Engaging students with self-assessment and tutor feedback to improve performance and support assessment capacity

Conor Thomas McKevitt
Dundalk Institute of Technology, conor.mckevitt@dkit.ie

Follow this and additional works at: http://ro.uow.edu.au/jutlp

RESEARCH ONLINE

Engaging students with self-assessment and tutor feedback to improve performance and support assessment capacity

Abstract

Assessment is one of the most important elements of student life and significantly shapes their learning. Consequently, tutors need to ensure that student awareness regarding assessment is promoted. Students should get the opportunity to practise assessing work and receive tutor feedback so that they might improve on both the work and their assessment of it. The purpose of this paper is to investigate how student engagement with criteria, exemplars, self-assessment, and feedback influenced students' performance, their assessment capacity, and also how students experienced the process. A mixed methods approach was used. Students' performance and assessments were established using a rubric that included 5 criteria each evaluated using 5 point likert scale linked to descriptors. A thematic analysis of the focus group resulted in two themes. The findings show that overall students' performance in the assignment significantly improved between draft and final submissions. Students' assessment of their work significantly differed to the tutor's on some criteria at both submissions but in opposite directions on one criterion between both submissions. The focus group found that the rubric guided students to produce their draft while tutor feedback guided them to improve on it. However, these findings require further investigation. The following recommendations ensue from the research and should assist student development concerning assessment. Tutors should give students an opportunity to assess work and also see tutor's assessment of that work using the same criteria. Also, tutors should provide constructive feedback during an assignment.

Introduction

Assessment is the single most important aspect of a student's academic life (Gibbs 2010) and it should direct their learning. Students who understand the assessment process may learn better (Price et al. 2011, p.485). The present research aims to guide students to perform better using self-assessment and feedback. The approach this study takes to assessment is both transparent and formative. Students are provided with explicit guidelines for how they will be assessed. The tutor and students use an identical marking rubric detailing the evaluation criteria to assess the work (Figure 2). This gives students an opportunity to assess their own work and observe how the tutor judges the work using the same rubric. They are required to submit a self-assessed draft piece of work to the tutor, who provides feedback. They are then given ample time to submit a self-assessed final piece of work that will be graded for the module. The participants are third-year students enrolled in a module taught by the author within the Humanities department in DkIT, a third-level Institute of Technology in the northeast of the Republic of Ireland. The research carried out is intended to be applicable within teaching situations where at least one form of mid-module or continuous assessment is used.

The main purpose of assessment is to evaluate what the student is learning. In considering this more closely I take the perspective of the student, who may require some time to realise what is expected for the task. This expectation changes during the assignment process. Taking this into account I ask questions that a student facing assessment might ask during the process and use the literature to answer each question. The purpose is to investigate what, if any, approach to assessment can benefit the assessment process and student learning.

What I need to do? What should it look like?

For students to comply with assessment requirements they need criteria that describe the task and a sample of the proposed quality (Rust 2002). Students have reported that criteria and example standards are useful for knowing what is expected in an assessment (Bell et al. 2013). Sadler (2005, 2009b) calls this a holistic approach to assessment that benefits the student. While criteria and example standards are more effective when used together, practising onlywith explicit criteria is also beneficial (Payne & Brown 2011). Studies show that students who receive criteria and example standards, engage in marking using criteria and receive explanations of marking develope a more complete understanding of how to fulfill task requirements (Rust et al. 2003; Payne & Brown 2011; Hendry et al. 2012). Indeed, the absence of a process where tutors and students could discuss the marking of example standards resulted in students' inability to develop knowledge to improve (Handley & Williams 2011, p.104). Hendry et al. (2012) found that "[t]eacher-led marking and discussion of exemplars in class results in increased student understanding of standards and higher achievement" (p.149), and that students had difficulty understanding criteria in the absence of example standards. In contrast to these studies, Wimhurst and Manning (2013) showed that students marking example standards and giving explanations for their marks, even in the absence of detailed criteria and marking workshops, were more successful in their own assessments than students who did not do the same marking activity. Taken together, these findings suggest that practice assessing with criteria combined with some tutor contribution can improve student understanding of assessment and student performance, and may improve the ability to self-assess.

Could I guide myself?

Self-assessment is an essential skill for effective learning (Carless et al. 2011; Boud et al. 2013), and could help develop assessor judgement. However, it is a skill that must be learned (Lew et al. 2010). Students should become assessors and regulators of their own work; this should be a principle of education (Sadler 1989, 2009a; Carless et al. 2011), as it could aid

them in improving both the work and their judgement of it (Smith et al. 2013). Therefore, tutors should stimulate more self-assessment in students (Nicol & MacFarlane-Dick 2006; Orsmond & Merry 2011). Sendziuk (2010) calls for tutors to use rubrics to aid self-assessment for students. Students who self-assessed using a rubric before submitting their work found that it improved both their learning and their work (Andrade & Du 2007). An opportunity to practise assessing is required for students to become more experienced as assessors in general (Bloxham et al. 2011).

Studies have found that students and tutors judge work differently. Lew et al. (2010) found that first-year students did not assess themselves at the same level in their first semester as their tutors assessed them; in a follow-up study they found that student accuracy actually deteriorated in the second semester. They concluded that students are poor judges of their own learning process. Boud et al. (2013) showed that with time, over at least three semesters, students' self-assessments did get more accurate in relation to the tutors' assessments of them. This study used data from an online database on which both tutors and students assessed student work against criteria. The study did not report what year of study the students were in at the time of the self-assessments. However, it is clear that they had had at least three semesters' practice in self-assessing. What is significant in this case is that students were able to see how the tutor had assessed the work. The study found that students' judgements of their own work do converge with tutors' over time, as long as they can see how the tutors assessed the work. Based on the fact that students' judgement improves the more judgements they make may suggest that practice at judging is leading to this improvement. Interestingly, both studies find that high achievers are more accurate in self-assessment and improve over time, but low achievers are not accurate in self-assessment and tend not to improve over time. Given these findings, it may take time for students to approximate the tutor's experience level in assessing performance. Also, students may derive greater benefit from sources of feedback outside themselves than they do from self-assessment. Boud et al. (2013) suggest that interventions that employ feedback for students on their assessment and engage students in exercises that will increase their knowledge of criteria and standards would benefit and develop self-assessment. Therefore, an approach that employs a self-assessment element with formative feedback that gives information not only on the work but the judgement of that work could help.

Taras (2003) used a model of self-assessment in which students engage in self-assessment of their work, receive tutor feedback and then take corrective action (Nicol & MacFarlane-Dick 2006). She asked final-year students to self-assess their work, then provided them with feedback that allowed them to understand and correct errors in their work of which they had previously been unaware (Taras 2003). These students were using tutor feedback to close the gap between their current work and a higher standard (Sadler 1989). In a similar study Sendziuk (2010) found that students who received tutor feedback and then had to provide feedback on their self-assessment became more critical as a result. In this case students were forced to actively engage with the tutor feedback and the assessment criteria simultaneously to develop their assessing skills. These findings illustrate that while feedback to oneself is worthwhile, tutor feedback on the same assessment and according to the same criteria is also valuable. In other words, tutor feedback at the right time is essential.

How do I know I am doing what is required? How could I improve?

Feedback has been singled out as the most influential element of the assessment process (Gibbs 2010; Carless et al. 2011; Ferguson 2011). Students report that they highly value feedback during the assessment process (Beaumont et al. 2011, p.684) because they recognise that it is important for student learning and development (Sadler 1989; Taras 2003; Poulos & Mahony 2008; Beaumont et al. 2011; Ferguson 2011). Students in focus groups reported that it was better to get feedback while drafting so that they could revisit their work and improve

(Pokorny & Pickford 2010). In a mixed-methods study that used both focus groups and a questionnaire, students reported that quality feedback was that which both helped them improve their work and was provided early enough for them to apply it (Beaumont et al. 2011). The major findings in a survey of students found that feedback that is related to clear and understandable criteria, provided in a timely fashion and personally specific is best (Ferguson 2011). These findings concur with Poulos and Mahony (2008), who found that students preferred timely feedback that had been written for them individually. Orsmond and Merry (2011) quantitatively analysed tutor feedback; they then interviewed the tutors regarding their intentions for the feedback and students regarding their perceptions and responses to that same feedback. They found a misalignment between what students wanted from feedback and what tutors were providing. They recommended that tutors should try to give feedback that can improve future assignments, and should guide students more on how to use such feedback effectively. For example, tutors could ask students what type of feedback they would like prior to providing it (Price et al. 2010). Carless et al. (2011) interviewed award-winning tutors in relation to the feedback strategies they employ, finding that an approach that develops the student as an assessor is most sustainable. Sadler (2010) further asserts that tutors have a responsibility to develop students' assessment capacities. Taken together, these findings show that timely feedback that is personal and understandable, supports improvement and fosters assessment ability is most effective for students. Feedback that helps the students "close the gap" between what they have done and what is expected from them is essential. However, it "can only be effective when the learner understands the feedback and is willing and able to act on it" (Price et al. 2010, p.279).

Practice in assessment (e.g. with criteria and example standards, assessing and self-assessing, engaging with feedback) can facilitate the development of assessor skills and should serve to answer the questions asked in this review. However, this must be facilitated by the tutor. The intervention described in this study is underpinned by the theory and findings outlined herein. Students were provided with example standards and a rubric containing the criteria that both they and the tutor discussed and used to assess the work. Students were asked to self-assess a draft piece of work on which they received timely feedback based on the criteria for how they could improve in the future. Students then submitted their final work, which was also self-assessed.

Research Questions

1. Does student performance on an assignment, as assessed by the tutor, improve between draft and final submission?
2. Does students' assessment of their own work differ to tutors' at both draft and final stage?
3. How do students experience the class interventions?

Methodology

A mixed-methods approach was used to best answer these questions; the specific method for each question is provided below

Sample

Third-year humanities students from two degree programs experienced the interventions outlined in Figure 1. Thirty-five students (59%) consented to let their rubric judgements (their own and the tutors') be used for the study, and five students (8%) who experienced each of the interventions consented to participate in the focus group. Both were convenient samples (Cohen et al. 2011). The methods used for each of the research questions will be discussed below.

Timeline for Assessment

The timeline in Figure 1 shows how the classroom assessment progressed during a 13-week module in the autumn semester of 2014.

Figure 1. Timeline for module assessment
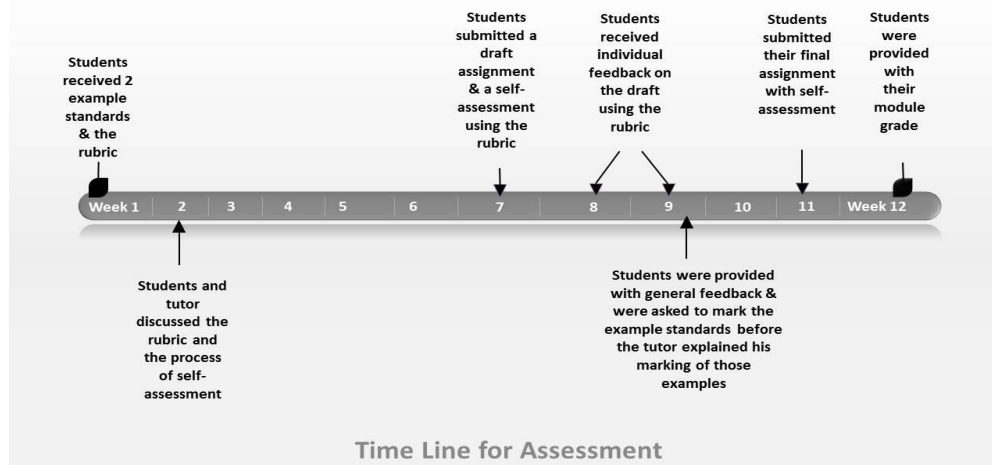
Questions 1 and 2

Tutors and students used the same rubric at both draft and final stage to assess performance on the assignment. Rubrics can enhance "reliable scoring of performance assessments" (Jonsson & Svingby 2007, p.141) and "can lead to a relatively common interpretation of student performance" (Reddy & Andrade 2010, p.442). However, their consistency is enhanced if they are "analytic, topic-specific, and complemented with example standards and/or rater training" (Jonsson & Svingby 2007, p.136). The rubric used in the current study was produced specifically for the purposes of the module based on the example standards used in class, and reviewed by a colleague prior to distribution. Figure 2 illustrates the evaluation criteria and achievement descriptors within the rubric. Each of the achievement descriptors relates to a point on a scale from 1 = not at all achieved to 5 = completely achieved. The use of Likert scales is similar to other self-assessment studies (Taras 2003; Lew et al. 2010; Boud et al. 2013). The rubric was discussed in class throughout the module.

Once the module was complete, three other tutors from different subject areas used it to assess three students' work at both draft and final stages. This was done for two reasons: to ensure the tutors agreed on how to assess using the rubric (Jonsson & Svingby 2007; Reddy & Andrade 2010) and to determine if the students' work, if independently assessed, improved between draft and final submission. Inter-rater reliability is a measure of the agreement among raters (in this case tutors) in terms of how they assess a student's performance on each of five different criteria. Tutors practised with the rubric by marking students' draft submissions (Jonsson & Svingby 2007; Hallgren 2012). Only their assessments for those students' final pieces of work were used to determine inter-rater reliability. An intra-class correlation (ICC) was used to establish inter-rater reliability (Field 2009; Hallgren 2012) by measuring the agreement of the tutors' assessments using each criterion in the rubric. For this study each of the three tutors assessed three different students' work, for a total of nine separate assessments. The author also independently assessed each of these nine assignments to avoid the case of the assessments being fully crossed. A one-way random ICC was chosen because each student's work was assessed by a different set of raters, with the author common to all. Absolute agreement was sought between the tutors and the author for each assignment

assessed. Table 1 shows the agreement levels between the tutors and the author on each of the five criteria (single measures) and between all tutors on each of the five criteria (average measures). Shrout and Fleiss (1979) suggest that the average agreements should be reported.

Cicchetti (1994) reports that ICC values are related to inter-rater reliability as follows: 0.40 and less is poor; 0.40 to 0.59 is fair; 0.60 to 0.74 is good; and 0.75 to 1.0 is excellent. The inter-rater reliabilities among the tutors in the current study was fair to good, with criterion four (relating to referencing) being poor. Discussion with the tutors uncovered that there was some discrepancy between tutors' approaches to referencing. Nevertheless, considering the modest amount of practice, and the lack of prior discussion of the rubric criteria, tutors' inter-rater reliability was reasonably good.

Figure 2. Marking rubric used by students and tutor at draft and final stage

Programme:			Module:		
Work is evaluated using the criteria below	*The quality of the Work is determined using the descriptions below – place a tick in the middle bottom of the description box which best describes your achievement for each of the evaluation criteria*				
	Completely Achieved	*Very Well Achieved*	*Well Achieved*	*Almost Achieved*	*Not at all Achieved*
Introducing and describing the topic using the literature (readings, both core and others) to interpret (shed light on) the impact this topic has had on society	Topic is clearly introduced with the relevant literature (core readings). It is then explained in depth using literature and illustrating its impact on society.	Topic is clearly introduced and defined using literature. The topic is then well described making reference to its possible impact on society.	Topic is clearly introduced and defined for the reader. An example of the topic is also provided.	Topic is mentioned briefly and a short description is provided.	The topic is not introduced or described in any way.
Comparing (<u>similarities in the literature</u>) & Contrasting (<u>differences/divergences in the literature</u>) what the literature is saying in relation to this topic	There is clear evidence that readings (both core and others) are being used together to confirm (verify) the academic opinion on this topic. There is also clear evidence that readings (both core and others) are being used, at least <u>once</u>, to debate (question) academic opinion on this topic.	It is evident that the core readings and at least one other reading is being used to illustrate academic opinion on this topic. There is no or little evidence of academic opinion being questioned in relation to the topic.	The core readings are mentioned briefly throughout, and show some comprehension in relation to the topic. However, there is no reference to any other reading or literature.	The core readings are barely mentioned. No evidence that they have been understood.	No attempt is made to compare or contrast any literature in relation to the topic in question.

Personal Reflection & Thinking in relation to this topic (what I have learned by having to think about this topic – before and after my reading/class)	It is clear from the outset that the writer has thought about and reflected on the topic. Some real life examples are provided. Reference is made to how they thought about this topic prior to the class and now, as a result of taking the class. The writing uses the readings to clarify (explain) their learning for the reader.	The writing provides some real life example of the topic illustrating that they have thought about it. They have not made clear reference to how they thought about this topic prior to the class and now. The way the writing uses the readings illustrates that they have thought about it.	There is little or no attempt to personally reflect on the topic. However, it is evident from reading the piece that the writer has reflected on the topic and made some references to the readings.	There is no evidence of personal reflection & little evidence of thinking in relation to the topic with only brief mention of core readings.	There is no reflection of a personal nature in relation to learning and the topic.
Referencing	Flawless referencing & bibliography.	Very good referencing & bibliography. Few minor inaccuracies	Adequate referencing & bibliography. But not fully accurate	Limited and highly inaccurate referencing & bibliography.	No referencing of bibliography
General Presentation	Outstanding Presentation & Writing skills – the work is excellent in terms of its coherence, syntax, spelling & grammar	Very good writing & presentation – logically structured using correct expression, syntax, spelling & grammar	Writing is acceptable but weaknesses are evident in structure, expression, syntax, spelling & grammar	Writing is marginally acceptable. Problems with presentation. Structure, presentation, expression, syntax, spelling & grammar are problematic	Poor writing throughout. Little structure. Expression, syntax, spelling & grammar poor.

Table 1. Intra-class correlations of tutors judging sample work

Intraclass Correlation Coefficient "Introduction of Topic"

	Intraclass Correlation	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.579	-.045	.885	3.750	8	9	.033
Average Measures	.733	-.094	.939	3.750	8	9	.033

Intraclass Correlation Coefficient "Comparing and Contrasting Literature"

	Intraclass Correlation	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.564	-.067	.880	3.583	8	9	.037
Average Measures	.721	-.145	.936	3.583	8	9	.037

Intraclass Correlation Coefficient "Reflection on Topic"

	Intraclass Correlation	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.918	.703	.981	23.500	8	9	.000
Average Measures	.957	.825	.990	23.500	8	9	.000

Intraclass Correlation Coefficient "Referencing"

	Intraclass Correlation	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.233	-.437	.750	1.607	8	9	.247
Average Measures	.378	-1.552	.857	1.607	8	9	.247

Intraclass Correlation Coefficient "General Presentation"

	Intraclass Correlation	95% Confidence Interval		F Test with True Value 0			
		Lower Bound	Upper Bound	Value	df1	df2	Sig
Single Measures	.515	-.135	.863	3.125	8	9	.055
Average Measures	.680	-.313	.927	3.125	8	9	.055

All one-way random effects model where people effects are random.

The boxplots in Figures 3, 4 and 5 illustrate that, in general, these tutors judged the final work to be marginally better than the draft work in over half the cases. This is encouraging, given that these tutors did not teach the subject and only had three students' work to assess.   The tutors had no knowledge of the research question for this study.

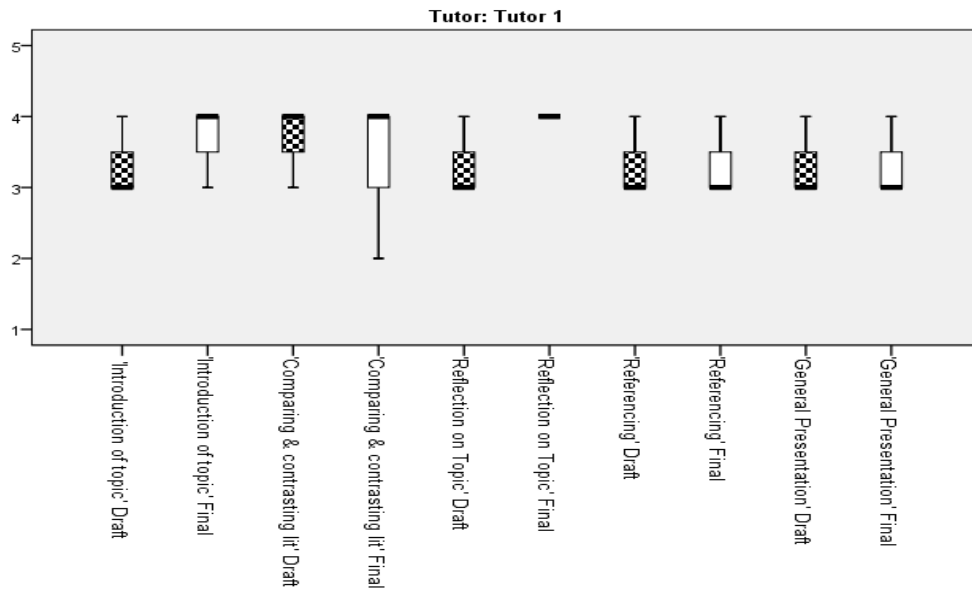Figure 3. Boxplots of Tutor 1's draft and final judgements of three students' work

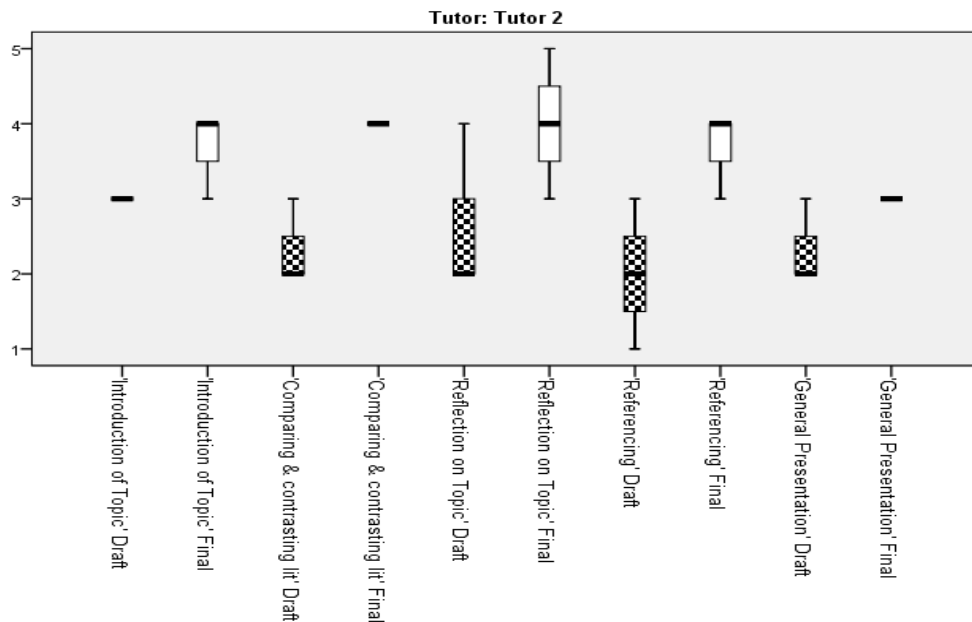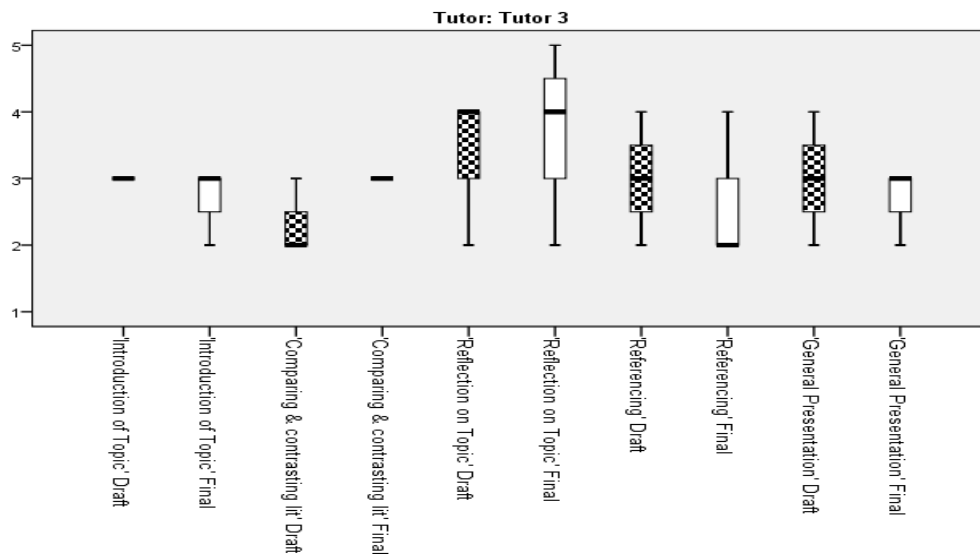Figure 4. Boxplots of Tutor 2's draft and final judgements of three students' work

Figure 5. Boxplots of Tutor 3's draft and final judgements of three students' work

A Wilcoxon signed ranks test was used to investigate if students' performance improved between the draft and final stages. The tutors' assessments of students' performance was used, rather than the students' self-assessment, as the tutors were the more experienced assessors (Lew et al. 2010; Boud et al. 2013). This allowed students' performance as assessed by the tutors at draft and final submission to be tested for significant difference. The tutors' assessment scores on a five-point Likert scale were used for each criterion. To investigate if students assessed their work differently from the tutors between the draft and final stages, Mann-Whitney U tests were used. This was done to determine if there were significant differences between the tutors' and students' assessments of the students' work. Both the tutors' and students' assessment scores on five-point likert scales for each criterion were compared at both the draft and final submission stages. These tests were used because the data were ordinal in nature (Cohen et al. 2011, p.606). The analysis was carried out using SPSS version 20.

Question 3

Focus groups are used for generating information on collective views, and the meanings that lie behind those views (Gill et al. 2008). Thus, they can let researchers study and understand a topic from the perspective of the group participants themselves (Wibeck et al. 2007, p.250). Studies have used focus groups to investigate similar phenomena. For example, Andrade and Du (2007) used focus groups to investigate self-assessment using a rubric. Poulos and Mahony (2008) and Pokorny and Pickford (2010) used them to investigate students' perception of feedback. Hendry et al. (2012) used them to investigate marking workshops. Other studies have used semi-structured interviews to investigate feedback in terms of how it is understood (Orsmond & Merry 2011) and student engagement (Price et al. 2010). The intervention in the current study took place in an environment where the group interaction was natural – the classroom – thus allowing participants to discuss their collective experiences of the class.

The questions focused on the students' experiences of the interventions. The group moderator was known to the participants and had no input into the module, and therefore was impartial. A group of five class members comprised the focus group. This is an acceptable size for focus groups that ensures the research question can be answered (Gill et al. 2008). The focus group

was digitally recorded and transcribed by the researcher. The data was subjected to a thematic analysis using the approach outlined by Braun and Clark (2006). The researcher conducted the analysis alone. The analysis was driven by the theory outlined earlier in the paper, as that theory underpinned what the focus-group members experienced in class and is central to the research question. The transcription was read several times to generate notes and preliminary thematic maps. The first stage involved coding all the data. The codes were then gathered together into similar groups according to the meanings deduced by the researcher from the students' experiences. These groupings were then collated under a theme, and thematic maps were generated. The themes were repeatedly checked and refined or collapsed under other, more prevalent, themes. This process was repeated until no more rational reduction of themes could take place. The analysed data was forwarded to the focus-group participants for verification.

Ethical concerns

This research was approved by the DkIT research ethics committee. All participants completed consent forms, which were stored securely. The research was explained to the students in the first week of their module, and they were told that their consent to participate in the study would be requested only after they received the grade for the module. All data was anonymised using codes and pseudonyms and stored securely.

Findings and Discussion

Does student performance on an assignment, as assessed by the tutor, improve between draft and final submission?

Table 2 shows significant statistical differences between draft and final submissions in all criteria except "Referencing".

Table 2. Comparison of draft and final student performance as assessed by tutor

Wilcoxon Signed Rank results on tutor assessment of student performance		
Final and draft (introduction of topic)	Z = -2.840	p < .05
Final and draft (comparing and contrasting literature)	Z = -4.258	p < .0005
Final and draft (reflection on topic)	Z = -3.911	p < .0005
Final and draft (referencing)	Z = -1.410	p = .159
Final and draft (general presentation)	Z = - 3.00	p < .05

Results show that generally students performed better on their final submission, as assessed by the class tutor. As outlined in the methodology, three independent tutors' assessments were used to determine reliability of the rubric criteria and to rule out unconscious bias on the part of the author specifically related to the research question. Focus-group data revealed that

students were quite clear that the specific feedback they received at draft stage affected how they approached the final submission:

Moderator: "What, if anything, impacted on the way you thought about and wrote your final draft?"

P4: "Basically the feedback from the rubric."

P3: "The feedback."

Others: "Yeah,"

However, a practice effect between draft and final submissions could lead to improvements in students using the rubric and also in their writing (Heiman 2002). Therefore, practice along with feedback is probably the most likely reason for the improvement in student performance.

Do students' assessments of their own work differ to tutors' at both draft and final stage?

Figures 6, 7, 9 and 10 illustrate that students' assessments of their work differed to tutors' for both draft and final submissions, with Figure 8 illustrating agreement between students and tutor on "reflection on topic".

Figure 6. Boxplots of students' and tutor's assessments of "introduction of topic" for draft and final submissions
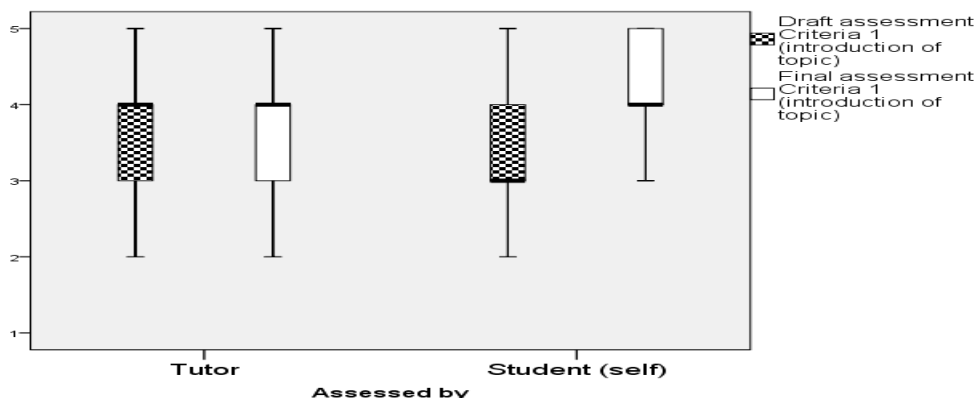

Figure 7. Boxplots of students' and tutors' assessments of "comparing and contrasting literature" for draft and final submissions
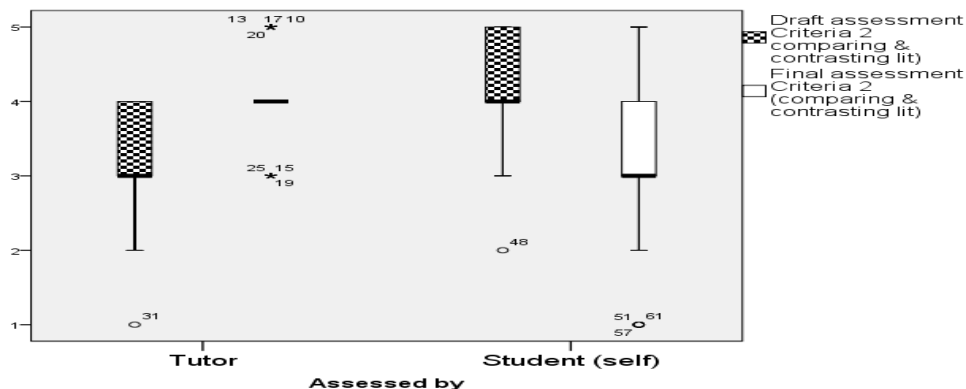
Figure 8. Boxplots of students' and tutors' assessments of "reflection on topic" for draft and final submissions
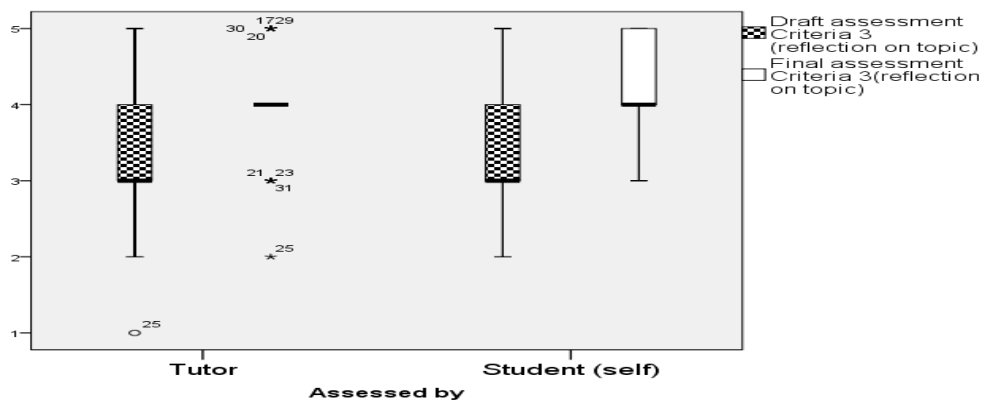

Figure 9. Boxplots of students' and tutors' assessments of "referencing" for draft and final submissions
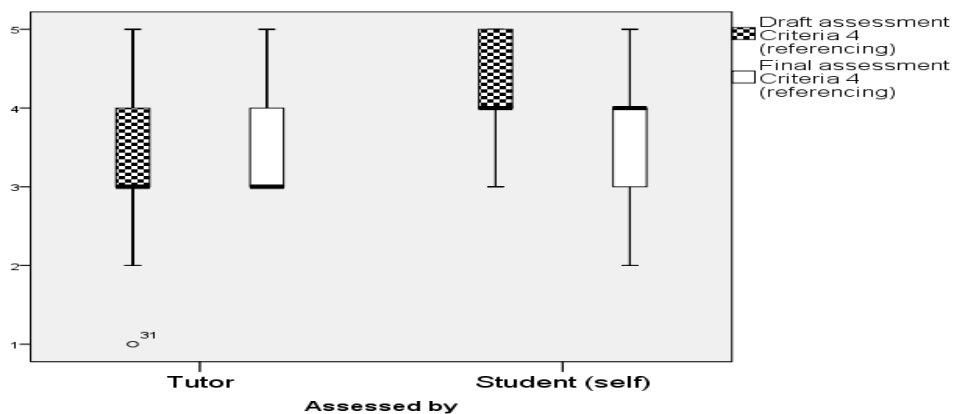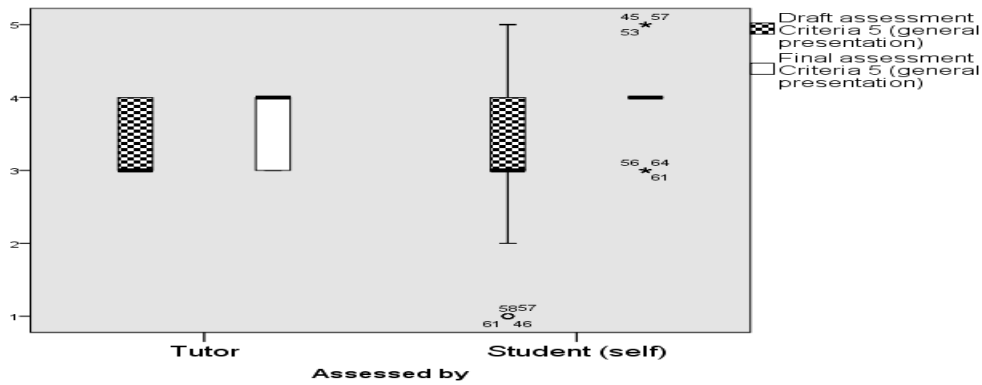

Figure 10. Boxplots of students' and tutors' assessments of "general presentation" for draft and final submissions

The Mann-Whitney U results in Table 3 show significant statistical differences between students' and tutors' assessment of performance at draft stage only for "comparing and contrasting literature" and "referencing".

Table 3. Mann-Whitney U results draft (tutor-assessed compared to self-assessed)

Mann-Whitney U test statistic results draft assessment		
Draft (introduction of topic)	U = 541	p = .962
Draft (comparing and contrasting literature)	U = 167	p < .0005
Draft (reflection on topic)	U = 517.5	p = .703
Draft (referencing)	U = 285	p < .0005
Draft (general presentation)	U = 528.5	p = .820

At draft stage, students' assessments of their performance were significantly higher than the tutors' for "comparing and contrasting literature" and "referencing", but aligned on the others. In general, this is contrary to what Lew et al. (2010) and Boud et al. (2013) have found. In contrast to their studies, students in this study discussed the assignment with the tutor. The rubric and example standards guided their expectations for the assignment prior to draft submission, as described in the focus group:

P1: "I didn't know what was expected until I got the rubric."

Moderator: "Okay."

P1: "Eh, until I got that and I could read through and then I knew what – as well as I know I said the sample wasn't great but it was like a guideline."

Moderator: "Right."

P1: "Mm so that's how I knew, and...."

Moderator: "And?"

P1: "Speaking with 'X' [tutor] as well."

Other: "Yeah."

Students reported being well informed on what to do for this assignment and said that they used this information to help them both write and assess their draft. This seems to have influenced their assessment at draft submission.

Table 4 shows a significant statistical differences between students' and tutors' assessment of performance at final stage for "introduction of topic", "comparing and contrasting literature", and "general presentation".

Table 4. Mann-Whitney U results final (Tutor assessed compared to Self-assessed).

Mann-Whitney U test statistic results final assessment		
Final (introduction of topic)	U = 387.5	p < .05
Final (comparing and contrasting literature)	U = 315	p < .005
Final (reflection on topic)	U = 515.5	p = .684
Final (referencing)	U = 537	p = .913
Final (general presentation)	U = 395.5	p < .05

At final stage students assessed their work higher than did the tutor on "introduction of topic" and "general presentation". However, students assessed their work lower than did the tutor on both "comparing and contrasting literature" (Figure 2) and "referencing" (Figure 4), although the difference was not statistically significant. The feedback provided at draft stage may explain what is happening in this case. At draft stage students generally assessed their performance on the criteria "comparing and contrasting literature" and "referencing" higher than the tutor. For each of the three remaining criteria, including "introduction of topic" and "general presentation", they generally agreed with the tutors' assessment of their performance. Both the tutors' and the students' assessment of performance was measured using the same five-point scale, allowing students to see immediately whether their assessment agreed with the tutors'. This direct numerical feedback, akin to Boud et al. (2013), probably forced students to recalibrate their judgements of "comparing and contrasting literature" and "referencing" in particular. It is therefore conceivable that they took note of the written feedback provided for these criteria. It is also conceivable that they underestimated their performance on their final submission as a result. Figures 7 and 9 illustrate such a change at final submission stage. The numerical feedback provided to students at draft stage affirmed their assessments of "introduction to topic" and "general presentation". As a result, students may not have taken note of the feedback pertaining to these criteria and overestimated their performance. Boud et al. (2013) conclude that tutors' feedback scores over time help students

recalibrate their judgement. However, in this case there was also written feedback. Students were clear that the feedback they received at draft stage affected the way they approached the final submission, but they did not indicate whether the numerical scores on the five-point scale were included. It appears that the numerical scores using the five-point scale constituted a large part of tutor feedback. It is therefore conceivable, but not conclusive, that students may recalibrate their assessments to meet their tutor's based on such feedback, as highlighted by Nicol and MacFarlane-Dick (2006). However, they only engage with written feedback if the tutor's assessment of their performance is significantly different to their own. This requires further research.

It is important to state that after the draft stage students were informed that each criterion would be weighted as follows: "introduction of topic" – 20%; "comparing and contrasting literature" – 30%; "reflection on topic" – 30%; "referencing" – 10%; "general presentation" – 10%. The weightings of 20% and 10% for "introduction of topic" and "general presentation" may have led students to think that these criteria were less important and therefore did not warrant further significant attention. It is also possible that the weightings of 30% for "comparing and contrasting literature" and 10% for "referencing" may have affected student assessment because of their perceived relative importance. These conclusions warrant further research.

How do students experience the class interventions?

Focus-group participants were questioned about their experiences of the class interventions. The group comprised 8% of the class and all had performed well in the assessment. Their responses were collected under two themes (Figure 11). Both themes were constantly referred to as being important in the context of the interventions. Guidance was found to be valuable in relation both to being able to practise with criteria to write and self-assess and to receiving formative individual feedback on their draft. The implication is that both forms of guidance helped the students achieve what was expected in the assessment by engaging them with the criteria through their self-assessed drafting and their receipt of individual feedback.

Figure 11. Thematic map of focus-group responses

| **Guidance as Practise with rubric (criteria) – drafting & self-assessment** | ⟷ | **Guidance as Formative Individual Feedback** |
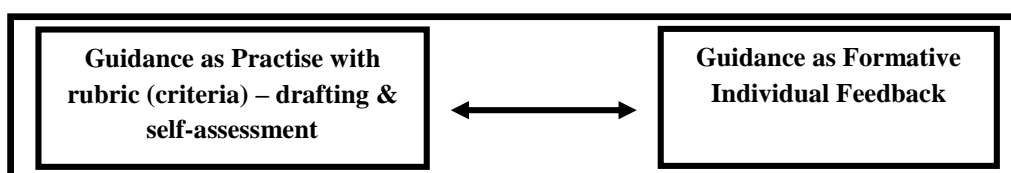
Guidance as practice with rubric

Students were asked to submit a self-assessed draft using the rubric. This task guided students to actively think about and engage with assessment criteria. Their description of the process details how the rubric was used as a guide for both drafting and self-assessing:

P1: "Hmm, when you're writing it and then reading back over, say, the introduction was broken down into, like, little other heading that you should have in your introduction and then your reflective piece. Hmm, so you know yourself when you are reading it you are, like, 'I didn't hit that point so I have to go back over it and fix it up.'"

Students referred to the rubric throughout the assessment in a practical way to ensure they were doing what they needed to and to guide them in their work. This was prevalent throughout the focus-group discussion:

P3: "..[the] guideline [was] to see were you hitting or was the essay constructed in the right way. And, you know, and it gave for me – it helped me to kind of see, okay am I doing that, am I answering that, you know."

The rubric acted as a guide for the drafting process.

Guidance as formative individual feedback

Students submitted their self-assessed draft and received feedback from the tutor. The feedback was intended to let students know how they were doing and how they could improve:

P3: "He'd write out a bit –"

P5: "– of what could be done."

P3 (at same time) "– what could be done."

Moderator: "For each criterion."

P5: "Yeah."

Others: "Yeah."

P4: "Why he had the tick in, we'll say, 'just achieved', rather than, you know, 'well achieved', you know."

Other: "You'd know what to change."

P4: "He ticked, we'll say, the 'just achieved' box, and then a suggestion on what you should do to improve on that."

In general, students reported that the feedback enabled them to revisit their work and make improvements because they knew what to do:

Moderator: "Yeah, okay, so you were given written feedback on each criterion and then you went away and did what with it? Some of you – I assume from what you said that some didn't do anything with it."

P4: "Just improved on what he had suggested."

P3: "Yeah, just improved."

P4: "You improve on...."

Moderator: "So you tackled each criterion individually."

Others: "Yeah. Oh, yeah."

The individual nature of this feedback seemed to be important for students:

P2: "And it was specific to your essay as well, like, he would put, mm , in brackets, like the bit you put and say, 'You could rather put this way instead to make it better,' like it wasn't just a broad thing where everybody should do this and everybody should do that. It was specific to your essay."

Moderator: "Right – what's the benefit of that, then?"

P2: "Kinda feels like he's listening to you, it's more personal, like, and he's actually taking the time to look through your thing –" (P3: "Yeah.") "– and go through each part to make it better for you."

Moderator: "And does that have an effect on, mm, kinda what you put into the essay after –"

P2: "Mmm, it makes you want to do it more if he's putting the time and effort in to do that for you."

Others: "Yeah."

It seems that feedback that was individual in nature and helped students improve their draft work was important. Overall, students were guided initially by the rubric and then by the individual feedback on their work. Essentially, each element provided guidance to students on what was expected and how to accomplish it. It is important to say that the students who participated in the focus group appear to have engaged with the individual feedback. Students found that guidance was important for the assessment process, and reported that the requirement to self-assess the draft submission guided them to reflect simultaneously on the work and the criteria to meet the assessment conditions. Prior to receiving feedback, students used both the criteria and the example standards as a guide for drafting their work. The feedback brought their attention to mistakes they had not noticed and helped them improve their work. The students valued this individualised feedback; this echoes findings in other studies (e.g., Price et al. 2010; Ferguson 2011). However, the tutors were supported by the rubric in producing specific feedback for students, who were then able to make specific changes to improve their work on individual criteria. It seems the use of the same rubric for students and tutors facilitated a more straightforward feedback process because the rubric was so familiar. The relational dimension appeared to be important for students; this has been mentioned in other studies on feedback (Pokorny & Pickford 2010; Price et al. 2010). Students in this study perceived the relationship that exists between the tutor and students as a process of exchange whereby the tutor provided feedback to which the student felt obliged to respond, either because the tutor had gone to the effort for them individually or because the tutor might discover that the students had not made the suggested changes. What this implies is that the relationship, as perceived by the student, can influence both why and how a student might react to feedback from a particular tutor. The relational dimension is complex in nature and warrants further investigation, particularly regarding how feedback can motivate students. In general, guidance provided for the assessment process seems to have benefitted students.

Recommendations

The following recommendations should help students develop their performance on both their assignments and their assessment of assignments.

- Provide criteria, example standards and discussion relating to the assessment task as early as possible. This will assist students in drafting and assessing their work.
- Ask students to self-assess their work as part of the assessment. This will actively engage them in reflecting on their work.
- Provide an opportunity for feedback to the student on a draft prior to submission so that they might realise mistakes that the tutor uncovers and observe how the tutor assessed the work.
- Tutors should use the same criteria as the students and provide individual feedback that is specific in nature and focused on students' improvement. This

will make it easier for students to make the necessary changes and may motivate students to act on the feedback they receive.

Conclusion

The findings in this study indicate that students can use feedback to significantly improve their performance. Students are able to assess their performance in a similar way to the tutor using the same criteria. However, it seems that tutors' feedback, both numerical and written, could affect the assessment process along with, for example, weightings. Further research on these findings is necessary. In general, students found the assessment experience valuable in terms of the guidance provided by the rubric prior to feedback and the fact that feedback was specific about how they could improve. Also, the students' perception of their relationship with the tutor was found to be important in motivating students to respond to feedback. However, this warrants further investigation. It seems that while assessment is hugely influential for students, the approach to assessment implemented by tutors could be just as important. Therefore, when it comes to assessment tutors need to be conscious that what they do, and how they do it, could affect students' performance and the development of their assessor capacity.

References

Andrade, H & Du, Y 2007. Student responses to criteria-referenced self-assessment. *Assessment & Evaluation in Higher Education* vol. 32, no. 2, pp. 159-181.

Beaumont, C, O'Doherty, M & Shannon, L 2011. Reconceptualising assessment feedback: A key to improving student learning? *Studies in Higher Education*, vol. 36, no. 6, pp. 671-687

Bell, A, Mladenovic, R & Price, M 2013. Students' perceptions of the usefulness of marking guides, grade descriptors and annotated exemplars. *Assessment & Evaluation in Higher Education*, vol. 38, no. 7, pp. 769-788.

Bloxham, S, Boyd, P & Orr, S 2011. Mark my words: The role of assessment criteria in UK higher education grading practices. Studies in Higher Education, vol. 36, no. 6, pp. 655-670.

Boud, D, Lawson, R & Thompson, D G 2013. Does student engagement in self-assessment calibrate their judgement over time? *Assessment & Evaluation in Higher Education*, vol. 38, no. 8, pp. 941-956.

Braun, V & Clarke, V 2006. Using thematic analysis in psychology. *Qualitative Research in Psychology*, vol. 3, no. 2, pp. 77-101.

Carless, D, Salter, D, Yang, M & Lam, J 2011. Developing sustainable feedback practices. *Studies in Higher Education*, vol. 36, no. 4, pp. 395-407.

Cicchetti, D V 1994. Guidelines, criteria, and rules of thumb for evaluating normed and standardized assessment instruments in psychology. *Psychological Assessment*, vol. 6, no. 4, pp. 284.

Cohen, L M, Mannion, L L & Morrison, K 2011. *Research methods in education* (7th ed.), Routledge, New York.

Ferguson, P 2011. Student perceptions of quality feedback in teacher education. *Assessment & Evaluation in Higher Education*, vol. 36, no. 1, pp. 51-62.

Field, A 2009. *Discovering statistics using SPSS* (3rd ed.), Sage Publications, London.

Gibbs, G 2010. *Using assessment to support student learning*, Met Press, Leeds, UK.

Gill, P, Stewart, K, Treasure, E & Chadwick, B 2008. Methods of data collection in qualitative research: Interviews and focus groups. *British Dental Journal*, vol. 204, no. 6, pp. 291-295.

Hallgren, K A 2012. Computing inter-rater reliability for observational data: An overview and tutorial. *Tutorials in Quantitative Methods for Psychology*, vol. 8, no. 1, p. 23.

Handley, K & Williams, L 2011. From copying to learning: Using exemplars to engage students with assessment criteria and feedback. *Assessment & Evaluation in Higher Education*, vol. 36, no. 1, pp. 95-108.

Heiman, G W 2001. *Understanding research methods and statistics: An integrated introduction for psychology* (2nd ed.), Houghton Mifflin, Boston.

Hendry, G D, Armstrong, S & Bromberger, N 2012. Implementing standards-based assessment effectively: Incorporating discussion of exemplars into classroom teaching. *Assessment & Evaluation in Higher Education*, vol. 37, no. 2, pp. 149-161.

Jonsson, A & Svingby, G 2007. The use of scoring rubrics: Reliability, validity and educational consequences. *Educational Research Review*, vol. 2, no. 2, pp. 130-144.

Lew, M D, Alwis, W A M & Schmidt, H G (2010). Accuracy of students' self-assessment and their beliefs about its utility. *Assessment & Evaluation in Higher Education*, vol. 35, no. 2, pp. 135-156.

Nicol, D J & MacFarlane-Dick, D 2006. Formative assessment and self-regulated learning: A model and seven principles of good feedback practice. *Studies in Higher Education*, vol. 31, no. 2, pp. 199-218.

Orsmond, P & Merry, S 2011. Feedback alignment: effective and ineffective links between tutors' and students' understanding of coursework feedback. *Assessment & Evaluation in Higher Education*, vol. 36, no. 2, pp. 125-136.

Payne, E & Brown, G 2011. Communication and practice with examination criteria. Does this influence performance in examinations? *Assessment & Evaluation in Higher Education*, vol. 36, no. 6, pp. 619-626.

Pokorny, H & Pickford, P 2010. Complexity, cues and relationships: Student perceptions of feedback. *Active Learning in Higher Education*, vol. 11, no. 1, pp. 21-30.

Poulos, A & Mahony, M J 2008. Effectiveness of feedback: The students' perspective. *Assessment & Evaluation in Higher Education*, vol. 33, no. 2, 143-154.

Price, M, Carroll, J, O'Donovan, B & Rust, C (2011). If I was going there I wouldn't start from here: A critical commentary on current assessment practice. *Assessment & Evaluation in Higher Education*, vol. 36, no. 4, pp. 479-492.

Price, M, Handley, K, Millar, J & O'Donovan, B 2010. Feedback: All that effort, but what is the effect? *Assessment & Evaluation in Higher Education*, vol. 35, no. 3, pp. 277-289.

Reddy, Y M & Andrade, H 2010. A review of rubric use in higher education. *Assessment & Evaluation in Higher Education*, vol. 35, no. 4, pp. 435-448.

Rust, C 2002. The impact of assessment on student learning: How can research literature practically help to inform the development of departmental ssessment strategies and learner-

centred assessment practices? *Active Learning in Higher Education*, vol. 3, no. 2, pp. 145-158.

Rust, C, Price, M & O'Donovan, B 2003. Improving students' learning by developing their understanding of assessment criteria and processes. *Assessment & Evaluation in Higher Education*, vol 28, no. 2, pp. 147-164.

Sadler, D R 1989. Formative assessment and the design of instructional systems. *Instructional Science*, vol. 18, no. 2, pp. 119-144.

Sadler, D R 2005. Interpretations of criteria-based assessment and grading in higher education. *Assessment & Evaluation in Higher Education*, vol. 30, no. 2, pp. 175-194.

Sadler, D R 2009a. Grade integrity and the representation of academic achievement. *Studies in Higher Education*, vol. 34, no. 7, pp. 807-826.

Sadler, D R 2009b. Transforming holistic assessment and grading into a vehicle for complex learning. In Joughin, G (ed.), *Assessment, learning and judgement in higher education*, Springer Netherlands, Houten, pp. 1-19.

Sadler, D R 2010. Beyond feedback: Developing student capability in complex appraisal. *Assessment & Evaluation in Higher Education*, vol. 35, no. 5, pp. 535-550.

Sendziuk, P 2010. Sink or swim? Improving student learning through feedback and self-assessment. *International Journal of Teaching and Learning in Higher Education*, vol. 22, no. 3, pp. 320-330.

Shrout, P E & Fleiss, J L 1979. Intraclass correlations: Uses in assessing rater reliability. *Psychological Bulletin*, vol. 86, no. 2, pp. 420-428.

Smith, C D, Worsfold, K, Davies, L, Fisher, R & McPhail, R 2013. Assessment literacy and student learning: The case for explicitly developing students' "assessment literacy". *Assessment & Evaluation in Higher Education*, vol. 38, no. 1, pp. 44-60.

Taras, M 2003. To feedback or not to feedback in student self-assessment. *Assessment & Evaluation in Higher Education*, vol. 28, no. 5, pp. 549-565.

Wibeck, V, Dahlgren, M A & Öberg, G 2007. Learning in focus groups: An analytical dimension for enhancing focus group research. *Qualitative Research*, vol. 7, no. 2, pp. 249-267.

Wimshurst, K & Manning, M 2013. Feed-forward assessment, exemplars and peer marking: Evidence of efficacy. *Assessment & Evaluation in Higher Education*, vol. 38, no. 4, pp. 451-465.