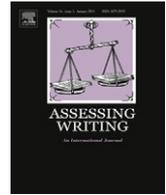




Contents lists available at [ScienceDirect](#)

Assessing Writing



Academic tutors' beliefs about and practices of giving feedback on students' written assignments: A New Zealand case study

Jinrui Li*, Roger Barnard

Department of General and Applied Linguistics, The University of Waikato, New Zealand

ARTICLE INFO

Keywords:

Assessment
Grades
Feedback
Academic writing
Tutor beliefs
New Zealand

ABSTRACT

Teachers' feedback on students' written work is an important aspect of pedagogy. However, theoretical views differ on what constitutes 'good' feedback, both among applied linguists and academics in other disciplines. In-depth research needs to be carried out into the contextual difficulties of evaluating and assessing academic assignments, and the awarding of grades, especially by those who are relatively inexperienced in this work. This article reports aspects of a case study which explored the beliefs and practices of a group of untrained and inexperienced part-time tutors in a New Zealand university. Data were collected from a preliminary survey, individual interviews, 'think aloud' and stimulated recall sessions, and focus group meetings. Extracts from the collected data are presented and discussed. The findings indicate that these tutors initially stated their belief that the purpose of providing feedback was to assist the students to improve their academic writing skills; however, it emerged that their primary concern was to justify the grades that they awarded. It is suggested that using a multi-method approach to data collection can bridge the gap between theoretical perspectives on what constitutes 'good' feedback and what tutors actually believe and do in their everyday work.

© 2011 Published by Elsevier Ltd.

1. Introduction

Feedback on students' written assignments is fundamental to students' understanding of good writing (Inoue, 2005, 2007). The three primary roles of feedback according to Joughin (2008) are to

* Corresponding author.

E-mail address: J1207@waikato.ac.nz (J. Li).

support the learning process, to judge current achievement, and to maintain disciplinary and professional standards. Joughin, however, suggests that there is a tension between the three roles: feedback may not support improvement, judgment may not be fair, and disciplinary standards are often unclear or even confusing to students. This is because disciplinary standards are essentially context-free since they are derived from theoretical perspectives as to what constitutes 'good writing' within the disciplinary tradition, whereas supporting learning processes is perforce dependent on the specific setting – the aims of specific programmes, the genre and content of the set assignments, the linguistic and educational background of the learners, and the academic identity of those responsible for providing feedback. Thus an assessor evaluating a piece of writing has to occupy a middle ground when actually forming a judgment on any student's work. This raises an ethical dimension to the provision of feedback and assessment which, as Hamp-Lyons (2002) has pointed out, "has been relatively unexplored" (p. 14).

Perhaps the key contextual factor is the purpose of the programme, whether the focus is on composition in itself, or on writing an assignment for a particular academic discipline. Applied linguists evidently have a closer and more informed interest in the construct of writing than do academics in disciplines such as the social or physical sciences. One branch of applied linguistics concerns itself with second language learners, and the discussions about feedback in this branch tend to focus on corrective feedback, usually with a narrow focus on lower order concerns (Keh, 1990) such as syntax, lexical choice and spelling, although higher order matters such as paragraph and text structure also receive attention. Here the debate is on whether or not such corrective feedback is actually effective (Ferris, 1997, 2006) or ineffective (Truscott, 1996, 2007) and issues relating to the difficulty of awarding grades for written work are less salient in the literature. Those applied linguists who focus on composition programmes for first language learners such as university students tend to consider the role of feedback less as corrective and more as formative in developing and motivating the learners' skills as effective writers. However, as Hamp-Lyons (2007) has also pointed out, "it is clear that in many situations writing *assessment* is taking over from writing *instruction*" (p. 6 – emphasis in original). Thus increasing attention is being paid to the issue of grading work. For example, Connors and Lunsford (1993) conducted a discourse analysis of comments on 3,000 marked papers and found that more than 80% of the comments indicated a judgmental tone: overall, the feedback was grade-driven.

After reviewing the history of writing assessment in terms of three phases, Hamp-Lyons (2001) pointed to an emerging 'fourth generation', which would encompass technological, humanistic, political and ethical aspects, arguing that an assessor has the responsibility "to use all means available to make any language test she or he is involved in as fair as possible" (p. 124). Subsequently, Carless (2006, p. 230) suggested "assessment dialogues" between tutors and students on the understanding of the content and roles of criteria, the expectations of tutors and students on the assessment, and opportunities of further assessment on the improved assignment or double marking. Similarly, Inoue (2005) proposed a student-centred or "community-based" (p. 222) model of assessment which emphasised the participation of the learning community in the whole process of assessment, including negotiation of common criteria of assessment. This model was challenged by Elbow (2006) who argued that good writing criteria varied even within one community and multi-values of good writing should be respected. Inoue agreed with Elbow that multi-values of good writing did exist in any community, but emphasised the need for communal standards of good writing for the sake of grading. However, such a student-centred model is problematic in terms not only of the difficulty students might face in articulating, negotiating and designing criteria but also of the match between students' beliefs of good writing and those of discourse communities (Orsmond, Merry, & Reiling, 2000).

Rust, O'Donovan, and Price (2005) have pointed to the difficulty of student-centred models in large classes, which is the norm in undergraduate programmes across the wide range of university disciplines. Bailey and Garner (2010) collected interview data from 48 teachers across departments in a faculty of a British university. The institutional policies of giving timely assessment, which should be both formative and evaluative, were analysed in the light of contextual factors such as large class size and heavy workloads. The authors found that the teachers believed that feedback was intended to inform students about their expectations and evaluation of the assignment and also used as 'feedforward' for students to make improvement and increase marks in future work. However, these beliefs tended to be overshadowed by their understanding that feedback had to satisfy various "institutional

requirements, procedures and priorities” (p. 195). Thus the teachers felt that they, rather than the institutional system, were often the target of criticism when feedback was ineffective.

Even within one discipline, studies have found that different values are used by assessors. For example, *Read, Francis, and Robson (2005)* found large variation among the comments and grades on two sample essays assessed by 50 historians. Not surprisingly, in studies of writing across disciplines, more divergences about values and criteria have been found (*Becher, 1989; Becker, 1991, 1994*) which not only influence teachers’ beliefs and practices (*Smeby, 1996*) but also learning outcomes. *Lea and Street (2000)* revealed that lecturers and tutors were “mainly influenced by specific conceptualizations of their own disciplines or subject area in their assessments of students’ writing”, and special attention was given to “structure and argument” (p. 39).

Several studies have pointed to the importance of the academic identity, experience and proficiency of those providing feedback (*Pula & Huot, 1993; Wolfe, 1997; Wolfe, Kao, & Ranney, 1998*). Differences in the linguistic background and professional experience of assessors contribute to their divergent beliefs of writing and assessment, which in turn results in their divergent practices of assessment (*Santos, 1988; Weigle, 1999*). Much of the literature in this area suggests that those primarily responsible for assessing undergraduate assignments have a relatively low academic status, and indeed may still be (graduate) students themselves, having received little or no explicit training in how to do the work. It seemed that the tutors in *Lea and Street’s (2000)* study had some general beliefs of good writing such as clarity, logical structure and argument despite differences of assignments, criteria, and disciplines – a point also made by *Nesi and Gardner (2006)*. However, these tutors could not describe explicitly “what a well-developed argument looks like in a written assignment” (*Lea & Street, 2000*, p. 39). In the absence of formal training, assessment practices carried out by individual tutors may be based on their personal “implicit criteria” or divergent understandings of the “explicit criteria” (*Rust et al., 2005*, p. 232), or their “intuition in action” (*Anson, 2006*, p. 104).

Despite such divergences, convergent beliefs about assignment and assessment have been found in some general areas. For example, *Connors and Lunsford (1993)* and *Ivanic, Clark, and Rimmershaw (2000)* indicated that students’ writing was regarded as a final product rather than work in progress, and that tutors read assignments for the purpose of grading; therefore, their feedback was used mainly to justify the grade. *Ivanic et al. (2000)* also found that subject tutors pointed out more negative than positive aspects and evaluated assignments against the expected answer which actually played the role of marginalising students from membership of the academic community; almost no feedback given by the tutors indicated an engagement of ongoing dialogue with students.

2. Finding out what teachers think and believe

The above overview of significant research-based literature has indicated that the provision of feedback is a fundamental element in the teaching–learning relationship, and it is necessary to explore the mental constructs and experiential contexts which guide teachers’ provision of feedback. The relationship between teachers’ beliefs and instructional practices has increasingly attracted educational researchers’ attention. In general, such research is based on three major assumptions: (1) teaching is largely influenced by teacher cognition; (2) teaching is guided by teachers’ thoughts and judgments; and (3) teaching constitutes a high-level decision-making process (*Isenberg, 1990*). As *Clark and Peterson (1986)* have claimed, teaching is “substantially influenced and even determined by teachers’ underlying thinking” (p. 255). However, it has long been recognised that individuals’ thinking processes and belief systems cannot merely be observed or measured, but instead must be inferred from what individuals say – and this has conventionally been investigated by attitudinal questionnaires.

Over the years, surveys within mainstream educational research have been carried out to identify teachers’ beliefs, employing a wide range of cognitive constructs and diverse operational definitions. *Borg (2006)* has argued that in order to ensure inter-study reliability, there is a need for a shared terminological framework. In the current absence of such, the view will be taken in this paper that *attitudes* elicited by surveys are the surface expression of underlying values, beliefs and knowledge. However, such attitudes cannot fully represent deeper mental constructs for various reasons such as: the limited choices any questionnaire imposes on the respondents; an individual’s lack of explicit awareness of the underlying conceptual framework; an internal contradiction between and within

belief categories; and/or a simple inability, or unwillingness, to convey these to another person. It is also likely that what a teacher believes or knows may not always correlate highly with his or her professional practice. Thus, investigations into teachers' beliefs should be balanced by consideration of their actual behaviour in planning and executing professional activities such as giving feedback, and the extent of the convergence or divergence between beliefs and practice.

3. The study

Informal conversations between the first author and Faculty tutors led her to undertake a contextualised study into the issue of the provision of feedback, of which this article reports one aspect. 'Feedback' in this study refers to both written comments and grades. The broad research question that guided the wider study was: What do the participants believe constitutes 'good' feedback, and to what extent does their practice converge with their beliefs? However, only the following sub-question will be the main focus of this article: What importance do the participants attach to the awarding of grades when giving feedback?

3.1. *Setting and participants*

This study was undertaken with tutors employed in several departments within the Faculty of Arts at a New Zealand university. The participants were, with two exceptions, temporarily employed as tutors on a part-time basis, whose main task was to provide feedback on some of the written assignments submitted by undergraduate students. The tutors were themselves students at the time, most of them taking graduate courses within their respective disciplinary areas. The extent of their working experience as tutors varied considerably; one had only begun this work in the current semester, others had been so employed for three or four years. All were paid on an hourly basis, with an assumption that there was a set time allowed for the grading and provision of feedback of each assignment; this time allocation varied between departments, but all of the participants reported that the time allocation was insufficient for them to do the work to their satisfaction.

3.2. *Data collection procedures*

After obtaining formal Human Research Ethics approval for the entire project, preliminary data were collected by means of an online survey. Emails were sent to the 52 tutors employed across the Faculty, requesting them to complete the questionnaire, containing both closed- and open-ended items, anonymously. The questionnaire also requested some basic biodata, and sought the respondents' informed consent to take part in the subsequent phases of the project. Data from 28 completed questionnaires were analysed to identify baseline categories to inform subsequent individual interviews.

Sixteen volunteers from eight departments were invited for interview at a time and place convenient to them to elicit their beliefs about the nature of good writing and effective feedback, and the sources of those beliefs. The semi-structured interviews lasted between 20 and 30 min; each was audio-recorded and transcribed, and a summary sent to each interviewee for respondent validation. The data from these interviews indicated the extent to which the individuals conformed to the general attitudes emerging from the survey and allowed the interviewer to explore more fully the sources of their knowledge and beliefs.

Nine of these participants subsequently agreed to allow the researcher to sit beside them, and audio-record, while they formulated their feedback on one or two anonymised students' assignments; the assignments and the tutors' written comments were both visible to the researcher. During these sessions, they were encouraged to verbalise their thought processes, that is, to 'think aloud' (Diab, 2005; Ericsson & Simon, 1984; Hyland, 2003), either through audible private speech or else by using the researcher as a silent auditor. Only occasionally would the researcher make minimal verbal prompts such as "Oh", "Yes", showing her attentiveness. As expected, the quantity and quality of intelligible verbalisation during these 'think aloud' sessions varied among the participants, and the researcher made field notes as she observed the tutor's formulation of feedback.

These field notes constituted the main stimulus for the Stimulated Recall session (Gass & Mackey, 2000) which followed, in most cases, immediately after the previous 'think aloud' session. During these discussions, each of which lasted between 30 and 45 min, the researcher invited the tutor to reflect on points of interest that occurred to either of them, and especially the extent of con/divergence between what the tutor had said that s/he believed in the interview and what actually occurred in the 'think aloud' session. As with the interviews, participants were sent summaries of transcripts, which again they all confirmed regarding accuracy and completeness.

The final data were collected several weeks later by inviting some of the tutors to meet in two focus group sessions, each lasting about 50 min. The primary purpose of these meetings was to broaden and deepen the researcher's understanding of the tutors' collective beliefs and knowledge about the provision of feedback, but it was also felt that the participants benefitted from sharing their experiences and co-constructing possible solutions to some problem areas that arose.

3.3. *Data analysis procedures*

The data were subjected to grounded analysis both manually and by using the software programme NVivo8 (Bazeley, 2007). The great advantage of this version of NVivo over its predecessors is the ability to upload sound files as well as text. Thus the collected audio data could be directly transcribed, and then managed, coded and categorised and aligned with all other inputted print data.

4. Findings

4.1. *Interview data*

Beliefs about what constituted 'good' feedback centred on three issues: (1) to help students to improve their future writing; (2) the importance of providing positive comments alongside the impact of negative feedback on students; and (3) the need to justify the grade eventually decided. (It should be noted that the quotations below accurately record what the participants actually said, and no attempt has been made to tidy up syntax, etc. Pseudonyms have been used.)

All participants said that they believed feedback was given for the purpose of helping the students to improve their writing. For example:

For improvement. It is not to say 'this is what you've done wrong'. It is so that people can improve on everything. That is why you don't give feedback on exams because by that point it is too late. (Emma)

To improve their grades. Things like improving their writing, improving their intellect, improving content, and to also improve themselves as people. (George)

Secondly, with regard to affective factors, all the tutors, reflecting on their own experiences as students when reading negative or positive feedback, considered that the feedback should be positive:

One of my supervisors could be very harsh, and he just wrote "Yuk" next to the sentences or paragraphs that he didn't like. I think it was little bit harsh so I think it probably influenced me to be more positive and try to put more positive type of feedback. (Cecile)

The comment that the lecturer gave me I thought you know a bit offensive I thought, like 'you've done not good enough and your English isn't good and you need help on your English' on everything. I thought it was quite offensive. (Mark)

Because all my lecturers pointed out good and bad points. They never ever pointed all negative points. They never said all the things are bad and then full stop. They always used words like "This could improve your grade". (George)

Such previous experiences influenced their present work in giving feedback to students who had difficulty in writing. For example, Helen believed "If I mentioned everything that was wrong, it would too negative for him . . . But unfortunately if you did not mention it, they'll think it was OK." However,

in a similar case, George gave lengthy feedback on most of the negative aspects, even though the student might be unhappy to receive it: “I probably wrote as much as he hated it on ways he could improve it.” This sensitivity towards students’ feelings was a common theme running through the interviews:

Maybe the only difficulty is that I don’t want to be too negative. I want to be positive even if they had got it wrong. I don’t feel like I should be the one to say ‘you got it completely wrong’.
(Jo)

Thirdly, all interviewees considered the awarding of a grade to be an integral element of their feedback. For example, Mark gave feedback because it would help students “get a better mark”; Simon said that good “written feedback explains how and why you got the grade”; George believed the feedback he gave “was a grade and then what they could do to improve that grade”. There were, however, different beliefs about how and what to grade: Mia, Mark and George mentioned that they tended to mark higher than other tutors. The tutors also had different opinions on whether they marked down language. Mark would mark down errors if they appeared repeatedly; George would not mark down for ‘minor’ mistakes. Despite this, they all believed they should try to be consistent with the lecturers or the senior tutors in the way of giving feedback and marking. For example, three tutors in the same department mentioned they had marking meetings every week. According to Cecile, cross-marking was used among tutors to ensure consistency and “enough and the right feedback”; Mia had the experience of marking the same assignment with other tutors to ensure the same level of marking; Emma exchanged opinions with the other tutor on the same course. However, all participants mentioned some areas of inconsistency in their feedback. For example, Helen believed the first assignment should be marked more leniently, and indeed said she was lenient in marking. She also mentioned her experience of re-marking because she marked lower than the lecturer. Mia felt that students like easy marking and so she gave higher grades than other tutors. Mark believed he marked higher, but he also said that if he was too lenient, students would not make any improvement. Jo was not sure whether to mark down or lower in case part of the answer was right. At the bottom of the scale, she would refer the assignment to the senior tutor rather than fail it herself.

It became increasingly clear during the set of interviews that grading the work was the most challenging aspect of providing feedback, and this was reflected in the think aloud and stimulated recall sessions.

4.2. *Think aloud data*

All nine participants in the think aloud sessions made positive and negative comments about both the content and formal aspects of the assignments. Six of the tutors made explicit remarks on clarity of expression, and the three interviewees who had emphasised the importance of good referencing also thought aloud about this while working through the students’ assignments. Three others who did not mention referencing in the interviews commented on this aspect in their practice. Interestingly, although in their interviews all the tutors emphasised the importance of content, five participants in the think aloud sessions spent much more time, and provided more feedback, on formal issues. All participants used ticks as positive in-text feedback; six of them used encouraging words in their feedback such as “Good”, “Excellent” “Well done”; and two others (Mark and Cecile) used the students’ names in their overall comments, and inserted ‘smiley face’ icons to be more positive.

So I just look at what he has got. Er, he hasn’t given the information, (xxx)¹ quite a bit of a big section (xxx). I’m going to give him 1.5 out of 6, he’s missing quite a lot of work. Er, just checking his grammar and spelling. I am going to take off a quarter from the grammar, he’s made a couple of small mistakes. He didn’t keep it all in past tense. So I take a quarter of (xxx) with half of the

¹ By its nature any private speech which audibly manifests itself tends to be syntactically fragmented, phonologically abbreviated and agglutinated, and externally incoherent (Vygotsky, 1934, p. 30), and so what the tutors said to themselves in the think aloud sessions was not always intelligible. Thus, in several transcripts, unclear utterances are indicated by “x”.

mark of that anyway. So I just ended up at 14 and 1.25 out of 20 and I just write “Good” and a smiley face. (Cecile)

Seven of the tutors provided overall comments. One other (George) only wrote “Great work” on one assignment he marked and made no overall comment on the other, while Cecile pointed out one negative aspect of one assignment, and wrote merely “Good job” on the other. The length of the overall comments of the other participants varied from fewer than ten words to more than fifty. Three participants (Helen, Simon, and Mia) expressed frustration with negative points in assignments, and there were fewer comments on positive aspects than on negative points, which were usually about referencing, syntax and lexis. For example, Mark wrote: “More social science terms and concepts”, implying that the student should use appropriate terminology, although he gave no specific advice on how to improve the work.

As will be illustrated below, all the tutors paid particular attention to what might be an appropriate grade, although sometimes they did not make a final decision. For example, Jo had been instructed not to write down the grade on the assignment, and Emma would give a grade later on the basis of comparison of a group of assignments. All of them used either marking schedules – of varying complexity – or model answers to guide their decision-making: “I’m going to check the marking guide and see if there is anything she could have mentioned which she missed out” (Cecile). They often hesitated to make a decision on what grade should be given where there were choices of whether to fail or not to fail, or when the mark might fall between two grades, as when Simon muttered to himself: “It is nineteen out of thirty. That is like maybe C + or something. So it is a B.” Three participants (Simon, Helen and Joe) showed a tendency to mark higher when there was the possibility to mark lower, and there was explicit relief when they came across a good assignment:

I love students’ work like that. Just make me so happy because then I don’t have to mark them down. (George)

I can probably say this one is probably the best essay I have read so far and they followed the criteria that have been laid out for them, and just looking at the essay and looking at the marking schedule and I’ll (probably?) give them a B. (Mia)

Despite the often limited amount of written feedback on the assignment, the process of reaching a conclusion about the grade to award involved considerable thought and generated much private speech. For example:

I’ll put it in the C range. Maybe in the, in the, you know, you have to break some of the, may be around, around 58, C, breaking it down: quality of writing 20, be half way there, organisation. . . (Anna)

Ah, so she gave, she gave me a really good, er, really well written essay, er, on the border, I probably, er, she’s missing, maybe with the social science terms and concepts. Er, I would possibly give an A-? (Mark)

Er, they got two wrong there, so I’ll give them five out of it six. Good. I’ll give them three out of point five, er, OK, xx. Yes. Yes. Yes. No, that is wrong. Er, it is x. Yes. Yes. No. No. Yes. Yes. No. No. Not xx. Yes. Yes. Yes. Yes. Good. There we go. It is a key word ‘non renewable’. Just put down here. Yeah. Yeah. Yeah. Yeah. OK. So no. Yes. Yes, both, both yeah, and this, both, so that. Yeah. Yes. Yes. Yes. Yeah. Yeah. Good. Ok now, yeah, that is good. Yeah. Good. Good. Good. Right now, get it done. So this person got fifteen points five for that. (George)

And the effort he is putting so is really good. OK, all these questions are just personal opinion, so I’m not interested in them, to be honest. I just skim over them very briefly because it is just opinion. So I am just doing them very quickly because their answers are either going to be right or wrong. OK, and this page she hasn’t done anything. Which I think ok why hasn’t she finished? But because she has run out of time, ok so she didn’t answer 1, 2, 3, 4, 5, 6, 7, 8, 9 question, ten questions. So given that she answered this question really well to the best she could, I am still going to give her a very good mark because she almost finished. (Jo)

On the whole, there was a general level of convergence between what the tutors said in their interviews and what they did in practice. As mentioned above, attention was paid to both positive and negative aspects, although there was a tendency to focus more on matters of form rather than substance. The feedback which was provided mainly comprised ticks and one- or two-word in-text comments with or without an overall summary – and a grade. Therefore, it seemed that feedback was used more to justify the grade than to suggest improvements for subsequent assignments.

4.3. *Stimulated recall data*

Many comments in the subsequent stimulated recall sessions focused on the tutors' explanation of how they graded the students' work. For example, with reference to the work just assessed:

Well, that was a very good essay. Good essays are easy to mark. You read them and you think "Yes! I got it" you know. And I find it easier to write positive comments than to write negative comments. I find writing negative comments, I don't want to be too negative. (Helen)

Yeah. I mean there are times if I was extensive, here and here [pointing at in-text feedback], I would just write, "See the essay" and I leave that, but this is more like a justification of the grade, and it is similar. This is similar format for everybody. (Simon)

However, it was not always easy to decide between two possible grades, for example whether an assignment should be given a C+ or a B–:

So I mean in terms of the marking schedule, his quality of writing [in this assignment] has gone down, and then his content, you know, is up there, and organisation isn't as er, 58 [the final percentage that she wrote on the assignment]. (Anna)

I mean I probably want to give it a B, but I would talk to him [the lecturer] whether, how much weight he is going to put on the lack of referencing really. Seems it is possible that some lecturers would fail it, but, but I need to talk to him about it so that we will be consistent across the course. (Helen)

As noted above, in their interviews and think aloud sessions, the participants mentioned the difficulty of maintaining grade consistency across the assignments they marked. The issue was reflected in the stimulated recall sessions, where, for example, Mia said that she marked her assignments in batches of ten and compared the grades she awarded with others "just to make sure I didn't mark the first one too harshly." Jo mentioned she might not be able to be consistent between the first and last assignments because of inexperience or fatigue:

I think when I am at the start, er, when I am marking, you haven't been doing it very long, and you can, you have the effort, and you can be (to) write something, but maybe on the last one, I wouldn't do that. Which is kind of unfair to the student, but it is just kind of what happens I think. (Jo)

To sum up, in the stimulated recall sessions the tutors mainly focused on the assignments they had just marked, but also reflected on their previous experience, their implicit theories of marking, and difficulties they had met. It seemed that, influenced by the impact that low grades would have on their students, tutors tended to give higher grades, and they rarely failed an assignment. They tried to justify the grades they awarded by adhering as far as possible with the lecturers' criteria, but were often left in academic doubt and ethical uncertainty. They tried to be consistent and provide quality assessment but this was mitigated by a variety of factors, especially their inexperience, the shortage of time, and the amount of guidance they could get from their lecturers.

4.4. *Focus group data*

Each of the two focus group meetings lasted just over 50 min and was audio-recorded without the researcher being present. Participants were asked to talk about a number of issues. For the sake

of brevity, and with reference to the research question specific to this article, the following extracts focus on the awarding of grades when giving feedback. Both groups spent longer on the first topic – marking – than on any other, and issues about marking frequently recurred when they discussed the other topics. Overall, the two groups spent 10 and 11 min discussing marking.

In the initial discussion about marking, Helen pointed out reasons for not failing students, with which the other two participants agreed.

Helen: I tended to avoid giving, yeah, I probably still wouldn't, I didn't in fact fail anyone without talking to him. Yeah I wouldn't, I don't mind giving an A+ at all but I don't like failing people unless, partly because, you know, I am not sure, and partly because I kind of think lecturers (they are going to be aware) they will complain.

Maria: Actually (xxx)² some interesting comments that I talked to, we marked our own paper, we should (xxx) in A semester, and, but I talked to the lecturer about it. And she said she tended try not to fail people for the reasons (xxx) but er, they are very, should be very aware of how that might affect the actual, particular this year, the actual (xxx) of the university. Er, social effects.

Helen: I think there are all sorts of commercial (aspects). They want them to come back, don't they?

Maria and George: Yeah.

The tutors agreed that they tried not to fail a student because “if you fail a course you have to pay it and do it again” (Mia). They hoped their students could be encouraged to do better in the next assignment if they got a pass rather than a fail grade. However, participants were not sure whether this was a good professional practice or not:

Frank: I think I am quite worried I was a bit too lenient while I was marking because actually I haven't failed anyone yet, but maybe I should always tell myself, like, I don't know, like maybe if I try and get them to realise the kind of strength they can kind of work from that, like give people like really low grade but never. . .

Mia: A bare pass to encourage. . .

Frank: Yeah a bare pass and say “Look, I should have failed [you] but I passed you, just work on your strong points. I think maybe in my mind maybe give them a bit of kick on, you know”.

Participants were cautious and tried to maintain consistent standards when grading assignments, and it seems that there was a common practice of checking previously marked assignments to make sure they were being fair:

Mia: You go back and do some of the first ones. You get enough after you have finished to make sure they are on the same stage?

Frank: Yeah. I usually mark them more than once and then just quickly run through them all again just because sometimes I gave someone a B then read it again and just think, ‘Oh, no. They actually had some good points’, so I didn't know they actually, they did very well.

Besides the above discussion, the conversation turned to grading when participants talked about the other topics, for example in regard to empathy:

² The nature of a free-flowing discussion between people leads to rapid overlapping between interlocutors, and hesitation, redundancy and internal incoherence within individual turns – this is represented by “X” in the abstracts. In the first of these, Maria's main point was that she tried not to fail people for a variety of reasons, including the negative impact on the university such as falling enrolments.

Mia: Again, do you find, like, knowing people in the class? I found that a lot of them they want to be like friends, trying to get relationship with you.

Henry: Do you find it probably better in terms of you can give them a low grade or something?

Mia: No I try not (Laugh)

There was general agreement on the usefulness of re-reading the assignments and making adjustments to the grades in order to be consistent:

George: I marked 38 assignments and then I've gone back after I had marked them and sort of changed some of the marks. So that they like, you know, more stratified thing. We got told they had to be more stratified than . . .

Helen: Yeah? Em.

George: So like such. . .

Helen: Yeah I know what you mean.

George: In comparison with the other people, you know. So like this paper here is fantastic piece of work. You couldn't find one error so they should get an A+, however, the other people who are given an A+ before they've got a lot more errors so I might blabla mark them down. Or at least change the number, so I put 25 to 27, made A+ and then I gave them 26 to 25.

Helen: Yeah. I think probable within the grade. Yeah, I would grade, you know, an A than that. Yeah and I think it is important because if you don't have that sort of flexibility, then you don't know. . .

In summary, the focus group participants spent a great deal of time talking about grading. This was not only in the opening topic, but was repeatedly returned to in their discussion of the other issues. The tutors tried to be consistent across the assignments they marked and shared a common practice of checking back on previously marked assignments to make sure they were being fair. They agreed that they did not want to fail students because this would diminish students' motivation to study, would affect the tutorial relationships, and have financial implications. There was agreement that their students would be encouraged to do better in the next assignment if they got a higher rather than a lower grade. However, they were not sure whether this was sound professional and ethical practice. Sometimes they felt confused when receiving conflicting instructions or advice from different lecturers, even within one department. Consequently they felt they lacked authority in providing feedback because they were unclear of the overall course design and the professional standards of giving feedback. Thus, when giving feedback, and especially when awarding grades, they relied upon their own experience as students when receiving feedback on assignments from their lecturers, and their often limited practical experience as assessors and markers.

5. Discussion

The major finding of this study was that these tutors' primary concern in giving feedback was less that of seeking to improve the students' writing skills and more that of justifying – to themselves, to their students, and to their academic superiors – the award of a specific grade for the assignments to hand. This concurs with findings from previous empirical studies investigating teachers in various disciplines (e.g. [Bailey & Garner, 2010](#); [Orrell, 2006](#)) and points to the need for theoretical perspectives on the nature of 'good' feedback to take account of the contextual constraints and opportunities, and the ethical dilemmas, facing those responsible for actually providing feedback.

We acknowledge the limitations inherent in this small-scale case study of tutors' beliefs and practices relating to the provision of feedback on students' academic writing, primarily, the limited number of participants and the restricted setting in which the data were collected. Generalisations beyond the specific setting could not be, and were not, expected. That the investigation was carried out by a single researcher was also a limitation, but this was mitigated by two considerations. Firstly, the researcher was herself a postgraduate student, and this facilitated the establishment and maintenance of an easy rapport with the participants. Secondly, while being in this way a cultural insider, the fact that the

first author was studying in an unrelated department (Applied Linguistics) helped to avoid any sense of a professional conflict of interest within the setting.

The strength of this in-depth case study lies in the application of multiple methods of data collection, the triangulation of which provided contextualised insights into how the participants perceived and carried out their work. The value of such insights is best judged by those readers who work in relatable contexts.

Many previous studies of teacher cognition have relied heavily on survey techniques and, while these might enable a cross-sectional elicitation of a large number of self-reported attitudes, these attitudes could not be probed in any depth. The value of questionnaires, such as that used in the present study (although only mentioned here in passing) lies in providing baseline constructs for further investigation and identifying participants for subsequent phases of a project. Thus, semi-structured interviews can, and in this case did, enable the reported attitudes to be explored in more depth, and with a particular focus on the participants' actual working context. The data thus derived indicated that there was individual convergence with what was generally reported in the survey but also, importantly, highlighted issues which were not salient in the questionnaire responses. Chief among these was the tutors' concern with awarding grades to the work they had to mark, and this concern tended to dominate their thinking while actually formulating feedback on specific assignments. This innovative capturing of 'cognition in flight' (Vygotsky, 1978) through think aloud procedures while formulating feedback was perhaps the most revealing element of the study. The impact was strengthened by the immediate reflection of these practices through stimulated recall as the two sessions very closely linked cognition and practice. Finally, the focus group meetings confirmed that the issue of grading was a serious matter for the participants. These meetings also served to build a sense of a community of practice among tutors in different departments via the sharing of concerns, experiences and ideas.

It may be noted that opportunities for professional co-construction among inexperienced and relatively low-status (hourly-paid) university teachers is as rare as any form of professional development before employing them in the tutorial role. If a university considers that appropriate feedback on students' assignments is an essential pedagogical function, we suggest that attention should be paid to raising the tutors' theoretical understanding of key issues, and their awareness of what is considered good professional practice in the provision of feedback. Also, we suggest that future studies into teachers' beliefs would benefit from a multi-method approach to data collection in specific contexts so that the findings from such empirical investigations can inform discipline-based standards of what constitutes 'good' feedback on written assignments.

Acknowledgments

The authors wish to thank the tutors for their participation, and to acknowledge the contributions made by Dr. Rosemary de Luca and Professor John F. Fanselow for their comments on earlier drafts of this article.

References

- Anson, C. M. (2006). Assessing writing in cross-curricular programs: Determining the locus of activity. *Assessing Writing*, 11 (2), 100–112.
- Bailey, R., & Garner, M. (2010). Is the feedback in higher education assessment worth the paper it is written on? Teachers' reflections on their practices. *Teaching in Higher Education*, 15 (2), 187–198.
- Bazeley, P. (2007). *Qualitative data analysis with NVivo*. London: Sage.
- Becher, T. (1989). *Academic tribes and territories: Intellectual enquiry and the cultures of disciplines*. Buckingham, UK: Buckingham University Press.
- Becker, H. J. (1991). When powerful tools meet conventional beliefs and institutional constraints. *The Computing Teacher*, 18 (8), 6–9.
- Becker, H. J. (1994). How exemplary computer-using teachers differ from other teachers: Implications for realizing the potential of computers in schools. *Journal of Research on Computing in Education*, 26, 291–320.
- Borg, S. (2006). *Teacher cognition and language education: Research and practice*. London; New York: Continuum.
- Carless, D. (2006). Differing perceptions in the feedback process. *Studies in Higher Education*, 31 (2), 219–233.
- Clark, C. M., & Peterson, P. L. (1986). Teachers' thought processes. In: M. C. Wittrock (Ed.), *Handbook of research and teaching* (3rd ed., pp. 255–296). New York: Macmillan.
- Connors, R. J., & Lunsford, A. A. (1993). Teachers' rhetoric comments on student papers. *College Composition and Communication*, 44 (2), 200–223.

- Diab, R. L. (2005). Teachers' and students' beliefs about responding to ESL writing: A case study. *TESL Canada Journal*, 23 (1), 28–43.
- Elbow, P. (2006). Do we need a single standard of value for institutional assessment? An essay response to Asao Inoue's "community-based assessment pedagogy". *Assessing Writing*, 11 (2), 81–99.
- Ericsson, K. A., & Simon, H. A. (1984). *Protocol analysis: Verbal reports as data*. Cambridge, MA: The MIT Press.
- Ferris, D. (1997). The influence of teacher commentary on student revision. *TESOL Quarterly*, 31, 315–339.
- Ferris, D. (2006). Does error feedback help student writers? New evidence on the short- and long-term effects of written error correction. In: K. Hyland & F. Hyland (Eds.), *Feedback in second language writing: Context and issues* (pp. 81–104). New York: Cambridge University Press.
- Gass, S. M., & Mackey, A. (2000). *Stimulated recall methodology in second language research*. Mahwah, NJ: Lawrence Erlbaum.
- Hamp-Lyons, L. (2001). Fourth generation writing assessment. In: T. Silva & P. K. Matsuda (Eds.), *On second language writing* (pp. 117–128). Mahwah, NJ: Lawrence Erlbaum.
- Hamp-Lyons, L. (2002). The scope of writing assessment. *Assessing Writing*, 8 (1), 5–16.
- Hamp-Lyons, L. (2007). Editorial. *Assessing Writing*, 12 (1.), 1–9.
- Hyland, F. (2003). Focusing on form: Student engagement with teacher feedback. *System*, 31, 217–230.
- Inoue, A. B. (2005). Community-based assessment pedagogy. *Assessing Writing*, 9 (3), 208–238.
- Inoue, A. B. (2007). A reply to Peter Elbow on "Community-Based Assessment Pedagogy". *Assessing Writing*, 12 (3), 228–233.
- Isenberg, J. P. (1990). Teachers' thinking and beliefs and classroom practice. *Childhood Education*, 66, 322–327.
- Ivanic, R., Clark, R., & Rimmershaw, R. (2000). What am I supposed to make of this? The messages conveyed to students by Tutors' written comments. In: I. M. R. Lea & B. Stierer (Eds.), *Student writing in higher education: New contexts* (pp. 47–65). Buckingham, UK/Philadelphia, PA: SRHE & Open University Press.
- Joughin, G. (2008). *Assessment, learning and judgement in higher education*. London: Springer.
- Keh, C. (1990). Feedback in the writing process: A model and methods for implementation. *ELT Journal*, 44 (4), 294–304.
- Lea, M. R., & Street, B. V. (2000). Student writing and staff feedback in higher education: An academic literacies approach. In: M. R. Lea & B. Stierer (Eds.), *Student writing in higher education: New contexts* (pp. 31–46). Buckingham, UK: The Society for Research into Higher Education and Open University Press.
- Nesi, H., & Gardner, S. (2006). Variation in disciplinary culture: University tutors' views on assessed writing tasks. In: R. Kiely, P. Rea-Dickens, H. Woodfield, & G. Clibbon (Eds.), *Language, culture and identity in applied linguistics* (pp. 99–118). London: BAAL/Equinox.
- Orrell, J. (2006). Feedback on learning achievement: Rhetoric and reality. *Teaching in Higher Education*, 11 (4), 441–456.
- Orsmond, P., Merry, S., & Reiling, K. (2000). The use of student derived marking criteria in peer and self-assessment. *Assessment and Evaluation in Higher Education*, 25 (1), 23–38.
- Pula, J. J., & Huot, B. A. (1993). A model of background influences on holistic raters. In: M. M. Williamson & B. A. Huot (Eds.), *Validating holistic scoring for writing assessment: Theoretical and empirical foundations* (pp. 237–265). Cresskill, NJ: Hampton Press.
- Read, B., Francis, B., & Robson, J. (2005). Gender, 'bias', assessment and feedback: Analyzing the written assessment of undergraduate history essays. *Assessment and Evaluation in Higher Education*, 30 (3), 241–260.
- Rust, C., O'Donovan, B., & Price, M. (2005). A social constructivist assessment process model: How the research literature shows us this could be best practice. *Assessment and Evaluation in Higher Education*, 30 (3), 231–240.
- Santos, T. (1988). Professors' reactions to the academic writing of nonnative-speaking students. *TESOL Quarterly*, 22 (1), 69–90.
- Smeby, J. (1996). Disciplinary differences in university teaching. *Studies in Higher Education*, 1, 69–79.
- Truscott, J. (1996). The case against grammar correction in L2 writing classes. *Language Learning*, 46 (2), 327–369.
- Truscott, J. (2007). The effect of error correction on learners' ability to write accurately. *Journal of Second Language Writing*, 16 (4), 255–272.
- Vygotsky, L. S. (1934). *Myshlenie i rech': Psikhologicheskie issledovaniya*. [Thinking and speech: Psychological investigations.]. Moscow and Leningrad, Russia: Gosudarstvennoe Sotsial'no-Ekonomicheskoe Izdatel'stvo.
- Vygotsky, L. S. (1978). *Mind in society* (M. Cole, Trans.). Cambridge, MA: Harvard University Press.
- Weigle, S. C. (1999). Investigating rater/prompt interactions in writing assessment: Quantitative and qualitative approaches. *Assessing Writing*, 6 (2), 145–178.
- Wolfe, E. W. (1997). The relationship between essay reading style and scoring proficiency in a psychometric scoring system. *Assessing Writing*, 4, 83–106.
- Wolfe, E. W., Kao, C. W., & Ranney, M. (1998). Cognitive differences in proficient and nonproficient essay scorers. *Written Communication*, 15, 465–492.

Jinrui Li was a lecturer in English and translation at a Chinese university before starting her doctoral research at the University of Waikato, New Zealand. Her research project explores teachers' beliefs and practices of giving feedback on students' written assignments.

Roger Barnard is a Senior Lecturer in applied linguistics at the University of Waikato, New Zealand. Prior to 1995, he worked in senior ELT positions in Britain, Europe and the Middle East. He prefers to research and write collaboratively, and is currently (co)editing two volumes of case studies and co-authoring several articles.