

NEWS On June 5th, Trevor Holster, J. Lake and Bill Pellowe will give a paper on the Peer Assessment add-on for MOARS at the [36th Language Testing Research Colloquium](#) (4-6 June 2014) in Amsterdam.

Title: Many-faceted Rasch analysis of peer-assessment as a diagnostic tool

Trevor A. Holster (Fukuoka Women's University, Japan)
J. W. Lake (Fukuoka Women's University, Japan)
William R. Pellowe (Kinki University Fukuoka, Japan)

Venue: VU University Amsterdam, The Netherlands

Date: Thursday 5 June

Language proficiency frameworks can guide curriculum planners and classroom teachers, exemplified by Hadley's (2001) accessible introduction to the ACTFL framework as the basis for instructional planning. The use of rubrics in instruction requires learners to both understand the rubric and to be able to assess the strengths and shortcomings of their own work relative to the rubric though, presupposing that learners have sufficient metalinguistic knowledge to understand the rubric, an assumption that Tokunaga (2010) challenges. Previous research has supported the formative use of peer assessment but found peer assessors to be inconsistent in their interpretation of rating rubrics (Cheng & Warren, 2005; Farrokhi, Esfandiari, & Schaefer, 2012; Mok, 2011; Saito, 2008). However Toppings (1998) notion of "learning by assessing" holds that interaction with the rubric during peer assessment can drive learning suggesting that instructional explanations based on proficiency frameworks may be less effective than using peer assessment as a mechanism to improve students understanding of the rubric. Additionally, fit analysis of peer assessors' interpretation of the rubric can guide remedial instruction by identifying rubric items that students struggle to interpret.

Peer assessment was piloted in academic writing classes at a Japanese women's university (n = 24). Students assessed each others' essays using a 9-item rubric and entered their ratings into an on-line database. Many-faceted Rasch analysis found general agreement in rank ordering between students' ratings and teachers' ratings, but that students tended to rate

holistically, did not use the full range of the rating scale, and were much more lenient than teachers. The rating patterns provide evidence that students were unable to interpret the rubric clearly rather than simply interpreting it differently than teachers, meaning that as well as being unable to provide diagnostic feedback to each other, students were unlikely to understand feedback from teachers. By anchoring the difficulty of the rubric items against teacher ratings using the Facets software package, the most misfitting items when rated by peer assessors could be identified, these being "Introduction", "Thesis statement", and "Conclusion". In their second essay, all rubric items improved substantively, but "Thesis statement" was the only item showing substantively and statistically significant improvement greater than overall improvement.

These pilot results indicated points of weakness in the instructional materials and also suggested that the observed gains in proficiency were more likely due to practice and learning by assessing than to instruction and feedback. Revised instructional materials were produced to address the problematic rubric items and are being operationally piloted during the second semester of 2013 (n = 105). Results due in February 2014 will confirm or disconfirm whether the revised instruction improved student understanding of the rubric and how this affected gains between the first and second essays produced by students.

Cheng, W., & Warren, M. (2005). Peer assessment of language proficiency. *Language Testing*, 22(1), 93-121. doi: 10.1191/0265532205lt298oa

Farrokhi, F., Esfandiari, R., & Schaefer, E. (2012). A many-facet Rasch measurement of differential rater severity/leniency in three types of assessment. *JALT Journal*, 34(1), 79-101.

Hadley, A. O. (2001). *Teaching language in context* (3rd ed.). Boston: Heinle and Heinle.

Mok, J. (2011). A case study of students' perceptions of peer assessment in Hong Kong. *ELT Journal*, 65(3), 230-239. doi: 10.1093/elt/ccq062

Saito, H. (2008). EFL classroom peer assessment: Training effects on rating and commenting. *Language Testing*, 25(4), 553-581. doi: 10.1177/0265532208094276

Tokunaga, M. (2010). Metalinguistic knowledge of university students. *Kyushu Sangyo University Language Education and Research Center Journal*, 5, 74-95.

Topping, K. J. (1998). Peer assessment between students in colleges and universities. *Review of Educational Research*, 68(3), 249-276. doi: 10.3102/00346543068003249