

Student Self-Assessment Re-Assessed

Katharina Weiss
Deree - The American College of Greece, Greece

Abstract

In a student-centered classroom, learners have to be actively involved both in learning and assessment, which in itself needs to become a learning tool. Therefore, students need to understand and be given the opportunity to apply assessment criteria themselves. Through self-assessment of their writing, they enhance their self-awareness and become autonomous learners capable of self-improvement and meta-cognition (Liang 2014, Nielsen 2012).

Self- and peer assessment are helpful tools that have been discussed in the literature, but the reliability of self-assessment is still debated (see for example Birjandi and Tamjid 2012, Matsuno 2009, Poehner 2012). The present study adds to the existing research by offering data that is not based on observation, but stems from a comparison of self- and instructor assessment where both parties used the same specific assessment rubrics. Assigning numerical values to the rubrics allowed for quantitative results. The data was collected in four classes of students in a course called 'Introduction to Academic Writing' at Deree - The American College of Greece. The outcome of the study did not confirm expectations with regard to reliability of self-assessment, and recommendations for future rubric-based studies are included. Self-assessment should be used as a formative and diagnostic learning tool, especially for weaker students, to foster development of learner autonomy.

Introduction

In an article devoted specifically to self-assessment methods in writing instruction, Nielsen points out the 'limitations in publications on self-assessment' concluding that most of the work published in this area is 'descriptive in nature' (2012: 13). As a result of this observation, Nielsen argues that more reliable measurements must be found and she lists as one of the techniques to be tested: 'the use of rubrics for scoring criteria clarity' (2012: 13). This is the focus of the present paper, in which research undertaken with regard to the reliability of student self-assessment is offered. While not being a highly scientific study, the research presented here is based not on observation alone, but on numerical evidence gathered several times in the course of a semester and thus also allowing for comparison over time. Specifically, the results of this study were obtained by comparing teacher assessment with student self-assessment. To make comparison possible, the same form with specific assessment criteria was used for both teacher and students and descriptors were turned into a numerical scale to permit objective measurements.

While the usefulness of self-assessment is not disputed in the literature, ways to measure its effectiveness are still to be found. The aim of this paper is, then, to add to the existing research on self-assessment by offering quantitative results, which will help colleagues in the field of writing (as well as other subjects) to determine the reliability of rubric-based self-assessment and, if they plan to use it, help them decide what parameters to take into consideration.

The study presented here was carried out at Deree, a division of The American College of Greece, which is located in Athens and accredited by both the New England Association of Schools and Colleges (NEASC) in the US and the Open University of Great Britain. The American College of Greece was founded in 1875 in Asia Minor and is, today, the oldest US-

accredited college in Europe. While including many international learners, its student body consists mainly of Greek students for whom English is typically a second language. This is one of the reasons why three consecutive academic writing courses are required at the undergraduate level. The present research was conducted in the first of these courses called 'Introduction to Academic Writing'. In this course, students read and discuss a variety of non-academic articles that serve as a basis for writing a thematic summary, a critical response essay (CRE) and an extended critical response essay (ECRE) over the course of the semester.

Literature Review

The need to support student-centered learning from an early age all the way through higher education is hardly disputed today, and neither is the concept that assessment is to be considered and used as a learning tool for students rather than a task undertaken by the instructor alone. However, while peer assessment has been popular and a good amount of literature on this subject exists, self-assessment so far has not received the same level of attention. And while it is clear from the point of view of learning theories that both self- and peer assessment are very useful, self-assessment is more debatable or 'somewhat idiosyncratic', as Matsuno (2009: 75) puts it. In some instances, peer assessment has been found to work better than self-assessment. Matsuno (2009) detected less bias in peer raters than in self-raters, while Birjandi and Tamjid (2012) refer to studies in which peer assessment produced better results than self-assessment due to the social interaction it requires. Additionally, Kirby and Downs (2007) suggest that peer assessment should precede self-assessment to give students the opportunity to apply objective criteria, first to others' work, and then to their own.

The most comprehensive studies of self-assessment are still the ones undertaken by Falchikov and Boud (1989). The conclusions in their meta-analysis of self-assessment studies in higher education, along with their critical analysis of quantitative studies, remain both unique in their scope and fundamental to self-assessment research. In a more recent article, Nielsen (2012) offers a useful framework that encompasses the theoretical basis for self-assessment and the various learning theories that underlie it. She focuses on the teaching of writing and also includes an inventory of '[s]trategies for effective implementation of self-assessment methods' which offers good practical advice for any instructors interested in introducing self-assessment into their courses (2012: 9). In her conclusion she notes that:

Inclusion of self-assessment methods in the assessment of writing is likely to foster growth in student writing ability and transfer to future writing tasks. In addition, numerous theoretical models support self-assessment's benefits to writing, the development of critical thinking and the fostering of positive learner behaviours (2012: 13).

This summarizes the crucial advantages that students will gain from doing any kind of self-assessment.

A number of studies focus on both peer and self-assessment as tools to enhance student-centered and autonomous learning. Cowan and Creme (2005) include both types, but while they comment positively on self-assessment, the focus of their article is mainly on peer review. An important aspect that their work raises, however, and one that should be included in future research on the topic, is that instructors also profit and learn from such methods. As they see it: '[the process of peer- and self-review] gave tutors an opportunity to understand better their students' approach to their writing and to evaluate the usefulness of the feedback they gave to students' (Cowan and Creme 2005: 113). Birjandi and Tamjid (2012), in their comparative study of the roles of self-, peer and teacher assessment, have a different focus that confirms the importance of making the classroom more student-centered, shifting away from merely teacher-based assessment methods. While they find that in the cultural setting of Iran teacher focus is of great importance to students, they also conclude that 'both self- and peer assessment, accompanied by teacher assessment, have the potential to improve EFL learners' writing performance', because they can help students 'improve their metacognition' (Birjandi and Tamjid 2012: 529). Similarly, Meusen-Beekman, Joosten-ten Brinke and Boshuizen conclude

in a more recent major study with sixth grade students that both 'peer- and self-assessments have a positive impact on self-regulation' and both lead to 'an increase of positive motivational beliefs' (2016: 134).

There is strong and convincing agreement in literature on the view that self-assessment is useful as long as it is intended for formative, rather than summative, assessment (Nielsen 2012). Matsuno (2009: 75) finds specifically that self-assessment is 'of limited utility as a part of formal assessment' and, therefore, 'it is difficult to recommend using self-assessment for formal grading' (Matsuno 2009: 95). Dearnley and Meddings (2007: 337) found that students do not like to give grades to themselves, as they consider this to be the role of the instructor who possesses the necessary experience, and that there is too much anxiety involved in the process. Andrade (2007) defines the important difference between self-evaluation and self-assessment. As she sees it: 'Self-assessment is formative – students assess their work in progress to find ways to improve their performance. Self-evaluation, in contrast, is summative – it involves students giving themselves a grade' (2007: 60). She reports on research findings which show that students reacted negatively when self-evaluation was used as part of their final grade, i.e. was used as a summative assessment. In contrast, according to her own research, the benefits students listed when doing self-assessment for formative purposes 'included improved ability to focus on key elements of an assignment, increased effectiveness in identifying strengths and weaknesses in their work, and higher motivation' (2007: 60).

More recent research shows that the capability to self-assess is to be built up carefully and developed as a skill over time with the help of the instructor. Thus, Liang (2014) develops a three-stage pedagogy that leads from teacher modeling, to guided peer assessment, to independent self-assessment, which is considered the most difficult stage. Poehner also points out that 'the low correlations [between self-assessment and other measurements] observed raise questions concerning the knowledge and experience necessary to meaningfully self-assess' (2012: 611) and suggests a step-by-step system that enhances self-assessment by means of mediation by teachers. Nidus and Sadler introduce another three-step model in K-12 education in which students are trained in 'how to make careful observations of their writing, set goals, and monitor their own progress – skills that are the backbone of college and career readiness' (2016: 62) – thus making clear the importance of applying self-assessment strategies diachronically.

Preparation and Method Used

Based on my own observations in the classroom, and the evidence in the relevant literature, there can be little doubt that peer evaluation is very productive, and I use it at various stages in all of my classes. To test, however, the reliability of self-assessment as an additional tool in my writing classes, I decided to conduct an experimental study. I am convinced that, just like peer assessment, self-assessment strengthens students' self-awareness, self-confidence, and learner autonomy, helps them to better understand and follow the writing process, and helps them to develop transferable skills. I wanted to understand, however, whether self-assessment could actually be reliable in any measurable way.

I added a student self-assessment component in four of my beginner's courses in two fall semesters, two years apart. The students in these WP 1010 classes are almost exclusively freshmen and in their first semester at college level. Even though I did not specifically ask for feedback, it was quite clear from their reactions that they had not done any kind of self-assessment previously, much less in a writing class.

Assessment and feedback are streamlined in the Writing Program at Deree. For each of the three essays they are required to write, students submit a first draft for formative purposes, and a final draft, which is graded. Instructor feedback on the first drafts consists of short written comments on the essays themselves, and more extensive typed feedback on what was done well and what needs improvement mainly in regard to higher order concerns. Additionally, students also receive feedback on a detailed form with rubrics. On this form there are more than twenty evaluation criteria which refer to aspects such as: 'responding to the prompt',

'focus', 'support', 'structure', 'use of language', and 'professionalism issues'. Checkmarks are given for the five ranges: 'very good,' 'good', 'workable', 'weak', and 'not applicable' (see Appendix 1).

For my research on self-assessment I used this specific checklist, since it offers the most objective criteria and makes comparisons possible. Thus, I had students self-assess the first draft of their essays in class, and I then used the same form to give them my feedback. This seemed an obvious way to obtain comparable numerical data. It also allowed comparison of results over time, specifically over a semester.

To better prepare students for this exercise, and to familiarize them with the jargon inevitably used, we went through the form in class to clarify the meaning of each criterion. In order to give students more time to become familiar with this specific type of assessment, and to give them more exposure to these concepts which were also, of course, referred to in each class session, I did not have them self-assess their first essay, but used the form myself to give them feedback. Additionally, in order for students to understand what characterizes good writing, they were required to work in class on revision exercises and were shown examples of writing to which they were asked to apply the criteria. They also conducted peer reviews, which gave them extensive exposure to the concepts and terminology used in academic writing. All these activities, so I hoped, would help students to have a clear understanding of the rubrics on the form used for their self-assessment.

While making it very clear that the self-assessment was used for formative purposes only, without any influence on their final grade, I also informed students that their self-assessments would be used – in totally anonymous form – for a study I was undertaking.¹ I hoped that letting them know that they were participating in my research would work, at least for some students, as an incentive to cooperate well and fill in the forms carefully and thoughtfully.

When collecting the data, I assigned the number (1) to the descriptor 'weak', (2) to 'workable', (3) to 'good', and (4) to 'very good', and was thus able to find the matches as well as differences between student and instructor assessment overall, noticing particularly the degree to which a student evaluated him-/herself overall better or worse than the instructor's evaluation. The forms contain a great number of items and thus gave a plethora of comparable material, an analysis of which, in its entirety, is beyond the scope of this article. That is the reason why, in this paper, I have chosen to present overall numbers. A study that looks at individual items may be done in a future evaluation. Here, however, I also include the results obtained by comparing the assessed rubrics to the summative grades assigned to the students for the final version of their essay. This allowed further insights into whether student achievement as defined by their grades has any correlation with their ability to self-assess. The tables (1-8) in Appendix 2 present all of these results for the four classes and the two essays.

While a total of 71 students were registered in the classes, comparable material was obtained in regard to the CRE for 50 students and material for the ECRE for 37 students. The missing evidence is due to various circumstances such as students having withdrawn from the course, student absence from class the day we did the self-assessments, or, what was mostly the case, students simply not having filled out the forms completely and/or carefully.

Expectations

Being able to write a good academic essay requires various skills, such as applying a certain structure, using source material appropriately, and writing in a proper tone. It means, in other words, being aware of certain concepts and being able to consciously and purposefully apply them. The level of this awareness, so I assumed, would be reflected in the self-assessment.

¹ The institution granted the author permission to conduct this project involving her students.

My main expectations in regard to the self-assessment, therefore, were as follows:

- a) Good students (defined so by the grades they received for their final draft), would do well on their self-assessment, which, in this specific context, means that the students' self-assessment is very close to the instructor's assessment. These students, therefore, would have more matches with the instructor assessment and there would be a smaller average numerical difference between teacher and student assessment overall.
- b) I was not sure what outcomes to expect in regard to students receiving lower grades on their final essays. Students might be well aware of their essays' shortcomings and thus have a self-assessment that closely matches the assessment carried out by the instructor. However, they could also lack understanding of what does not work in their essays and overestimate their effectiveness in various areas, thereby undermining the reliability of their self-assessment. Another possibility could be that students receiving lower grades are aware of their weaknesses but are not prepared to admit this in a self-assessment and that therefore – again – greater discrepancies between student and instructor findings can be observed.
- c) I expected that there would be an overall improvement in the students' self-assessment from the second to the third essay. This is because students had had more practice in doing the self-assessment and they had had more exposure to the necessary concepts through my own specific feedback and through exposure to these concepts in class over time. In other words, I assumed that self-assessment should become more reliable over time and the more it is practiced.

Results

Contrary to expectations, no clear cut evidence resulted from the comparisons of student and instructor assessment forms, and a considerable number of exceptions make it hard to draw valid and sound conclusions from this study as to the reliability of rubric-based self-assessment.

The following is a description of the results in regard to the expectations listed earlier:

- a) The first expectation, i.e. that higher achieving students would be able to assess their essays more "correctly", was, on the whole, confirmed in three classes in regard to the CRE, while the results for one class did not really allow any such conclusion (Appendix 2, Table 4). In three classes (Appendix 2, Tables 1-3), students receiving higher grades on their final draft had, on average, more matches with the instructor assessment in comparison to weaker students.

As far as the average numerical difference is concerned, a few good students tended to overrate their capabilities somewhat in comparison to what the instructor found. On the other hand, several very good students underestimated their work and were too strict in their self-assessment, resulting in the instructor giving them better assessments than they gave themselves.

However, with the exception of possibly one class (Appendix 2, Table 5), the evaluation of the ECRE showed no correlation between grades and matches in student and teacher assessment (Appendix 2, Tables 5-8). In other words, a higher number of student and teacher matches did not unambiguously correlate to higher grades on the essay or the other way round. The same is true for the correlation of the average numerical difference between the instructor's assessment and the student self-assessment, i.e. it is not possible to draw clear conclusions and the results do not correspond to expectations.

- b) Results show that weak students do not have the ability to assess themselves well. In general, they have fewer matches with the instructor assessment than higher achieving

students have. What is quite clearly confirmed in this study, however, in the evaluation of the results from both essays and in all classes, is that weaker students tend to overestimate their abilities and overvalue their work somewhat or even greatly. This is evident in the fact that these students gave themselves in most cases higher points on the assessment than the instructor did.

- c) The expectation in regard to improvement of student self-assessment over time, i.e. from one essay to the next, was not confirmed. Evidence for this is not presented here as individual students were not to be identified. However, the data I have available allows comparison for 34 students and indicates that the number of matches between instructor and student assessment decreased for 17 students, while four had the same number of matches for both essays, and only 13 students actually had more matches. In regard to the average numerical difference, the picture is only slightly better. Exactly half of all students had a larger difference in their second assessment, while the other half diminished the difference to the instructor assessment, thus indicating that the two values came closer and students' self-assessment was more in tune with the instructor's assessment and can therefore be considered more reliable.

Discussion of Results

One of the outcomes of this study is that good students have a more realistic view of their capabilities and limitations than weaker students do. The fact that stronger students sometimes rate themselves too low was also found by Matsuno who accredited this outcome to cultural factors, i.e. Japanese students' tendency 'to display a degree of modesty' (2009: 94). While possibly more pronounced in Japanese culture, this tendency may well be more generally a human tendency and therefore explain the agreement found here. This, as well as the following outcome of my study, confirms one of Boud and Falchikov's findings in their critical analysis which they summarize as follows:

The general trend in these studies suggests that high achieving students tend to be realistic and perhaps underestimate their performance while low achieving students tend to overestimate their achievements probably to a greater extent than the underestimation (1989: 541).

These same findings were confirmed by Boud, Lawson and Thompson (2015) once again.

Indeed, the present study also confirms results for low achieving students and it is rather striking that weaker students overestimate themselves to the extent that they do. This might imply that these students believe they have understood and can apply certain concepts, while their essay writing demonstrates otherwise. It might also imply, as indicated earlier, that weaker students know about their problems but are embarrassed to hand in a low assessment. While both factors may come into play, the first explanation seems more likely. What is especially intriguing though is the fact that, in my experience, even weak students can do very good peer assessments. This seems to indicate that they have understood concepts, but are not able to apply them in their own writing. This should be a topic for further research. However, rather than simply concluding that weaker students' self-assessments are often unreliable, this outcome should give teachers the incentive to use self-assessment in the future as a diagnostic tool to help identify areas where students need to be given specific support. Instructors can sit down with lower-achieving students and discuss possible reasons for major differences in assessment, thus helping them identify areas for improvement. Self-assessment can help instructors understand where students struggle. Rubric-based self-assessment can help both students and instructors identify specific concepts that were not understood or applied, and self-assessment can thus be used as a learning tool for both.

The last finding in this study, that there was no or little improvement in self-assessment over time, i.e. from the first self-assessed essay to the next in the course of the semester, was unexpected, because common sense would allow one to assume that the more students practice self-assessment the better they get at it. However, in a recent long-term study on

student improvement in self-assessment over time, Boud, Lawson and Thompson found important evidence indicating that 'ability level did have an effect on students' accuracy of judgment' (2015: 51). Indeed, and rather surprisingly, the authors found that over a period of five years:

[t]he high ability group significantly underestimated their grades on all stages of assessment. However, the mid-range group was significantly higher than the tutors at the beginning task but by the end task there was no significant difference between themselves and the tutors (Boud, Lawson and Thompson 2015: 51).

However, the authors also found that lower achieving students showed 'no improvement in their judgments over time' and this means 'that these students are at risk in terms of both their academic performance and their competency to self-assess' (2015: 51). This finding is worrying indeed and future research should be conducted on the reason as to why this happens, i.e. why weaker students do not have the capability to self-assess – even with practice over time. Until an answer is found, in these instances instructors have to step in, and self-assessment can be used as a diagnostic tool, based on which student and instructor together can find solutions for existing problems.

All in all, then, the data obtained in this rubric-based study of self-assessment did not confirm expectations. However, problems more likely lie with the arrangement of the research, rather than with self-assessment itself. I remain, therefore, utterly convinced of self-assessment's usefulness and plan to continue using it while trying to find new ways to measure its outcomes.

Limitations of this Study and Re-Assessment

Obviously, there are a number of limitations to this study. One of them is the rather restricted amount of comparable material obtained, considering the number of students per class. As the high grades in the essays show, mainly students with good abilities completed all self-assessment forms carefully and completely. Thus, there is undoubtedly a bias in the outcomes.

Other limitations are the missing definition as to 'what degree of agreement or correspondence between student and teacher was regarded as acceptable' (Boud and Falchikov 1989: 545) and the fact that all assessment criteria on the form with the rubrics were treated as equal, thus neglecting a possibly necessary distinction between more important criteria like essay structure and less important ones like issues of professionalism. Criteria such as 'logic and relevance of ideas' should obviously carry more weight than 'proofread, spellchecked, presentable essay' (Appendix 1).

In the present study, I also did not look at gender or age differences. This was because, in the case of the former, I did not consider gender important in this context, and, in the case of the latter, my students were all college freshmen. Writing on gender and self-assessment, Boud and Falchikov conclude that '[s]tudies of gender differences remain inconclusive' (1989: 543), while Andrade and Du were surprised to find out that, contrary to their expectations, there was a 'lack of clear evidence of gender differences in responses to academic self-assessment' (2007: 169). To my knowledge, no further study since has come up with reliable results about possible disparities in regard to gender and self-assessment. As far as age differences are concerned, Falchikov and Boud conclude that '[s]enior students taking introductory courses appear not to self-assess significantly better than do first-year students' (1989: 425), thus indicating that age is not an influential factor.

While there are certainly more limitations to the present study, the last one that must be referred to here is that self-assessment is not measured effectively enough when students are given only a form with criteria to complete without any follow up on actual implementation of the results and measuring of – hopefully improved – writing outcomes. Therefore, re-assessing my own approach to self-assessment, I offer the following recommendations:

Rubric-based self-assessment is useful as specific concepts that can be quite clearly defined give a basis for comparison. The number of rubrics used should, however, be limited so as not to overtax students.

Contrary to what was done in this study, and to what Nielsen (2012) suggests, self-assessment does not have to be an in-class activity. Giving students ample time to do carefully structured self-assessment activities at home should give more accurate and thoughtful results to the advantage of both students and instructors. This, of course, requires very careful preparation and explanation of the self-assessment exercise and extensive teacher support in class.

Students should be involved in creating rubrics (Andrade 2007). Amongst the important strategies for implementing self-assessment, Nielsen also lists this component with the rationale that:

Working with students to create a shared understanding of good writing provides an opportunity for students to learn, through exploration, the qualities that will be expected in their assignments [...] Students are also more likely to use self-assessment methods correctly because they will have a greater understanding of the process, having taken part in their development (2012: 10).

Above all, it is crucial for instructors to realize how important preparation and guidance for self-assessment are. While the very nature and name of it seems to imply that self-assessment is a task that is up to the students and undertaken by them alone, it cannot be stressed enough that instructor involvement is essential and that the building of self-assessment skills needs to be carefully structured. This is emphasized by Falchikov and Boud who stress that self-assessment should 'be regarded as a skill and, as such, needs to be developed' (1989: 426). Poehner describes specific steps undertaken in L2 language learning applying a 'kind of cooperative self-assessment' (2012: 621) closely involving a teacher in the process. He underlines 'that learner efforts to self-assess must be carefully mediated as they move from a *cooperative* to an *independent* mode of self-assessment' (2012: 610).

Equally crucial is following up on the actual self-assessment exercise. Students have to be held accountable for their findings and the implementation of corresponding revisions in their future work. Giving students the chance to act upon their self-assessment will lead to both better writing outcomes and increased motivation (Meusen-Beekman, Joosten-ten Brinke and Boshuizen 2016).

Even if many studies on self-assessment, including the present one, may not offer clear cut or expected findings, and even if research results are difficult to obtain because assessing self-assessments remains a major challenge, especially in writing courses where it means measuring improvements in the process of writing, it is of great importance to have students at all levels do self-assessment and have them do it often. Nearly all studies on self-assessment, no matter what results they find, end with the recommendation, based on findings as well as teaching experience, that self-assessment is fundamental in (writing) classrooms because it furthers important and transferable skills as well as meta-cognition. McMillan and Hearn put it like this:

Self-assessment could mean that students simply check off answers on a multiple-choice test and grade themselves, but it involves much more than that [...] [It is] a process by which students 1) monitor and evaluate the quality of their thinking and behavior when learning and 2) identify strategies that improve their understanding and skills (2008: 40).

This aspect of self-assessment, that is, the emphasis on training not in essay writing or any other specific subject or skill, but training in meta-cognition, in thinking about how and why students do what they do, makes any exercise in self-assessment valuable.

Conclusion

In this paper, research on rubric-based self-assessment is presented which offers results that are not based on outcomes of classroom observation alone. Having students fill in the same form that their instructor uses for feedback on a first draft allows for objective and numerical comparison of results. As expected, and confirming previous research findings, well-performing students' self-assessment is closer to teacher assessment than that done by weaker students. Some very good students tend to underestimate their abilities, while weaker students overestimate their skills, thus having greater differences to the instructor's assessment both in number of matches as well as in overall numerical difference. The result that especially low achieving students do not seem to be able to self-assess adequately is worrying and will need further research. In the meantime, rubric-based self-assessment can be used to identify areas where weaker students need special support. The expected clear improvement in self-assessment over time, i.e. in the course of a semester from one essay to the next, was not confirmed and, overall, self-assessment did not seem reliable in the way it was measured.

While this outcome may have to do with the limitations of the study, the required careful preparation of students for their self-assessment had taken place. Therefore, recommendations for future self-assessment studies include limiting the number of criteria students are asked to apply, giving students enough time to do the self-assessment, and involving students in the creation of the criteria themselves with careful guidance by the instructor at all stages. Also, follow-up on the self-assessment is necessary. Students should be held responsible for applying results of their self-assessment in future writing tasks, thus allowing monitoring of the actual process of writing and outcomes of self-assessment.

In spite of the difficulties associated with measuring the effectiveness of student self-assessment, its continued application for formative purposes is important. As supported by learning theories and a good number of studies, self-assessment requires students to think and develop meta-cognition. It fosters autonomous learning, increases motivation, and supports, with continuous practice, the development of transferable skills.

References

- Andrade, H. (2007) 'Self-Assessment Through Rubrics'. *Educational Leadership* [online] 65 (4), 60-63. available from <<http://www.ascd.org/publications/educational-leadership.aspx>> [9 June 2016]
- Andrade, H., and Du, Y. (2007) 'Student Responses to Criteria Referenced Self-Assessment'. *Assessment and Evaluation in Higher Education* 32 (2), 159-181
- Birjandi, P., and Tamjid, N.H. (2012) 'The Role of Self-, Peer and Teacher Assessment in Promoting Iranian EFL Learners' Writing Performance'. *Assessment and Evaluation in Higher Education* 37 (5), 513-533
- Boud, D., and Falchikov, N. (1989) 'Quantitative Studies of Student Self-Assessment in Higher Education: A Critical Analysis of Findings'. *Higher Education* 18 (5), 529-549
- Boud, D., Lawson, R., and Thompson, D. (2015) 'The Calibration of Student Judgement through Self-Assessment: Disruptive Effects of Assessment Patterns'. *Higher Education Research and Development* 34 (1), 45-59
- Cowan, J.K., and Creme, P. (2005) 'Peer Assessment or Peer Engagement? Students as Readers of Their Own Work'. *Learning and Teaching in the Social Sciences* 2 (2), 99-119
- Dearnley, C.A., and Meddings, F.S. (2007) 'Student Self-Assessment and its Impact on Learning—A Pilot Study'. *Nurse Education Today* [online] 27, 333-340. available from <<http://dx.doi.org/10.1016/j.nedt.2006.05.014>> [18 May 2017]
- Falchikov, N., and Boud, D. (1989) 'Student Self-Assessment in Higher Education: A Meta-Analysis'. *Review of Educational Research* 59 (4), 395-430
- Kirby, N.F., and Downs, C.T. (2007) 'Self-Assessment and the Disadvantaged Student: Potential for Encouraging Self-Regulated Learning?' *Assessment and Evaluation in Higher Education* 32 (4), 475-494
- Liang, J. (2014) 'Toward a Three-Step Pedagogy for Fostering Self-Assessment in a Second Language Writing Classroom'. *The CATESOL Journal* [online] 26 (1), 100-119. available from <<http://www.catesoljournal.org>> [29 June 2016]
- Matsuno, S. (2009) 'Self-, Peer-, and Teacher-Assessments in Japanese University EFL Writing Classrooms'. *Language Testing* 26 (1), 75-100
- Meusen-Beekman, K.D, Joosten-ten Brinke, D., and Boshuizen, H.P.A. (2016) 'Effects of Formative Assessments to Develop Self-Regulation Among Sixth Grade Students: Results from a Randomized Controlled Intervention'. *Studies in Educational Evaluation* 51, 126-136
- McMillan, J.H., and Hearn, J. (2008) 'Student Self-Assessment: The Key to Stronger Student Motivation and Higher Achievement'. *Educational Horizons* [online] 87 (1), 40-49. available from <<https://eric.ed.gov/?id=EJ815370>> [11 October 2016]
- Nidus, G., and Sadder, M. (2016) 'More Than a Checklist'. *Educational Leadership* [online] 73 (7), 62-66. available from <<http://www.ascd.org/publications/educational-leadership.aspx>> [9 June 2016]
- Nielsen, K. (2012) 'Self-Assessment Methods in Writing Instruction: A Conceptual Framework, Successful Practices and Essential Strategies'. *Journal of Research in Reading* 37 (1), 1-16
- Poehner, M. (2012) 'The Zone of Proximal Development and the Genesis of Self-Assessment'. *The Modern Languages Journal* 96 (4), 610-622

Appendix 1

(Form reprinted here by permission of the Writing Program at Deree - the American College of Greece.)

WP 1010: Introduction to Academic Writing

Analytical Rubric for **Critical Response Essay – First Draft**

Name: _____ Topic/Title: _____

EVALUATION CRITERIA	ASSESSMENT RANGE				
	Very Good	Good	Workable	Weak	N/A
Responding to the Prompt					
Understanding of the source's ideas					
Responding to a text using another text					
Focus					
Introduction places the topic in a wider context, and leads to thesis					
In the introduction identify the text engaged with (e.g. article title, author, date, publication)					
Focused summary of the key points of the main text					
Thesis expresses a response to the main idea of the text and relevant secondary ideas					
Sustained discussion of chosen aspect of topic in body paragraphs: Unity					
Consistent point-of-view in representing the ideas of the source					
Conclusion effectively restates the text's main points/ideas, while emphasizing an understanding of the topic					
Support					
Introduction of second source (used to critique the main text); provide accurate summary of the second source's main idea and its full bibliographic information					
Second source used in specific and convincing ways to support the writer's views of the main text					
Quotations, summaries and paraphrases appropriately and meaningfully integrated (e.g. introducing-framing-explaining or blending)					
Other supports used in specific and convincing ways (e.g. from personal observation or experience)					
In-text documentation of evidence					
Structure					
Logic and relevance of ideas					
Paragraph development (main idea and evidence /examples)					
Conceptual/transitional links between paragraphs					
Links between ideas within paragraphs (coherence)					
Use of Language					
Sentence variety, conciseness					
Appropriate use of certain verbs for summary / quotation / paraphrase					
Tone					
Grammar and Syntax					
Professionalism Issues					
Proofread, spellchecked, presentable essay					

Within the required word range		YES		NO	
Work Cited List (MLA style)					
Appropriate submission		Hard Copy	Blackboard	Turnitin	
Writing specifications (title, page numbers, double-spaced, 12-pt. font, appropriate info cited in the right- or left-hand corner, including word count)		Comments:			
Deadline	On time submission	Late Work delivered:			

NB: Also refer to the summary comments your instructor has made on the cover sheet and on your paper.

Appendix 2

Results from the comparison between teacher and student self-assessment for Critical Response Essay (CRE) and Extended Critical Response Essay (ECRE) for four classes.

Explanations:

- In the above evaluation criteria (4) points were assigned for 'very good', (3) for 'good', (2) for 'workable' and (1) for 'weak'.
- Matches: Number of times student and instructor assessment coincided.
- Grade on essay: Final grade student received for the specific essay.
- Difference avg.: overall numerical difference between assessment by instructor and by student. A negative value indicates that the instructor gave less points; a positive value that the instructor gave more points than the student assigned to him/herself.
- Higher points by instructor: Indicated by a + sign: How many more points than the student, in total across all rubric categories, the instructor assigned to the student's work.
- Higher points by student: Indicated by a – sign: How many more points than the instructor, in total across all rubric categories, the student assigned to her/his work.

Table 1: Class 1, CRE

Matches	Grade on essay	Difference avg.	Higher points by instructor	Higher points by student
5	F	-1.208	2+	31-
7	B	-0.667	1+	21-
7	C+	-0.250	7+	13-
8	A	-0.333	2+	15-
8	B	+0.087	9+	7-
8	C+	-0.208	7+	12-
9	B	+0.542	17+	4-
9	A	-0.130	6+	9-
11	A-	+0.200	4+	8-
12	B	-0.455	2+	9-
13	A-	+0.208	8+	3-
18	A-	+0.042	4+	1-

Table 2: Class 2, CRE

Matches	Grade on essay	Difference avg.	Higher points by instructor	Higher points by student
2	F	-1.250	2+	17-
2	C	-1.458	2+	32-
3	F	-1.000	4+	28-
3	C+	-0.250	3+	16-
5	C	-0.917	4+	26-
6	A-	-1.458	1+	36-
6	C	-1.167	1+	27-
8	B	-0.375	7+	16-
9	A	-0.167	5+	10-
16	A	+0.083	4+	4-

Table 3: Class 3, CRE

Matches	Grade on essay	Difference avg.	Higher points by instructor	Higher points by student
2	C	-1.083	4+	30-
3	A	+0.792	23+	1-
3	C+	-0.333	6+	17-
4	F	-1.000	0+	24-
4	F	-0.958	0+	32-
5	B	-1.333	4+	20-
5	A-	-0.667	6+	18-
6	B	-0.500	12+	5-
7	C+	+0.167	5+	18-
7	A-	-0.542	9+	10-
8	A	+0.708	19+	3-
11	B	+0.083	4+	9-

Table 4: Class 4, CRE

Matches	Grade on essay	Difference avg.	Higher points by instructor	Higher points by student
2	B	-0.667	2+	18-
3	A	-0.792	3+	22-
6	C+	-0.750	1+	19-
6	A-	+0.417	15+	5-
7	B+	-0.417	6+	16-
7	A	+0.125	13+	10-
7	B	-0.583	6+	20-
8	C+	-0.500	3+	14-
9	C	-0.583	3+	17-
10	B+	-0.333	4+	12-
10	A	-0.500	2+	14-
10	A-	+0.500	17+	5-
11	B+	-0.125	6+	9-
12	A	-0.292	2+	9-
13	B+	-0.083	5+	7-
13	A	+0.250	9+	4-

Table 5: Class 1, ECRE

Matches	Grade on essay	Difference avg.	Higher points by instructor	Higher points by student
4	B	-0.696	5+	21-
6	C	-0.913	2+	23-
6	A-	+0.435	14+	4-
6	C	-1.087	2+	27-
9	A	-0.174	5+	9-
10	C	+0.087	8+	6-
13	A	-0.217	3+	8-
14	B+	-0.217	3+	8-
15	B+	0.000	4+	4-

Table 6: Class 2, ECRE

Matches	Grade on essay	Difference avg.	Higher points by instructor	Higher points by student
3	F	-1.087	1+	26-
4	C+	-1.261	0+	29-
4	C	-1.391	0+	32-
4	B	-0.696	1+	22-
5	B	-0.783	3+	21-
5	C	-1.000	0+	24-
9	C+	-0.652	3+	18-

Table 7: Class 3, ECRE

Matches	Grade on essay	Difference avg.	Higher points by instructor	Higher points by student
3	A	+0.304	15+	8-
5	F	-1.087	2+	27-
5	A-	-0.043	10+	11-
6	A	+0.913	22+	1-
9	A	+0.609	15+	1-
10	B+	-0.130	6+	9-
10	A-	-0.087	8+	10-
10	A-	-0.478	4+	15-

Table 8: Class 4, ECRE

Matches	Grade on essay	Difference avg.	Higher points by instructor	Higher points by student
5	A	+0.174	8+	12-
6	B+	-0.043	5+	6-
6	A-	+0.913	23+	2-
7	C	-0.478	4+	13-
8	B	+0.174	11+	7-
9	A	+0.261	10+	4-
10	A-	+0.348	11+	3-
10	B	-0.391	4+	13-
11	A	-0.391	1+	10-
12	B	+0.130	8+	5-
12	A	+0.043	6+	5-
14	A	+0.087	5+	3-
17	A	+0.087	4+	2-