

**Peer Assessment in Translation for Language Learning: R23201 2014-15**  
**Jerome Carroll**  
**TRI-CAMPUS LANGUAGE TEACHING CONFERENCE**  
**UoN, 9-15 JULY 2015**

**Peer Assessment background**

Background comments re peer assessment are taken from F. Dochy, M. Segers and D. Sluijsmans, 'The use of self-, peer and coassessment in higher education: A review', *Studies in Higher Education*, 24:3 (1999), 331-350. This presents a survey of 63 articles on self-, peer- and co-assessment.

The main focus of the survey is on validity, fairness, accuracy of peer assessment.

Re issues of validity and accuracy, it has to be said that as many articles indicate the invalidity or inaccuracy of peer assessment (e.g. Dancer & Dancer 1992) as those that indicate its validity or accuracy (e.g. Topping 1998):

**Peer assessment is inaccurate:**

'The results of a study by Orsmond et al. (1996) are less positive regarding accuracy of peer assessment. [...] Orsmond et al. found that there was 18% agreement between students and tutor, with 56% of the students overmarking and 26% of the students undermarking.'

'Stefani (1992) reported 14% complete agreement between students and tutor, 58% of students overmarking and 28% undermarking. Peer marks (both in the case of under- and overmarking), however, did differ by less than 10 percentage points from the tutor.'

**Peer assessment is accurate:**

'Oldfield & Macalpine (1995) investigated the competence of students in making assessments. [...] Each task was assessed by the peer group and compared with the assessment of the lecturer. Results show high correlations between student marks and lecturers' marks for individual essays and presentations.'

'Fry (1990) describes a study in which the tutor introduced peer marking. The tutor first marked the scripts of the students and then handed them over to the students. The tutor asked the students to mark each others' work according to a marking scheme. The agreement between the tutor marks and the students' marks was generally very high.'

'Fry's findings are confirmed by Rushton, et al. (1993), who developed a computerised peer assessment tool. [...] The class was split into groups of three or four students. Each group member used a peer assessment 'window' to mark the others' work. Contrary to expectations, the marks awarded by the peers were remarkably similar to those

awarded by the tutors, suggesting that peer and teacher assessments were equally reliable.’

These issues of accuracy/validity/fairness are important – but my concern was not to introduce peer assessment as a method of assessment, albeit an alternative and equally fair, accurate and valid means of assessment – rather to introduce it as a tool for learning.

This touches on Dochy et al’s view of alternative assessments like PA as an important components of an ‘era of assessment’, as distinct from the ‘era of testing’, in which:

- student is an ‘active person who shares responsibility’
- assessment is a ‘tool for learning’ (After, 1997; Dochy & McDowell, 1997)
- use of methods that encourage ‘meta-cognitive competencies such as self-reflection and self-evaluation’

The survey by Dochy et al does register this issue of learner responsibility: under the concept of ‘effects’:

- ‘Keaten et al. (1993) report that peer assessment is a practice that can foster high levels of responsibility among students, requiring that the students be fair and accurate with the judgments they make regarding their peers.’
- In the Orsmond et al. (1996) study, the students filled in a questionnaire which showed that 76% of them thought that ‘the peer assessment had make them think more, and work in a more structured way’ (p. 243)

### **Particular challenges of translation for language learning**

Why is the responsibility and reflection that comes with PA so important for translation for language learning? In my view there are particular challenges that PA might address, and three interrelated reasons why I wanted to introduce an element of peer assessment into this module:

- firstly, I discerned a certain level of ignorance among students about the specifics of the marking process, even if they had access to formal assessment criteria;
- secondly, I wanted to get them to think about the connection between published marking criteria and how we actually mark their work;
- thirdly, I wanted to instil in students the need to concentrate on every word as they undertake assessments, whether these are take-home assessments where they have lots of time, or controlled conditions exams, in which they have barely an hour to complete a similar task – i.e. not thinking more so that the assess more accurately/reliably, but so that they perform tasks more consciously and therefore better

### **Module background**

The module in which the peer assessment pilot was run is the team-taught Y4 translation module, R23201 Translation from German 2, which is one of two core language modules that Y4 students of German take. The other core module, called Essay and Oral, focuses on free writing and spoken language. The translation module uses translation as a vehicle for language learning. In the 2014-15 session there were 75 students on the module, of whom all took part in the PA, and 65 responded to the questionnaire.

### **Assessment of the module**

The module also uses translation tasks as the mode of assessment of the module, which comprises three elements: one take-home coursework, one in-class translation test, and one end-of-semester exam. There is gradual withdrawal of support, insofar as students may use dictionaries and on-line dictionaries in the first (take-home) coursework, they may use print dictionaries in the in-class test, but have no dictionaries in the exam. Each translation text is circa 150 words long, and translation in the assessments is always from German into English.

### **Marking translations**

The marking of the assessment is guided by the CLAS assessment criteria for translations into English (Appendix A). Markers employ a system of mistake points and bonus points. Deviation from accepted versions are penalised with a range of mistake points, from half-point to one and a half point, depending on the severity of the mistake in comprehension or rendering. For example, a punctuation errors or stylistic lapses which do not greatly affect the rendering of the original are penalised with a half-point error. Clear vocab errors, where the word or individual phrase chosen leads to a clear loss of meaning of the original, are penalised with a single point error. Errors that range over several words or a clause, where one error may be seen to lead to or be connected with the other errors, are penalised with a one-and-a-half point error. Bonus points (of either half or whole point value) are awarded, not for renderings that are simply acceptable and without error, but where particular fluency is achieved whilst retaining faithfulness to the original meaning.

The errors scores that are generated by deducting bonus marks from mistake points are then converted to marks expressed as a percentage. CLAS assessment criteria for translations into English are consulted at this point to make sure that the class marks awarded appropriately reflect the performance.

### **The PA process**

The 90-word text for the translation exercise which was to be peer-assessed was issued in Week 2, and students were instructed to bring in 3 copies of their version of it into class in Week 3. These were then exchanged in groups of 4, to be marked and discussed in Week 4.

### **Marking instructions**

The following instructions were given to students, together with a copy of the CLAS assessment criteria for translation into English:

Please mark the translations by your 3 peers in the following manner:

- Major errors, i.e. those that clearly distort the meaning of the original – these you should double underline
- Minor errors, i.e. those that are more stylistic and do not greatly effect the meaning, or punctuation errors – these you should single underline

You should also put a dotted line under rendered elements that seem not quite ideal but which you would not penalise

- Also, award bonuses – denoted with a tick – for exceptionally fluent and accurate renderings of words, phrases or whole sentences

Add up the tally of each and note it at the bottom of the text.

Prepare for a discussion in Week 4 of what you allowed, penalised (with a minor or major error) and awarded a bonus to.

Also look at the attached Assessment Criteria, with a view to thinking and discussing how we ‘translate’ the distribution of errors into a class mark.

(See Appendix B: Week 4 (Science) Peer Assessment Instructions and Appendix A: CLAS Marking Criteria for Translations into English)

### **The questionnaire**

Class teachers discussed the marking experience with students, though no data was gathered from these informal discussions. Rather, each student was given a questionnaire to fill in (see Appendix C).

Several questions related to students’ awareness of the assessment criteria and their confidence in applying the assessment criteria. Several questions related to the PA process itself. In this presentation I’ll focus primarily on the three questions relating to student’s self-perception before and after peer assessment. These addressed three key aspects of students engagement and understanding of assessment: understanding of assessment criteria, confidence in judging own work, finally confidence in ability to improve their performance. The responses were measured on a 5 point Likert scale, analysed with the Wilcoxon Signed Ranks test. This tests the difference between two

related samples – for instance before and after, as here - in order to establish whether the distribution of differences between the two is statistically significant.

Open qualitative comments are analysed thematically.

### The Results: quantitative results

#### Before peer assessment

Questionnaire responses show that:

- Most students had come across assessment criteria (48 YES; 17 NO)
- Majority of students were not using assessment criteria (18 YES; 47 NO)

#### Comparisons before and after peer assessment

#### Students' understanding of assessment criteria

Figure 1 below shows that students' understanding of assessment criteria before and after the peer assessment exercise was not affected. The Wilcoxon Signed-Rank test confirms what the frequencies show.

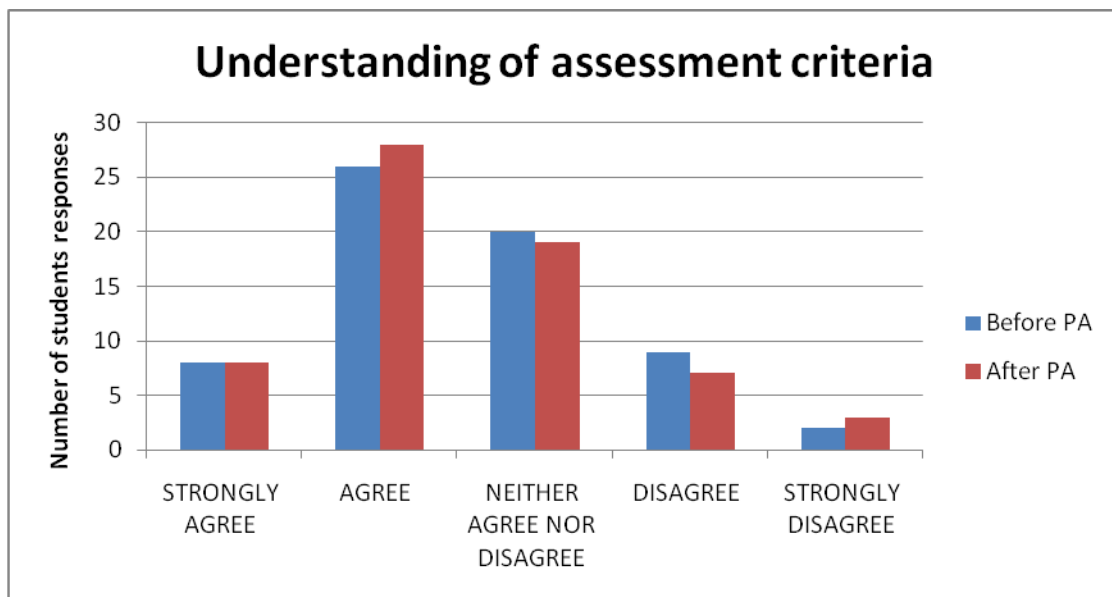


Fig. 1 Student ratings of their understanding of assessment criteria before and after peer assessment

#### Students' perception of their ability to judge their own translations before and after peer assessment

Figure 2 below shows a slight increase in the level of confidence reported by students after peer assessment in relation to their ability to judge their own translation work. The Wilcoxon Signed-Ranks test compared the ratings before and after by each individual students.

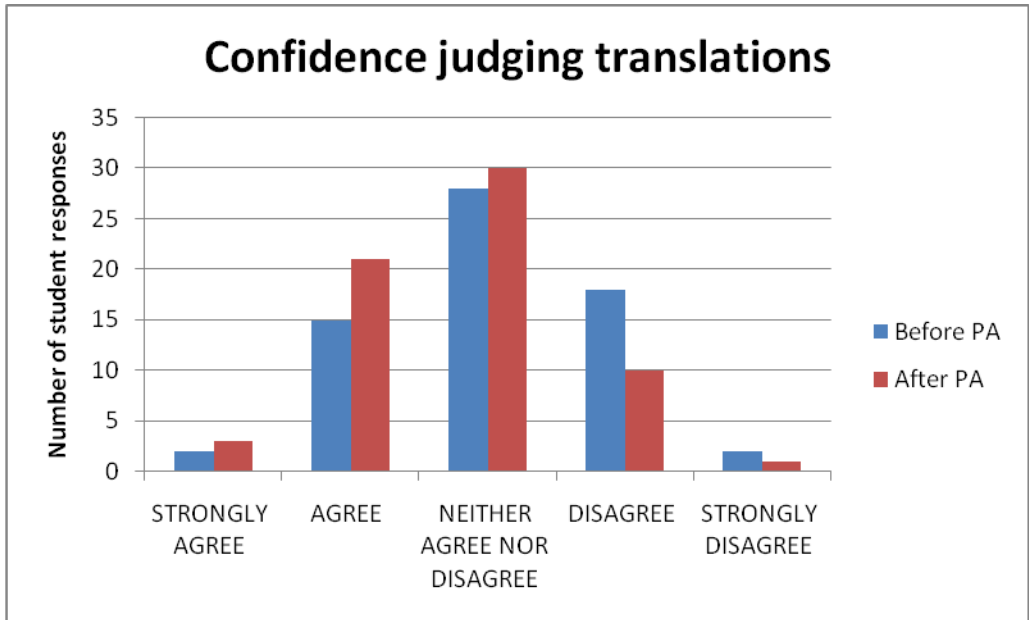


Fig. 2 Self-reported confidence in judging own translation work before and after peer assessment

Figure 3 below shows the group averages and the error bars. The test result confirms that the change in self-perception is statistically significant. The Z-value is -1.9746. The p-value is 0.04884. The result is significant at the  $p \leq 0.05$  level of significance.

Fig. 3 Average level of confidence in judging translations before and after peer assessment

**Students’ perception of their ability to improve their performance**

Figure 4 below shows a slight increase in the level of confidence reported by students after peer assessment in relation to their ability to improve their performance. The Wilcoxon Signed-Ranks test compared the ratings before and after by each individual students.

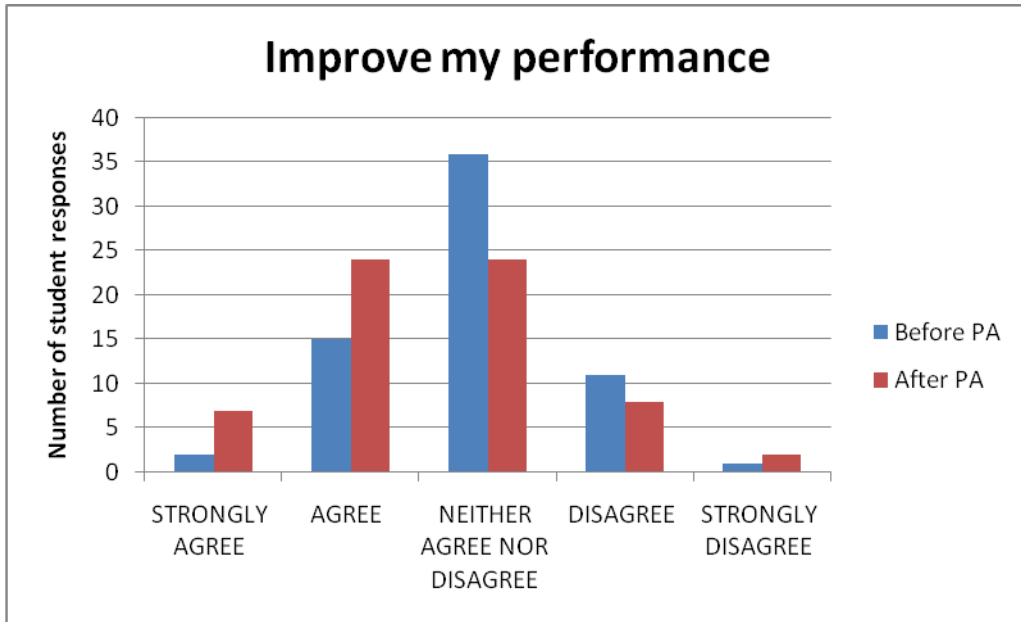


Fig. 4 Self-reported confidence in ability to improve performance in the module before and after peer assessment

Figure 5 below shows the group averages and the error bars. The test result confirms that the change in self-perception in this area is statistically significant. The Z-value is -2.528. The p-value is 0.0114. The result is significant at the  $p \leq 0.05$  level of significance.

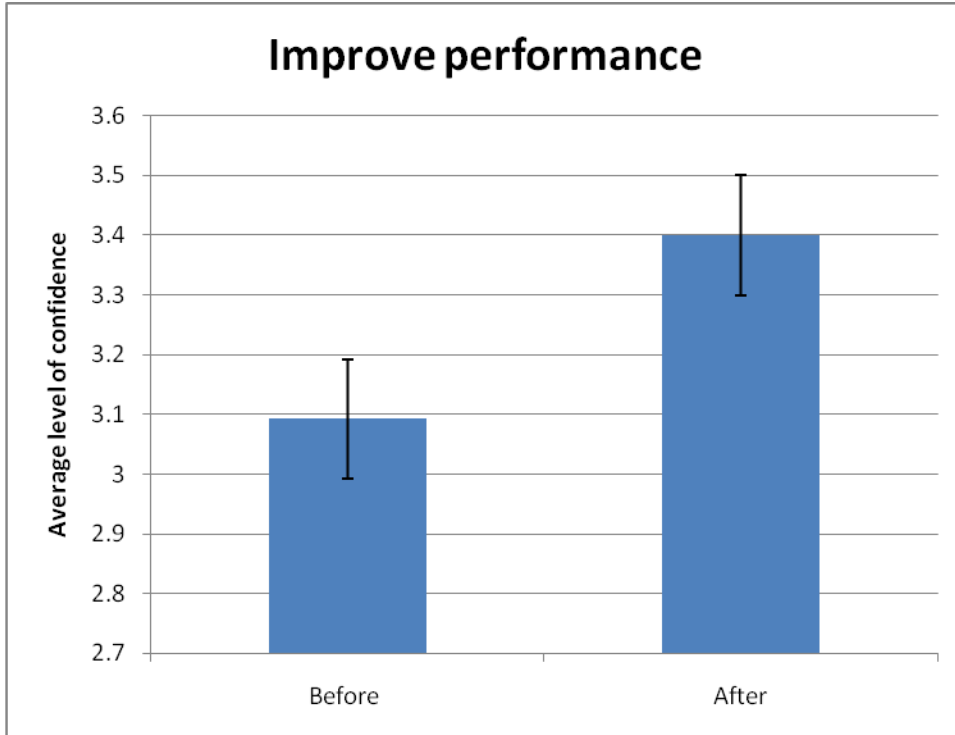


Fig. 5 Average level of confidence to improve performance before and after peer assessment

### **Perceptions of value of the task**

34 of 65 would do the task again – half the students would do the task again

14 of 65 would not do it again

17 of 65 did not reply to this question

### **The Results: Qualitative results**

Qualitative results were collated in the form of open comments on value of the task. Open comments were invited in three of the 21 questions Here I have compiled all of the remarks in the open comments sections of the questionnaire, in response to questions 18, 19 and 21. N.b. that 17 of 65 did not answer these questions.

18. Having completed the marking for this exercise, would you change the marks you had initially awarded?

#### **List changes to marking and reason for this below:**

- I found more errors
- I gave bonuses too readily
- I should only bonuses for really good, idiomatic renderings
- I gained more of a sense of what counts as major/minor errors
- I realised I had been too harsh/would be less harsh
- I wasn't strict enough

19. Did this exercise help you understand how you can improve your translations in future? If so, please describe what you plan on doing that you weren't considering prior to this exercise:

#### **Negative comments**

15 of 65 students responded negatively to this question

Subjective, difficult, unjustified judgements

- It just made me worry because it seems subjective
- It's a little embarrassing having others judge your work
- Very difficult to judge eg the difference between good and average
- I feel we didn't know enough to judge others
- I didn't feel confident about applying the criteria accurately

Timing

- Maybe have it not so early in the module

#### **Positive comments**

33 of 65 students responded positively to the question asking them if the task had helped them to improve their translation in the future. The comments mostly referred to focusing on specific aspects of their translation (fluency, accuracy, vocabulary,



different types of errors). Four of these 33 comments related explicitly to understanding the marking process and enhanced self-assessment.

#### Different perspective

- It was useful to see opinions/translations of others
- It helped to see different approaches/ways of phrasing things/vocab choices
- I think I can recognise my mistakes/where I will lose marks better

#### Concentration on detail

- I will pay more attention to detail, tenses, particular phrases
- I will concentrate on translating each and every word
- I will consider every different meaning of a word before writing it down
- I will learn more vocab/spend more time thinking about individual items of vocab

#### Comprehension

- I will take more time to read the original text
- I will reread the original text more before beginning translation

#### Accuracy

- It helped me to realise that a clunky yet semantically accurate rendition of a text is better (sometimes) than a flowing piece that lacks exact meaning
- I'll stick closer to the German text
- I will avoid letting a desire for fluency get in the way of accuracy

#### Fluency

- I will concentrate on fluency
- I will make sure my translation sounds idiomatic
- I'll pay more attention to revising my translation after a cool-off period
- I will take a more idiomatic approach
- It helped me to think about my English phrasing
- I will not worry too much about sticking closely to SL
- I will think about fluency to gain bonus points
- I will focus on restructuring sentences to make them sound more English

#### Consciousness of marking process

- It was very useful to understand how our translations will be marked
- Better understanding of what things you lose marks for and what you get penalised for
- Showed the importance of understanding what was major/minor error
- Helped me see mistakes I hadn't been aware of before

21. What could be more helpful to you in understanding assessment criteria?

Requests for:

- concrete examples of marked work at the different levels
- an actual mark from the teacher/a text with real marking by the lecturer
- more practise/peer assessments

### **Summary and conclusion**

Peer assessment was not seen to contribute to students' understanding of assessment criteria.

Peer assessment did contribute to improving students perception of their ability to judge their own work and to improve it.

Peer assessment can support students in a way that goes beyond knowledge of the assessment criteria, to understanding about how those criteria may be applied.

### **Module results**

The module average in semester 2, 2014-15 was higher than in previous years, and semesters, though not by enough to draw any reliable conclusions:

S1 2011/12 64
S2 2011/12 63
S1 2012/13 59
S2 2012/13 60
S1 2013/14 62
S2 2013/14 64
S1 2014/15 60
S2 2014/15 65

Fig. 6 Module averages since 2011-12.

### **Appendixes**

- a) CLAS Assessment Criteria for Translation into English
- b) Week 4 (Science) Peer Assessment Instructions
- c) Questionnaire