Open Research Online

The Open University's repository of research publications and other research outputs

What types of essay feedback influence implementation: Structure alone or structure and content?

Conference or Workshop Item

oro.open.ac.uk

What types of essay feedback influence implementation:

Structure alone or structure and content?

Denise Whitelock[a],*, Alison Twiner[a], John T. E. Richardson[a], Debora Field[b], Stephen Pulman[b]

[a] *Institute of Educational Technology, The Open University, Walton Hall, Milton Keynes MK7 6AA, UK*

[b] *Department of Computer Science, University of Oxford, Parks Road, Oxford OX1 3QD, UK*

---

* Corresponding author. Tel: +44 1908 653777.

E-mail address: Denise.Whitelock@open.ac.uk

ABSTRACT

Students approach educational courses with varying levels of experience and understanding, and so need appropriate support to inform them of expectations and to guide their learning efforts. Feedback is critical in this process, so that learners can gauge how their performance aligns with expectations, and how they can improve their efforts and attainments. This study focused on the effects of providing different types of feedback on participants' written essays, as well as on participants' motivations for learning using measures of motivation and self-efficacy. In terms of research questions, it was important to ascertain whether participants performed differently in subsequent essays after receiving feedback on structure alone or on structure and content; whether their self-reported levels of motivation and attitudes to learning were related to essay performance; and whether the difference in type of feedback affected their self-reported levels of motivation and attitudes to learning. Findings revealed no significant difference in marks between those receiving feedback on structure alone and those receiving feedback on structure and content, which is surprising and deserves further exploration. Even so, using feedback to highlight certain structural elements of essay writing can have a lasting positive impact on participants' future essay performance.

Keywords: Essay structure, Feedback, Motivation, Self-efficacy, Self-reports

1. Introduction

People come to educational courses and learning tasks with varying levels of experience and understanding. They therefore need support in their courses to inform them of expectations and to guide their learning efforts. They also need feedback on their performance, so that they can gauge how their performance aligns with expectations, and how they can improve their efforts and attainments. This is particularly important when supporting distance and online learners, who may have little or no face-to-face contact with instructors or peers. This paper focuses on the effects of providing different types of feedback electronically on participants' written essays, as well as participants' motivations for learning using measures of motivation and self-efficacy. In terms of the research questions, it was important to ascertain whether participants performed differently in subsequent essays after receiving feedback on structure alone, or structure and content. The research team also set out to investigate whether this difference in type of feedback affected participants' self-reported levels on a standardised questionnaire regarding motivation and attitudes to learning, and whether their self-reported levels on these measures were related to their essay performance.

Such a study regarding feedback and performance linked to motivations for learning and self-efficacy has considerable implications for supporting students to improve their work, and also supporting students to *believe* that they can improve their academic work—no small feat for learners who may often feel isolated and stretched trying to squeeze study around other commitments and demands on their time. This work is intended to offer some illumination on the kinds of feedback given to written academic essays that students pay attention to, find useful and ultimately implement in future writing efforts, and how this interacts with their motivation and self-efficacy. Such findings could be of substantial benefit to students and instructors alike. This paper begins by reviewing how the literature portrays

the key issues of feedback, including aspects related to motivation and self-efficacy.

1.1. Feedback

Feedback has been a popular topic of educational research for some decades, and it is largely agreed that feedback is central to learning (Black & Wiliam, 1998). Some researchers have however argued that the positive effects of feedback are not guaranteed (Kluger & DeNisi, 1996), and so it is important that research continues to investigate how feedback can be offered in ways that support improvements in students' learning (understanding of topics) as well as performance (marks achieved). Chickering and Gamson (1987) listed "gives prompt feedback" and "encourages active learning" as two of their seven principles of good practice for undergraduate education. Therefore by this argument facilitating students to take ownership of and reflect on their work, through provision of feedback at a point where they can build on it in subsequent tasks, should have the most impact on students' understanding of the requirements of academic writing.

Nelson and Schunn (2009) carried out a correlational analysis of 1,073 segments of peer review feedback that had been given to undergraduate students on writing tasks. In terms of making use of feedback, Nelson and Schunn proposed that understanding problems raised within feedback about one's own work was a critical factor in implementing suggestions. They continued to explore this potential, in stating that understanding was more likely where those giving feedback provided solutions, specified where identified problems occurred, and summarised performance. Nelson and Schunn identified feedback as involving motivation, reinforcement, and information. They addressed five features of feedback: summarisation; specificity; explanations; scope (i.e., local versus global); and affective language (praise, inflammatory, and mitigating language). The first four are cognitive features, whereas the

latter is an affective feature. It is these five features of feedback that were drawn on in the present study in determining the types of feedback to offer participants on their draft essays.

Nelson and Schunn proposed that there are "mediators" that operate between the provision of feedback features and implementation of suggestions. The authors addressed the mediators of "understanding feedback" and "agreement with feedback". They suggested that cognitive feedback features were most likely to influence understanding, but that affective features were more likely to influence agreement and hence implementation. Nelson and Schunn's results therefore showed how understanding feedback is critical to implementing suggestions from feedback. Thus, it is important in course design that consideration is given to how to increase the likelihood that feedback is understood, if students are to make use of it in current and future work—to learn from it (and improve performance) by understanding it, rather than just improving one-off performance by blind implementation.

Lee (2014) argued that feedback research needed to address real-life writing tasks and contexts in order to have pedagogical value. Hawe and Dixon (2014) adopted a case study approach to investigate "best practice" regarding feedback among teachers in New Zealand. They concluded that students needed to be given authentic opportunities to assess and edit their own work during its production to gain a better understanding of what constitutes a "good" piece of writing. This would increase the likelihood of them being able implement suggestions made in feedback and to learn from the feedback given both to themselves and to other students. Baker (2014) drew attention to the distinction between "feedback" (to be used to inform future writing) and ("grading" as a final evaluative mark allocated to work already done). Even so, she reported that in practice the terms were often used interchangeably. Baker had adopted an ethnographic approach at one U.S. university and provided an in-depth analysis of teachers' views and approaches to giving feedback. However, she said little about what was actually covered in the feedback (such as content, structure, or grammar).

The latter issue has been discussed specifically with regard to the learning of English as a second language (ESL). In secondary education in Hong Kong, Lee (2005) found that, in spite of receiving feedback, students were concerned that they would make the same errors again, indicating that in this context the main focus of feedback was on error correction rather than on content or organisation. Lee suggested that students might not be so error-focused if feedback were focused more on content and organisation as opposed to errors. However, in subsequent research in a similar educational context, Lee (2011) identified a persisting focus on error correction, resulting in an approach that emphasised "testing" rather than "teaching". Hyland (2001) had similarly observed the predominance in ESL practices of feedback as "evaluation" rather than "education". Ferris, Brown, Liu, and Stine (2011) argued that it was important to note the similarities between the feedback given to students writing in English as their first language and the feedback given to those writing in English as a second language. Such considerations led Hyland (2010) to argue that we must address students' response to feedback as well as the feedback itself, which is very much the focus of the present study.

Relevant to the idea that feedback should be aimed at "education" rather than solely "evaluation", S.-L. Wang and Wu (2008) considered feedback in terms of knowledge of results, knowledge of correct response, and elaborative feedback. They commented:

Research suggests that feedback is one of the most significant sources of information helping individual students to correct misconceptions, reconstruct knowledge, support metacognitive processes, improve academic achievement, and enhance motivation (Clark & Dwyer, 1998; Foote, 1999; Warden, 2000; Zimmerman & Martinez-Pons, 1992). (p. 1591)

From this it is apparent that feedback can play a key role in enhancing students' motivation to

study and to improve their work. Therefore it is appropriate now to turn attention more specifically to issues of motivation.

1.2. Motivation

Schraw (2010) proposed that self-regulated learning (SRL) consisted of knowledge, metacognition, and motivation. Banyard, Underwood, and Twiner (2006) also outlined that, whilst SRL refers to learners' goals and knowledge of strategies to work towards those goals, it is highly susceptible to external influences. So how can external influences be presented to facilitate students' SRL efforts? In the current study, the aim was to guide the external influences perceived by participants by giving feedback that would support them in considering the requirements of academic writing in the context of their own work and suggest where there might be areas for development. The onus was therefore on the participants to incorporate this feedback within their subsequent essay writing. Related to motivation is the concept of self-efficacy, which will now be addressed.

1.3. Self-efficacy

Self-efficacy refers to students' judgments of how well they can do a task (Bandura, 1993). Much research on self-efficacy refers back to Bandura's original definition and work (e.g., Alkharusi, Aldhafri, & Alnabhani, 2013; Tang & Tseng, 2013; van Dinther, Dochy, Segers, & Braeken, 2014). García Duncan and McKeachie (2005) offered a definition of self-efficacy that incorporated "expectancy for success (which is specific to task performance) and judgments of one's ability to accomplish a task and confidence in one's skills to perform a task" (p. 119). In a survey of the use of electronic information resources among university

students in Nigeria, Tella, Tella, Ayeni, and Omoba (2007) also defined self-efficacy as whether an individual feels capable of doing a particular task or succeeding in a current situation. In other words, self-efficacy is a matter of the perception of one's own ability in a perceived context, rather than a measure of ability itself. Tella et al. concluded:

> The results indicate that self-efficacy and the use of electronic information jointly predict and contribute to academic performance; that respondents with high self-efficacy make better use of electronic information and have better academic performance; that a correlation exists among self-efficacy, use of electronic information and academic performance. (Abstract)

Thus, raising students' self-efficacy is an important aim. Tang and Tseng (2013) carried out a survey of online distance learners and reported:

> Distance learners who have higher self-efficacy for information seeking and proficiency in information manipulation exhibited higher self-efficacy for online learning. Moreover, students with high self-efficacy demonstrated superior knowledge of digital resources selection. Students who have low self-efficacy with regard to information seeking were more likely to express interest in learning how to use the library resources, although learning techniques for database searching was the exception. (p. 517)

Such findings have strong implications for the support that can be offered to enable students to increase their self-efficacy, by offering opportunities for them to review, reflect on and improve their work and so build on progressive successes, rather than receiving only final-

grade evaluations with little scope to learn from mistakes. This is particularly important in distance learning contexts, where there is minimal or no opportunity for face-to-face interaction between tutors and students or among students themselves.

Shrestha and Coffin (2012) reported on a study to support distance learning students' academic writing exercises. Instructors on the course provided support and feedback to students via email, in response to successive drafts of assignments. However, the study and the detailed feedback interactions described by Shrestha and Coffin only involved two students. It is likely, therefore, that this was a very valuable provision, but one that would be difficult to resource on a large scale.

In stark contrast in terms of research design, C.-H. Wang, Shannon, and Ross (2013) conducted structural equation modelling of survey data from 256 graduate and undergraduate students taking online courses with one U.S. university. They reported a series of effects:

> Students with previous online learning experiences tended to have more effective learning strategies when taking online courses, and hence, had higher levels of motivation in their online courses. In addition, when students had higher levels of motivation in their online courses, their levels of technology self-efficacy and course satisfaction increased. Finally, students with higher levels of technology self-efficacy and course satisfaction also earned better final grades. (p. 302)

This again shows a finding of a cumulative effect for online distance learners, in terms of students using effective learning strategies, having high levels of motivation, self-efficacy and course satisfaction, and achieving better final grades. Much of this therefore comes down to the individual learner's stance toward their study. However, much can also be done by instructors and course designers to support students in selecting appropriate learning

strategies and understanding task requirements and in enabling students to feel that they can do well. It is this aim for feedback that is addressed in the study reported here.

1.4. Research questions and analyses

In the current study, participants completed adapted versions of the Motivated Strategies for Learning Questionnaire (MSLQ: Pintrich, Smith, Garcia, & McKeachie, 1991, 1993) before and after writing two essays. They were given qualitative feedback (based on the marks awarded but not including the marks awarded) after submitting each essay. One group of participants received feedback only on the structure of their essays, while the second group received feedback on both the structure and the content of their essays.

The first research question was:

- Does the kind of feedback that the participants receive for the first essay influence the marks that they obtain on the second essay?

The fact that the marks awarded for the second essay might be different from the marks awarded for the first essay is in itself not surprising, because the essays were written on different topics (although the marking criteria were not topic-specific). This question asks whether the difference was greater for the participants who received feedback on the structure and the content of their essays than for the participants who only received feedback on the structure of their essays. The latter would entail an interaction between the effect of groups and the effect of essays (first versus second) on the marks awarded.

The second research question was:

- Do the scores the participants obtain on the pre-test questionnaire predict the marks that they receive for their essays?

An additional question here is:

- Is the regression between the questionnaire scores and the essay marks the same regardless of whether the participants receive feedback only on the structure of their essays or feedback on both the structure and content of their essays?

Preliminary tests were carried out to check for interactions between the pre-test questionnaire scores and the effect of groups on the marks awarded to test for the homogeneity of the regression between the pre-test questionnaire scores and the essay marks.

The third research question was:

- Do the scores that participants obtain on the post-test questionnaire differ, depending on the kind of feedback that they receive for their essays?

The final research question was:

- Do the marks that participants receive for their essays predict the scores that they obtain on the post-test questionnaire?

A complication is that, if the marks depended on their scores on the pre-test questionnaire, the marks might simply serve as a mediator between the pre-test scores and the post-test

scores. This point can be addressed by asking whether the marks awarded for the essays predicted their post-test scores when the effects of the pre-test scores on the latter had been statistically controlled. Once again, an additional question is:

- Is the regression between the marks awarded for the essays and the scores on the post-test questionnaire the same regardless of whether the participants received feedback only on the structure of their essays or feedback on both the structure and content?

A preliminary test was carried out to check for an interaction between the essay marks and the effect of groups on the post-test questionnaire scores to test for the homogeneity of the regression between the essay marks and the post-test questionnaire scores. Similar tests were carried out to check for interactions between the pre-test questionnaire scores and the effect of groups on the post-test scores to test for the homogeneity of the regression between the pre-test scores and the post-test scores.

2. Method

2.1. Participants

Ninety-one participants were recruited from a subject panel maintained by colleagues in the Department of Psychology consisting of people who were interested in participating in online psychology experiments. Some of them were current or former students of the University, but others were just members of the public with an interest in psychological research.

2.2. Materials

The MSLQ consists of 81 statements in 15 scales that measure motivational variables and learning strategies on particular courses. Respondents are asked to rate each statement on a 7-point scale from 1 for "not at all true of me" to 7 for "very true of me". Pintrich et al. (1991, 3) commented that the 15 scales were "designed to be modular" so that they "can be used together or singly". The present investigation used items drawn from the first six scales, which are concerned with respondents' motives and attitudes. Pintrich et al. (1993) explained these scales as follows:

> The motivational scales are based on a general social-cognitive model of motivation that proposes three general motivational constructs . . . : (1) expectancy, (2) value, and (3) affect. Expectancy components refer to students' beliefs that they can accomplish a task. Two expectancy-related subscales were constructed to assess students' (a) perceptions of self-efficacy and (b) control beliefs for learning. Value components focus on the reasons why students engage in an academic task. Three subscales are included in the MSLQ to measure value beliefs: (1) intrinsic goal orientation (a focus on learning and mastery), (2) extrinsic goal orientation (a focus on marks and approval from others), and (3) task value beliefs (judgments of how interesting, useful, and important the course content is to the student). The third general motivational construct is affect, and has been operationalised in terms of responses to the test anxiety scale, which taps into students' worry and concern over taking exams. (p. 802)

The MSLQ was intended for use with students taking formal courses. For the present

investigation, references to "this class" or "this course" were reworded to refer to "this exercise". One item from the Intrinsic Goal Orientation scale and four items from the Test Anxiety scale could not be adapted, leaving 26 items. These were phrased in the present or future tense in the pre-test questionnaire and in the present or past tense for the post-test questionnaire. Sample items are shown in Table 1. Both surveys were hosted online on secure websites.

(Insert Table 1 about here)

2.3. Procedure

Communication with the participants was solely online. They were assigned alternately to two groups. (Those in Group 1 were provided with feedback only on the structure of their essays; those in Group 2 were provided with feedback on both the structure and content of their essays.) Each participant was asked to complete the pre-test version of the MSLQ online, for which they were allowed 2 weeks. They were then asked to "write an essay on human perception of risk" of between 500 and 1,000 words, for which they were also allowed 2 weeks.

Two of the authors who were academic staff with considerable experience in teaching and assessment marked the submitted essays using an agreed marking scheme and without reference to the groups to which participants had been assigned. The marking scheme is shown in Table 2. If the difference between the total marks awarded was 20 percentage points or less, essays were assigned the average of the two markers' marks. Discrepancies of more than 20 percentage points were resolved by discussion between the markers.

(Insert Table 2 about here)

The participants were provided with qualitative feedback on their first essays based on

the marks awarded (but they were not given the marks themselves). Those in Group 1 received feedback based on the marks awarded against Criteria 1, 2, 5, 6, and 7. Those in Group 2 received feedback based on the marks awarded against all 10 criteria. They were then asked to "write an essay on memory problems in old age" of between 500 and 1,000 words, for which they were again allowed 2 weeks. Finally, they were provided with feedback on their second essays and asked to complete the post-test version of the MSLQ online. Participants who completed both questionnaires and submitted both essays were rewarded with an honorarium of 40 pounds sterling in Amazon vouchers.

3. Results

Of the 91 participants who were invited to complete the pre-test questionnaire, 76 responded, but two only provided incomplete data. Of these 76 participants, 42 submitted Essay 1, of whom 38 submitted Essay 2. Of the latter 38 participants, all completed the post-test questionnaire. They consisted of eight men and 30 women; they were aged between 23 and 65 (mean age = 42.3 years); and 19 participants were in each group.

3.1. Marks awarded for essays

The correlation coefficients between the marks initially awarded by the two markers were .85 for Essay 1 and .82 for Essay 2. The discrepancy between the two markers was more than 20 percentage points for just one essay, and this discrepancy was resolved by discussion between the markers. The mean final mark for Essay 1 was 56.8 (SD = 14.7), and the mean final mark for Essay 2 was 63.3 (SD = 12.4).

A mixed-design analysis of variance was carried out on the average mark that was

awarded to the participants who submitted two essays. This employed the within-subjects variables of essays (Essay 1 versus Essay 2) and marking criteria (1–10) and the between-subjects variables of group (Group 1 versus Group 2). Post hoc tests were carried out to identify the marking criteria on which any significant differences in marks had arisen.

There was no significant difference in the overall marks awarded to participants in the two groups, $F(1, 36) = 2.83$, $p = .10$, partial $\eta^2 = .07$. The mean mark awarded for Essay 2 (63.1) was significantly higher than the mean mark awarded for Essay 1 (57.3), $F(1, 36) = 4.89$, $p = .03$, partial $\eta^2 = .12$. However, the interaction between the effects of groups and essays was not significant, $F(1, 36) = .40$, $p = .53$, partial $\eta^2 = .01$, implying that the difference between the marks awarded for the two essays was similar for participants in the two groups.

The main effect of criteria was statistically significant, $F(9, 324) = 39.70$, $p < .001$, partial $\eta^2 = .52$, which is unsurprising since different numbers of marks were awarded against the 10 criteria. There was a significant interaction between the effect of groups and the effect of criteria, $F(9, 324) = 2.50$, $p = .01$, partial $\eta^2 = .07$, implying that the two groups showed a different pattern of marks across the different criteria. There was also a significant interaction between the effect of essays and the effect of criteria, $F(9, 324) = 6.69$, $p < .001$, partial $\eta^2 = .16$, implying that the two essays showed a different pattern of marks across the different criteria. Finally, there was a significant three-way interaction between the effects of groups, essays and criteria, $F(9, 324) = 2.05$, $p = .03$, partial $\eta^2 = .05$.

Post hoc tests showed that the difference between the marks awarded to the two groups was only significant on Criterion 8 (Definition), where Group 2 obtained a higher mean mark (5.70) than did Group 1 (3.21), $F(1, 36) = 7.24$, $p = .01$, partial $\eta^2 = .17$. However, this was qualified by a significant interaction between the effect of groups and the

effect of essays, $F(1, 36) = 5.12$, $p = .03$, partial $\eta^2 = .13$. For the participants in Group 1, the

mean mark awarded on Criterion 8 was higher on Essay 2 (4.03) than on Essay 1 (2.40).

However, for the participants in Group 2, the mean mark awarded on Criterion 8 was higher

on Essay 1 (6.50) than on Essay 2 (4.90). In short, there was no evidence that providing

feedback on both the structure and content of an essay led to higher marks on a subsequent

essay than providing feedback on the first essay's structure alone.

Finally, the difference between the marks awarded to the two essays was significant

on Criterion 4, $F(1, 36) = 12.86$, $p = .001$, partial $\eta^2 = .26$, on Criterion 7, $F(1, 36) = 19.82$, p

$< .001$, partial $\eta^2 = .36$, and on Criterion 9, $F(1, 36) = 5.94$, $p = .02$, partial $\eta^2 = .14$. The

mean mark awarded on Criterion 4 (Evidence) was higher for Essay 2 (11.12) than for Essay

1 (8.97). The mean mark awarded on Criterion 7 (References) was also higher for Essay 2

(7.22) than for Essay 1 (4.01). However, the mean mark awarded on Criterion 9 (Written

presentation) was higher for Essay 1 (7.93) than for Essay 2 (7.33).

3.2. Questionnaire scores

The participants were assigned scores on the pre-test and post-test questionnaires by

calculating the mean response to the constituent items in each scale. Table 3 shows the means

and standard deviations of these scores, together with the relevant values of Cronbach's

(1951) coefficient alpha as an estimate of reliability. The latter were broadly satisfactory on

conventional research-based criteria (see Robinson, Shaver, & Wrightsman, 1991).

(Insert Table 3 about here)

The correlation coefficients between corresponding scales in the pre-test

questionnaire and the post-test questionnaire were highly significant and ranged between

+.46 and +.74, except in the case of Extrinsic Goal Orientation, $r = +.31$, $p = .06$.

A doubly multivariate analysis of variance was carried out on the scores that the participants obtained on the pre-test and post-test questionnaires. This used the between-subjects variable of groups and the within-subject comparison between the pre-test and post-test scores. This analysis found that on Self-Efficacy the post-test scores were significantly higher than the pre-test scores, $F(1, 36) = 4.19$, $p = .04$, partial $\eta^2 = .10$. However, there were no other significant differences between the pre-test scores and the post-test scores, no significant differences between the scores obtained by the two groups, and no significant interactions between these two effects. Thus, the two groups were similar in terms of their MSLQ scores both at the beginning and at the end of the study, and the increase in scores on Self-Efficacy occurred regardless of the kind of feedback they had received.

3.3. Using the pre-test scores to predict the essay marks

A univariate analysis of variance was carried out using the between-subjects variable of group, the scores on the pre-test questionnaire as covariates, and the average essay mark as the dependent variable. A preliminary analysis included the interactions between the scores on the pre-test questionnaire and the effect of group to check for homogeneity of regression. This found no significant interactions between the scores on the pre-test questionnaire and the effect of groups on the average essay mark, $F(1, 24) \leq 1.47$, $p \geq .24$, partial $\eta^2 \leq .06$, implying homogeneity of regression between the two groups. This analysis also found no significant difference between the two groups in terms of their average essay mark when their pre-test scores were taken into account, $F(1, 24) = .84$, $p = .37$, partial $\eta^2 = .03$. Accordingly, the difference between the two groups was ignored in the main analysis.

This used the scores on the pre-test questionnaire as covariates and the average essay mark as the dependent variable. The average essay mark was significantly predicted by the scores on the Task Value scale, $B = +6.93$, $F(1, 31) = 5.61$, $p = .02$, partial $\eta^2 = .15$, and by the scores on the Control of Learning Beliefs scale, $B = +4.37$, $F(1, 31) = 4.21$, $p = .04$, partial $\eta^2 = .12$, but not by the scores on any of the other four scales. Thus, those participants who produced higher scores on Task Value and Control of Learning Beliefs in the pre-test questionnaire tended to obtain higher marks for their essays.

3.4. Using the pre-test scores and the essay marks to predict the post-test scores

A multivariate analysis of variance was carried out using the between-subjects variable of group, the scores on the pre-test questionnaire, and the average essay mark as covariates and the scores on the post-test questionnaire as dependent variables. A preliminary analysis included the interactions between the scores on the pre-test questionnaire and the effect of group and the interaction between the average essay mark and the effect of group to check for homogeneity of regression. This found that there was a significant interaction between the scores on Test Anxiety on the pre-test questionnaire and the effect of groups on the scores on the post-test questionnaire, $F(6, 17) = 6.60$, $p = .001$, partial $\eta^2 = .70$. In particular, this interaction was significant for the scores on Control of Learning Beliefs in the post-test questionnaire, $F(1, 22) = 26.78$, $p < .001$, partial $\eta^2 = .55$, and the scores on Self-Efficacy on the post-test questionnaire, $F(1, 22) = 5.00$, $p = .04$, partial $\eta^2 = .19$, but not on the other four scales of the post-test questionnaire. Separate analyses carried out on the two groups showed that the pre-test scores on Test Anxiety were negatively correlated with the post-test scores on Control of Learning Beliefs in Group 1, $B = -.36$, $F(1, 12) = 11.73$, p

$= .005$, partial $\eta^2 = .49$, but were positively correlated with the post-test scores on Control of Learning Beliefs in Group 2, $B = +.65$, $F(1, 12) = 16.93$, $p = .001$, partial $\eta^2 = .59$. A similar pattern was evident in the post-test scores on Self-Efficacy, but neither of the groups demonstrated a significant regression coefficient in this case. None of the other interaction terms was significant in the preliminary analysis, implying homogeneity of regression between the two groups in other respects, including the regression between the average essay mark and the post-test scores. The latter interaction terms were dropped from the main analysis.

This used the between-subjects variable of group, the scores on the pre-test questionnaire and the average essay mark as covariates and the scores on the post-test questionnaire as dependent variables. The statistical model included the interaction between the scores on Test Anxiety on the pre-test questionnaire and the effect of groups. This remained significant for the scores on Control of Learning Beliefs in the post-test questionnaire, $F(1, 28) = 24.38$, $p < .001$, partial $\eta^2 = .47$. The mean score obtained by Group 1 on this scale (5.30) was significantly higher than the mean scores obtained by Group 2 (5.03), $F(1, 28) = 24.53$, $p < .001$, partial $\eta^2 = .47$. Finally, there was a positive relationship between the participants' average essay mark and their scores on Extrinsic Goal Orientation in the post-test questionnaire, $B = +.06$, $F(1, 28) = 6.93$, $p = .01$, partial $\eta^2 = .20$. In other words, even when the participants' scores on the pre-test questionnaire had been statistically controlled, those who obtained higher marks for their essays subsequently tended to produce higher scores on Extrinsic Goal Orientation in the post-test questionnaire.

4. Discussion

This paper has reported on a study addressing the effects of offering feedback regarding structure or regarding structure and content on written academic essays. The study also focused on participants' reported levels of self-efficacy and motivation for such learning tasks, using an adapted version of the MSLQ. Our intention was that feedback given on a first essay would be implemented and supportive in writing a second essay. We also hypothesised that there would be a difference in marks related to the type of feedback participants received—on structure or on structure and content—as well as a relationship between marks received and reported levels of motivation and self-efficacy.

Based on the existing literature, there can be reciprocal relationships between motivation, self-efficacy, and improving academic performance, whereby enhancements in the former often correspond with improvements in the latter (e.g., Tella et al., 2007). The current study therefore uniquely set out to measure aspects of motivation and self-efficacy before and after the study activities—writing two essays on which participants were given qualitative feedback (but not numerical marks)—and to see whether providing different types of feedback affected participants' reported levels of motivation and self-efficacy, as well as the essay marks they achieved. Analysis of the collected data revealed a number of interesting findings.

First, there was no evidence that providing feedback on both structure and content, compared to just structure, led to a higher mark being achieved on the second essay. For both groups however, marks allocated for the criteria regarding use of evidence and references significantly increased on Essay 2, suggesting that feedback received on Essay 1 may have encouraged participants to focus on these aspects in their second essay.

With regard to the participants' MSLQ responses, both groups had similar scores for

their pre-test and post-test questionnaires, apart from values on the self-efficacy scale, which showed significantly higher scores on the post-test regardless of group. This suggests that all participants increased their self-efficacy by the end of the study, regardless of the type of feedback received.

Linking the pre-test scores and essay marks, no evidence was found of an effect of the type of feedback received. It was apparent, however, that those participants who scored more highly on the task value and control of learning beliefs scales in their pre-test questionnaires also tended to achieve higher marks in their essays. Linking pre-test scores, essay marks and post-test scores, evidence was found that those who scored more highly on their essays also reported having higher extrinsic goal orientation on the post-test questionnaire.

5. Conclusion and implications

Together these findings tell a complicated but interesting story. For instance, participants' marks on their second essays were similar whether they received feedback on essay structure alone or on both structure and content. However, both groups received higher marks on their second essays with regard to use of evidence and references (key aspects of essay structure highlighted in the feedback) after receiving feedback on their first essay. This is a very important finding, as use of evidence and references are key components in writing academic essays, and the feedback provided may have supported participants in improving their efforts and subsequently their performance toward these aspects. It can only be speculated that participants' higher self-reported levels of self-efficacy in their post-test questionnaire may have been related to this improved understanding and performance. As self-efficacy is such an important factor in people feeling they can do a task they set themselves, this is a crucial indicator that such feedback may have a key role to play in

helping participants to understand task requirements and how they can improve. This is supported by the finding that those participants who scored more highly on task value and control of learning beliefs in their pre-test questionnaires also scored more highly on their essays, suggesting that there was a relationship between participants who wrote better essays and those who felt they were in control of, saw the value of, and understood their efforts toward the essay-writing activities.

The lack of significant difference in marks between those receiving feedback on structure alone and those receiving feedback on structure and content is perhaps surprising and deserves further exploration. On the basis of this project it can however be concluded that using feedback to highlight certain structural elements of essay writing, in particular use of evidence and references, can have a lasting positive impact on participants' future essay performance. This is significant for all efforts to support the perception of feedback as "education" rather than just "evaluation" (Hyland, 2001; Lee, 2011), or as "advice for action" (Whitelock, 2010), as a means to help participants to improve their future work rather than simply as a mechanism for marking past work.

References

Alkharusi, H., Aldhafri, S., & Alnabhani, H. (2013). The impact of students' perceptions of assessment tasks on self-efficacy and perception of task value: A path analysis. *Social Behavior and Personality, 41*, 1681–1692. doi:10.2224/sbp.2013.41.10.1681

Baker, N. L. (2014). "Get it off my stack": Teachers' tools for grading papers. *Assessing Writing, 19*, 36-50. doi:10.1016/j.asw.2013.11.005

Bandura, A. (1993). Perceived self-efficacy in cognitive development and functioning. *Educational Psychologist, 28*, 117–148. doi:10.1207/s15326985ep2802_3

Banyard, P., Underwood, J., & Twiner, A. (2006). Do enhanced communication technologies inhibit or facilitate self-regulated learning? *European Journal of Education, 41*, 473–489. doi:10.1111/j.1465-3435.2006.00277.x

Black, P., & Wiliam, D. (1998). Assessment and classroom learning, *Assessment in Education, 5*, 7–74. doi:10.1080/0969595980050102

Chickering, A. W., & Gamson, Z. F. (1987). Seven principles for good practice in undergraduate education. *American Association of Higher Education Bulletin, 39*(7), 3–7. Retrieved from http://www.aahea.org/aahea/articles/sevenprinciples1987.htm

Clark, K., & Dwyer, F. M. (1998). Effect of different types of computer-assisted feedback strategies on achievement and response confidence. *International Journal of Instructional Media, 25*, 55–63.

Cronbach, L. J. (1951). Coefficient alpha and the internal structure of tests. *Psychometrika, 16*, 297–334. doi:10.1007/BF02310555

Ferris, D., Brown, J., Liu, H. S. & Stine, M. E. A. (2011). Responding to L2 students in college writing classes: Teacher perspectives. *TESOL Quarterly, 45*, 207-234. doi:10.5054/tq.2011.247706

Foote, C. J. (1999). Attribution feedback in the elementary classroom. *Journal of Research in Childhood Education, 13*, 155–166. doi:10.1080/02568549909594736

García Duncan, T., & McKeachie, W.J. (2005). The making of the Motivated Strategies Learning Questionnaire. *Educational Psychologist, 40*, 117–128. doi:10.1207/s15326985ep4002_6

Hawe, E. M. & Dixon, H. R. (2014). Building students' evaluative and productive expertise in the writing classroom. *Assessing Writing, 19*, 66-79. doi:10.1016/j.asw.2013.11.004

Hyland, F. (2001). Dealing with plagiarism when giving feedback. *ELT Journal, 55*, 375-381. doi:10.1093/elt/55.4.375

Hyland, F. (2010). Future directions in feedback on second language writing: Overview and research agenda. *International Journal of English Studies, 10*, 171-182.

Kluger, A. N., & DeNisi, A. (1996). The effects of feedback interventions on performance: A historical review, a meta-analysis, and a preliminary feedback intervention theory. *Psychological Bulletin, 119*, 254–284. doi:10.1037/0033-2909.119.2.254

Lee, I. (2005). Error correction in the L2 writing classroom. What do students think? *TESL Canada Journal, 22*(2), 1-16. doi:10.18806/tesl.v22i2.84

Lee, I. (2011). Working smarter, not working harder: Revisiting teacher feedback in the L2 writing classroom. *Canadian Modern Language Review, 67*, 377-399. doi:10.3138/cmlr.67.3.377

Lee, I. (2014). Editorial. Feedback in writing: Issues and challenges. *Assessing Writing, 19*, 1-5. doi:10.1016/j.asw.2013.11.009

Nelson, M. M., & Schunn, C. D. (2009). The nature of feedback: How different types of peer feedback affect writing performance. *Instructional Science, 37*, 375–401.

doi:10.1007/s11251-008-9053-x

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W.J. (1991). *A manual for the use of the Motivated Strategies for Learning Questionnaire (MSLQ).* Ann Arbor, MI: University of Michigan, National Center for Research to Improve Postsecondary Teaching and Learning.

Pintrich, P. R., Smith, D. A. F., Garcia, T., & McKeachie, W. J. (1993). Reliability and predictive validity of the Motivated Strategies for Learning Questionnaire (MSLQ). *Educational and Psychological Measurement, 53*, 801–813. doi:10.1177/0013164493053003024

Robinson, J. P., Shaver, P. R., & Wrightsman, L. S. (1991). Criteria for scale selection and evaluation. In J. P. Robinson, P. R. Shaver, & L. S. Wrightsman (Eds.), *Measures of personality and social psychological attitudes* (pp. 1–16). San Diego, CA: Academic Press.

Schraw, G. (2010). Measuring self-regulation in computer-based learning environments. *Educational Psychologist, 45*, 258–266. doi:10.1080/00461520.2010.515936

Shrestha, P., & Coffin, C. (2012). Dynamic assessment, tutor mediation and academic writing development. *Assessing Writing*, *17*, 55–70. doi:10.1016/j.asw.2011.11.003

Tang, Y., & Tseng, H. W. (2013). Distance learners' self-efficacy and information literacy skills. *Journal of Academic Librarianship, 39*, 517–521. doi:10.1016/j.acalib.2013.08.008

Tella, A., Tella, A., Ayeni, C. O., & Omoba, R. O. (2007). Self-efficacy and use of electronic information as predictors of academic performance. *Electronic Journal of Academic and Special Librarianship, 8*(2). Retrieved from http://southernlibrarianship.icaap.org/content/v08n02/tella_a01.html

Van Dinther, M., Dochy, F., Segers, M., & Braeken, J. (2014). Student perceptions of

assessment and student self-efficacy in competence-based education. *Educational Studies, 40*, 330–351. doi:10.1080/03055698.2014.898577

Wang, C.-H., Shannon, D. M., & Ross, M. E. (2013). Students' characteristics, self-regulated learning, technology self-efficacy, and course outcomes in online learning. *Distance Education, 34*, 302–323. doi:10.1080/01587919.2013.835779

Wang, S.-L., & Wu, P.-Y. (2008). The role of feedback and self-efficacy on web-based learning: The social cognitive perspective. *Computers and Education, 51*, 1589–1598. doi:10.1016/j.compedu.2008.03.004

Warden, C. A. (2000). EFL business writing behaviors in differing feedback environments. *Language Learning, 50*, 573–616. doi:10.1111/0023-8333.00141

Whitelock, D. (2010). Activating assessment for learning: Are we on the way with Web 2.0? In M. J. W. Lee & C. McLoughlin (Eds.), *Web 2.0-based-E-Learning: Applying social informatics for tertiary teaching* (pp. 319–342). Hershey, PA: IGI Global. doi:10.4018/978-1-60566-294-7.ch017

Zimmerman, B. J., & Martinez-Pons, M. (1992). Perceptions of efficacy and strategy use in the self-regulation of learning. In D. H. Schunk & J. L. Meece (Eds.), *Student perceptions in the classroom* (pp. 185–207). Hillsdale, NJ: Lawrence Erlbaum.

Table 1

Changes of wording in sample MSLQ items for pre-test and post-test questionnaires.

Intrinsic Goal Orientation scale

Original wording: In a class like this, I prefer course material that really challenges me so I can learn new things.

Pre-test questionnaire: In an exercise like this, I prefer tasks that really challenge me.

Post-test questionnaire: In an exercise like this, I prefer tasks that really challenge me.

Extrinsic Goal Orientation scale

Original wording: Getting a good grade in this class is the most satisfying thing for me right now.

Pre-test questionnaire: Getting a good grade in this exercise is the most satisfying thing for me.

Post-test questionnaire: Getting a good grade in this exercise was the most satisfying thing for me.

Task Value scale

Original wording: I think I will be able to use what I learn in this course in other courses.

Pre-test questionnaire: I think I will be able to use what I learn in this exercise in other situations.

Post-test questionnaire: I think I will be able to use what I learned in this exercise in other situations.

Control of Learning Beliefs scale

Original wording: If I study in appropriate ways, then I will be able to learn the material in this course.

Pre-test questionnaire: I will be able to master the material needed for this exercise.

Post-test questionnaire: I was able to master the material needed for this exercise.

Self-Efficacy for Learning and Performance scale

Original wording: I believe I will receive an excellent grade in this class.

Pre-test questionnaire: I believe I will receive an excellent grade for this exercise.

Post-test questionnaire: I believe I received an excellent grade for this exercise.

Test Anxiety scale

Original wording: When I take a test I think about how poorly I am doing compared with other students.

Pre-test questionnaire: I will think about how poorly I am doing compared with the other participants.

Post-test questionnaire: I thought about how poorly I was doing compared with the other participants.

Note. Original wording quoted from Pintrich et al. (1991, pp. 9–15).

Table 2

Marking scheme for essays.

Criterion	Maximum marks
1. Introductory paragraph sets out argument.	10
2. Concluding paragraph rounds off discussion.	10
3. Argument is clear and well followed through.	10
4. Evidence for argument in main body of text.	20
5. All paragraphs seven sentences long or less.	5
6. Word count between 500 and 1,000 words.	5
7. Award 5 for two or three references, 10 for four or more.	10
8. Provides a clear and explicit definition of risk or memory.	10
9. Extensive vocabulary, accurate grammar and spelling.	10
10. Understanding of practical issues, innovative proposals.	10

Table 3

Means, standard deviations, and values of coefficient alpha on pre-test and post-test

questionnaires.

Scale	No. of items	Pre-test questionnaire			Post-test questionnaire		
		M	*SD*	Alpha	*M*	*SD*	Alpha
Intrinsic Goal Orientation	3	5.55	.97	.70	5.50	0.89	.65
Extrinsic Goal Orientation	4	4.32	1.19	.78	4.40	1.17	.74
Task Value	6	5.16	.81	.79	5.41	0.94	.87
Control of Learning Beliefs	4	4.97	1.07	.75	5.13	0.98	.64
Self-Efficacy	8	4.78	.91	.92	5.01	0.91	.88
Test Anxiety	1	3.37	1.73	—[a]	2.87	1.71	—[a]

[a]Coefficient alpha cannot be calculated for scales consisting of a single item.

Highlights

- Feedback on students' written essays is critical to their academic learning.

- Feedback on structure and content was compared with feedback on structure alone.

- Students given the two kinds of feedback obtained similar marks on a second essay.

- Both groups obtained higher marks on their use of evidence and on referencing.

- Marks were correlated with scores on task values and control of learning beliefs.