meant to focus on? Do I spend too much time on a digression? Have I sufficiently addressed another topic?" For example, if this paper is uploaded to Glosser, the top 3 topics that would emerge are Use Glosser, Automated feedback, and Writing in engineering. An author can then look at a particular topic and see a list of all sentences that pertain to that topic. Finally, authors writing collaboratively can get a sense about who has been the primary contributor for each topic, as a name is listed next to each topic.

D. Keywords—Knowledge Map

The gloss comes in the form of a knowledge map style visualization of the key concepts identified in the document and the relationships between them. The concepts or keywords displayed are drawn from within the document and from an *ad hoc* thesaurus. These concepts make up the nodes. Linking lines between the nodes represent relationships between the concepts, and are identified via sentences in which both concepts co-occur. By hovering over a linking line, a user can see (a) which sentences provide the evidence for this link, and (b) who the primary author is of this connection (sentences). This visualization has the effect of showing, at a glance, the concept or concepts that are central to a written work (since related concepts visually cluster around central points), as well as identifying stray concepts that are not well integrated into the document. Hence, this visualization can also help detecting amorphous work that is lacking in any central ideas. Moreover, the authorship information can indicate the semantic contributions that users make to a document. The interface is adjustable by the user, so authors can drag concepts around in order to create a map that is most meaningful to them. In the future, a clustering algorithm will be added to allow concepts to be automatically grouped based on 'ownership'.

New techniques are being developed to automatically create concept maps (Villalon and Calvo, 2009) that would have a hierarchy of concepts and eliminate the need for a manually generated thesaurus.

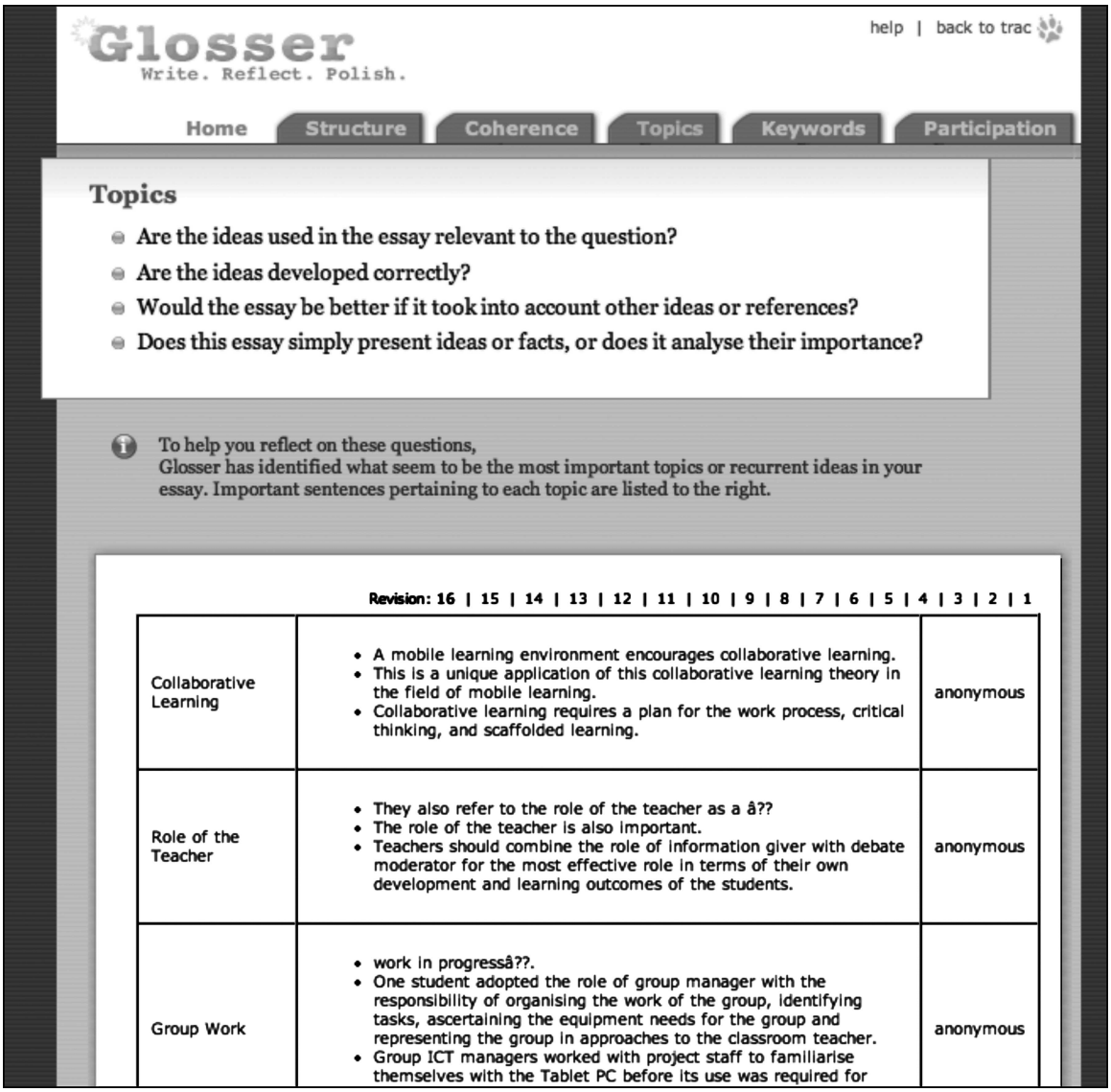

Figure 2. The topics section. Key topics identified in the writing are listed at left. Associated sentences and the primary author of that topic are listed to the right.

VI. Method

The purpose of this evaluation is twofold. First, it is to be used to improve understanding of student conceptions regarding feedback in relation to writing tasks, so this can inform the way we structure and scaffold feedback for writing activities. Second, we have argued that the development of software for the purposes of student learning should be informed by an analysis of the student experience (Calvo et al., 2007), so that the support provided by such software is more likely to meet students' needs.

A. Participants and Course Context

Glosser was used by undergraduate students in a third-year software engineering course at the University of Sydney. Over a period of six weeks, students ($N = 46$) were asked to write a project proposal for the development of an e-business. In writing the proposal, the students were asked to place themselves in the role of a technical manager. There is considerable value in writing activities in which engineering students can imagine themselves in a professional role (Yalvac et al., 2007).

The first version of the project proposal had to be submitted in Week 4 of the semester, and the second version in Week 6. During the period of the writing tasks, students had access, but were not required, to use Glosser. In Week 2, faculty and staff from the Learning Centre (who provide support for teaching writing) provided students with a writing workshop designed to improve their academic writing proficiency, particularly for the type of proposal writing required for the assignment. During the workshop, students were also introduced to Glosser and useful strategies they could adopt in order to use it.

In Week 4, each student was assigned to review another group's proposal, and asked to use Glosser to support the reviewing process. The proposals and reviews had to be written in a wiki system (Trac) and each wiki page had a link to Glosser. The students alternated between their own texts, the text they had to review, and Glosser as they assessed the strength and weaknesses of the writing and ideas.

Thus, students received feedback from peers, from their tutor, and from Glosser at various stages of the process. They were able to use the Glosser feedback, not only to prompt reflection on their own work, but also to support them in reflecting on another group's work. Students submitted their texts to the tutor who gave them a final mark based on the quality of the writing and ideas.

B. Data Collection and Analysis

To evaluate the contribution of Glosser to the student experience of feedback in their writing process, students were interviewed individually by the researchers. All students were aware, from early in the course, that they would be asked to report on their experiences and that the participation in the research was voluntary. All provided informed consent to the project as approved by the relevant university ethics committee. From those who consented, about half ($N = 22$) were randomly selected for the interviews discussed here. Each interview was approximately 20 minutes long. Each student was asked the following questions:

Conceptions of feedback

1. What was the purpose of tutor feedback in your writing task?

Conceptions of the automated feedback from Glosser

2. What was the purpose of Glosser in your writing task?

Perceptions of automated feedback in general

3. Do you think meaningful automated feedback is possible?

The analysis of the transcribed interviews used a phenomenographic method, which involves the development of qualitatively different categories of student conceptions (Marton and Booth, 1997). In this method the questions above are triggers for the student to elicit their views and they can be followed with further questions from the interviewer. The line of questioning can continue until the interviewer feels that the subject has provided enough information to later form the categories.

The transcripts were analyzed by two researchers and discussed with a third. Each researcher independently read the student responses to the question. After individual reflection they met to discuss the key themes and main groupings in the student data.

The researchers agreed on categories and representative quotations. Their structure was influenced by the SOLO taxonomy (Biggs, 1999). In a second meeting, the classifications produced by each where discussed, and agreement reached on the few (< 10 percent) that had originally been classified differently.

VII. Results

The main outcomes of the study encompass qualitative variations in the way students report thinking about tutor feedback, feedback from Glosser, and automated feedback in general. For the purposes of the analysis, student conceptions of feedback are divided into (1) cohesive conceptions: those that relate feedback to the development of learning through writing, and (2) fragmented conceptions: those indicating that feedback is little more than ideas for task completion or grammatical accuracy. Student perceptions of automated feedback are identified as positive or negative, depending on their orientation to whether or not they feel they could learn from feedback provided by an automated system. For each section of the analysis, representative quotes are used as samples of the discourse used by students in each category.

A. Conceptions of Feedback

The interview transcripts suggested qualitatively different conceptions of feedback. One group of responses tended to reveal the use of alternate perspectives from feedback to improve understanding. When asked about the purpose of feedback in writing, they responded:

> … it helps us get a different view of our project, of what we are doing, giving us an outsider's idea of things we can change, things we can improve on. (Student 12)

> It's shaping ideas but also you can test ideas out and they might get bounced back at you, like maybe this doesn't work, or this or that, and you can sound ideas out. That to me is the main purpose of the feedback. (Student 5)

> Generally feedback is to get views or thoughts that give more room for improvement, what I can do better, what might actually be bad so I can rethink something… rethink the points, what should we change if we change anything, sort of tie back into the purpose or the system of the PSD…I think feedback on both writing style and the content. (Student 6)

These quotes display an awareness of feedback as a way of getting different perspectives and improving understanding of the topic and of one's written communication. It is through the feedback that they are able to look at their text in a new light, to reconsider some of their ideas and how they are expressed.

In contrast, another group of students tended to conceive of feedback as predominately about task and document completion.

> That's so the student knows what is expected of that assignment. (Student 9)

> I think that's the purpose of it, to keep them on track. (Student 11)

> The feedback from the tutor was pretty much just what we should put for each part. (Student 4)

These types of responses focus on external aspects of the experience, such as completing the task as easily as possible. An absence of the value of feedback to bring a fresh perspective on the task was a noticeable aspect of the fragmented conceptions. This category of conception seemed to focus more on feedback as a regulatory mechanism, ensuring compliance.

B. Conceptions of Automated Feedback from Glosser

The second question focused on students' conceptions of automated feedback through Glosser. One group of students reported conceptions that indicated an awareness of how automated feedback could improve understanding of both academic style and the communicativeness of texts.

> Glosser is essentially a program that gives you an evaluation on your structure, the coherence and the overall readability of your actual text and it gives you pointers where you can edit and where you can change a certain piece of text…to

better enable us to carry across information between ourselves and the tutor when we write, to improve our writing skills obviously, for a better way of communicating through written language. (Student 20)

Basically I think it's for the flow of the writing. I think that's a good word for it, just linking points from point to point and not talking about or writing things that don't relate to the thing that you previously said, stuff like that…For helping with our e-business proposal, so we can use it to determine what we're talking about, and what the community is talking about and if we're doing it correctly. (Student 3)

These extracts suggest an awareness of how Glosser is able to improve the communicability of the texts, both in terms of the ideas being discussed and the way they are expressed. An awareness of the textual dimension of the feedback is prominent, but it does not dominate the contribution of the feedback to ideas, meaning making, and communicability, which is a combination of disciplinary ideas and their written expression.

In contrast, another group of students provided responses which focused on predominately structural aspects of their texts.

I mean like structure, the right structure for writing, it should be grammatically correct or it should be coherent or something like that. (Student 9)

I was looking at Glosser itself, like it's basically structured the whole essay—it's got something like—I guess it marks your comprehension in grammar, in a way. (Student 14)

I think the idea behind doing that was if Glosser said it was okay it was probably better than what we had initially. We're just trying to get marked before being marked, because we're aware that the markers would probably look at Glosser as well, so if you get Glosser to give an output you probably have a better chance of getting a better mark. (Student 1)

Conceptions classified as fragmented tended to conceive of Glosser as primarily being about checking surface features of the text for the purpose of getting more marks. These ideas of Glosser were similar to the concepts reported about feedback more generally. There is little awareness of the value of learning arising from the feedback; rather, the focus is on superficial aspects of the text and the experience.

Interestingly this version of Glosser did not provide any feedback at the surface level of writing. Spelling and grammar checkers were not included as forms of feedback, due to the poor learning gain results shown in studies discussed earlier. A few students (e.g., Student 14) who described automated feedback in these terms were perhaps highlighting their expectations on what Glosser would be about, and not on their actual experiences.

C. Perceptions of the Value of Automated Feedback in General

Automated feedback tools, particularly for experiences of writing, are still experimental and not well understood by researchers, teachers, or students. The perceptions that students have of automated feedback, and the potential value they can provide, were investigated to see if they were related to the student's experience of writing.

The perceptions that students reported could be divided into positive and negative responses to the potential of automated feedback. The interviews revealed that some students held a positive perception.

Yes, I think it's possible, yes...I could see how a machine could actually help. (Student 4)

A tool like Glosser could provide some feedback but it won't give you all the feedback you need to create a perfect document. (Student 2)

Another group of students tended to think automated feedback for writing was not possible. Mistrust of machines was a key theme that arose in these perceptions.

No, I don't trust software tools. It's something that I use—it might be useful for mathematical problems, something that is exact, but for writing, no. I don't trust software. (Student 9)

We don't really trust in words being provided by Glosser and that's one of the things that's been holding us back in using it. (Student 6)

D. Associations Between Conceptions and Achievement

The following tables present some quantitative analyses of the qualitative categories. Table 1 presents the frequencies of the classification of categories. Tables 2 and 3 present associations between the qualitatively different aspects of the conceptions, and Table 4 shows associations with academic achievement.

Table 1 can be read as three columns and three rows. Column 1 provides a description of the frequencies being presented in column 2. Column 3 indicates that there were some misconceptions about the purpose of feedback and the purpose of Glosser in the students' experience of writing.

Conceptions of feedback (Q1)	Frequency	Percentages
Cohesive	13	59%
Fragmented	9	41%
Conceptions of the Automated Feedback from Glosser (Q2)		
Cohesive	11	50%
Fragmented	11	50%
Perceptions of Automated Feedback in general (Q3)		
Positive	11	50%
Negative	11	50%

Table 1. Frequencies of students' conceptions of feedback in writing.

Conceptions of Feedback	Conceptions of the automated feedback from Glosser		
	Cohesive	Fragmented	Total
Cohesive	10	3	13
Fragmented	1	8	9
Total	11	11	22

$\chi^2(1,22) = 9.2$, $p = 0.004$ (Fisher Exact test)

Table 2. Conceptions of feedback and of automated feedback from Glosser.

Student conceptions of automated feedback from Glosser	Perceptions of Automated Feedback		
	Positive	Negative	Total
Cohesive	8	3	11
Fragmented	3	8	11
Total	11	11	22

$\chi^2(1,22) = 4.5$, $p = 0.043$ (Fisher Exact test)

Table 3. Conceptions of the automated feedback from Glosser and perceptions of automated feedback in general.

Conceptions of Feedback	Final writing assignment mark			
	N	Mean#	SD	Effect Size
Cohesive	12	18.0	3.4	
Fragmented	9	13.7	4.3	
t-Test		2.5*		1.26

$n = 21$, $^*p < 0.05$, #; mark out of 25

Table 4. Relationship between conceptions and achievement.

Table 2 summarizes the relationships between conceptions of feedback and conceptions of the automated feedback provided by Glosser.

Table 2 shows a strong and statistically significant relationship between the quality of conceptions of feedback in writing and the conceptions of the automated tool ($\chi^2(1,22) = 9.2$, $p = 0.004$). Students with cohesive/fragmented conceptions of feedback tended to also have cohesive/fragmented conceptions of automated feedback.

Table 3 shows a strong and statistically significant relationship between the quality of conceptions students had of the tool and their general perceptions regarding automated feedback. Cohesive conceptions of Glosser tended to be positively related to perceptions of the potential of automated systems to provide useful feedback on writing.

Table 4 presents the relationships between student conceptions of feedback and achievement, as measured by their grades for the writing activity. Students who reported a conception of feedback classified as cohesive tended to be rated more highly in their writing tasks. No significant relationship was found between conceptions of feedback from Glosser and achievement.

E. Limitations

This is an exploratory study, one designed to help identify emerging uses of automated feedback for students and teachers in writing tasks in order to help inform the users and the software developers in creating the automated system. It was conducted on a relatively small population sample and looked at the key conceptions being reported. Further research that includes approaches to using Glosser would be useful, particularly if it involves larger populations and a variety of specialized disciplines in engineering.

VIII. Discussion and Conclusions

This study was designed to investigate engineering students' experiences of feedback on their writing. To do so, 22 students from a third year software engineering course were interviewed about their experiences with feedback from both the tutor and from the automated system Glosser as they wrote an e-business project proposal. By understanding the variations in student conceptions of feedback, the authors provide evidence on how to better design feedback tools for student learning.

The three hypotheses posed by the study coalesce into an important insight for the design process. Our first hypothesis aimed to establish if there are any variations between students in their conceptions of feedback (both human and automated), and to characterize the structure of those variations. The results from the qualitative analysis suggest that there are indeed variations, and that these could be generally categorized as coherent and fragmented. These categories can be mapped to those found in other studies on the conceptions about writing (Hounsell, 1984). Following Marton and Booth's analysis (1997), cohesive views could be seen as leading to a *deep approach to learning* where the focus is on writing as a search for meaning, while fragmented conceptions lead to *surface approaches* that focus on "the discourse contained in the text" (Marton and Booth, 1997, p. 28).

This study also shows that students were likely to have the same category of conceptions for both traditional and automated feedback (either cohesive for both or fragmented for both).

The second hypothesis was that there would be variation in how students perceived the value of automated feedback. Students in the population sample demonstrated both positive and negative perceptions of *automated* feedback in equal numbers but these perceptions are not predictive of their conceptions of feedback in general.

These two results are of great importance for developers of automated feedback. Students view automated and human feedback as fulfilling the same needs (either cohesive or fragmented, Table 2), and they have similar grouping of expectations (as shown in Table 3). Deep learners (of whom we want more) see feedback as way of learning about the topic; therefore, on the one hand, the trigger questions and the document visualization and summarization features should explicitly focus on this aspect, emphasizing these values. On the other hand, discourse level features are the hardest to identify computationally, and will for some time be inaccurate. Perhaps the system should make explicit the inherent imprecision of the computer software so that its feedback can be taken as a reader's perspective on the document, and not as normative. Shallow learners (of whom we want less) focused on communication aspects of the writing, particularly the surface features of the composition.

Glosser (or any other automated feedback tool) could provide accurate feedback on some of these features (e.g., spelling), and the accuracy of such feedback may improve their perceptions of what computers can do. However this would also likely reinforce the misconception that this is the type of learning that they should aim for. Our results have shown that these fragmented conceptions are strongly related with low achievement.

The third hypothesis was that there would be significant associations between qualitatively different conceptions and achievement. The results described identified statistically significant associations between cohesive conceptions of feedback and better academic achievement. In contrast, no significant associations between perceptions of *automated* feedback and academic achievement were identified.

These outcomes are significant in themselves, as they offer a description of the way students report thinking about feedback in writing, and thus a framework and impetus for future studies designed around larger sample sizes that could lead towards results that are more statistically generalizable. These results and those from future studies can begin to inform new technology and pedagogical design strategies for the provision of feedback that take into account the importance and impact of student conceptions.

This study suggests that students would benefit from more explanation about the purpose of feedback, and, specifically, about the purpose of automated feedback. The relationship between the two sets of conceptions also suggests that improving student conceptions of automated feedback may be enough to help broaden student understanding more generally. Conversely, the results suggested that students who held an impoverished conception of feedback, tended to also have an impoverished conception of Glosser. This points to a software design technique for the structure of the Glosser system that could be augmented to include a section on orientating students to the role and value of feedback in their experience of writing at university and how automated feedback is meant to complement other forms of feedback.

This study also suggests that engineering faculty who wish to use tools such as Glosser (or even CPR in the U.S.) need to work carefully to frame and value the tools in ways that support deeper learning. This may require faculty development programs on how to provide feedback on academic writing, particularly when using automated tools like Glosser.

The outcomes of this study have pointed the way for future studies. For the purposes of triangulation, future research could investigate what students are able to tell us about their experience of writing along with more quantitative, survey-based methodologies, to complement their accounts of writing experiences at university. The introduction of innovative technologies such as Glosser needs to be carefully evaluated if we are to unpack the challenges involved in such experiences for students, teachers and software designers.

Acknowledgments

This research was supported under the Australian Research Council's Discovery Projects funding scheme (projects number DP0665064 and DP0986873). The project was approved by the Human Research Ethics Committee at the University of Sydney. The authors would like to thank the anonymous reviewers and the Editors for their extensive, useful, invaluable feedback to the different versions of this paper.

References

Andeweg, B.A., and J.C. De Jong. 1996. Helpviewer or textbook? The case of "Ganesh Helper." *Journal of Technical Writing and Communication* 26 (4): 419–34.

Beals, T.J. 1998. Between teachers and computers: Does text-checking software really improve student writing? *English Journal, National Council of Teachers of English* 87 (1): 67–72.

Biggs, J.B. 1999. *Teaching for quality learning at university: What the student does*. Buckingham, UK: Open University Press.

Bonanno, H., and J. Jones. 1997. Measuring the academic skills of university students—the MASUS procedure, a diagnostic assessment. Australia: L. A. C. P. University of Sydney.

Bonanno, Helen, and Janet Jones. 2000. *The MASUS procedure, measuring the academic skills of university students: a diagnostic assessment*. [Sydney]: University of Sydney, Student Services.

Calvo, Rafael A., Robert A. Ellis, Nicholas Carroll, and Lina Markauskaite. 2007. eLearning software development processes with the student experience of learning. In *The transformed university: The scholarship of teaching and learning in practice*, eds. A. Brew and J. Sachs. Australia: Sydney University Press

Connolly, P., and T. Vilardi. 1989. *Writing to learn mathematics and science*. New York: Teachers College Press.

DeNisi, Angelo S., and Avraham N. Kluger. 2000. Feedback effectiveness: Can 360-degree appraisals be improved? *Academy of Management Executive* 14 (1): 129–39.

Dong, A. 2004. Quantifying coherent thinking in design: A computational linguistics approach. In *Design computing and cognition* 2004, eds. J.S. Gero. Dordrecht, The Netherlands: Kluwer Academic Publishers.

Ellis, R.A. 2004. Student approaches to learning science through writing. *International Journal of Science Education* 26 (15): 1835–54.

Ellis, Robert, and Rafael A. Calvo. 2006. Discontinuities in university student experiences of learning through discussions. *British Journal of Educational Technology* 37 (1): 55–68.

Ericsson, P.F., and R.H. Haswell, eds. 2006. *Machine scoring of student essays: Truth and consequences*. Logan, Utah: Utah State University Press.

Flower, L. 1994. *The construction of negotiated meaning: A social cognitive theory of writing*: Carbondale, IL: Southern Illinois Univ Press.

Foltz, Peter W., Sara Gilliam, and Scott A. Kendall. 2000. Supporting content-based feedback in on-line writing evaluation with LSA. *Interactive Learning Environments* 8 (17): 111–27.

Foltz, P.W., W. Kintsch, and T.K. Landauer. 1998. The measurement of textual coherence with latent semantic analysis. *Discourse Processes* 25 (2–3): 285–307.

Gouli, Evangelia, Agoritsa Gogoulou, Kyparisia A. Papanikolaou, and Maria Grigoriadou. 2005. An adaptive feedback framework to support reflection, guiding and tutoring. In *Advances in web-based education: Personalized learning environments*, ed. G.M.a.S. Chen. Hershey, PA: IDEA Group Publishing

Haswell, Richard. 2006. The complexities of responding to student writing; or, looking for shortcuts via the road of excess. *Across the Disciplines*. http://wac.colostate.edu/atd/articles/haswell2006.cfm (last accessed, August 2010).

Hayes, J.R., D.M. Bajzek, J. Brooks, B. Reyes, N. Hallinen, and E.R. Steinberg. 2007. Developing an online writing tutor to improve technical-writing skills in engineering and science students. *Studies in Writing* 21:107–124.

Herrington, A. J. 1981. Writing to learn: Writing across the disciplines. *College English* 43 (April): 379–87.

Hounsell, D. 1987. Essay writing and the quality of feedback. In *Student learning: Research in education and cognitive psychology*, ed. J.T. Richardson, M.W. Eysenck, and D.W. Piper. Milton Keynes, UK: Society for Research into Higher Education and Open University Press.

Hounsell, D. 1984. Essay Planning and Essay Writing. *Higher Education & Research* 3 (1): 13–31.

Hounsell, D. 1997. Contrasting conceptions of essay writing. In *The experience of learning: Implications for teaching and studying in higher education*, eds. F. Marton, D. Hounsell, and N.J. Entwistle. Edinburgh: Scottish Academic Press.

Hyland, K., and F. Hyland. 2006. Feedback on second language students' writing. *Language Teaching* 39 (2): 83–101.

Kakkonen, T., N. Myller, and E. Sutinen. 2004. Semiautomatic evaluation features in computer-assisted essay assessment. In *Proceedings of the Computers and Advanced Technology in Education Annual Conference*. Kauai, Hawaii.

Keh, Claudia. 1990. Feedback in the writing process: a model and methods for implementation. *English Language Teaching Journal* 44 (4): 294–304.

Keith, Z.K. 2003. Validity of automated essay scoring systems. In *Automated essay scoring: A cross-disciplinary perspective*, eds. J.M.D. Shermis and J. Burstein. Hillsdale, NJ: Lawrence Erlbaum Associates.

Kreth, M. L. 2000. A survey of the co-op writing experiences of recent engineering graduates. *IEEE Transactions on Professional Communication* 43 (2): 137–52.

Landauer, Thomas K. 2003. *Automatic essay assessment. Assessment in education: Principles, policy & practice* 10 (3): 295–308.

Landauer, T.K., D.S. McNamara, S. Dennis, and W. Kintsch, eds. 2007. *Handbook of latent semantic analysis*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Lavelle, E. 1993. Development and validation of an inventory to assess processes in college composition. *British Journal of Educational Psychology* 63 (3): 489–99.

Levin, Tamar, and Tili Wagner. 2006. In their own words: Understanding student conceptions of writing through their spontaneous metaphors in the science classroom. *Instructional Science* 34 (3): 227–78.

Leydens, J.A. 2008. Novice and insider perspectives on academic and workplace writing: Toward a continuum of rhetorical awareness. *IEEE Transactions on Professional Communication* 51 (3): 242–63.

Liu, Meihua, and George Braine. 2005. Cohesive features in argumentative writing produced by Chinese undergraduates. *System* 33 (4): 623–36.

Lizzio, A., K. Wilson, and R Simons. 2002. University students' perceptions of the learning environment and academic outcomes. *Sudies in Higher Education* 27 (1): 27–52.

Marton, F., and S. Booth. 1997. *Learning and awareness*. Hillsdale, NJ: Lawrence Erlbaum and Associates.

McNamara, D.S., Scott A. Crossley, and Philip M. McCarthy. 2010. Linguistic features of writing quality. *Written Communication* 27 (1): 57–86.

Narciss, S., and K. Huth. 2004. How to design informative tutoring feedback for multimedia learning. In *Instructional design for multimedia learning*, eds. H.M. Niegemann, D. Leutner and R. Brunken. New York: Waxmann.

O'Rourke, Stephen T., and Rafael A. Calvo. 2009. Analysing semantic flow in academic writing. In *Artificial intelligence in education*. Brighton, UK: IOS Press.

Prosser, Michael, and Carolyn Webb. 1994. Relating the process of undergraduate essay writing to the finished product. *Studies in Higher Education* 19 (2): 125–38.

Prosser, M., and K. Trigwell. 1999. *Understanding learning and teaching: The experience in higher education*. Buckingham, UK: Society for Research into Higher Education & Open University Press. .

Ramsden, P. 2002. *Learning to teach in higher education*. 2nd. ed. London: Routledge.

Shermis, M.D., and J. Burstein, eds. 2003. *Automated essay scoring: A cross-disciplinary perspective*. Hillsdale, NJ: Lawrence Erlbaum Associates.

Shute, Valerie J. 2008. Focus on formative feedback. *Review of Educational Research* 78 (1): 153–89.

Silyn-Roberts, H. 1998. Using engineers' characteristics to improve report writing instruction. *Journal of Professional Issues in Engineering Education and Practice* 124 (1): 12–16.

Tenopir, Carol, and Donald Ward King. 2004. *Communication patterns of engineers*. Hoboken, NJ: Wiley-IEEE Press.

Thiesmeyer, E.C., and J.E. Thiesmeyer. 1990. *Editor*. New York: Modern Language Association.

Villalon, Jorge, and Rafael A. Calvo. 2009. Concept extraction from student essays, towards concept map mining. In *9th IEEE International Conference on Advanced Learning Technologies (ICALT)*. Riga, Latvia.

Wade-Stein, D., and E. Kintsch. 2004. Summary Street: Interactive computer support for writing. *Cognition and Instruction* 22 (3): 333–62.

Weaver, Melanie R. 2006. Do students value feedback? Student perceptions of tutors' written responses. *Assessment & Evaluation in Higher Education* 31 (3): 379–94.

Wiemer-Hastings, P., and A.C. Graesser. 2000. Select-a-Kibitzer: A computer tool that gives meaningful feedback on student compositions. *Interactive Learning Environments* 8 (2): 149–69.

Witte, Stephen P., and Lester Faigley. 1981. Coherence, cohesion, and writing quality. *College Composition and Communication* 32 (2): 189–204.

Yalvac, B., H.D. Smith, J.B. Troy, and P. Hirsch. 2007. Promoting advanced writing skills in an upper-level engineering class. *Journal of Engineering Education* 96 (2): 117–24.

Young, P. 2000. 'I might as well give up': Self-esteem and mature students' feelings about feedback on assignments. *Journal of Further and Higher Education* 24 (3): 409–18.

AUTHORS' BIOGRAPHIES

Rafael A. Calvo is a senior lecturer in the School of Electrical and Information Engineering at the University of Sydney specializing in learning systems and director of the Learning and Affect Technologies Engineering (Latte) group. He has a Ph.D. in Artificial Intelligence applied to automatic document classification and has worked at Carnegie Mellon University, and Universidad Nacional de Rosario, and as a consultant for projects worldwide. Rafael is author of numerous publications in the areas of Web engineering and learning systems, recipient of four teaching awards, Senior Member of IEEE and Associate Editor of *IEEE Transactions on Affective Computing*.

Address: School of Electrical and Information Engineering, Bldg J03, the University of Sydney, NSW 2006, Australia; telephone: (+61) 29351.8171; e-mail: rafael.calvo@sydney.edu.au.

Robert A. Ellis, Ph.D., is associate professor and director of eLearning at the University of Sydney Australia. He has operational oversight of the University's strategy for eLearning supporting over 46,000 students and 3,000 academic and general staff in 16 faculties.

This role includes policy writing and advice, strategic planning and management, and benchmarking activities with international universities in the United Kingdom and Australia. To support this role, Robert is the current recipient of three large Australian Research Council Grants investigating blended learning in higher education with Professor Peter Goodyear at the University of Sydney and Professor Michael Prosser of the University of Hong Kong.

Address: The University of Sydney, Office of the Deputy Vice-Chancellor (Education) Sydney, NSW 2006, Australia; e-mail: robert.ellis@sydney.edu.au.